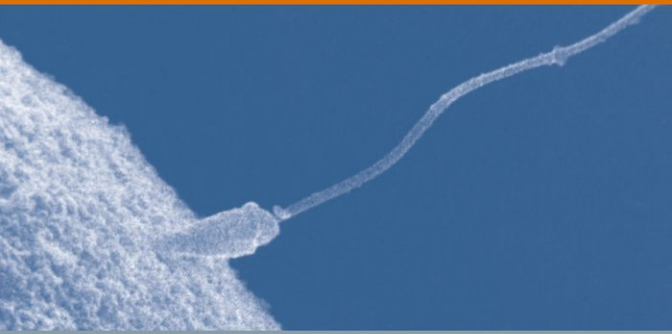


Reproductive and Perinatal Epidemiology

EDITED BY
GERMAINE M. BUCK LOUIS
ROBERT W. PLATT

OXFORD

Reproductive and Perinatal Epidemiology

This page intentionally left blank

Reproductive and Perinatal Epidemiology

Edited by

Germaine M. Buck Louis
Robert W. Platt

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Library of Congress Cataloging-in-Publication Data

Reproductive and perinatal epidemiology / [edited by]
Germaine Buck Louis and Robert Platt.
p. ; cm.
Includes bibliographical references.
ISBN 978-0-19-538790-2
1. Obstetrics. 2. Perinatology. 3. Epidemiology. I. Louis, Germaine Buck.
II. Platt, Robert.
[DNLM: 1. Genital Diseases, Female—epidemiology. 2. Epidemiologic Methods.
3. Infant, Newborn, Diseases—epidemiology. 4. Pregnancy Complications—epidemiology.
WP 140 R4244 2011]
RG526.R47 2011
618.2—dc22
2010025520

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

Dedicated to our spouses, sons, and daughters,
and Summer Institute students from whom we have learned much…

This page intentionally left blank

Preface

Reproductive and perinatal epidemiology has been a longstanding subspecialty of epidemiology, and is characterized by a relatively small but dedicated group of practicing epidemiologists. In 2002, the *Society for Paediatric and Perinatal Research (SPER)* surveyed its membership regarding the availability and composition of formal graduate courses in reproductive and perinatal epidemiology, along with issues pertaining to the recruitment and retention of students in this area. The survey revealed that approximately 40% of academic institutions represented in the study reported formal graduate course offerings in either/both reproductive and perinatal epidemiology. Two key findings emerged from the survey that were relevant to the development of the summer institute: there was considerable heterogeneity in the content, format, and frequency of course offerings; and an urgent need existed to expand options for formalized training in the field, particularly given the absence of such courses in most institutions. The Summer Institute in Reproductive and Perinatal Epidemiology was one global attempt to help build the next generation of reproductive and perinatal epidemiologists.

The origin of the Summer Institute began in 2004, when representatives from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) and the Institute of Human Development, Child and Youth Health (IHDCYH) met at NICHD to discuss strategies for developing formal training programs in reproductive and perinatal epidemiology. There was unanimous support for developing an intensive training program, and the following year was spent recruiting faculty who were willing to develop an integrative curriculum. In addition, each year, an esteemed reproductive or perinatal epidemiologist is selected to be the keynote speaker for the Institute. In 2005, the inaugural Summer Institute for Reproductive and Perinatal Epidemiology was held in Woods Hole, Massachusetts. At the time this book went to press, 120 students from throughout the world representing 14 countries and 49 academic institutions have participated in the Summer Institute.

In planning the curriculum for the Summer Institute, the faculty recognized the absence of a suitable textbook in reproductive and/or perinatal epidemiology designed from a curriculum perspective. This prompted the faculty to develop curriculum for the Institute that met the following criteria: (a) coverage of key reproductive and perinatal outcomes inclusive of each topic's descriptive and analytic epidemiology; (b) an international perspective for the presentation of research findings; and (c) identification of key methodologic issues underlying each endpoint. Suffice to say, the overarching goal of the Summer Institute is to provide students with an appreciation of the epidemiology of successful and unsuccessful

human reproduction and development, largely from a public health perspective, along with important measurement and study design issues related to particular outcomes. The faculty collaborated to remove redundancy and to reinforce methodologic issues shared by two or more outcomes. Course evaluations completed by students were used to further refine the Institute's curriculum, which essentially became the content of this book, enhanced by contributions from keynote speakers. This book is a compilation of the faculty and students' efforts in perfecting the Summer Institute's curriculum for sharing with others interested in the field. The authors remain indebted to and appreciative of the students who inspired the origins and development of this text.

ACKNOWLEDGMENTS

The authors wish to acknowledge the teaching provided by two other summer institute faculty Drs. Mark Klebanoff and Michael Kramer, and keynote speakers Drs. Nigel Paneth and Jon Tyson for their participation in the summer institute.

This page intentionally left blank

Contents

Contributors

Germaine M. Buck Louis, Ph.D., M.S.
Division of Epidemiology, Statistics & Prevention Research
Eunice Kennedy Shriver National Institute of Child Health and Human Development National Institutes of Health
Rockville, Maryland

William D. Fraser, M.D., M.Sc.
Department of Obstetrics and Gynecology
University of Montreal Faculty of Medicine
Montreal, Quebec
Canada

Mary L. Hediger, Ph.D.
Epidemiology Branch
Division of Epidemiology, Statistics & Prevention Research
Eunice Kennedy Shriver National Institute of Child Health & Human Development
National Institutes of Health
Rockville, Maryland

K.S. Joseph, M.D., Ph.D.
Departments of Obstetrics and Gynaecology
School of Population and Public Health
University of British Columbia
Women's Hospital and Health Centre of British Columbia
Vancouver, British Columbia
Canada

Robert W. Platt, Ph.D.
Departments of Pediatrics and of Epidemiology, Biostatistics, and Occupational Health
McGill University
Montreal, Quebec
Canada

David A. Savitz, Ph.D.
Departments of Community Health and Obstetrics and Gynecology
Brown University
Providence, Rhode Island

Enrique F. Schisterman, Ph.D., M.A.
Epidemiology Branch
Division of Epidemiology, Statistics, and Prevention
Eunice Kennedy Shriver National Institute of Child Health and Human Development
National Institutes of Health
Rockville, Maryland

Martha M. Werler, M.P.H., Sc.D.
Slone Epidemiology Center
Boston University
Boston, Massachusetts

Michelle A. Williams, Sc.D.
Department of Epidemiology
University of Washington School of Public Health
Seattle, Washington

Jun Zhang, M.D., Ph.D.
Division of Epidemiology, Statistics and Prevention Research
Eunice Kennedy Shriver National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, Maryland
Reproductive and Perinatal Epidemiology

This page intentionally left blank

Reproductive and Perinatal Epidemiology

This page intentionally left blank

1

Introduction to Reproductive and Perinatal Epidemiology

GERMAINE M. BUCK LOUIS AND ROBERT W. PLATT

To date, no universally accepted definitions exist for reproductive and/or perinatal epidemiology, nor is there agreement among epidemiologists regarding how best to characterize this field. For example, does reproductive and perinatal epidemiology comprise one or two epidemiologic subspecialties? Should they be taught as one or two separate courses? How far into the life course should the sequelae of reproductive and perinatal events be followed?

For purposes of this textbook, the Faculty for the Summer Institute of Reproductive and Perinatal Epidemiology recognize reproductive and perinatal epidemiology as one subspecialty with slightly different but complementary research domains. To encourage dialogue regarding the definition and conceptual paradigm for the field, the Faculty defines reproductive and perinatal epidemiology for the purposes of this book as the study of the distribution, determinants, and sequelae of reproductive and/or perinatal processes and events. This definition is consistent with the more general definition of epidemiology as defined by John Last (2001) in his textbook *A Dictionary of Epidemiology*. The Faculty further recognizes that the field is essentially defined by reproductive and perinatal outcomes (and not exposures, per se), including some that are in fact "markers" of earlier untoward events that escaped measurement, as discussed later in this chapter. Study outcomes for reproductive epidemiology may be conceptualized as falling into one of two research domains—fecundity and fertility—whereas study outcomes for perinatal epidemiology may be similarly conceptualized in two domains—pregnancy and the neonate. As such, pregnancy represents a unique time period within the reproductive time period and, thereby, presents researchers with the option of study factors that affect pregnancy outcomes or the ability to assess the impact of pregnancy outcomes on later-onset adult disease. Irrespective of research domain or outcome selected for study, a host of exposures including genetics, environmental factors, or clinical treatment (e.g., high oxygen for preterm infants, assisted reproductive technologies) are of interest to reproductive and perinatal epidemiologists.

Paradigm for reproductive and perinatal epidemiology

Figure 1.1 illustrates a paradigm for reproductive and perinatal epidemiology from a life course perspective. Such a perspective is important, given the highly interrelated and conditional nature of human reproduction and development. Three important assumptions underlie this paradigm. First, it recognizes both maternal and paternal influences on human reproduction and development, consistent with their couple-dependent nature. As such, a myriad of paternally, maternally, or parentally mediated exposures during vulnerable periods or windows of human development are of interest to reproductive and perinatal epidemiologists. The importance of parental exposures during these windows cannot be overstated in light of a rapidly evolving, albeit controversial, literature demonstrating the permanent reprogramming of the fetus despite the absence of overt pathology. Partially fueled by the literature on endocrine disruption, of late there is growing recognition that parentally mediated periconception or in utero exposures that do not cause structural malformations per se may still affect fetal programming, with implications across the lifespan (Hertzman 1995; Gluckman and Hanson 2004). These subtle changes may manifest as susceptibilities later in life or overt disease. The relation between the in utero environment and adult health has gained stature within epidemiology and across disciplines, giving rise to a hypothesized in utero origin for adult diseases. These hypotheses have been referred to as *fetal programming* or the fetal origins of adult disease (Barker and Osmond 1986; Lucas et al. 1999) and, more recently, the developmental origins of health and disease (Barker 1993). A classic example of an in utero exposure affecting human fecundity and adult health is diethylstilbestrol (DES), which has been associated with effects in

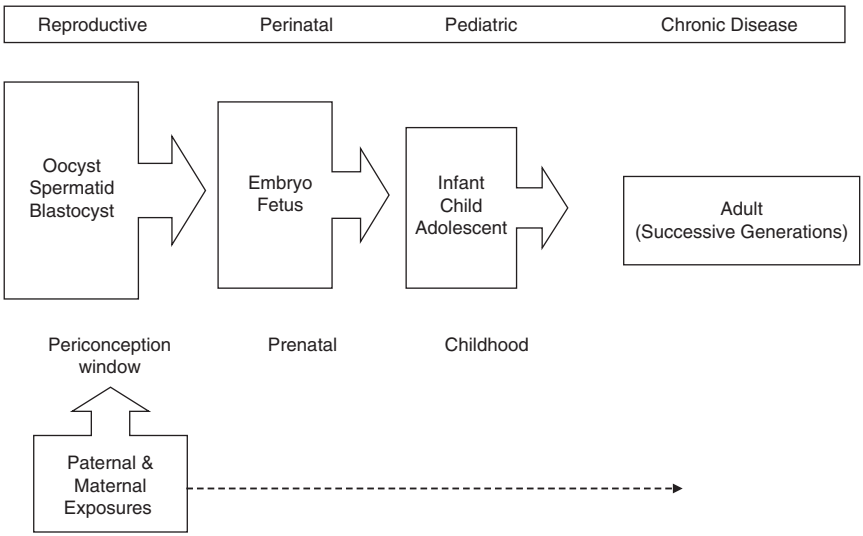


Figure 1.1 Reproductive and perinatal epidemiology–life course perspective.

both males and females (e.g., Gill et al. 1979; Brouwers et al. 2006). Of late, experimental evidence from mice demonstrates the effect of one hormonally active estrogenic environmental chemical—bisphenol A (BPA)—on fecundity over successive generations or the so-called *grandmaternal effect* (Susiarjo et al. 2007). For example, low-dose BPA exposure affected oocyte development in unborn fetuses that later were found to have chromosomally abnormal eggs and embryos as adults. These data support the need to consider in utero effects not only across the lifespan but, possibly, over generations.

A second assumption underlying our conceptual paradigm is the interrelated and conditional nature of human reproduction and development. Successful human reproduction requires the completion of parentally mediated events and processes that are highly interrelated and time sensitive. As such, perinatal endpoints are dependent upon reproductive processes, for without pregnancy there can be no birth or neonate. Reproductive outcomes can be studied from varying paradigms, such as independent outcomes, exposures, or as part of the pathway to pregnancy-related outcomes; similarly, pregnancy outcomes can be studied as independent outcomes, exposures, or intermediates on causal pathways between exposures and child and adult outcomes. Causal diagrams, formalized as directed acyclic graphs (DAGs) (Pearl 1995), are helpful approaches for conceptualizing the paradigm from which fecundity is to be considered when attempting to identify reproductive factors that may influence perinatal outcomes in the context of both measured and unmeasured confounders. For example, couple fecundity, as measured by time-to-pregnancy (TTP), is often used as an endpoint for assessing reproductive toxicants (Baird et al. 1986). But, in fact, a shorter TTP has been associated with a greater likelihood of twinning (Basso et al. 2004; Ferrari et al. 2007) whereas a longer TTP has been associated with miscarriage (Axmon and Hagmar 2005), low birth weight (Basso and Baird 2003), preterm delivery (Henriksen et al. 1997), birth defects (Zhu et al. 2006), and developmental delays (Zhu et al. 2009). However, Cooney and colleagues (2006) failed to observe a relation between conception delay and decrements in gestation or birth weight when controlling for potential confounders such as gravid health conditions, suggesting that conception delay may mediate the relation between health conditions and outcome. The equivocal nature of findings for TTP and pregnancy outcomes requires concerted attention to the conceptual paradigm for study and the selection and inclusion of study variates. The use of causal diagrams may be helpful and are further discussed in Chapter 15.

A third key assumption underlying our conceptual paradigm is the well-known clustering of reproductive and perinatal outcomes (Buck Louis et al. 2006; Hutcheon and Platt 2008a). Our paradigm illustrates that, although each pregnancy may be unique, similar exposures, histories, and risk profiles may present for subsequent pregnancies, particularly if the couple comprises the same set of partners. In the past decade, a number of important papers have empirically demonstrated the potential bias that arises from estimation procedures that do not address the clustering of reproductive outcomes, while other authors have offered new approaches for efficiently addressing this issue. We discuss these issues in Chapter 16.

SPECTRUM OF ENDPOINTS

Reproductive and perinatal epidemiology encompasses a spectrum of endpoints, many of which can be studied either in their normal and/or abnormal state. Figure 1.2 illustrates endpoints by the two research domains underlying reproductive epidemiology and by whether an individual or couple is needed for study. We define *fecundity* as the biologic capacity of males and females for reproduction, irrespective of pregnancy intentions, and *fertility* as demonstrated fecundity, as evident by live births. Fecund men are capable of sexual intercourse and ejaculation, and they have semen quality consistent with fertilizing ability. Fecund women menstruate and ovulate on a consistent basis and are capable of conception and implantation. Fertile men are defined as having fathered a live birth, and fertile women as having delivered a live birth; stillbirths are sometimes included as well.

As Figure 1.2 reflects, the study of female fecundity includes one or more of the following endpoints: onset and progression of puberty, menstruation, ovulation, sexual libido, gynecologic health and disorders, conception, implantation, and menopause. Factors considered in the study of male fecundity may include: onset and progression of puberty, semen quality, sexual libido, urologic health and disorders, paternal determinants of conception and implantation, and andropause. When conception or pregnancy is the study endpoint, the ideal unit is the couple, as illustrated in Figure 1.2. Fecundity impairments also can be studied, and these include conception delays, infecundity, and early pregnancy loss. Fertility endpoints include live (and sometimes still-) births, multiple births, and the ratio of male-to-female births, defined as the *secondary sex ratio*. Recognizably, pregnancy

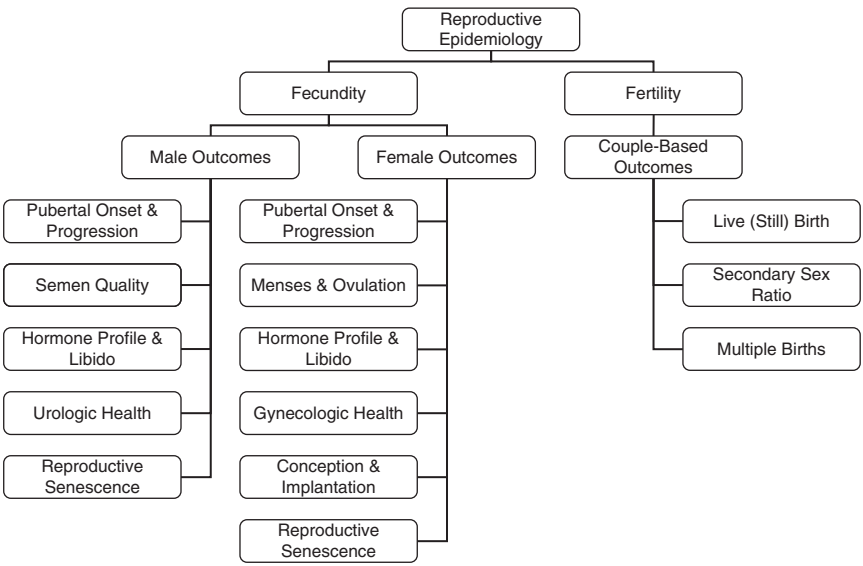


Figure 1.2 Spectrum of reproductive endpoints.

and live birth are outcomes of interest to both reproductive and perinatal epidemiologists, with the latter focusing on health- or disease-related issues impacting the neonate. A more detailed discussion of fecundity and fertility appears in Chapter 3.

The perinatal period has many definitions, reflecting the perspectives of various disciplines ranging from subspecialty clinical perspectives, such as maternal-fetal medicine and neonatology to population-level perspectives inclusive of epidemiology, public health, and demography (with the latter three reflecting a global rather than individual focus). Part of the evolving definitions reflect advances in clinical medicine and public health, given the dramatic improvements in infant survival at the lower limits of gestation and our improved ability to capture fetal demise. For the purposes of this book, we define the perinatal period as the interval before and after delivery, irrespective of whether labor was spontaneous or induced. The perinatal period may be narrowly defined as the time period just before, during, and after birth, or more broadly based upon the lower limits of fetal viability through varying chronological ages following delivery (Lamb and Siegel 2004). Figure 1.3 illustrates a spectrum of perinatal endpoints by timing of occurrence or diagnosis, recognizing that some endpoints are dependent upon chronologic age for diagnosis (e.g., developmental milestones and disabilities). As the figure reflects, many perinatal endpoints focus on pregnancy, including reductions in gestation or fetal growth, pregnancy complications, or multiple births,

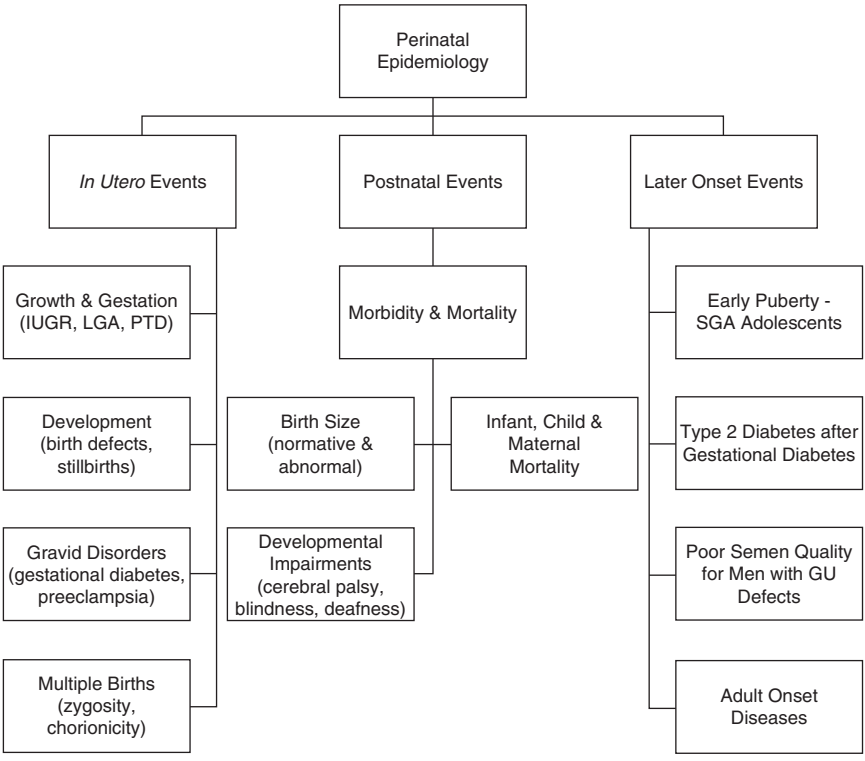


Figure 1.3 Spectrum of perinatal endpoints.

whereas a number of other outcomes occur during the neo- or postnatal periods (e.g., birth size, morbidity, and mortality). Increasingly, epidemiologists recognize that in utero and postnatal events may be associated with later-onset diseases, most notably the association between fetal growth and cardiovascular disease (Barker et al. 1993), and more recently, a spectrum of later-onset adult disease, such as the risk of type 2 diabetes among women with gestational diabetes (Kim et al. 2002) or poorer semen quality among men born with genital-urinary malformations such as cryptorchidism (Boisen et al. 2004) and hypospadias (Virtanen et al. 2001). These emerging associations underscore the interrelatedness of human reproduction and development, including epigenetically induced outcomes or the so-called early origins of health and disease.

OVERVIEW OF METHODOLOGIC CHALLENGES

A number of methodologic nuances underlie reproductive and perinatal epidemiology, and these issues need to be considered when designing reproductive and perinatal epidemiologic research. These include the choice of study population and appropriate unit of analysis, correlated or dependent outcomes, the "hidden data" problem (Hutcheon and Platt 2008b; Paneth 2008), recurrent or clustered events, and outcomes as predictors or causal intermediates. A brief description of each follows, along with more thorough discussions in Chapters 15 and 16.

Choice of study population and appropriate unit of analysis

Defining the referent and study population is a basic tenet of the epidemiologic method that is largely driven by the research question. As noted in the section "Spectrum of Endpoints," a spectrum of reproductive and perinatal outcomes is of interest to the field and impacts the choice of study population. For example, researchers interested in pubertal onset and development or gynecologic disorders would most likely focus on prepubescent children or women of reproductive age, respectively. However, when couple-dependent outcomes, such as pregnancy, are of interest, it may be insufficient to focus on only men or only women, thus necessitating the need to target and recruit couples, as only a few TTP studies have done (Zinaman et al. 1996; Bonde et al. 1998). For study designs focusing on birth defects or other adverse perinatal outcomes that utilize dyad or triad designs, mothers, fathers, and infants may represent the study population (Vermeulen et al. 2009). Thus, choice of study population may require sampling on individuals, couples, or families. Even within populations, sampling and design issues may arise. For example, cohorts of pregnant women are typical study cohorts when assessing perinatal events, but investigators may be interested in targeting subgroups of pregnant women, such as women at high risk for gravid diseases (e.g., hypertensive women) or adverse perinatal outcomes (e.g., women carrying multiples).

Along with deciding upon the study population, it is important to determine a priori the unit of analysis, so that data collection can be designed accordingly

and to ensure sufficient statistical power. Many reproductive and perinatal epidemiologic studies have hierarchical data structures denoting the various levels of data. For example, TTP studies may sample on couples interested in becoming pregnant but recognize the need for couple- (both partners) and individual- (each partner) level data along with a finer unit for an individual partner, such as menstrual cycles or even calendar days. Among couples undergoing assisted reproductive technologies, the unit of analysis can extend beyond couples to treatment cycle–level data to the number of oocytes retrieved, fertilized, or implanted (Buck et al. 2004a). The same may occur with multiple pregnancies per woman (Buck Louis et al. 2006); epidemiologists must take care in defining the cohort of interest (e.g., primiparous women versus all pregnancies per women) and recognize the potential issues that may arise. A similar issue arises for perinatal epidemiologists working with twins or higher-order pregnancies. Historically, epidemiologists restricted study populations to women with singleton pregnancies, sometimes to circumvent having to address a varying unit of analysis when interested in pregnancy outcomes or given the biologic immaturity of multiples in comparison to singletons. Fortunately, a host of new modeling techniques developed in the statistical literature in recent decades, such as generalized linear mixed models (Breslow and Clayton 1993) and generalized estimating equations (Liang and Zeger 1986), and advances in Bayesian computation (Gelman et al. 2003), enable investigators to design research that includes all pregnant women, irrespective of plurality of birth, thus allowing study populations to more closely resemble their targeted populations. Suffice to say, specification of the research question in relation to the target study population will help inform about choice of study population and unit(s) of analysis needed. A more in-depth discussion of the unit of analysis is presented in Chapter 16.

Correlated or dependent outcomes

Closely linked to the unit of analysis is the correlation that arises between observations that are not independent of each other, such as observations taken from multiple pregnancies from the same woman. Reproductive and perinatal epidemiologists have long recognized the correlation or dependent nature of pregnancy outcomes in that women tend to repeat pregnancy outcomes in successive pregnancies. For example, women with a prior history of an ectopic pregnancy (Levin et al. 1982), a pregnancy loss or preterm birth (Carr-Hill and Hall 1985), a low-birth-weight infant (Khoury et al. 1989), or an infant with a birth defect (Lie et al. 1994) are at increased risk of experiencing that same outcome in future pregnancies. Also, female fecundability or TTP has been reported to cluster within women, meaning that women who conceive quickly tend to do so over successive pregnancy attempts (Basso et al. 1997). The considerable body of evidence supporting the dependent nature of pregnancy outcomes is the basis for obstetricians' querying women about prior reproductive history upon initiation of prenatal care. In fact, many perinatal scoring systems include prior history of adverse outcomes as a criterion (Wall 1988) and may result in affected women being considered for high-risk pregnancy protocols.

During the past decade, considerable research has addressed the dependent nature of reproductive outcomes (Scheike et al. 1999; Buck Louis et al. 2006; Hutcheon and Platt 2008a) and cautioned against past approaches, such as restricting samples to include only one pregnancy per woman or treating prior history as a confounder (Weinberg 1993). In fact, an entire issue of *Paediatric and Perinatal Epidemiology* (Ananth 2007) was devoted to the topic of recurrent risk. From a biologic perspective, multiple data points from a single woman or man tend to be more alike than those from a random pair of women or men. In statistical terminology, this is known as a *correlated* or *dependent-data problem*, which is a common characteristic of longitudinal data (Diggle et al. 2002), and a key feature of reproductive and perinatal epidemiology. Currently, several statistical models are well suited for analyzing correlated data structures, such as hierarchical, multilevel, or Bayesian models. A more in-depth description is provided in Chapter 16.

Hidden data

A unique challenge in reproductive and perinatal epidemiology is our inability to identify the referent population—the so-called *hidden* or *missing-data problem*. For much of reproductive epidemiology, conception is the intended denominator, yet, in the absence of longitudinal biomarker capture, a rather large (and relatively unknown) proportion of conceptions are unmeasurable prior to implantation or following implantation. Conception cohorts are the denominator for much of the field, yet, we rely on denominator proxies, such as human chorionic gonadotropin–detected pregnancies or women presenting for prenatal care. Paneth (2008) has described the hidden-data problem in perinatal epidemiology, particularly when focusing on fetal deaths, building upon earlier work that identified fetuses in utero as the appropriate denominator for research focusing on fetal deaths or stillbirths (Yudkin et al. 1986). This controversial avenue of research has been collectively referred to as the *fetuses-at-risk paradigm*, and has been extended to postnatal birth outcomes as well, as noted in Chapter 13 (Joseph et al. 2003; Platt et al. 2004).

The concept of hidden data also affects other perinatal outcomes. For example, in the study of fetal growth, birth weight is readily measured and gestational age can be estimated. However, birth weight and gestational age represent the cumulation of growth processes throughout pregnancy. As a result, the complete trajectory of fetal growth, with all its biological processes, remains unobservable, or hidden, from most researchers. Such missing or hidden data may introduce bias when exposures affect both fetal growth and gestation (Hutcheon and Platt 2008b). The fetuses-at-risk and hidden-data problem are discussed in Chapters 14 and 16.

Interval censoring and truncation

Censoring is typically defined in epidemiology as the loss of participants over the course of the study, so that the exact time of an outcome's occurrence is unknown.

Right censoring occurs when the endpoint occurs after loss to follow-up, so that it is only known that the event had not occurred by the time of the last follow-up. For example, in TTP studies, a couple who enroll but are lost to follow-up before becoming pregnant would be right-censored. *Left censoring* occurs when the endpoint is known to have occurred before a specific follow-up time, but the exact time is unknown. Conception is an example of a (usually) left-censored variable; if a woman presents already pregnant, it is known that conception occurred prior to that date, but the precise time is unknown. Censoring requires close monitoring of attrition and withdrawal rates to ensure that its sources are unrelated to the study outcome. Withdrawal rates for prospective pregnancy studies with preconception enrollment range from 4% to 62% (Buck Louis et al. 2004b), so a considerable percentage of couples will not have a TTP and will be right-censored, as described above. As a discipline, reproductive and perinatal epidemiology needs to be concerned about *interval censoring*, or simultaneous left and right censoring, in which the time of the event is known to fall within an interval. For example, frequently, we do not know the exact timing of embryonic or fetal death; rather, we know only a date when the embryo presumably was alive (i.e., that there was a positive pregnancy test) and the time a woman noted vaginal bleeding, which may be several weeks following in utero demise of the embryo or fetus.

Truncation, referring to truncated time windows of observation in which the endpoint may or may not occur, is more of an acute concern in reproductive and perinatal epidemiology than in chronic disease epidemiology. Left truncation occurs when follow-up starts after another event (not necessarily the endpoint) has occurred, such that it is not known whether the endpoint occurred prior to the start of follow-up. Examples of truncation and interval censoring include TTP and pregnancy loss. Time-to-pregnancy studies typically enroll women or couples upon discontinuation of contraception, but all too often, an unknown period exists during which the couple is at risk for pregnancy but not necessarily enrolled in the study; this time is left truncated (in that pregnancy and loss may have occurred, but will be undetected in the study). This situation may arise when a couple discontinues contraception, engages in sexual intercourse without any form of contraception, but does not enroll in the study for several days, weeks, or months.

Pregnancy loss or miscarriage studies are other examples of endpoints with possible truncation and censoring issues. As with TTP studies, conception may be unmeasured for some or all women in prospective pregnancy studies; thus, women enter the prospective pregnancy study at varying time intervals, including after becoming pregnant or after already having experienced an unrecognized pregnancy loss. Further complicating this truncation is the considerable uncertainty underlying the exact timing when embryonic or fetal demise occurs or the duration of follow-up during pregnancy. This latter phenomenon is interval censoring: it is known that the event (fetal demise) occurred, but only within a window of time. In sum, truncation and censoring in pregnancy studies imply the inability to observe all time (initiation for truncation and completion for censoring) at risk for the outcome. Thus, reproductive and perinatal epidemiologists need to be cognizant of interval censoring and truncation in the design and analysis of their research.

OUTCOMES AS PREDICTORS AND CAUSAL MEDIATION

An exciting aspect of reproductive and perinatal epidemiology is the ability to design etiologic research aimed at delineating causal mechanisms for a particular outcome, followed by research in which that same endpoint is used as a potential cause or predictor of other (adverse) reproductive or perinatal outcomes; it is said that the endpoint may mediate the causal pathway between the upstream exposure and the reproductive outcome (Pearl 1995). Reproductive and perinatal epidemiologists often study "intermediates" in a causal pathway as study outcomes, in part, given our inability to identify etiologic mechanisms that arise earlier in the causal pathway. For example, preterm birth is the subject of a voluminous body of research, but in fact is a proxy for as-yet unmeasured events. Similarly, much research focuses on preeclampsia that arises during pregnancy. This gravid disease may have an in utero origin despite its first appearance during pregnancy and therefore be an intermediate in the pathway to later-onset diseases such as metabolic syndrome or type 2 diabetes. However, such an approach may assume a greater etiologic understanding of the outcomes than is actually known, with implications for its use in prediction models. For example, a handful of prospective cohort studies have attempted to identify lifestyle factors that affect female fecundability as measured by TTP. Although several factors have been associated with reduced fecundability or a longer TTP, giving the impression that we understand the determinants of female fecundity, Axmon and colleagues (2006) reported that only 14% of the variation in retrospectively reported TTP was explained by these so-called risk factors. Despite considerable uncertainty regarding the determinants of female and male fecundity, TTP is increasingly used as a predictor of pregnancy outcomes, including preterm (Henriksen et al. 1999) and low-birth-weight (Williams et al. 1991) deliveries and birth defects (Zhu et al. 2006). In fact, these findings are interpreted by some as suggesting that women with impaired fecundity or those requiring a longer TTP are at risk for many adverse outcomes, including treatment with assisted reproductive technologies. The challenge facing reproductive and perinatal epidemiologists will be the specification of the causal pathway for reproductive and perinatal outcomes and, concomitantly, model specification. Recently, investigators have suggested the use of causal diagrams, or directed acyclic graphs (DAGs), for determining how best to utilize reproductive history when interested in unbiased estimates of an effect (Howards et al. 2007). Chapter 15 addresses this issue further and underscores the importance of such considerations for reproductive and perinatal epidemiologic research and other life-course epidemiologic research that is grounded within a periconceptional or in utero origin-of-disease framework.

In sum, these are but a few of the many methodologic issues that require careful planning when designing reproductive and perinatal epidemiologic research, and each will be discussed further. These challenges are, however, in keeping with the dependent nature of human reproduction and development, the strong role of behavior in the context of biology (for example, when couples behave to place themselves at risk for pregnancy), and the important role of prior history in understanding the etiology of adverse outcomes or in models aimed at their prediction.

REFERENCES

Ananth, C.V., ed. 2007. Successive pregnancy outcomes: A decade of progress. *Paediatr Perinatal Epidemiol* 21(Suppl 1):1–65.

Axmon, A. and Hagmar, L. 2005. Time to pregnancy and pregnancy outcome. *Fertil Steril* 84:966–74.

Axmon, A., Rylander, L., Albin, M., and Hagmar, L. 2006. Factors affecting time to pregnancy. *Hum Reprod* 21:1279–84.

Baird, D.D., Wilcox, A.J., and Weinberg, C.R. 1986. Use of time to pregnancy to study environmental exposures. *Am J Epidemiol* 124(3):470–80.

Barker, D.J. and Osmond, C. 1986. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet* 1:1077–81.

Barker, D.J., Gluckman, P.D., Godfrey, K.M., Harding, J.E., Owens, J.A., and Robinson, J.S. 1993. Fetal nutrition and cardiovascular disease in adult life. *Lancet* 341(8850): 938–41.

Basso, O., Olsen, J., Bisanti, L., Bolumar, F., Kuppers-Chinnow, M., the European Study Group on Infertility and Subfecundity. 1997. Repeating episodes of low fecundability: A multicentre European study. *Hum Reprod* 12:1448–53.

Basso, O. and Baird, D.D. 2003. Infertility and preterm delivery, birthweight, and caesarean section: A study within the Danish National Birth Cohort. *Hum Reprod* 18:2478–84.

Basso, O., Christensen, K., and Olsen, J. 2004. Fecundity and twinning. A study within the Danish National Birth Cohort. *Hum Reprod* 19:2222–26.

Boisen, K.A., Kaleva, M., Main, K.M., Virtanen, H.E., Haavisto, A.M., Schmidt I.M., et al. 2004. Difference in prevalence of congenital cryptorchidism in infants between two Nordic countries. *Lancet* 363:1264–69.

Bonde, J.P., Hjollund, H., Jensen, T.K., Ernst, E., Kolstad, H., Henriksen, T.B., et al. 1998. A follow-up study of environmental and biological determinants of fertility among 430 Danish first pregnancy planners: Design and methods. *Reprod Toxicol* 12:19–27.

Breslow, N.E. and Clayton, D.G. 1993. Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 88:9–25.

Brouwers, M.M., Feitz, W.F.J., Roelofs, L.A.J., Kiemeney, L.A.L.M., de Gier, R.P.E., and Roeleveld, N. 2006. Hypospadias: A transgenerational effect of diethylstilbestrol? *Hum Reprod* 21:666–69.

Buck, G.M., Schisterman, E.F., Dukic, V.M., and Schieve, L.A. 2004a. Research hurdles complicating the analysis of infertility treatment on child health. *Hum Reprod* 20:12–18.

Buck, G.M., Lynch, C.D., Stanford, J.B., Sweeney, A.M., Schieve, L.A., Rockett, J.C., et al. 2004b. Prospective pregnancy study designs for assessing reproductive developmental toxicants. *Environ Health Perspect* 112:79–86.

Buck Louis, G., Dukic, V., Heagerty, P.J., Louis, T.A., Lynch, D.C., and Ryan, L.M. 2006. Analysis of repeated pregnancy outcomes. *Stat Meth Med Res* 15:103–26.

Carr-Hill, R.A. and Hall, M.H. 1985. The repetition of spontaneous preterm labour. *Br J Obstet Gynecol* 92:921–28.

Cooney, M.A., Buck, G.M., Wenyu, S., Rice, M.M., and Klebanoff, M.A. 2006. Is conception delay a risk factor for reduced gestation or birthweight? *Paediatr Perinatal Epidemiol* 20:201–209.

Diggle, P.J., Heagerty, P.J., Liang, K.-Y., and Zeger, S.L. 2002. *Analysis of longitudinal data*. New York: Oxford University Press.

Ferrari, R.M., Cooney, M.A., Vexler, A., Liu, A., and Buck Louis, G.M. 2007. Time to pregnancy and multiple births. *Hum Reprod* 22:407–13.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 2003. *Bayesian data analysis*, 2nd ed. Boca Raton: Chapman & Hall.

Gill, W.B., Schumacher, G.F., Bibbo, M., Straus, F.H. 2nd, and Schoenberg, H.W. 1979. Association of diethylstilbestrol exposure in utero with cryptorchidism, testicular hypoplasia and semen abnormalities. *J Urol* 122:36–39.

Gluckman, P.D. and Hanson, M.A. 2004. Living with the past: Evolution, development, and patterns of disease. *Science* 305:1733–36.

Henriksen, T.B., Baird, D.D., Olsen, J., Hedegaard, M., Secher, N.J., and Wilcox, A.J. 1997. Time to pregnancy and preterm delivery. *Obstet Gynecol* 89:594–99.

Henriksen, T.B., Baird, D.D., Olsen, J., Hedegaard, M., Secher, N.J., and Wilcox, A.J. 1999. Time to pregnancy and preterm delivery. *Obstet Gynecol* 89(4): 595–99.

Hertzman, C. 1995. The biologic embedding of early experience and its effects on health in adulthood. *Ann NY Acad Sci* 896:85–95.

Howards, P.P., Schisterman, E.F., and Heagerty, P.J. 2007. Potential confounding by exposure history and prior outcomes: An example from perinatal epidemiology. *Epidemiology* 18:544–51.

Hutcheon, J.A. and Platt, R.W. 2008. The impact of past pregnancy experience on subsequent perinatal outcomes. *Paediatr Perinatal Epidemiol* 22:400–408.

Hutcheon, J.A. and Platt, R.W. 2008. The missing data problem in birth weight percentiles and thresholds for "small-for-gestational-age." *Am J Epidemiol* 167(7):786–92.

Joseph, K.S., Liu, S., Demissie, K., Shi Wu, W., Platt, R.W., Ananth, C.V., et al. 2003. A parsimonious explanation for intersecting perinatal mortality curves: Understanding the effect of plurality and of parity. *BMC Pregnancy Childbirth* 3(1): 3.

Khoury, M., Calle, E., and Joesoef, R. 1989. Recurrence of low birthweight in siblings. *J Clin Epidemiol* 42: 1171–78.

Kim, C., Newton, K.M., and Knopp, R.H. 2002. Gestational diabetes and the incidence of type 2 diabetes: A systematic review. *Diabetes Care* 25(10):1862–68.

Lamb, V.L. and Siegel, J.S. 2004. Health demography. In: *The methods and materials of demography*, 2nd ed., pp. 341–70. J.S. Siegel, D. Swanson, and H.S. Shryock, eds. Boston: Elsevier Academic Press.

Last, J.M. 2001. *A dictionary of epidemiology*, 4th ed., p.62. New York: Oxford University Press.

Levin, A.A., Shoenbaum, S.C., Stubblefield, P.G., Zimicki, S., Monson, R.R., and Ryan, K.J. 1982. Ectopic pregnancy and prior induced abortion. *Am J Public Health* 72: 253–56.

Liang, K.-Y. and Zeger, S.L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73(1): 13–22.

Lie, R.T., Wilcox, A.J., and Skjerven, R. 1994. A population based study of the risk of recurrence of birth defects. *NEJM* 331:1–4.

Lucas, A., Fewtrell, M.S., and Cole, T.J. 1999. Fetal origins of adult disease–the hypothesis revisited. *BMJ* 319:245–49.

Paneth, N. 2008. Invited commentary: The hidden population in perinatal epidemiology. *Am J Epidemiol* 167(7):793–96.

Pearl, J. 1995. Causal diagrams for empirical research. *Biometrika* 82(4): 669–88.

Platt, R.W., Joseph, K.S., Ananth, C.V., Grondines, J., Abrahamowicz, M., and Kramer, M.S. 2004. A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death. *Am J Epidemiol* 160(3): 199–206.

Scheike, T.H., Petersen, J.H., and Martinussen, T. 1999. Retrospective ascertainment of recurrent events: An application to time to pregnancy. *JASA* 94:713–25.

Susiarjo, M., Hassold, T.J., Freeman, E., and Hunt, P.A. 2007. Bisphenol A exposure in utero disrupts early oogenesis in the mouse. *PLoS Genetics* 3(1):e5.

Vermeulen, S.H., Shi, M., Weinberg, C.R., and Umbach, D.M. 2009. A hybrid design: Case-parent triads supplemented by control-mother dyads. *Genetic Epidemiol* 33: 136–44.

Virtanen, H.E., Kaleva, M., Haavisto, A.M., Schmidt, I.M., Chellakooty, M., Main, K.M., et al. 2001. The Barth rate of hypospadias in the Turku area in Finland. *APMIS* 109: 96–100.

Wall, E.M. 1988. Assessing obstetric risk. A review of obstetric risk-scoring systems. *J Fam Practice* 27:153–63.

Weinberg, C.R. 1993. Toward a clearer understanding of confounding. *Am J Epidemol* 137:1–8.

Williams, M., Goldman, M.B., Mittendorf, R., and Monson, R. 1991. Subfertility and the risk of low birthweight. *Fertil Steril* 56(4): 668–71.

Zhu, J.L., Basso, O., Obel, C., Bill, C., and Olsen, J. 2006. Infertility, infertility treatment, and congenital malformations: Danish national birth cohort. *BMJ* 333(7570):679.

Zhu, J.L., Basso, O., Obel, C., Hvidtjorn, D., and Olsen, J. 2009. Infertility, infertility treatment and psychomotor development: The Danish National Birth Cohort. *Paediatr Perinatal Epidemiol* 23:98–106.

Zinaman, M.J., Clegg, E.D., Brown, C.C., O'Connor, J., and Selevan, S.G. 1996. Estimates of human fertility and pregnancy loss. *Fertil Steril* 65:503–509.

2

Sensitive Windows of Human Reproduction and Development

GERMAINE M. BUCK LOUIS

OVERVIEW OF HUMAN REPRODUCTION AND DEVELOPMENT

Successful human reproduction and development is dependent upon men and women undergoing a series of highly interrelated and conditional processes, many of which delineate critical or sensitive windows of human development. A *critical window* of human development has been defined as a period marked by cellular proliferation and development, and increasing metabolic capabilities of the developing organism. As a time-sensitive interval, exposure(s) during a critical window may disrupt or interfere with the physiology of a cell, tissue, or organ (Morford et al. 2004). Often, exposures during these periods will result in permanent and irreversible effects (Calabrese 1986), including imprinting-related effects. *Sensitive windows* may or may not be inclusive of critical windows, but if they do, they typically extend the period of susceptibility and result in nonstructural defects. These may include functional or subtle effects that may not manifest until later in life. For example, exposures that affect the number and quality of oocytes, whose number is presumed to be determined in utero, may not manifest until later in life and will have important implications for a woman's ovarian reserve. Ben-Shlomo and Kuh (2002) have eloquently articulated the need to distinguish critical and sensitive windows, given that exposures during the latter may still adversely affect development and the appearance of adult-onset diseases, although possibly with reduced magnitude.

Critical and sensitive windows are also important for understanding the role of epigenetics in human reproduction and development, since early embryonic development is regulated by both genetic and epigenetic mechanisms. Unlike genetic mechanisms that depend upon the DNA code, epigenetic mechanisms are not DNA sequenced–based (Lucifero et al. 2004). Epigenetic changes may be defined as molecular alterations in gene expression or phenotype that do not produce a change in DNA sequence and are an example of effects stemming from the differential expression of parental alleles following epigenetic modification of genes in the germline and embryo (Latham 1999; Ferguson-Smith and

Surani 2001; Li 2002a). Epigenetic effects are mediated by modifications of chromatin structure that may silence, enhance, or modulate gene expression. DNA methylation or histone modifications are two aspects of epigenetics that have received considerable study to date.

A host of environmental factors interact with the genome and may disrupt epigenetic programming (e.g., chemicals, diet, teratogens). Some of these effects are presumed to be transitory, whereas others are passed through the germline to the offspring (Owen and Segars 2009). Epigenetic mechanisms warrant couple-based approaches for assessing environmental exposures during critical or sensitive windows when couple-dependent outcomes, such as pregnancy, are of interest. Epigenetic effects arising at the gamete or embryo stage of human development are postulated to be associated with a couple's use of assisted reproductive technologies (ART). For example, Beckwith-Wiedemann syndrome, Angelman syndrome, and retinoblastoma are three imprinting disorders that have been associated with ART (DeBaun et al. 2003; Gicquel et al. 2003; Moll et al. 2003).

Human beings are inefficient reproducers in comparison to other mammals (Stevens 1997; Norwitz et al. 2001). Understanding the reasons for inefficient human reproduction is difficult, given our limited access to couples and our ability to measure exposures and responses during critical and sensitive windows of human development. To address critical data gaps in our understanding of human reproduction (and its inefficiency), better methods are needed for identifying two important denominators for much of our work—ovulation and conception. Available methods for measuring ovulation and conception are still imprecise, except in subgroups of the population, such as couples undergoing fertility-related treatments, in whom the presence or absence of ovulation and/or conception can be identified. The extent to which findings from such subpopulations are generalizable to the general population remain to be established. Perhaps some of this inefficiency may reflect the quality of our methods for measuring timed intercourse and ensuring pregnancy or our inability to capture time-varying exposures that affect conception probabilities. In addition, we are just beginning to appreciate the purported ability of the developing organism to repair some damage (developmental plasticity). Greater understanding of the mechanisms underlying critical and sensitive windows of male and female reproduction and development will better inform researchers on how best to collect data and biospecimens and, ultimately, answer the remaining questions about the pathophysiology of adverse reproductive and perinatal outcomes.

For the purposes of this text, we identify six critical and sensitive windows for human reproduction and development that may have implications for health across the lifespan: folliculogenesis, spermatogenesis and steroidogenesis, menstruation, ovulation and the fertile window, implantation window, and the intrauterine environment. These windows are either directly or indirectly dependent, at least initially, upon the hypothalamus-pituitary-gonadal (HPG) axis, which is the primary hormonal pathway responsible for reproduction. Briefly, the hypothalamus, located in the base of brain above the pituitary gland, releases gonadotrophin-releasing hormone (GnRH) in a pulsatile manner that stimulates the anterior

pituitary via blood flow through the hypothalamic-pituitary portal vessels to release the gonadotropin hormones, follicle-stimulating hormone (FSH) and luteinizing hormone (LH). The gonadotrophins stimulate the gonads to release steroid hormones. In females, estrogen and progesterone are secreted; in males, testosterone. The remainder of this chapter presents a brief description of each window, along with critical data gaps to encourage further research.

FOLLICULOGENESIS

The follicle is the functional unit of the ovary and comprises the oocyte and its layers of granulosa cells that nourish the oocyte until ovulation. *Folliculogenesis* refers to the production and development of the oocyte across its stages of development: primordial, primary, early secondary, late secondary, tertiary, and preovulatory (Gonzalez-Bulnes and Veiga-Lopez 2007). Although oocytes originate in utero commencing between 4 and 5 months postconception as primordial follicles, their maturation remains arrested until the first meiotic division, which occurs just prior to ovulation (Peters 1970). Meiosis is the process by which diploid germ cells (oogonia or spermatogonia) reduce by half their number of chromosomes in preparation of fertilization (Eppig et al. 2004). During the ovarian cycle, meiosis resumes in response to LH, although the mechanisms are incompletely understood, resulting in the oocyte completing its first meiotic division. Following puberty, a given number of primordial follicles are recruited daily from a pool of resting ovarian follicles. Follicular growth is dependent on FSH, and the process of recruiting and selecting the antral (dominant) follicle for ovulation takes approximately 4 to 6 months (Zeleznik 2004). Selection of the dominant follicle occurs during the follicular phase of the ovarian cycle; selection of more than one follicle is believed to be the basis for dizygotic twinning (Lambalk et al. 1998).

Although the development of the oocyte is relatively long, commencing during fetal life and ending shortly before ovulation in the postpubescent girl, the lifespan of the mature fertilizable oocyte is short, possibly a few hours. Exposures during folliculogenesis may damage or reduce the pool (or quality) of available oocytes and thereby affect female fecundity. The duration of folliculogenesis is determined by a number of factors, including those originating in utero (quality and/or number of oocytes, number of oocytes ovulated) or during the woman's lifetime, such as exposures to ovarian toxicants or ovulation induction treatments (Scott et al. 1990; Hansen et al. 2003). Of late, methods for empirically estimating ovarian reserve as a measure of female fecundity have arisen and include measuring inhibin B and anti-Müllerian hormone (De Vet et al. 2002; Fanchin et al. 2003) when high-resolution ultrasonography—considered to be the gold standard—is not available (Wallace and Kelsey 2004). Although considerable insight has been gained regarding folliculogenesis, many questions remain, including the identification of the molecular mechanisms, such as the oocyte-specific transcription factors, that control the growth and selection of the oocyte.

SPERMATOGENESIS AND STEROIDOGENESIS

Spermatogenesis and *steroidogenesis* refer to the testicular processes responsible for the production of gametes (spermatozoa) and sex hormones (androgens), respectively. A number of interrelated processes define spermatogenesis, including the replication of stem cells, meiotic division of spermatocytes, and the transformation of haploid germ cells (spermatids) into sperm or spermiogenesis. These processes occur in the seminiferous tubules located in the testes and result in the repackaging of the paternal genome (Krawetz 2005). The HPG axis controls spermatogenesis following stimulation of the hypothalamus and pituitary to release FSH and LH. Follicle-stimulating hormone acts directly on the seminiferous epithelium, whereas LH exerts its effect by stimulating testosterone secretion from the Leydig cells. Steroidogenesis refers to the enzymatic reactions that result in the production of androgens, and it occurs in the interstitial compartment of the testes. As with folliculogenesis, there is considerable variability in the processes controlling spermatogenesis and steroidogenesis, possibly arising from differences in the onset and progression of puberty, the male's ability to produce quality sperm, and changes in androgen production or synthesis or the so-called andropause (Vermeulen 1993).

There is considerable intra- and intervariability in semen quality and, as yet, no one marker of semen quality is highly predictive of male fecundity and fertility. Recent evidence suggests that paternal spermatozoal factors activate the zygotic genome (Miller et al. 1999; Parrington 2000; Ostermeier et al. 2002, 2005;), thus generating new theories about paternally mediated influences on human development. For example, what testis-specific histones or proteins underlie the packaging of the paternal genome, and what environmental factors may affect these processes?

MENSTRUATION

Age at menarche is defined as the onset of menstruation and is considered a developmental milestone for girls, representing the lower bound of the female's reproductive life. Conversely, menopause or the absence of menses for 12 consecutive months represents the upper bound of the female's reproductive life. A period of subfecundity or lower probability of conception typically follows the onset of menarche and precedes the onset of menopause.

The *menstrual cycle* is typically described as having three phases—bleeding, proliferative, and secretory—reflecting the changes that are occurring in the endometrium. The term *ovarian cycle* is sometimes interchanged with the menstrual cycle, but technically refers to changes in the ovary and is defined as having two phases—follicular and luteal. Ovulation is the event that separates the proliferative and secretory phases of the menstrual cycle, and the follicular and luteal phases of the ovarian cycle. Figure 2.1 illustrates the menstrual and ovarian cycles in relation to the hormonal mechanisms underlying their control. Estradiol concentrations rise in the proliferative phase and the inherent negative feedback mechanism underlying the HPG axis reduces FSH during the later stage of this phase.

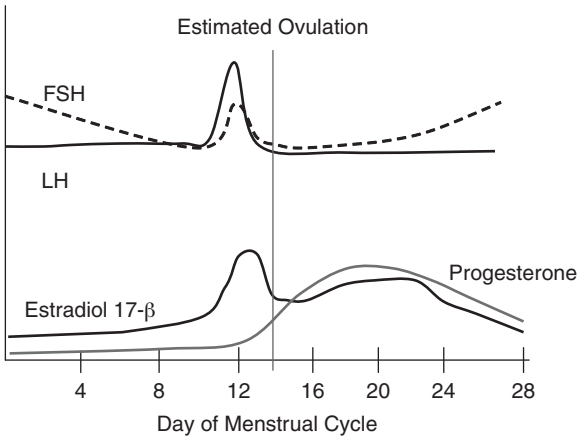


Figure 2.1 Hormonal profiles during the menstrual cycle.

Rising estradiol concentration during the proliferative phase has a positive feedback on the production of LH and is believed responsible for triggering the LH surge. Following ovulation, the ruptured follicle or corpus luteum begins secreting progesterone, estradiol, and 17-α-hydroxyprogesterone. Progesterone and estradiol release exerts a negative feedback on gonadotropin secretion and, thereby, serum levels gradually decline. As our ability to capture time-varying exposures and covariates during this interval improves, it may be possible to identify adverse (including subtle) effects during this sensitive window.

The menstrual cycle begins on day 1, which is typically defined as the first day of bleeding and ends at day 1 of the next cycle. Measurement using this definition requires two assumptions, albeit based on limited population data. The assumptions are: (1) women accurately recognize and report menstrual bleeding, and (2) no other bleeding occurs within the menstrual cycle. Prospective longitudinal data suggest otherwise for both assumptions: cycles vary within women (Treloar et al. 1967; Vollman 1977; Fehring 2006) and distinct menstrual bleeding patterns may exist (Mikolajczyk et al. 2010).

Although it is not infrequent to encounter references to the mean cycle length or duration of bleeding in the clinical literature, it is important to note that some of these estimates have not utilized statistical methodologies suitable for correlated outcomes, such as menstrual cycles. For example, a mean menstrual cycle of 28.9 ± 3.4 days (with 95% of cycles between 22 and 36 days) and a mean duration of bleeding of 5.8 ± 2.9 days (with 95% of cycles having lengths between 3 and 8 days) was reported in a longitudinal study of women utilizing a fertility monitor that tracked estrone-3-glucuronide (E_3G) and LH in an attempt to avoid pregnancy (Fehring et al. 2006). The estimated mean follicular length was 16.5 ± 3.4 days (with 95% cycles having follicular length of 10–22 days), whereas the mean luteal phase was 12.4 ± 2 days (with 95% cycles between 9 and 16 days). None of these estimates, however, was derived from statistical methods capable of addressing the correlated menstrual cycles within women. Although much of the available

research relies upon retrospectively collected menstrual cycle data, its validity has received only limited study. Small and colleagues (2007) reported that 43% of women reported menstrual cycle lengths to be 2 or more days longer than their mean observed lengths, as captured with daily diaries. Only moderate (κ coefficient 0.33) agreement was observed in another study, with menstrual cycle length being overestimated by approximately 0.7 days (95% confidence interval [CI] 0.3–1.0) based on self-reports (Jukic et al. 2007). Bachand and colleagues (2009) assessed the reliability of menstrual cycle length data in a small sample of women and reported 93% agreement between retrospective and prospective diary data when categorizing menstrual cycles at 35 days or less versus more than 35 days. Such a crude dichotomy may be insufficient for detecting smaller changes in menstrual cycle length associated with various exposures. At a minimum, these combined findings support prospective measurement when possible and the need for cautious interpretation of results, particularly for modest changes in menstrual cycle characteristics or their meaning for female fecundity. Even greater caution is warranted when relying upon retrospective reporting for menstrual cycle data, given the inherent measurement error associated with recall.

OVULATION AND THE FERTILE WINDOW

The expulsion of the oocyte from an ovarian follicle is defined as ovulation. To date, no biomarker of ovulation has been found; hence, transvaginal daily ultrasonography is considered the gold standard for ovulation detection. Several other proxies of ovulation are used by clinicians, women, and researchers and include calendar-based methods, in which one counts backward from the date of expected menstruation, or the tracking of basal body temperature, cervical mucus, or daily hormonal profiles. With regard to the latter proxy, various approaches have been utilized, including (a) the ratio of urinary estrogen to progesterone metabolites (E/P algorithm) (Baird et al. 1991); (b) peak serum LH (Stewart et al. 1993); (c) midcycle peak urinary FSH (Li et al. 2002b), and (d) the rapid rise of urinary progesterone metabolites (pregnanediol-3-glucuronide (PdG)-rise algorithm) during the early luteal phase (Chen et al. 2005).

In the 1970s, the World Health Organization attempted to determine the relation between hormone markers and ovulation, and reported the surge in LH to be the best marker of impending ovulation, with the rise in estradiol signaling the beginning of the fertile period (WHO 1980a,b). Subsequently, E_3G was reported to be a predictive urinary estrogen metabolite for the start of the fertile window (Adlercreutz et al. 1982; Branch et al. 1982). The accuracy of using various hormonal profiles indicative of ovulation was evaluated by Li and colleagues (2002b) using daily ultrasonography to identify follicular collapse. All methods for estimating ovulation were observed to have error; however, the distributions of errors differed by method.

The term *fertile window* implies that there is a specific interval during the menstrual cycle when fertilization may occur. A menstrual cycle at risk for pregnancy requires vaginal-penile intercourse during this window, or insemination or in vitro

fertilization if the couple is undergoing fertility-related treatment. The absence of a biomarker for ovulation, coupled with considerable variability within and across women's cycles, makes it difficult to accurately define the fertile window in terms of a one-size-fits-all approach. With the advent of rapid immunoassay test sticks for the detection of LH and human chorionic gonadotropin (hCG), women are better able to time intercourse and to identify pregnancy early in gestation, respectively. These technologies afford epidemiologists new tools for population-based research and in estimating conceptions.

Presumably, the length of the fertile window is determined by the lifespan of the gametes within the female reproductive track relative to the time of ovulation. Again, the fertile window is a couple-dependent sensitive window that is likely to vary within and across couples. Survival of spermatozoa is affected by the quality of the semen, characteristics of the woman's cervical mucus, and xenobiotic agents to which the gametes may be exposed. Preparation of the endometrium for implantation is critical and is under the control of estrogen and LH. Currently, it is estimated that the oocyte has a short period of viability after ovulation.

Although some authors have reported the fertile window to comprise 5 days, preceding and including the day of ovulation when based upon the ratio of urinary estrogen to progesterone among presumably fecund women (Wilcox et al. 1995), other investigators have reported much more variability in the length of the window, ranging from less than 1 to more than 5 days among couples seeking care, when utilizing serial vaginal ultrasonography in combination with postcoital and sperm mucus-penetration tests (Keulers et al. 2007). Fehring and colleagues (2006) estimated the fertile window among women attempting to avoid pregnancy by using a fertility monitor that was found to be accurate in identifying ovulation in relation to the gold standard of daily ultrasonography (Behre et al. 2000). The authors estimated the fertile window to range from day 3 through day 29 of the menstrual cycle. Among participating women, 66% of cycles had all days of the 6-day fertile phase within days 13–20 of the menstrual cycle. However, 34% of women had cycles with fertile windows that varied by more than 7 days; fewer than 1% of women had cycles that varied by 14 or more days. Together, these studies suggest that fecund women may have less variation than do women with impaired fecundity who seek medical care.

Lynch and colleagues (2006) summarized studies attempting to estimate the day-specific probabilities of conception. The fertile window ranged from 7 days before ovulation to 4 days following ovulation across seven authors. Specifically, authors relying on basal body temperature to estimate ovulation reported a range in the fertile window of from 4 to 7 days before ovulation, to 1 to 4 days following ovulation (Barrett and Marshall 1969; Schwartz et al. 1980; Royston 1982; Colombo and Masarotto 2000). Using the peak day of cervical mucus secretion as the proxy for ovulation, the fertile window was reported to range from 5 to 7 days before ovulation to 3 to 4 days following ovulation (Colombo and Masarotto 2000; Stanford et al. 2002; Stanford et al. 2003). Using the ratio of urinary pregnanediol 3-glucoronide to estrone 3-glucuronide with or without the LH surge, respectively, the fertile window was defined as 5 days before and the day of ovulation only (Wilcox et al. 1995; Dunson et al. 2001).

IMPLANTATION WINDOW

Fertilization may be defined as the fusion of a haploid sperm and egg. This process requires the sperm cell (comprising mostly DNA) to burrow through the egg's outer shell or zona pellucida, where it enters the internal cytoplasm of the egg (ooplasm) and locomotes the male DNA (half the number of chromosomes) and combines with the female half. This process is estimated to occur within 3–4 hours, at which time a complete set of 46 chromosomes is created. The one-celled embryo then begins to cleave, becoming a two-cell embryo at approximately 22–28 hours following fertilization, a four-cell the following day, and an eight-cell around day 3. Figure 2.2 illustrates these early embryonic stages that accompany the preimplantation window.

The early conceptus is known as a *morula* until it reaches the eight-cell stage, at which time it is known as a blastocyst upon forming an outer layer of cells (trophectoderm) that eventually becomes the placenta and inner cells, including the embryonic stem cells, that become the embryo. The blastocyst implants in the endometrium approximately 7 days following fertilization (Hertig 1975). The factors facilitating or impeding its occurrence remain largely unknown, in part, given the difficulty in visualizing this process.

Under the influence of ovarian steroids, the uterus undergoes structural changes in preparation for implantation, including changes in glandular epithelium and underlying stroma and vasculature (Nikas et al. 1995). Two forms of cellular differentiation are believed responsible for implantation, including one giving rise to the trophoblast and the other to the receptive endometrium (McLaren 1984). Embryonic signaling is necessary for both the process of implantation and the maternal recognition of pregnancy (Edwards 1994). Endometrial receptivity or the implantation period spans a few days following ovulation to several days

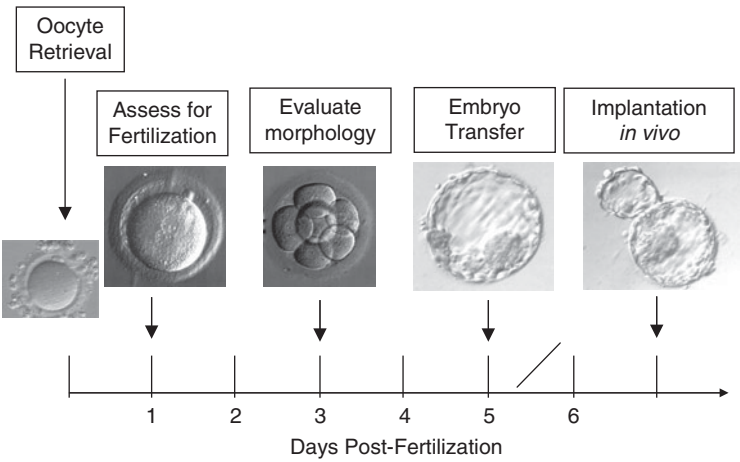


Figure 2.2 Early embryonic stages accompanying the pre-implantation window. Courtesy of Aidita James, Ph.D. of the ART Institute of Washington, Walter Reed ART Program.

before menstruation, although reasons for such variation are largely unknown. The molecular basis for the receptive state of the endometrium for implantation is poorly understood, although it has been hypothesized that critical levels of estrogen are needed for regulating the window of uterine receptivity for implantation (Ertzeid and Storeng 2001). Estrogen levels within a very narrow range are critical for transforming uterine receptivity to a refractory state, suggesting that the uterus is extremely sensitive to estrogen levels with respect to implantation. The implantation window is believed to be open longer in relation to cycles with lower estrogen levels, but open for a shorter period of time for cycles with higher levels (Ma et al. 2003). As we move forward in understanding successful human reproduction, it will be necessary to jointly consider the fertile and implantation windows, rather than considering them as separate entities. For example, assessing the day of fertilization within the fertile window relative to day of implantation may shed light on mechanisms associated with pregnancy maintenance or loss.

Successful implantation requires completion of a series of timed and interrelated events requiring dialogue between the receptive endometrium and the intrusive blastocyst (Tabibzadeh et al. 1998). To date, a host of factors are being considered to understand how the endometrium becomes receptive for implantation (e.g., cytokines, integrins, heat shock proteins, tumor necrosis factor). Less attention has been paid to the implantation window in comparison to the fertile window, from both a biostatistical or epidemiologic perspective, which is surprising given the high rate of pregnancy loss during this interval in both unassisted and assisted reproduction. In part, this may reflect the degree of invasiveness if endometrial biopsies are required and other methodologic or logistical challenges in obtaining repeated biospecimens or sonographic measurements. Little is known about the mechanisms underlying the signaling between the morula/blastocyst and endometrium, in part given our inability to visualize or to obtain biospecimens that coincide with this interval, although increasing recognition of its importance has spawned considerable molecular research.

In one of the few prospective cohort studies aimed at identifying the implantation window relative to ovulation, the initial rise in urinary hCG hormone occurred on days 6–12 following ovulation for women whose pregnancy resulted in a live birth, in contrast to an initial rise on days 7–11 days for women experiencing pregnancy losses (Wilcox et al. 1999). The estimated day of implantation for the two groups of women was 9.1 and 9.2 days following ovulation, respectively. Another study comprising 21 fertile women from whom endometrial specimens were obtained in relation to the LH surge reported a 6-day implantation window (Sarani et al. 1999). Suffice to say, concerted epidemiologic investigation delineating both the fertile and implantation windows, particularly in relation to each other, is critically needed.

IN UTERO ENVIRONMENT

The in utero environment refers to the site of human development within the uterine cavity; it commences with implantation and ends with delivery.

Following implantation of the blastocyst, the three embryonic germ layers—endoderm, mesoderm, and ectoderm—begin development. The endoderm is the innermost layer of the embryo from which the linings of the cavities and coverings of the internal organs are formed. The mesoderm layer eventually gives way to the bone, connective tissue, muscle, blood, vascular and lymphatic tissue, and the pleurae of the pericardium and peritoneum, whereas the ectoderm layer gives rise to the nervous system, sensory organs, epidermis and epidermal tissue, and mucous membranes of the mouth and anus. Organ formation is highly timed and interrelated and includes cellular proliferation, differentiation, migration, and apoptosis. The many interactions between cells and these processes underscore the susceptibility of the embryo or fetus to intrauterine exposures. Embryology has long recognized the role of critical windows for organogenesis, as well as the ensuing structural malformations resulting from teratogens during this window (as discussed in Chapter 10).

From a life course epidemiologic perspective, pregnancy is a sensitive window for children and adults, particularly if one accepts the early-origins-of-disease hypothesis, including imprinting-related effects (Owen and Segars 2009). Although reproductive and perinatal epidemiology has long recognized the importance of exposure timing during gestation and risk of birth defects, we still need concerted work to identify the critical and sensitive windows for nonstructural defects, such as functional changes in organ systems, behavior, or a host of other subtle and long-term outcomes. A more complete description of pregnancy—both its normal and abnormal states—is presented in Chapters 4 and 5.

In summary, considerable heterogeneity in male and female fecundity and fertility exists and understanding its origin may benefit from more careful attention to the sensitive windows underlying reproduction and development, as well as in designing methodologies for capturing time-varying exposures and occult outcomes. Such etiologic research is critical for developing prediction models that can be used for clinical populations or the public at large.

REFERENCES

Adlercreutz, H., Brown, J., Collins, W., et al. 1982. The measurement of urinary steroid glucuronides as indices of the fertile period on women. *J Steroid Biochem* 17: 695–702.

Bachand, A.M., Cragin, L.A., and Reif, J.S. 2009. Reliability of retrospectively assessed categorical menstrual cycle length data. *Ann Epidemiol* 19:501–503.

Baird, D.D., Weinberg, C.R., Wilcox, A.J., McConnaughey, D.R., Musey, P.I., and Collins DC. 1991. Hormonal profiles of natural conception cycles ending in early, unrecognized pregnancy loss. *J Clin Endocrinol Metab* 72:793–800.

Barrett, J.C., and Marshall, J. 1969. The risk of conception on different days of the menstrual cycle. *Population Studies* 23:455–461.

Behre, H.M., Kuhlage, J., Gabner, C., Sonntag, B., Schem, C., Schneider, H.P.G., and Nieschlag, E. 2000. Prediction of ovulation by urinary hormone measurements with the home use ClearPlan ® Fertility Monitor: Comparison with transvaginal ultrasound scans and serum hormone measurements. *Hum Reprod* 15:2478–2482.

Ben-Shlomo, Y. and Kuh, D. 2002. A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol* 31(2):285–293.

Branch, C.M., Collins, P.O., and Collins, W.P. 1982. Ovulation prediction: Changes in the concentrations of urinary estrone-3-glucuronide, estradiol-17 beta-glucuronide and estriol-16 alpha-glucuronide during conceptional cycles. *J Steroid Biochem* 16: 345–347.

Calabrese, E.J. 1986. Sex differences in susceptibility to toxic industrial chemicals. *Br J Ind Med* 43(9):577–579.

Chen, C., Wang, X., Wang, L., Yang, F., Tang, G., Xing, H., et al. 2005. Effect of environmental tobacco smoke on levels of urinary hormone markers. *Environ Health Perspect* 113(4):412–417.

Colombo, B. and Masarotto, G. 2000. Daily fecundability: First results from a new data base. *Demographic Res* 35:39 Online at www.demographic-research.org/volumes/vol3/5

DeBaun, M.R., Niemitz, E.L., and Feinberg, A.P. 2003. Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19. *Am J Hum Genet* 72:156–160.

DeVet, A., Laven, J.S.E., De Jong, F.H., Themmen, A.P.N., and Fauser, B.C.J.M. 2002. Anti-Mullerian hormone serum levels: A putative marker for ovarian ageing. *Fertil Steril* 77:357–362.

Dunson, D.B., Weinberg, C.R., Baird, D.D., Kesner, J.S., and Wilcox, A.J. 2001. Assessing human fertility using several markers of ovulation. *Stat Medicine* 20:965–978.

Edwards, R.G. 1994. Implantation, interception and contraception. *Hum Reprod* 9 (Suppl 2): 73–87.

Eppig, J.J., Vivieros, M.M., Marin-Bivens, C., and De La Fuente, R. 2004. Regulation of mammalian oocyte maturation. In: *The ovary*, 2nd ed., pp. 113–129. P.C.K. Leung and E.Y. Adashi, eds. Amsterdam: Elsevier Academic Press.

Ertzeid, G. and Storeng, R. 2001. The impact of ovarian stimulation on implantation and fetal development in mice. *Hum Reprod* 16:221–225.

Fanchin, R., Schaunaver, L.M., Righini, C., Frydman, N., Frydman, R., and Taieb, J. 2003. Serum anti-Mullerian hormone dynamics during controlled ovarian hyperstimulation. *Hum Reprod* 18:328–332.

Fehring, R.J., Schneider, M., and Raviele, K. 2006. Variability in the phases of the menstrual cycle. *JOGNN* 35:376–384.

Fergusonn-Smith, A.C. and Surani, M.A. 2001. Imprinting and the epigenetic asymmetry between parental genomes. *Science* 293:1086–1089.

Gicquel, C., Gaston, V., Mandelbaum, J., Siffroi, J.P., Flahault, A., and Le Bouc, Y. 2003. In vitro fertilization may increase the risk of Beckwith-Wiedemann syndrome related to the abnormal imprinting of the KCNQ1OT gene. *Am J Hum Genet* 72:1338–1341.

Gonzalez-Bulnes, A. and Veiga-Lopez, A. 2007. Revisiting concepts on ovarian folliculogenesis and function of the reproductive axis. In: *Novel concepts in ovarian endocrinology*, pp. 1–26 A. Gonzalez-Bulnes, ed. Kerala, India: Transworld Research Network.

Hansen, K.R., Morris, J.L., Thyer, A.C., and Soules, M.R. 2003. Reproductive aging and variability in the ovarian antral follicle count: Application in the clinical setting. *Fertil Steril* 80:577–583.

Hertig, A.T. 1975. Implantation of the human ovum: The histogenesis of some aspects of spontaneous abortion. In: *Progress in infertility*, 2nd ed., pp. 411–438. S.J. Behrman and R.W. Kistner, eds. Boston: Little, Brown.

Jukic, A.M.Z., Weinberg, C.R., Wilcox, A.J., McConnaughey, D.R., Hornsby, P., and Baird, D.D. 2007. Accuracy of reporting of menstrual cycle length. *Am J Epidemiol* 167:25–33.

Keulers, M.J., Hamilton, C.J.C.M., Franx, A., Evers, J.L.H., and Bots, R.S.G.M. 2007. The length of the fertile window is associated with the chance of spontaneously conceiving an ongoing pregnancy in subfertile couples. *Hum Reprod* 22:1652–1656.

Krawetz, S.A. 2005. Paternal contribution: New insights and future challenges. *Nat Rev Genetics* 6:633–642.

Latham, K.E. 1999. Epigenetic modification and imprinting of the mammalian genome during development. *Curr Top Dev Biol* 43:1–49.

Lambalk, C.B., DeKoning, C.H., and Braat, D.D.M. 1998. The endocrinology of dizygotic twinning in the human. *Mol Cell Endocrinol* 145:91–102.

Li, E. 2002. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* 3:662–673.

Li, H., Chen, J., Overstreet, J.W., Nakajima, S.T., and Lasley, B.L. 2002. Urinary follicle-stimulating hormone peak as a biomarker for estimating the day of ovulation. *Fertil Steril* 77:961–966.

Lucifero, D., Chaillet, J.R., and Trasler, J.M. 2004. Potential significance of genomic imprinting defects for reproduction and assisted reproductive technology. *Hum Reprod Update* 10:3–18.

Lynch, C.D., Jackson, L.W., and Buck Louis, G.M. 2006. Estimation of the day-specific probabilities of conception: Current state of the knowledge and the relevance for epidemiological research. *Paediatr Perinatal Epi* (Suppl.1); 20:3–12.

Ma, W.G., Song, H., Das, S.K., Paria, B.C., and Dey, S.K. 2003. Estrogen is a critical determinant that specifies the duration of the window of uterine receptivity for implantation. *PNAS* 100:2963–2968.

McLaren, A. The control of implantation. In: *In vitro fertilization and donor insemination. Proceedings of the Twelfth Study Group of the Royal College of Obstetricians and Gynaecology*, pp. 13–22. W. Thompson, D.N. Joyce, and J.R. Newton, eds. London: Royal College of Obstetricians and Gynaecology, 1984.

Mikolajczyk, R., Louis, G.B., Cooney, M., Lynch, C., and Sundaram, R. 2010. Heterogeneity of vaginal bleeding and its implications for measuring time in time-to-pregnancy. *Paediatr Perinatal Epidemiol* 24:24–30.

Miller, D., Briggs, D., Snowden, H., Hamlington, J., Rollinson, S., Lilford, R., and Krawetz, S.A. 1999. A complex population of RNAs exists in human ejaculate spermatozoa: Implications for understanding molecular aspects of spermiogenesis. *Gene* 237:385–392.

Moll, A.C., Imhof, S.M., Cruysberg, J.R., Schouten-van Meeteren, A.Y., Boers, M., and van Leeuwen, F.E. 2003. Incidence of retinoblastoma in children born after in-vitro fertilization. *Lancet* 361:309–310.

Morford, L.L., Henck, J.W., Breslin, W.J., and DeSesso, J.M. 2004. Hazard identification and predictability of children's health risk from animal data. *Environ Health Perspect* 112(2):266–271.

Nikas, G., Drakakis, P., Loutradis, D., Mara-Skoufari, C., Koumantakis, E., Michalas, S., and Psychoyos, A. 1995. Uterine pinopodes as markers of the "nidation window" in cycling women receiving exogenous oestradiol and progesterone. *Hum Reprod* 10: 1208–1213.

Norwitz, E.R., Schust, D.J., and Fisher, S.J. 2001. Implantation and the survival of early pregnancy. *N Engl J Med* 345:1400–1408.

Owen, C.M. and Segars, J.H. 2009. Imprinting disorders and assisted reproductive technology. *Sem Reprod Med* 27:417–428.

Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P., and Krawetz, S.A. 2002. Spermatozoal RNA profiles of normal fertile men. *Lancet* 360:772–777.

Ostermeier, G.C., Goodrich, R.J., Moldenhauer, J.S., Diamond, M.P., and Krawetz, S.A. 2005. A suite of novel human spermatozoal RNAs. *J Androl* 26:70–74.

Parrington, J., Lai, F.A., and Swann, K. 2000. The soluble mammalian sperm factor protein that triggers Ca^{2+} oscillations in eggs: Evidence for expression of mRNA(s) coding for sperm factor protein(s) in spermatogenic cells. *Biol Cell* 92:267–275.

Peters, H. 1970. Migration of gonocytes into the mammalian gonad and their differentiation. *Philos Trans R Soc Lond B Biol Sci* 259:91–101.

Royston, J.P. 1982. Basal body temperature, ovulation and the risk of conception, with special reference to the lifetimes of sperm and egg. *Biometrics* 38:397–406.

Sarani, S.A., Ghaffari-Novin, M., Warren, M.A., Dockery, P., and Cooke, I.D. 1999. Morphological evidence for the "implantation window" in human luminal endometrium. *Hum Reprod* 14:3101–3106.

Schwartz, D., MacDonald, P.D.M., and Heuchel, V. 1980. Fecundability, coital frequency, and the viability of ova. *Population Studies* 34:397–400.

Scott, R.T., Jr., Hofmann, G.E., Oehninger, S., and Muasher, S.J. 1990. Intercycle variability of day 3 follicle-stimulating hormone levels and its effect on stimulation quality in vitro fertilization. *Fertil Steril* 54:297–302.

Small, C.M., Manatunga, A.K., and Marcus, M. 2007. Validity of self-reported menstrual cycle length. *AEP* 17:163–170.

Stanford, J.B., White, G.L., and Hatasaka, H. 2002. Timing intercourse to achieve pregnancy: Current evidence. *Obstet Gynecol* 100(6):1333–1341.

Stanford, J.B., Smith, K.R., and Dunsonh D.B. 2003. Vulvar mucus observations and the probability of pregnancy. *Obstet Gynecol* 101:1285–1293.

Stewart, D.R., Overstreet, J.W., Nakajima, S.T., and Lasley, B.L. 1993. Enhanced ovarian steroid secretion before implantation in early human pregnancy. *J Clin Endocrinol Metab* 76:1470–1476.

Stevens, V.C. 1997. Some reproductive studies in the baboon. *Hum Reprod Update* 3: 533–540.

Tabibzadeh, S. 1998. Molecular control of the implantation window. *Hum Reprod* 4: 465–471.

Treloar, A.E., Boynton, R., Behn, B., and Brown, B. 1967. Variation of the human menstrual cycle through reproductive life. *Inter J Fertility* 12:77–126.

Vermeulen, A. 1993. Environment, human reproduction, menopause, and andropause. *Environ Health Perpsect* 101 (Suppl 2):91–100.

Vollman, R. 1977. *The menstrual cycle*. Philadelphia: W.B. Saunders.

Wallace, W.H. and Kelsey, T.W. 2004. Ovarian reserve and reproductive age may be determined from measurement of ovarian volume by transvaginal sonography. *Hum Reprod* 19:1612–1617.

WHO Task Force on Methods for Determination of the Fertile Period. 1980a. Temporal relationships between ovulation and defined changes in the concentration of plasma estradiol-17β, luteinising hormone, follicle stimulating hormone and progesterone. I. Probit analysis. *Am J Obstet Gynaecol* 138:383–390.

WHO Task Force on Methods for Determination of the Fertile Period. 1980b. Temporal relationships between ovulation and defined changes in the concentration of plasma estradiol-17β, luteinising hormone, follicle stimulating hormone and progesterone. II. Histologic dating. *Am J Obstet Gynaecol* 139:886–895.

Wilcox, A.J., Weinberg, C.R., and Baird, D.D. 1995. Timing of sexual intercourse in relation to ovulation: Effects on the probability of conception, survival of the pregnancy, and sex of the baby. *N Engl J Med* 333:1517–1521.

Wilcox, A.J., Baird, D.D., and Weinberg, C.R. 1999. Time of implantation of the conceptus and loss of pregnancy. *N Engl J Med* 340:1796–1799.

Zeleznik, A. 2004. Dynamics of primate follicular growth: A physiological perspective. In: *The ovary*, 2nd ed., pp. 45–53. Amsterdam: Elsevier Academic Press.

3

Fecundity and Fertility

GERMAINE M. BUCK LOUIS

OVERVIEW OF HUMAN FECUNDITY AND FERTILITY

Understanding human fecundity and fertility requires recognition of the many conceptual and methodologic challenges underlying this avenue of study. These challenges include working toward universal definitions for fecundity and fertility, facilitating the synthesis of research findings for populations, moving toward a couple-dependent etiological or behavioral focus, measuring longitudinal time-varying biological and behavioral factors, and conceptualizing intermediates (e.g., ovulation) in the causal pathway to conception or birth, as discussed below. Despite these challenges, reproductive epidemiology is an exciting avenue of research and one with promise for answering the most basic questions regarding successful human reproduction and development.

Fecundity is defined as the biologic capacity for reproduction, irrespective of pregnancy intentions; fertility is demonstrated fecundity, measured by live births or sometimes stillbirths, as discussed in Chapter 1. These definitions are recognized by reproductive epidemiologists and demographers, but they are not universally accepted by all scientific disciplines or geographic regions. For example, the French translation of *fertilité* is fecundity, whereas *fécondité* is fertility. The demographer Corrado Gini (1924, 1926) is credited with being the first to define fecundity in response to concerns about declining fertility in Europe, which may be an indicator of declining fecundity. Gini defined fecundity as the likelihood of conceiving during a normal menstrual cycle with sexual relations and no contraception. This definition underscores that fecundity is a necessary but insufficient cause of fertility, given the important role of intercourse behavior in achieving pregnancy. In contemporary epidemiologic practice, time-to-pregnancy (TTP) is used as a proxy for estimating fecundability.

Clinical disciplines recognize fertility rather than fecundity and utilize a classification scheme—fertile, subfertile, or infertile—reflecting the length of time women/couples report trying to become pregnant (i.e., <6 months, 6–12 months, and >12 months, respectively). The definitions for subfertility or infertility are sometimes extended to 12 and 24 months, respectively, to accommodate couples who spontaneously conceived within 12–24 months at risk for pregnancy. Irrespective of the definitions used, categorization of couples by length of time

attempting pregnancy assumes that time is accurately known or recalled. The uncertainty about the variability of the fertile window, often noted to include the 5 days before and day of ovulation, within and across women (Lynch et al. 2006) can result in measurement error if couples unknowingly are not having frequent intercourse during the fertile window. In addition, only two studies have empirically evaluated the validity of self-reported TTP using prospective longitudinal methods as the gold standard. Good validity is reported when recall is limited to 3–20 months (Zielhuis et al. 1992); however, poor validity was observed when recall was approximately one decade later (Cooney et al. 2009). It is important to keep in mind that simply classifying couples as fertile, subfertile, or infertile ignores individuals not at risk for pregnancy (i.e., sexually inactive or currently contracepting) and those in the early stages of trying. The lack of universally accepted terminology impairs the synthesis of research over disciplines and geographic regions (Buck et al. 1997).

Historically, researchers have focused on female fecundity and fertility, most likely given the rather longstanding interest in pregnancy relative to the spectrum of fecundity outcomes available for study. Of late, growing interest in male fecundity and fertility has arisen in relation to the evolving body of evidence suggesting that male fecundity may be on the decline (Skakkebaek et al. 2001; Jorgensen et al. 2002). The rate of purported decline in male fecundity is suggestive of a role for environmental factors including in utero exposures (Jensen et al. 2004), possibly in the context of familial susceptibility. Other reasons for the renewed interest in male fecundity stems from recognition of the couple-dependent nature of many fecundity and all fertility outcomes, such as conception, pregnancy, or birth. A couple-based approach is in keeping with growing evidence supporting male-mediated effects in successful human reproduction that go beyond the spermatozoon simply transmitting the paternal haploid genome. For example, a positive association has been reported between sperm morphology and blastomere cleavage rates (Salumets et al. 2002), and the transmission of spermatogenic RNA to the oocyte at fertilization underscores a potential role in early embryonic development (Ostermeier et al. 2004). Hierarchical statistical methods allow the analysis of the complicated data structure underlying much of reproductive epidemiology.

Given the highly interrelated and time-sensitive windows underlying human reproduction, time-varying covariates are an essential aspect of study design, data collection, and analysis. Investigators interested in identifying exposures that adversely affect semen quality or ovulation will need to utilize data collection methods that support the longitudinal measurement of the study exposures during the sensitive windows for spermatogenesis and ovulation, as discussed in Chapter 2. For example, when interested in conception probabilities, it is important to capture both the timing and frequency of sexual intercourse during the fertile or critical window, along with other exposures. Olsen and colleagues (2005) define fecundity as a function of timing and frequency of sexual intercourse along with the reproductive capacity of the partners, thus underscoring the importance of biology and time-varying behavior during this sensitive window. Similarly, when studying early pregnancy loss, care must be taken to ensure that exposures are measured up

to the date of the event, and not subsequently, to avoid bias arising from reverse causality.

The absence of fecundity biomarkers suitable for population-based research has implications not only for precision in studying endpoints such as conception, implantation, and early loss, but also in specifying causal models. This is particularly an issue when attempting to specify intermediates in the causal pathway. For example, when specifying a causal model for conception, how does an investigator handle ovulation in the absence of a gold standard? Without a valid and reliable marker of conception, investigators often face the use of proxy measures for both the outcome as well as intermediate factors (see Chapter 15). The absence of a biomarker for ovulation and conception necessitates the use of proxies, except for select population subgroups such as couples undergoing assisted reproductive technologies (ART), in which ovulation can be directly measured through sonographic procedures or conception through in vitro techniques.

In sum, understanding human fecundity and fertility is essential for continuation of the species. And yet, we have limited empirical evidence, particularly at the population level, for understanding the determinants of human fecundity and fertility despite the inefficiency of human reproduction relative to other mammals (Stevens 1997; Viudes-de-Castro et al. 1997). In fact, recent discoveries challenge our thinking about some of the most basic tenets of human reproduction. For example, the longstanding premise that women are born with all their oocytes is now open to debate following a recent publication reporting that adult mouse ovaries have primitive germline cells that may give rise to new oocytes (Zou et al. 2009). Similarly, evidence from the mouse model suggests that female sexual development requires maintenance of the mammalian ovarian phenotype throughout adulthood by actively repressing the Sertoli cell-promoting gene *Sox9* (Uhlenhaut et al. 2009). As such, ovarian granulosa/theca cells may retain plasticity, allowing transdifferentiation into Sertoli/Leydig cells with ensuing testosterone production. These and other recent discoveries shall be important in planning future epidemiologic research aimed at understanding the underlying mechanisms responsible for human reproduction and development.

CONCEPTUAL PARADIGM FOR HUMAN FECUNDITY AND FERTILITY

Although the biologic basis for human fecundity arises in utero, the referent population for much of the available research is typically restricted to women of reproductive age or those aged 15–44 years. Crudely, this interval was meant to capture adolescents with established menstruation through women approaching menopause. However, the relevant interval is reported to be changing over time, given that secular patterns suggest that the reproductive lifespan may have increased from 36.1 to 37.7 years for women born between 1915–1919 and 1935–1939, respectively (Nichols et al. 2006). Age at menarche is reported to be declining in many developed countries (Herman-Giddens et al. 1997, 2001), accompanied by a faster rate of growth (Karpati et al. 2002). Combined, these two

physical changes have been labeled the secular trend in growth that may be indicative of better childcare (Samaras and Storms 2002). The natural age upon reaching menopause is approximately normally distributed, with most values between the ages of 40 and 58 years, and a mean of 51.4 years (North American Menopause Society 2002); however, a combination of genetic, reproductive, and lifestyle factors are reported to affect timing (Henderson et al. 2008). Although menopausal women are no longer fecund, male fecundity is reported to have no upper bound. Researchers who assess human fecundity from a couple-based approach have observed declines in conception probabilities associated with parental ages (Dunson et al. 2002). As we begin to understand the in utero origin of human fecundity and its impact on health status across the lifespan, as discussed below, the conceptual paradigm can be adapted to include ages outside the conventional range.

At the population level, it is challenging to conceptualize individuals or couples by fecundity status from either a cross-sectional or longitudinal perspective, with the latter approach including couples with completed reproductive performance. Figure 3.1 offers a conceptual paradigm for study that categorizes a population into three fecundity categories: fecund, impaired fecundity, and infecund. Various subcategories illustrate the considerable heterogeneity within each group. For example, fecund individuals may or may not be sexually active, which would affect their degree of risk for pregnancy. Sexual debut or the first engagement in sexual intercourse is an important aspect of fecundity, and must be

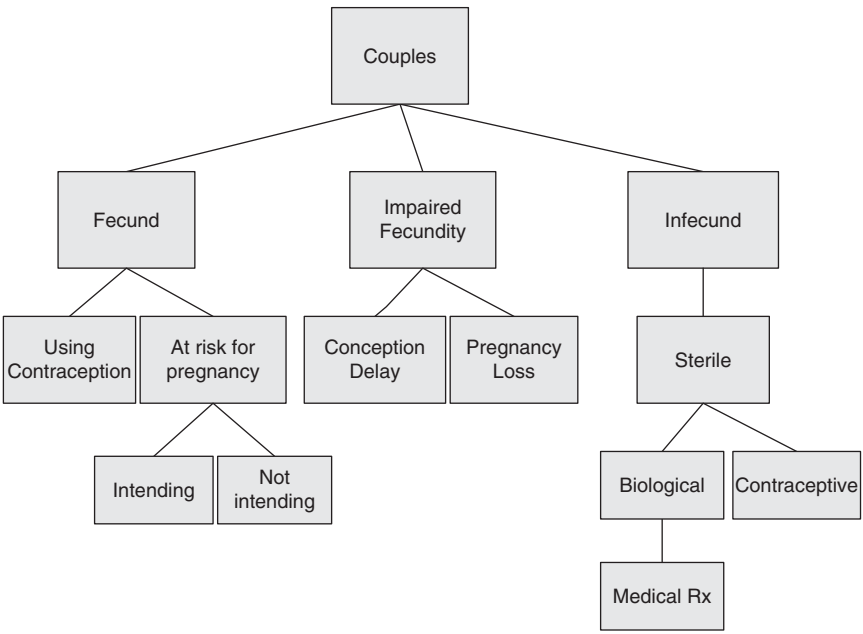


Figure 3.1 Conceptual paradigm for population fecundity.

considered when defining the lower bound for age of the study population. Approximately 70% of U.S. high school students reported being sexually active by age 18 years (Cavazos-Rehg et al. 2009). Sexual debut varied, however, by race/ethnicity. For example, 82% of African-American males reported being sexually active by age 17 years, compared with 74% of African-American females, 69% for Hispanic males, 59% of Hispanic females, 58% of Caucasian females, 53% of Caucasian males, 33% of Asian males, and 35% of Asian females.

The impaired fecundity category noted on Figure 3.1 includes individuals experiencing conception delays, typically defined as requiring more than 6 months for conception, and pregnancy loss irrespective of the duration of pregnancy at the time of recognized loss. Infecund individuals are incapable of reproduction irrespective of pregnancy intentions, either as a result of biologic (e.g., genetic syndromes) or iatrogenic (e.g., voluntary sterilization) causes. The heterogeneity within the fecundity subcategories is considerable as well. For example, some couples experiencing conception delays may truly have diminished fecundity, whereas others may have mistimed sexual intercourse with regard to the fertile window. Similarly, infecund couples may be medically assisted to become fecund, or the reverse may occur, in which formerly fecund couples undergo sterilization procedures that render them infecund. It is imperative to keep in mind the dynamic nature of human fecundity and its variability over individuals' or couples' reproductive lives. Another important consideration not demonstrated on the figure is the possible change in couple fecundity following a change in partner.

Figure 3.1 also illustrates the extreme difficulty in identifying couples either at risk for pregnancy or planning a pregnancy in the near future, which comprise the sampling framework for TTP studies. Although the exact percentage of couples planning a pregnancy within the next few months at any point in time is unknown at the population level, it is estimated to be approximately 1% in U.S. women (Buck et al. 2004). Elsewhere, 1.2% (95% confidence interval [CI] = 0.6%–1.9%) of French women aged 18–44 years reported planning attempts within 6 months (Slama et al. 2006), and 1.6% of Danish women (Bonde et al. 1998) were reported to be at risk for pregnancy at the population level. This underscores the limited external validity that may arise from TTP studies that rely exclusively upon volunteers.

MEASURING HUMAN FECUNDITY

Human reproduction is highly inefficient in comparison to other mammals, with a monthly fecundity rate or probably of pregnancy approximately 20% (Leridon and Spira 1984) along with a postimplantation pregnancy loss rate of approximately 30% (Wilcox et al. 1988). Moreover, daily sperm production in human males is approximately 4.4 million/gram of testis compared to three to four times the amount produced by bulls and stallions (Sharpe 1994). Fecundity is not a dichotomy, but rather a continuum (e.g., some couples conceive quickly, whereas others conceive after long trying periods with or without medical intervention). Tietze and colleagues (1950) were among the first to publish evidence supporting

fecundity as a continuum, and reported it is highest in the first few months that couples are attempting pregnancy, relative to later months. This pattern has been consistently reported across study populations and study designs, thus supporting the concept that the decline in conception probabilities is due to the most fecund couples being removed from the cohort along with the gradual selection of the least fertile couples (Leridon 2007).

Human fecundity can be assessed as a single endpoint or in combination with other endpoints, as discussed in Chapter 1. Given the interrelatedness of endpoints, investigators often design research capable of measuring a spectrum of endpoints either at baseline (e.g., pubertal status, gynecologic health, and hormonal profile) followed by longitudinal data collection to ascertain daily information on menstrual cycles, ovulation, and conception for couples at risk for pregnancy. Male fecundity studies may employ a similar strategy, collecting baseline information on urologic health status and hormonal profile while longitudinally collecting semen samples for evaluation. The interchangeability of fecundity and fertility endpoints for men and women reflects a relatively similar, yet distinct, developmental biologic origin largely under the influence of the hypothalamus-pituitary-gonadal axis, which regulates the cascade of processes underlying folliculogenesis, spermatogenesis, ovulation, and conception (see Chapter 2). For example, fecundity can be assessed in men (e.g., semen quality), women (e.g., ovulatory cycles), or couples (e.g., conception), as can fertility, as measured by births to both partners of the couple.

Hormone profiles, puberty, and sexual function

Although beyond the scope of this chapter, a number of clinical approaches are available for assessing hormonal profiles indicative of fecundity or possible impairments. A number of methods are available for assessing pubertal status, ranging from self-rating tools (e.g., Tanner Scales), hormonal profiles (e.g., gonadotrophins, such as luteinizing or follicle-stimulating hormone; or sex steroid hormones, such as testosterone), physical biomarkers (e.g., skeletal growth, body composition), and molecular and cellular biomarkers (e.g., bone growth and mineralization, Müllerian-inhibiting substance, spermaturia) (Rockett et al. 2004). Despite the various approaches, it remains challenging to rate pubertal onset, especially in the early stages.

Sexual function is another important endpoint receiving increased public health and clinical attention. Sexual dysfunction comprises disturbances in sexual desire and/or sexual response. Although age is a widely cited determinant of female sexual dysfunction or erectile dysfunction, its study in the context of other behavioral or psychosocial factors may provide a more complete understanding. For example, results from the Study of Women's Health Across the Nation (SWAN) Study suggest that sexual function and practices remain unchanged for pre- and perimenopausal women (Cain et al. 2003). Other researchers have reported that emotional well-being and relationship quality are more influential on sexuality than aging (Bancroft et al. 2003). The infamous Kinsey report was among the first data in the U.S. to discuss sexual dysfunction (Kinsey et al. 1953)

until completion of the U.S. National Health and Social Life Survey. In this later work, sexual dysfunction was observed to be more prevalent for women aged 18–59 years than in similarly aged men (i.e., 43% and 31%, respectively) (Laumann et al. 1999). Also, age was strongly predictive of sexual dysfunction, being associated with younger-aged women but older-aged men, particularly for erectile dysfunction. Marital status and emotional and stress problems were associated with dysfunction, thus underscoring the importance of assessing age in a broader context. The prevalence of erectile dysfunction among men aged 40–70 years in the Massachusetts Male Aging Study was reported to be 35% and associated with age, health status, and emotional function (Feldman et al. 1994).

Semen quality

Male fecundity is typically assessed through semen analysis, which can be supplemented with a physical examination and/or hormonal evaluation. Fecund men produce good-quality semen, which in turn is conditional upon an endocrine system capable of supporting spermatogenesis and spermatozoa production, an intact ductal system with normal epididymal function and accessory glands function, and an intact nervous system for ejaculation. The noninvasive aspect of semen collection permits implementation at the population level, as is currently being done for military conscripts or in a country-level semen-quality surveillance program (Jensen et al. 2002).

From a research perspective, semen can be conceptualized to contain three components: sperm, seminal plasma, and xenobiotic agents or exogenous substances. Seminal plasma contains fluids from the accessory glands that add nutrients to sperm, and these glands include the prostate, seminal vesicles, and Cowper's gland. Semen can be analyzed with regard to a host of measurements depending upon measurement tool, but typically include semen volume, sperm number, and concentration, motility, and morphology. Computer-assisted sperm analysis (CASA) is one method that provides a computerized measurement of sperm motility, function, and structure. The recent publication of the fifth edition of the WHO Laboratory Manual for the Examination of Processing of Human Semen (2010) provides reference values for assessing human semen based upon fertile men whose partner became pregnant within 12 months. Reference values are partially dependent upon the study population and underlying characteristics, method utilized for semen analysis, and the process for calculating reference intervals (Boyd 2010). Examples of the recently released WHO lower (5th centiles) reference limits for semen quality include volume 1.5 mL (95% CI 1.4–1.7), sperm concentration 15×10^6 per ejaculate (95% CI 12–16), total motility 40 (95% CI 38–42), and morphology of 4% normal forms (95% CI 3–4). It is important to note, however, that men with values below the lower reference limits are not necessarily infecund nor will they necessarily be clinically diagnosed with infertility.

At the population level, sperm concentration below 40×10^6/mL was associated with a reduction in the probability of pregnancy per a given menstrual cycle (Bonde et al. 1998). To date, however, no single measure of semen quality is

predictive of fertility (Knuth et al. 1988), thus prompting investigators to search for new markers, such as DNA fragmentation, which has been linked to lower in vivo fertility, poor ART outcomes, recurrent pregnancy loss, and childhood morbidity (Bungum et al. 2007; Carrell et al. 2003; Lewis et al. 2005).

Semen quality varies, underscoring the importance of collection and laboratory techniques in its quantification. In a prospective study of 20 healthy men, sperm concentration had the highest within- and between-subject variation, whereas vitality had the lowest (Alvarez et al. 2003). Irrespective of the method used for quantifying semen and its constituents, a number of factors are reported to affect semen quality. These include period of abstinence (ideally 2 days), frequency of ejaculation and sexual intercourse, type of collection method (e.g., coitus interruptus or silastic collection devices), and medical history such as infections, illnesses, or injuries that may affect scrotal temperature or spermatogenesis. Given the many purported factors affecting semen quality, including intraindividual variation, some clinicians and researchers require two samples. However, recent papers have suggested that one semen sample may be sufficient for epidemiologic purposes (rather than diagnostic purposes) in assessing male fecundity, assuming attention to relevant covariates purported to affect semen quality (Rylander et al. 2009; Stokes-Riner et al. 2007). Historically, multiple semen samples per male have been simply averaged, rather than using hierarchical modeling to address the inherent correlatedness of multiple samples from the same man. As such, extreme caution is needed in interpreting data from studies that did not incorporate hierarchical approaches.

Age is a consideration in assessing male fecundity, although reportedly not to the extent observed for female fecundity. Androgen production and spermatogenesis continues over the male's lifespan, although not necessarily at the same capacity. Testes, accessory glands, and the epididymis age, along with an increase in germ cell degeneration and a decrease in spermatogenesis, Leydig cell function, and blood testosterone (Feldman et al. 1994; Pasqualotto et al. 2008). Secular and regional variations in semen quality also have been reported (Auger et al. 1995; Fédération CECOS et al. 1997), underscoring the importance of considering environmental factors (Jensen et al. 1995).

Last, when assessing potential reproductive and/or developmental toxicants in relation to semen quality, however defined, it is important to keep in mind that a semen sample represents exposures initiated approximately 3 months earlier. This interval includes approximately 22 days for sperm production followed by about 25 days for maturation and about 12–26 days of passage through the epididymis. As such, semen collection needs to be timed to reflect the research question and, specifically, the type (acute/chronic) and timing of environmental exposures.

Time-to-pregnancy

Ovulation and conception comprise the biologic critical window for female fecundity, with extension to implantation and birth in its broadest sense. Ovulation may be viewed as a measure of female fecundity despite the absence of a biomarker appropriate for research at the population level. In contrast, ovulation can be

directly visualized or confirmed among women receiving infertility treatment. Fortunately, a number of relatively good proxy measures of ovulation have been demonstrated to be appropriate at the population level. This includes TTP, when conceptualized as a functional measure of a couple's fecundity, irrespective of whether couples are the unit of analysis.

Time-to-pregnancy is defined as the number of calendar months or menstrual cycles required to become pregnant from the initiation of "trying," whether determined by the detection of human chorionic gonadotrophin (hCG) hormone, sonography, or other clinical means. Time-to-pregnancy is not synonymous with time to conception, given the relatively large and uncertain percentage of losses that occur between fertilization and implantation, even among couples undergoing assisted reproductive technologies. Women who become pregnant quickly are often assumed to be more fecund than women requiring longer periods of time. This assumption may be overly simplistic, since it ignores the role of male fecundity in establishing a couple's TTP, the potential for mistimed intercourse, and errors in recall if retrospectively ascertained.

Gini (1924) first defined TTP as the probability of conception for women exposed to unprotected sexual intercourse in the absence of lactational anovulation, pregnancy, or sterility. He later commented that fecundity is important but insufficient for fully describing fecundity (Gini 1926). Fecundity and its correlate fecundability has long been the focus of demography, and has made its eventual way into epidemiology. One of the first uses of TTP was in the U.S. Collaborative Perinatal Project that enrolled 48,197 women from 12 centers who contributed 54,390 pregnancies for observation (Niswander 1972; Buck et al. 2006). Women were queried about each pregnancy they contributed to the study and specifically asked if it were unplanned or planned and, if planned, the number of months required to become pregnant. Baird and colleagues (1986) later demonstrated that TTP was a sensitive endpoint for assessing potential reproductive toxicants by further operationalizing it in terms of two questions: Did you become pregnant during your first cycle of trying? If not, how many cycles did it take you?

The TTP distribution is typically characterized as having a reverse J-shaped appearance, with the highest probability being in the first cycles and declining thereafter, with an upward rise at approximately month/cycle 12. This upward rise often reflects censoring (see Chapter 16) at this cutpoint. Figure 3.2 illustrates the TTP distribution by type of measurement, including prospective longitudinal measurement (the gold standard) and retrospectively reported TTP for the same cohort of women obtained approximately a decade later (Buck Louis et al. 2009; Cooney et al. 2009). Figure 3.3 reflects the conditional TTP distribution for the same cohort by the two methods of ascertainment. The two distributions vary, with more prospectively measured pregnancies occurring after 6 months of trying than obtained by retrospective report. At six and 12 cycles, approximately 76% and 94% of women were pregnant based upon prospective measurement compared with 85% and 95% based upon retrospective report. The cumulative six- and 12-cycle pregnancy rates for the few other prospective cohort studies that followed women up through 12 cycles ranged from 81% to 90% (Gnoth et al. 2003) and 92% to 95% (Wang et al. 2003).

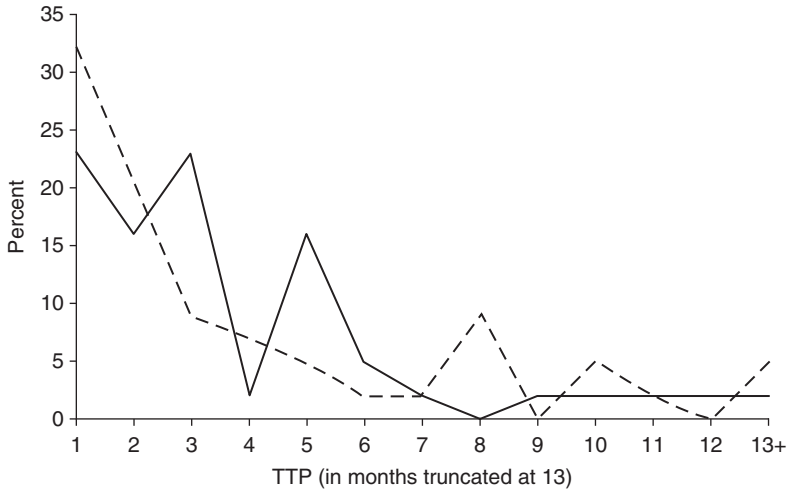


Figure 3.2 Time-to-pregnancy distributions by prospective (gold standard in dashed line) and retrospective (solid line).

The peak probability of pregnancy often reported for cycle 1 most likely is a function of left censoring, in that couples may have been trying prior to enrolling in a study at cycle 1. This peak probability also represents two groups of couples: (1) those who become pregnant without a full cycle under observation in a prospective cohort study (pregnant within 1–2 weeks or cycle 0) and (2) those who become pregnant in the first fully observed cycle (cycle 1). Recently completed prospective cohort studies that can separate these two potential types of couples may better estimate the probability of pregnancy in cycles 0 and 1. As such, the peak for the first fully observed cycle in a study may actually be reduced.

Regional differences in TTP have been reported and could not be explained by commonly cited fecundity determinants such as body mass index (BMI), age, or parity (Juul et al. 1999). As noted above, implicit in the definition of TTP is a woman/couple's ability to accurately recall the length of time if a prospective cohort design is not used, and identify the length of time (months or cycles) at risk for pregnancy, assuming sexual intercourse during the fertile window. For retrospectively ascertained TTP, digit preference is another consideration (Ridout and Morgan 1991), as are bi-directional reporting errors (Cooney et al. 2009). The reliability of TTP, however, has been shown to be good, including for long periods of recall (Joffe et al. 1993) or recall by the male partner (Coughlin et al. 1998; Nguyen and Baird 2005). With regard to the accuracy in specifying duration of active trying, the fertile window is still being defined, and it is likely to vary within and across women, as noted in Chapter 2. Lynch and colleagues (2006) summarized studies attempting to estimate the day-specific probabilities of conception and reported the interval ranged from 7 days before through 4 days after ovulation, depending upon the method used to identify the day of ovulation, given the absence of a biomarker for ovulation itself. The variation in the fertile window

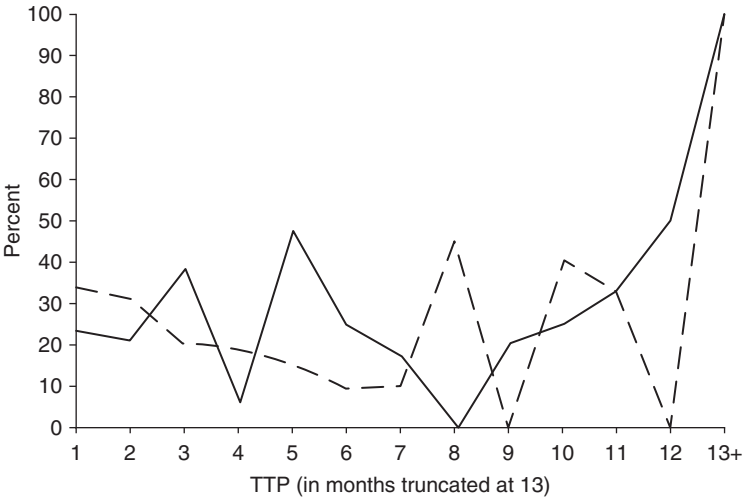


Figure 3.3 Conditional time-to-pregnancy distributions by prospective (gold standard in dashed line) and retrospective (in solid line) methods.

may impact TTP if couples unknowingly include cycles not at risk for pregnancy (intercourse outside the fertile window) as trying cycles.

Time-to-pregnancy, whether prospectively or retrospectively ascertained, is used as a measure of the couple's fecundability as measured by the probability of pregnancy per menstrual cycle. When attempting to identify reproductive and/or developmental toxicants, fecundability can be estimated for exposed women/couples relative to the unexposed. When estimating fecundity, it is important to address the time-varying frequency of intercourse and other exposures. Most of the available research has relied upon use of a discrete analog of the Cox model to estimate the fecundity odds ratio (FOR), which denotes the odds of becoming pregnant among exposed couples/women in comparison to unexposed couples/women, given all other covariates being equal. A FOR below one denotes reduced fecundability or a longer TTP, whereas an FOR above one denotes increased fecundability or a shorter TTP. More recent modeling approaches are discussed in the later section on prospective design.

Determinants of male and female fecundity

A number of genetic or medical conditions impact fecundity and are beyond the scope of this chapter, as are each of the many other (largely behavioral) factors reported in the literature. Perhaps one of the strongest biologic determinants of fecundity is age, in which an inverse relation has long been reported. Although the so-called female biological clock has captured the attention of the lay and scientific press for some time, a recent paper evaluated the evidence in support of a male biological clock. Specifically, male fecundity was observed to decline with age, as evident by age-related testicular changes, an inverse relation between aging and the genetic quality of sperm, and a longer time required for conception,

thus underscoring the importance of assessing couples (rather than individuals) when interested in fecundity or TTP (Pasqualotto et al. 2008), including the reduction in day-specific probabilities of conception in relation to male and female age (Dunson et al. 2002).

In addition to biologic determinants, much of the literature focusing on determinants of TTP has stemmed from retrospective studies that queried women about their lifestyles, viz, use of cigarettes, alcohol, and caffeinated beverages. In general, the effect size for much of the literature on lifestyle is a change in TTP by a few months or cycles. It is important to note that this effect size is within the error observed for retrospectively reported TTP, thus underscoring the importance of cautious interpretation of findings (Cooney et al. 2009). Table 3.1 summarizes some of the factors reported to affect TTP, including those that reduce fecundity as measured by a longer TTP. These include extremes in age and body mass index, various occupational and environmental exposures, diets high in trans-fatty acids or low glycemic foods, cigarette smoking, and to a lesser extent, consumption of alcoholic and caffeinated beverage (Schwartz et al. 1983; Menken et al. 1986; Augood et al. 1998; Dunson et al. 2002; Sharpe and Franks 2002; Hassan and Killick 2004; Leridon 2004; Hauser 2006; Buck Louis et al. 2006; Chavarro et al. 2007; Ramlau-Hansen et al. 2007).

Parity is observed to increase fecundity, as measured by a shorter TTP (Buck Louis et al. 2009). Similarly, mothers with naturally occurring twin pregnancies are reported to have shorter TTPs than do mothers of singletons (Basso et al. 2004; Ferrari et al. 2007). During the past few decades, a number of environmental chemicals, including those with endocrine-disrupting properties, have been associated with decreased female and male fecundity as well. These include polychlorinated biphenyl congeners or PCBs (Gesink Law et al., 2005; Buck Louis et al. 2009), 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene or DDE (Gesink Law

Table 3.1 Determinants of male and female fecundity: Examples

Determinants	*Male*	*Female*
Relatively consistent		
Age	↓	↓
Parity		↑
Extremes in body mass index	↓	↓
Occupational exposures	↓	↓
Environmental chemicals	↓	↓
Cigarette smoking	↓	↓
Suggestive		
Alcohol consumption	↓	↓
Caffeine consumption	↓	↓
Stress		↓
Anabolic steroids	↓	
Biking	↓	↓
Diet		↓

Female fecundity measured by TTP; male fecundity measured by TTP or semen quality. ↓ denotes decreased fecundity; ↑ denotes increased fecundity.

et al. 2005), dioxin (Eskenazi et al. 2010), polybrominated diphenyl ether or PBDE congeners (Harley et al. 2010), and mercury (Cole et al. 2006).

Many of the fecundity determinants noted above are associated with TTP as reported by women or semen quality, including in the few studies that considered couples' exposures. For example, a dose-dependent effect was observed between fecundity and male/female partner's smoking more than 15 daily cigarettes, male/female partner's tea or coffee consumption of more than 6 daily cups, male partner's consumption more than 20 weekly alcohol units, and female's BMI of greater than 25, with relative risks ranging between 1.4 and 1.9 (Hassan and Killick 2004). Overweight and obese men and women were more likely to be subfertile than were normal weight couples, based upon recalled BMIs and TTP at approximately 16 weeks of gestation among participants in the Danish National Birth Cohort Study (Ramlau-Hansen et al. 2007). Similarly, obesity has been associated with a reduction in successful ART procedures (Bellver et al. 2003; Fedorcsak et al. 2004). Despite the many attempts to identify fecundity determinants, little overall variation in TTP has been explained by the factors mentioned. Specifically, Axmon and colleagues (2006) reported that only 14% of the variation in TTP was explained by oral contraceptive usage prior to attempting pregnancy, menstrual cycle length, age, and parity in a random sample of women from the general Swedish population who were queried about lifestyle and TTP. None of the traditional lifestyle factors implicated with diminished fecundity was retained in models. Hopefully, future couple-based prospective cohort studies with preconception enrollment will help answer questions about male- and/or female-mediated effects on couple fecundity and fertility.

FECUNDITY STUDY DESIGNS AND METHODS

It is difficult to capture the exact date at which couples first become at risk for pregnancy or "start" trying to become pregnant, except perhaps for couples using certain contraceptives that require removal, such as an intrauterine device. This difficulty may reflect differences in couples' perceptions about whether they are at risk or in fact trying or the date of their official start date. Couples may be hesitant to start the clock for actual trying by allowing themselves to have some warm-up time, during which they are neither trying nor avoiding pregnancy, per se. The lack of uniformly accepted definitions for pregnancy planning also impacts measuring the initial planning period (Peterson et al. 1997). Operational definitions vary and have included terms such as *trying*, *planning*, *intending*, and *desiring*. The discordance in women's report of the intendedness of their pregnancies was 25% in one study (Kaufmann et al. 1997). Factors that significantly predicted discordant reports included age, marital status, household income, education, parity, and time since pregnancy.

For some time, epidemiologists interested in designing human fecundity studies had one of two study design options available for consideration. Such approaches included retrospective and prospective designs, each of which carries its traditional study strengths and limitations, and others more targeted to fecundity.

Other options have been proposed and include the current duration, historic prospective, and case-cohort designs. Thus, investigators have options to consider when designing research focusing on human fecundity. A brief description of each design, along with important methodologic considerations, is presented in Table 3.2. A more complete description of some novel designs is presented in Chapter 17.

Retrospective design

The more traditional approach for retrospective inquiry was to recruit a sample or population of pregnant women or women who had given birth for the retrospective collection of TTP. When conditioning on pregnancy or births, TTP estimates effective fecundity or that resulting in pregnancy or birth. More recently, study samples and populations have been expanded to include women/couples with unsuccessful pregnancy attempts, to minimize the under-representation of subfecund or infecund individuals in sampling frameworks conditional on pregnancy or live birth. Excluding couples not achieving pregnancy (or birth) can inflate fecundity or distort the relation between an exposure and TTP that results only in pregnancy or birth. Various sampling frameworks are available for retrospective designs including pregnancy-based samples (both current and past pregnancies); cross-sectional samples, as have frequently been used for study of occupational or environmental exposures; and population-based birth cohorts, in which members are questioned about reproductive history up to the time of the interview and then followed prospectively (Joffe 2005). One important caveat when relying upon retrospective report is the need to remove time that is not at risk for pregnancy from retrospective reports. This can be a difficult task both in terms of extending the length of the interview and in the ability of couples to identify time not at risk. Suffice to say, prompts such as time periods using contraception or periods of sexual inactivity should be used to ensure accurate estimates of TTP.

A number of possible biases have been suggested and discussed for retrospectively collected TTP. These include behavior modification, exposure time trends, planning, wantedness, pregnancy recognition, medical intervention, and the unhealthy-worker effect (Weinberg et al. 1994). Joffe and colleagues (2005) evaluated these potential biases and concluded that most could be addressed with proper study design and an analytic plan that included sensitivity analyses, such as varying the criteria to include accidental pregnancies, unsuccessful attempts, or pregnancy losses. It is important to consider these biases irrespective of study design, as was recently summarized (Slama et al. 2006).

The validity of retrospective TTP has been assessed in only two studies using a gold standard (prospectively measured TTP and pregnancy). Validity was good for short-term (3–20 months) recall of TTP among 100 participating women with a mean reduction of 0.6 months (Zielhuis et al. 1992). Long-term validity, however, was poor (17% exact agreement) when women were asked to recall TPP approximately a decade later (Cooney et al. 2009). Moreover, reporting errors were bidirectional in that women both over- and under-reported TTP. Reliability has been assessed by a number of investigators and has been shown to be high

Table 3.2 Comparison of study designs for human fecundity research: Considerations for selection

Retrospective	*Prospective*	*Current duration*	*Historic prospective*	*Case-Cohort*
• Complete or representative sampling possible with efforts to collect information on nonrespondents • Reliable TTP estimates • Recruit sample or population conditional on pregnancy, although can be adapted to include unsuccessful trying events (e.g., pregnancy-based studies, occupational cohorts) • Unable to collect preconception or timing-specific exposures • Selection and information bias are considerations • Analytic plan should include sensitivity analyses	• Intense recruitment of reproductive aged women or couples at risk for pregnancy or planning to be for prospective longitudinal data collection and ascertainment of pregnancy including losses and births • Valid TTP estimates • Daily collection of sexual intercourse, ovulation, menses, and time-varying covariates; capture pregnancy test results • Follow-up ideally through 12 cycles at risk for pregnancy • Feasible to recruit, but requires a large denominator to find couples at risk if population-based • If attrition low, valid estimates of fecundability • If convenient sampling is utilized, selection bias is possible	• Cross-sectional survey of couples at risk for pregnancy regarding duration of "current attempt" or time since stopping contraception to interview. Women typically are not followed-up; TTP is right-censored • TTP-like distributions can be generated • Follow-up of couple to monitor pregnancies is possible • Design excludes couples not waiting • Cannot necessarily differentiate TTP resulting in a pregnancy or not	• Representative sample of women historically report reproductive history including TTP for all pregnancy attempts regardless of outcome • Recall over long (uncertain) intervals introduces recall bias • Model defines success as TTP resulting in pregnancy, whereas others are censored if they quit trying or do not become pregnant	• Survey a cohort currently awaiting pregnancy at baseline along with monitoring of cohort and follow-up of pregnancies at specific times, such as in 6 or 12 months • Based upon Cox regression model with delayed entry into cohort • Ability to consider multiple endpoints
Baird et al. 1986; Joffe 2005	Bonde et al. 1998; Buck Louis et al. 2009	Keiding et al. 2002; Slama et al. 2006	Slama et al. 2006	Olsen and Andersen 1999

TTP, time-to-pregnancy

even after long periods of recall, for male partner reporting or across interview methods such as in-person or self-administered questionnaires (Baird et al. 1991; Zielhuis et al., 1992; Joffe et al. 1993). Most of the published TTP literature is based upon retrospective designs, including research that utilizes it as an outcome (e.g., identifying determinants of TTP) or as a predictor of pregnancy or other health outcomes. This most likely reflects ease of use, although cautious interpretation of effect sizes is required in light of potential threats to the validity of findings.

Prospective design

Prospective designs measures fecundability irrespective of pregnancy and its outcome. This design is the gold standard in that women, men, or couples are recruited at a milestone, such as upon discontinuing contraception or when declaring they are beginning to attempt pregnancy, and then prospectively followed for an a priori specified period of time. As such, prospective designs include all couples at risk and are not restricted to those achieving pregnancy or a live birth, as often is the case with retrospective approaches. The ideal length of follow-up for prospective designs is 12 menstrual cycles at risk for pregnancy. This should capture pregnancies within this interval and will include couples who are more or less fecund, as well as those who do not achieve pregnancy, and thus allows for assessment of the sterility fraction. The *sterility fraction* refers to the fraction of couples who cannot ever conceive. To distinguish sterile couples from those couples with delayed conception, one has to follow them for a sufficient period of time (at least 12 cycles) to avoid being confounded with conception delay. However, logistical or fiscal considerations often necessitate the need to follow couples through 6 months of attempting pregnancy, under the premise that most pregnancies will be captured. The research question will certainly impact the duration of follow-up selected for prospective inquiry.

Prospective designs are challenging for two key reasons: the difficulty of enrolling couples before or as they become at risk for pregnancy in real time, and the need to keep couples motivated and complying with study protocols, particularly if covariates are being asked on a daily basis by both partners of the couples. Cycle 1 is particularly challenging to define and not all investigators discuss what cycle 1 truly represents, irrespective of study design. For example, cycle 1 includes the so-called most fecund couples as they became pregnant in the very first cycle of purported trying, and it also includes couples with mistimed or miscounted trying periods. Thus, cycle 1 represents at a minimum two groups of couples—those who provide a full cycle of observations and those who become pregnant in the first few weeks of trying and do not provide a full cycle of observed data. In essence, investigators obtain more longitudinal data from the less fecund couples as the most fecund become pregnant earlier and complete data collection earlier while trying. This has prompted some investigators to consider categorizing this first cycle as cycle 0 (pregnant within the first few weeks of participation without a full menstrual cycle of observations) and cycle 1 (pregnant within the first fully observed cycle). As more data become available from this approach, it is likely

that the so-called peak in fecundability for cycle 1 may become much lower as more of the pregnant couples are moved to cycle 0. This consideration is relevant for retrospective designs as well.

The validity of prospectively measured TTP is quite good, particularly when sexual intercourse and bleeding data are captured daily, along with some biologic marker for estimating ovulation and pregnancy. Even with such daily-diary approaches, it can be challenging to delineate menstruation for all women, given the varying bleeding profiles within and between women (Mikolajczyk et al. 2010). Although investigators typically define a menstrual cycle as day 1 of bleeding to the next day 1 of bleeding, daily diaries typically reflect a considerable number of days with some bleeding noted, thus making it difficult to establish algorithms for defining menstruation from other forms of bleeding. Thus, defining a menstrual cycle as day 1 to day 1 of the next cycle is overly simplistic and ignores other bleeding within the cycle and, possibly, estimation of cycle length. Some investigators have suggested including bleeding intensity in the algorithm or using menstrual pictograms for measuring blood loss (Wyatt et al. 2001).

Current duration

The current duration design queries couples at risk for pregnancy, irrespective of pregnancy intentions, per se, about the duration of time for their "current attempt" or time since stopping contraception prior to the interview. In essence, time pertains to the interval between the beginning of unprotected intercourse and inclusion in the study (Slama et al. 2006). This design is flexible and can be used with a population-based or other representative sample of women. Also, the design is not restricted to women/couples who are actively attempting pregnancy. Given its cross-sectional nature, women are not typically followed, thus requiring TTP to be right-censored. This design estimates the proportion of couples not pregnant after a given number of months of unprotected intercourse. Duration is then used to estimate the effect of potential exposures on fecundity. When certain assumptions are met, the current duration approach can estimate TTP (Slama et al. 2006).

It is important to note that less fecund couples may be over-represented in this design, since the probability of being included in a cross-sectional study is positively associated with the duration of unprotected intercourse. This over-representation is attributed to length-biased sampling (Zelen 2004), which can be addressed in the analytic plan. This design also excludes couples becoming pregnant right away and, hence, with little to no waiting time.

Historic prospective

The historic prospective design selects a representative sample of women who are queried about their (historic) reproductive history, including TTP for all pregnancy attempts regardless of outcome, and then prospectively followed. As such, successful and unsuccessful pregnancy attempts can be included, although all are retrospectively reported. The analytic model defines pregnancy success as trying attempts resulting in pregnancy or censors women on becoming pregnant.

The potential limitations of this design are similar to any retrospective approach. Additional considerations for fecundity include the validity of retrospective TTP and other potential biases, such as recall bias, especially for longer periods of recall. In addition to utilizing representative sampling, another design strength is the ability to ascertain multiple TTPs for a woman/couple. As such, this design is useful for assessing the correlatedness of TTP within a woman.

Case-cohort

The case-cohort design is applicable when studying multiple outcomes that may have different etiologies without having to implement multiple, nested case-control studies. As applied to reproductive and perinatal epidemiology, the case-cohort design utilizes a cohort of women/couples currently awaiting pregnancy at some specified baseline and then monitors them for subsequent pregnancies at specified intervals. For example, a typical prospective TTP cohort study samples women/couples upon discontinuation of pregnancy and follows them through their trying attempts. In additional to measuring TTP, this cohort design measures other sensitive endpoints, such as pregnancy loss, infertility, or birth outcomes. Rather than conducting separate nested case-control studies, one for each of the three outcomes subject to sufficient power, an investigator could take a random sample of the initial TTP cohort, which will include "cases" for each of the three possible outcomes, plus controls. This random sample is supplemented by all remaining cases not selected in the initial random sample. This design utilizes the standard case-cohort analytic approach, and typically utilizes Cox regression with delayed cohort for assessing fecundity-related determinants. To date, utilization of this design in reproductive and perinatal epidemiology is limited, perhaps a reflection of a limited number of prospective TTP studies. Another methodologic limitation associated with this design is the limited modeling approaches for prospective longitudinal TTP measurement for case-cohort designs. Hopefully, novel analytic approaches will be forthcoming.

Statistical models

Demographers in the 1960s first suggested modeling variations in fecundability using either a Pearson-I or β distribution. Statistical models aimed at estimating fecundability, defined as the probability of conception among exposed women/couples relative to unexposed, have continued to evolve and build upon the early work of Barrett and Marshal (1969) and Schwartz and colleagues (1980). Subsequently, methodologic work developed models that began to address the biologic aspects of fecundability (Weinberg et al. 1994; Scheike and Jensen 1997; Stanford and Dunson 2007), particularly with regard to the timing and frequency of intercourse during the fertile window (Dunson and Stanford 2005). Analytic strategies employed by reproductive epidemiologists for assessing fecundity determinants largely include the discrete time-proportional probabilities model, in which the probability of fecundability is a function of covariates and (discrete) time. This approach estimates the conditional fecundability ratio (CFR) or the

probability of pregnancy per a given menstrual cycle for exposed divided by unexposed participants. A second analytic strategy is discrete conditional logistic regression, in which TTP is categorized as conception delay (TTP >6 months) or infertility (>12 months).

Noticeably absent until recently, however, are models capable of handling sexual behavior and other exposures without having to condition in any way on intercourse patterns. To this end, Kim and colleagues (2010) recently proposed a joint model that can accommodate the heterogeneity in intercourse behavior during the menstrual cycle when estimating conception probabilities while considering other covariates. Such approaches will help reproductive epidemiologists more precisely identify fecundity determinants and conception probabilities, in keeping with a biologically and behaviorally specified model more closely resembling human reproduction. A more complete discussion of analytic considerations is presented in Chapters 15–17.

MEASURING HUMAN FERTILITY

Fertility denotes live births, although sometimes stillbirths are included as well. As such, fertility is more straightforward to measure, relative to fecundity. The recording of births and deaths in populations has been a longstanding practice and is the basis for vital registries. Most countries make some attempt to enumerate live births, thus allowing for the secular comparison of fertility rates along with characteristics of mothers and offspring. In addition to meeting administrative needs, birth certificates continue to be used by researchers to address public health issues. Concerns remain about the validity and reliability of such data, especially when comparing populations. The validity and reliability of birth certificate data vary by the gold standard chosen (e.g., birth certificate, medical record, maternal report) and the type of data field, respectively. In general, sociodemographic variables have higher accuracy and reliability as recorded on birth certificates than do specific medical variates or behaviors during pregnancy, such as use of alcohol or tobacco (Northam and Knapp 2005).

A unique characteristic of fertility, especially relative to death, is that it is repeatable and can be linked with other outcomes, such as morbidity and mortality. In many developed countries, notable fertility trends during the 20th century include a declining age at first birth until approximately the latter two decades, and a smaller completed family size. The 21st-century fertility trends include a delayed age at first pregnancy for women, along with fewer children. These characteristics, coupled with a longer life expectancy, are but a few of the population forces resulting in an older population structure in many developed countries.

Fertility rates

Much of the fertility data for the United States is generated by the National Vital Statistics System, National Center for Health Statistics, Centers for Disease Control and Prevention. Fertility rates are defined as the number of live births per

1,000 women aged 15–44 years for a particular year or time period and/or population subgroup. The birth rate is defined as the number of live births per 1,000 population, as specified for a particular year and/or population subgroup. As such, fertility rates are quantified to the population at risk for pregnancy (women aged 15–44 years), whereas birth rates pertain to the entire population. These two measures provide different perspectives on the number of births relative to women or the population. With regard to the United States, the 2006 fertility rate for all racial groups was 68.5 but varied substantially by racial groups (i.e., 68.0 for whites, 72.1 for blacks, 63.1 for Native Americans, and 67.5 for Asians or Pacific Islanders) (Martin et al. 2009). Similarly, the overall 2006 U.S. birth rate was 14.2, but varied by race (i.e., 13.7 for whites, 16.8 for blacks, 14.9 for American Indians, and 16.6 for Asians or Pacific Islanders).

In the United States, fertility rates declined approximately 44% between 1960 and 2002, although with notable differences by age and race/ethnicity (Hamilton and Ventura 2006). Specifically, rates increased for women 30 years and older, resulting in a mean increase in the age of mother at first birth by approximately 4 years during this interval. The wantedness of births is periodically assessed. For example, the 2002 National Survey of Family Growth (NSFG), comprising 7,643 females aged 15–44 years, noted that 14% of recent births were reported unwanted at the time of conception (Chandra et al. 2005). Another key finding from this survey is that 64% of births occurred to married mothers.

Pregnancy terminations are another important aspect to consider when assessing fertility outcomes or rates. In the United States, abortion rates have declined since 1980, although considerable differences in usage are observed according to a woman's age and race/ethnicity, with higher rates for women aged 30–39 and non-Hispanic black women in comparison to their counterparts (Hamilton and Ventura 2006).

Another recent aspect of fertility is the growing proportion of infants conceived through infertility treatment. Assisted reproductive technologies are defined as the manipulation of oocytes and sperm outside the body to establish a pregnancy and includes in vitro fertilization (IVF), gamete intrafallopian transfer (GIFT), zygote intrafallopian transfer (ZIFT), frozen/donor embryo transfer (ET), and intracytoplasmic sperm injection (ICSI). The advent of ART occurred in 1969, with in vitro fertilization of human oocytes (Edwards et al. 1969), followed by the world's first birth in 1978 (Steptoe and Edwards 1978). Oocyte donation was introduced in 1984 (Lutjen et al. 1984), followed by other technologies such as cryopreservation and embryo storage (Mohr and Trouson 1985), preimplantation diagnosis of genetic defects (Handyside et al. 1990), and ICSI (Palermo et al. 1992). ART classification includes the following components: (a) use of own/donor eggs, (b) transfer of fresh/frozen embryos, and (c) transfer of fresh/frozen embryos into a gestational carrier. Approximately 0.3% of women in the United States have undergone ART, and approximately 4.6% report lifetime use of ovulation-stimulation agents (Chandra et al. 2005). Schieve and colleagues (2009) estimated that 3% to 7% of U.S. infants were conceived with non-ART infertility-related treatment, which is approximately two to six times higher than the percentage of live births conceived with ART. Thus, approximately 6% of U.S. births may be

conceived following any infertility-related treatment. This figure is comparable to the 6.2% of infants born after assisted reproduction in Denmark in 2002 (Andersen and Erb 2006).

Sex ratios

The primary (1°) sex ratio is defined as the number of male-to-female conceptions, whereas the secondary (2°) sex ratio is defined as the number of male-to-female live births. A sex ratio of 1 denotes a comparable number of male-to-female conceptions or births. Sex ratios are usually restricted to singleton births or include multiples only when certain conditions prevail and with appropriate caution, in that monozygotic twins are biologically the same sex whereas dizygotic twins can be either the same or different sex. The true 1° sex ratio is relatively unknown, given our inability to measure conceptions. However, a male excess is reported, along with a disproportionate loss of male conceptuses relative to females following conception to birth, resulting in approximately 102–106 male to 100 female births for a 2° sex ratio of 1.06 (Pyeritz 1998). A ratio over 1.00 denotes male excess, whereas a ratio below 1.00 denotes a female excess. An inverse relation has been reported between gestation and the sex ratio, ranging from an estimated 143/100 at week 5 through 116 at week 8 and 107 at week 12. Thus, males appear to be disproportionately lost from conception through the first year of life, although our ability to quantify this differential is inexact until birth. The 2° sex ratio is also expressed as a proportion in which the number of male births is divided by the total number of births to estimate the percentage of all births that are male.

The 1° sex ratio is genetically determined, although endogenous (e.g., parental gonadotropin and testosterone concentrations) and exogenous (e.g., endocrine-disrupting chemicals) factors may affect sex selection (Mocarelli et al. 2000; James 2004; Taylor et al. 2006). The 2° sex ratio is an endpoint used to quantify births by sex and to evaluate potential reproductive and/or developmental toxicants that may disproportionately affect one sex over the other. Although some authors have speculated that an excess of male conceptions is essential to ensure an excess of live-born males (Charnov and Bull 1989), alternative arguments suggest a preferential loss of female embryos, possibly due to an inherent developmental advantage of males (Krackow et al. 2003; Boklage, 2005).

A recent paper described trends in the 2° sex ratio for the U.S. population during the past 60 years (1940 to 2002) using birth certificates. Three significant transitions were observed. First, the 2° sex ratio declined between 1942 and 1959, followed by an increase between 1959 and 1971, and a decline between 1971 and 2002 (Mathews and Hamilton 2005). Characteristics associated with a 2° sex ratio below 1 were maternal age of less than 15 and more than 45 years, birth order of seven or more, and Japanese race/ethnicity.

A variety of factors have been studied in relation to the 2° sex ratio, including parental sociodemographic characteristics, lifestyle behaviors, hormonal milieu at conception, hyperemesis gravidarum, individual and population stressors, and environmental pollutants, as reviewed elsewhere (James 1987; Chahnazarian 1988;

Eliakim et al. 2000; Rosenfeld and Roberts 2004; Catalano et al., 2006). Still, few established risk factors have been substantiated at the population level, with the exception, perhaps, of parental ages and race. Given the 2° sex ratio's purported sensitivity to exogenous factors, some authors have suggested that it may be a useful public health measure, reflecting embryonic mortality early in development (Williams et al. 1995), or as a sentinel health event for population health (Davis et al. 1998).

Interpretation of the 2° sex ratio is subject to debate, especially in relation to the declining secular pattern observed in many developed countries during the past 50 years (Moller, 1996; Mathews and Hamilton 2005) and, most recently, in a First Nation Community (Mackenzie et al. 2005). Environmental agents, such as PCBs, have been associated with reductions in the 2° sex ratio (Weisskopf et al. 2003; Taylor et al. 2006), along with dioxin (Mocarelli et al. 2000), polychlorinated dibenzofurans, and polychlorinated dibenzo-p-dioxins (del Rio Gomez et al. 2002). Last, 2° sex ratios are reported to vary by type of ART used. For example, a notable female excess was observed for blastocyst-stage embryos conceived with ICSI in comparison to those without ICSI (Luke et al. 2009).

Multiple births

Multiple births denote the presence of two or more embryos or fetuses per pregnancy and include twins and higher-order (triplets and beyond) births. Twins comprise the majority (~94%) of multiples (Luke et al. 2005). Not all multiple births are the same, and the research question will drive to a large extent the type

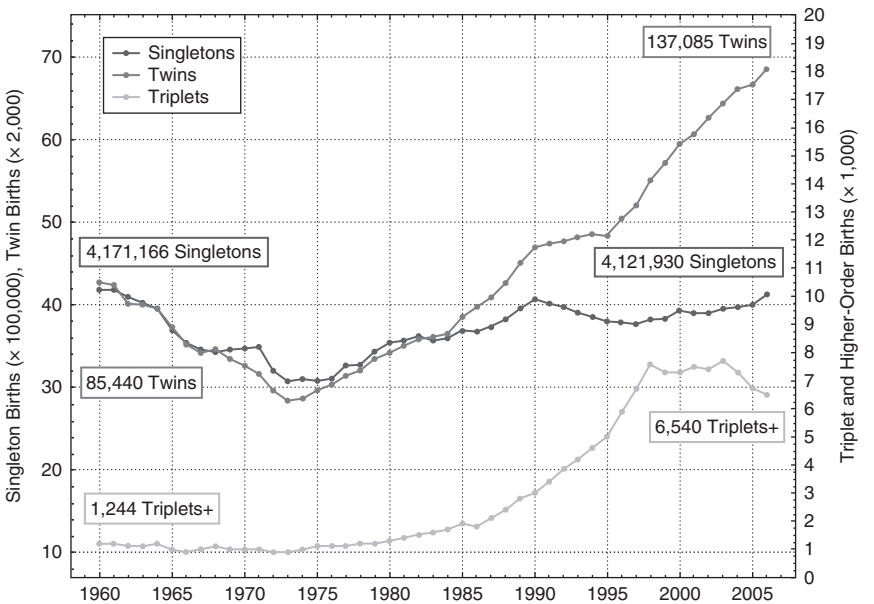


Figure 3.4 Number of live births by plurality, United States, 1960–2006. Courtesy of Dr. Barbara Luke, Michigan State University.

of multiples to be studied. When considering twins, it is important to consider both zygosity (monozygotic or dizygotic) and chorionicity (mono-, di-, or trichorionic), as the descriptive epidemiology varies along with fetal and maternal morbidity and mortality (O'Rahilly and Muller 1992). Monozygotic (MZ) twins are the result of a single fertilized embryo splitting early in development, whereas dizygotic (DZ) twins arise from the fertilization of two oocytes. Multiple births, or *multiples* as they are sometimes called, are typically less biologically mature than singletons, reflecting their earlier deliveries. As a result, they disproportionately contribute to preterm and low-birth-weight deliveries (Russell et al. 2003). Twins also have a somewhat slower postnatal course with regard to growth and development in comparison to singletons, thus necessitating the need for perinatal epidemiologists and clinicians to correct postnatal age for prematurity. Despite higher mortality and morbidity risk, twins provide a unique population for assessing heritable traits, which is tremendously important for assessing the genetic and/or environmental determinants of a spectrum of health outcomes. As such, multiple births are a fascinating component of fertility.

Multiple birth rates denote the number of live-born infants delivered in multiple gestation pregnancies per 1,000 live births. Multiple births of all orders have increased tremendously in the United States and elsewhere. For example, there has been an approximate 65% to 80% increase in twins since the 1980s, and an approximate 460% increase in higher-order births, of which 25%–30% is attributed to older mothers and 70%–75% is attributed to infertility treatment (Keith et al. 2000; Reynolds et al. 2003; Wright et al. 2005; Luke et al., 2005). Figure 3.4 illustrates the number of live births by plurality in the United States between 1960 and 2006. Both triplets and twins increased since the 1980s, although the number of triplets began to stabilize then decline in the mid-1990s, with a recent decline. Conversely, the number of twins continues to increase. These findings underscore the changing landscape of fertility in recent decades.

In the United States between 1997 and 2000, the proportion of multiple births attributable to ART increased from 11.2% to 13.6%, while the percentage attributable to spontaneous conception decreased from 69.9% to 64.5%. The twin rate for ART couples increased from 434.5 to 444.7 per 1,000 live births in 1997 and 2000, respectively, compared to 26.8 and 29.3 for the U.S. population during similar time periods (Reynolds et al. 2003). The triplet or higher birth rates for the United States in 1997 and 2000 were relatively stable at 1.7 and 1.8, respectively. Higher-order birth rates for ART declined during this interval, from 134.3 to 98.7, respectively. Thus, ART has tremendously impacted the multiple birth rate in the United States, in part due to the transfer of multiple embryos rather than single embryo transfer (SET) as performed in many other countries (ESHRE Campus Report 2001). Recent guidance from a Cochrane review suggested that consideration be given to the transfer of a single embryo whenever reasonable (Joint SOGC-CFAS 2008). Gravid health and perinatal risks also appear to vary by whether the pregnancy was aided by infertility treatments. In comparison to naturally conceived multiples, those conceived through ART and ovulation induction may be at increased risk for pregnancy complications (Jackson et al. 2004) and adverse perinatal outcomes (Jackson et al. 2004; Schieve et al. 2007).

The incidence of MZ twinning is relatively constant worldwide at approximately 3–4 per 1,000 births, with much of the variability in twinning stemming from DZ twinning rates (Bulmer 1970). Monozygotic twinning is believed to be a random embryological event except in the presence of ART. Growing evidence supports higher MZ twinning in ART rather than in spontaneous pregnancies, although mechanisms have yet to be determined (Aston et al. 2008).

A number of factors have been associated with DZ twinning and include geography, maternal age, race and ethnicity, nutrition, and family history for DZ pregnancies (Hemon et al. 1981; Bortolus et al. 1999; Hall 2003). Examples of varying risk include a doubling of the risk of DZ twins for a woman whose mother or sister had DZ twins (Parazzini et al. 1994). Dizygotic twinning is highest for blacks followed by whites and Asians (Bulmer 1970).

FECUNDITY AND HEALTH ACROSS THE LIFESPAN

During the past few decades, two conceptual changes in our thinking about human fecundity and fertility have arisen. First, we increasingly recognize that human fecundity and fertility has its basis in utero, with a host of exposures shown to adversely impact later fecundity outcomes. Some in utero exposures (e.g., bisphenol A [BPA] and diethylstilbestrol [DES]) have been shown to have transgenerational fecundity effects, as discussed in Chapters 1 and 2. The second change is an appreciation that fecundity and fertility have implications across the lifespan, and related impairments may be an early marker of later-onset disease. In truth, these conceptual changes are not necessarily new, as developmental biologists and demographers have long appreciated both scenarios. What is new, though, is the recognition on the part of the epidemiologist (and to a certain extent, clinicians) that research needs to be changed in such a manner as to permit our investigation of the origin of human fecundity and fertility in relation to heath across the lifespan.

Of late, two changes have occurred in our conceptual thinking about fecundity. One change is that fecundity arises in utero, and the other is a spectrum of possible health implications across the lifespan associated with fecundity-related impairments. With regard to the first conceptual change, growing evidence supports the testicular dysgenesis syndrome (TDS) hypothesis and, to a lesser extent, the ovarian dysgenesis syndrome (ODS) hypothesis. The TDS hypothesis posits that diminished male fecundity, as measured by semen quality or infertility care-seeking behavior, may be in the causal pathway to testicular cancer and, possibly, genital-urinary malformations in the next generation (Skakkebaek et al. 2001). Semen quality has been associated with mortality in a dose–response manner among 40,000 men seen at a sperm laboratory, thus suggesting its utility as a marker of overall male health (Jensen et al. 2009). The ODS hypothesis posits that women with fecundity disorders such as endometriosis and polycystic ovarian syndrome (PCOS) are at increased risk of later-onset adult diseases, such as autoimmune disorders and reproductive site cancers for the former group of women and type 2 diabetes, metabolic syndrome, and cardiovascular disease in the latter

group, as recently summarized (Louis and Cooney 2007). The effect of in utero exposures and subsequent ovarian function has been recognized for some time, since girls with restricted intrauterine growth have long been recognized to have altered ovarian function (de Bruin et al. 1998). As evidence continues to evolve in support of an in utero origin for endometriosis and PCOS (Cresswell et al. 1997; Missmer et al. 2004), it will be important for researchers to consider underlying causal pathways if we are to delineate the pathophysiology of fecundity impairments and their sequelae. Perhaps in the near future, fecundity will be an early marker of later-onset disease or health across the lifespan, in keeping with a life course epidemiologic research framework.

REFERENCES

Alvarez, C., Castilla, J.A., Martinez, L., Ramirez, J.P., Vergara, F., and Gaforio, J.J. 2003. Biological variation of seminal parameters in healthy subjects. *Hum Reprod* 18: 2082–88.

Andersen, A.N. and Erb, K. 2006. Register data on assisted reproductive technology (ART) in Europe including a detailed description of ART in Denmark. *Int J Androl* 29: 12–16.

Aston, K.I., Peterson, C.M., and Carrell, D.T. 2008. Monozygotic twinning associated with assisted reproductive technologies: A review. *Reproduction* 136: 377–86.

Auger, J., Kunstmann, J.M., Czglik, F., and Jouannet, P. 1995. Decline in semen quality among fertile men in Paris during the last 20 years. *New Engl J Med* 332: 281–85.

Augood, C., Duckitt, K., and Templeton, A.A. 1998. Smoking and female infertility: A systematic review and meta-analysis. *Hum Reprod* 13: 1532–39.

Axmon, A., Rylander, L., Albin, M., and Hagmar, L. 2006. Factors affecting time to pregnancy. *Hum Reprod* 21: 1279–84.

Baird, D.D., Wilcox, A.J., and Weinberg, CR. 1986. Use of time to pregnancy to study environmental exposures. *Am J Epidemiol* 124: 470–80.

Baird, D.D., Weinberg, C.R., and Rowland, A.S. 1991. Reporting errors in time-to-pregnancy data collected with a short questionnaire. Impact on power and estimation of fecundability ratios. *Am J Epidemiol* 133: 1282–90.

Bancroft, J., Loftus, J., and Long, J.S. 2003. Distress about sex: A national survey of women in heterosexual relationships. *Arch Sex Behav* 32: 193–208.

Barrett, J.C. and Marshall, J. 1969. The risk of conception on different days of the menstrual cycle. *Pop Stud* 23: 455–61.

Basso, O., Christensen, K., and Olsen, J. 2004. Fecundity and twinning: A study within the Danish National Birth Cohort. *Hum Reprod* 19: 2222–26.

Bellver, J., Rossal, L.P., Bosch, E., Zuniga, A., Corona, J.T., Melendez, F., et al. 2003. Obesity and the risk of spontaneous abourtion after oocytes donation. *Fertil Steril* 79: 1136–40.

Boklage, C.E. 2005. The epigenetic environment: Secondary sex ratio depends on differential survival in embryogenesis. *Hum Reprod* 20: 583–87.

Bonde, J.P., Hjollund, N.H., Jensen, T.K., Hjollund, N.H.I., Kolstad, H., Henriksen, T.B., et al. 1998. A follow up study of environmental and biologic determinants of fertility among 430 Danish first-pregnancy planners: Design and methods. *Reprod Toxicol* 12: 19–27.

Bonde, J.P., Ernst, E., Jensen, T.K., Hjollund, N.H., Kolstad, H., Henriksen, T.B., et al. 1998. Relation between semen quality and fertility: A population-based study of 430 first-pregnancy planners. *Lancet* 352: 1172–77.

Bortolus, R., Parazzini, F., Chatenoud, L., Benzi, G., Biachi, M.M., and Marini, A. 1999. The epidemiology of multiple births. *Hum Reprod Update* 5: 179–87.

Boyd, J.C. 2010. Defining laboratory reference values and decision limits: Populations, intervals, and interpretations. *Asian J Androl* 12: 83–90.

Buck Louis, G.M., Sever, L.E., Batt, R.E., and Mendola, P. 1997. Review of lifestyle factors and female infertility. *Epidemiology* 8: 435–41.

Buck Louis, G.M., Lynch, C.D., Stanford, J.B., Sweeney, A.M., Schieve, L.A., Rockett, J.C., et al. 2004. Prospective pregnancy study designs for assessing reproductive developmental toxicants. *Environ Health Perspect* 112: 79–86.

Buck Louis, G.M., Dukic, V.M., Heagerty, P.J., Louis, T.A., Lynch, C.D., Ryan, L.M., et al. 2006. Statistical issues in modeling pregnancy outcome data. *Stat Method Med Res* 15(2): 103–126.

Buck Louis, G.M., Lynch, C.D., and Cooney, M.A. 2006. Environmental influences on female fecundity and fertility. *Sem Reprod Med* 24: 147–55.

Buck Louis, G.M., Dmochowski, J., Lynch, C.D., Kostyniak, P.J., McGuinness, B.M., and Vena, J.E. 2009. Polychlorinated biphenyl concentrations, lifestyle and time-to-pregnancy. *Hum Reprod* 24: 451–58.

Bulmer, M.G. 1970. *The biology of twinning in man*. Oxford: Clarendon Press.

Bungum, M., Humaidan, P., Axmon, A., Spano, M., Bungum, L., and Erenpreiss, J. 2007. Sperm DNA integrity assessment in prediction of assisted reproductive technology outcome. *Hum Reprod* 22: 174–79.

Cain, V.S., Johannes, C.B., Avis, N.E., Mohr, B., Schocken, M., Skurnick, J., and Ory, M. 2003. Sexual functioning and practices in a multi-ethnic study of midlife women: Baseline results from SWAN. *J Sex Res* 40: 266–76.

Carrell, D.T., Liu, L., Peterson, C.M., Jones, K.P., Hatasaka, H.H., Erickson, L., and Campbell, B. 2003. Sperm DNA fragmentation is increased in couples with unexplained recurrent pregnancy loss. *Arch Androl* 49: 49–55.

Castilla, A.C., Martinez, L., Ramirez, J.P., Vergara, F., and Gaforio, J.J. 2003. Biological variation of seminal parameters in healthy subjects. *Hum Reprod* 18: 2082–88.

Catalano, R., Bruckner, T., Marks, A.R., and Eskenazi, B. 2006. Exogenous shocks to the human sex ratio: The case of September 11, 2001 in New York City. *Hum Reprod* 21: 3127–31.

Cavazos-Rehg, P.A., Krauss, M.J., Spitznagel, E.L., Schootman, M., Bucholz, K.K., Peipert, J.F., et al. 2009. Age of sexual debut among US adolescents. *Contraception* 80: 158–62.

Chahnazarian, A. 1988. Determinants of the sex ratio at birth: Review of recent literature. *Soc Biol* 35: 214–35.

Chandra, A., Martinez, G.M., Mosher, W.D., Abma, J.C., and Jones, J. 2005. Fertility, family planning, and reproductive health of U.S. women: Data from the 2002 National Survey of Family Growth. *Vital Health Stat* 23: 1–160.

Charnov, E.L. and Bull, J.J. 1989. The primary sex ratio under environmental sex determination. *J Theor Biol* 139: 431–36.

Chavarro, J.E., Rich-Edwards, J.W., Rosner, B.A., and Willett, W.C. 2007. Diet and lifestyle in the prevention of ovulatory disorder infertility. *Obstet Gynecol* 110: 1050–58.

Cole, D.C., Wainman, B., Sanin, L.H., Weber, J.P., Muggah, H., and Ibrahim, S. 2006. Environmental contaminant levels and fecundability among non-smoking couples. *Reprod Toxicol* 22: 13–19.

Cooney, M.A., Buck Louis, G.M., Sundaram, R., McGuiness, B.M., and Lynch, C.D. 2009. Validity of self-reported time to pregnancy. *Epidemiology* 26: 56–59.

Coughlin, M., LaPorte, R., O'Leary, L., and Lee, P. 1998. How accurate is male recall of reproductive information? *Am J Epidemiol* 148: 806–09.

Cresswell, J.L., Barker, D.J.P., Osmond, C., Egger, P., Phillips, D.I., and Fraser, R.B. 1997. Fetal growth, length of gestation, and polycystic ovaries in adult life. *Lancet* 350: 1131–35.

Davis, D.L., Gottlieb, M.B., and Stampnitzky, J.R. 1998. Reduced ratio of male to female births in several industrial countries: A sentinel health indicator? *JAMA* 279: 1018–23.

De Bruin, J.P., Dorland, M., Bruinse, H.W., Spliet, W., Nikkels, P.G., and Te Velde, E.R. 1998. Fetal growth restriction as a cause of impaired ovarian development. *Early Hum Dev* 51: 39–46.

del Rio Gomez, I., Marshall, T., Tsai, P., Shao, Y.S., and Guo, Y.L. 2002. Number of boys born to men exposed to polychlorinated byphenyls. *Lancet* 360: 143–44.

Dunson, D.B., Colombo, B., and Baird, D.D. 2002. Changes with age in the level and duration of fertility in the menstrual cycle. *Hum Reprod* 17: 1399–403.

Dunson, D.B. and Stanford, J.B. 2005. Bayesian inferences on predictors of conception probabilities. *Biometrics* 61: 126–33.

Edwards, R.G., Bavister, B.D., and Steptoe, P.C. 1969. Early stages of fertilization in vitro in human oocytes matured in vitro. *Nature* 221: 632–35.

Eliakim, R., Abulafia, O., and Sherer, D.M. 2000. Hyperemesis gravidarum: A current review. *Am J Perinatol* 17: 207–18.

ESHRE Campus Report. 2002. Prevention of twin pregnancies after IVF/ICSI by single embryo transfer halves the twinning rate without decrease in the ongoing pregnancy rate of an IVF/ICSI programme. *Hum Reprod* 17: 2626–31.

Eskenazi, B., Warner, M., Marks, A.R., Samuels, S., Needham, L., Brambilla, P., and Mocarelli, M. 2010. Serum dioxin concentrations and time to pregnancy. *Epidemiology* 21: 224–31.

Fédération CECOS, Auger, J., and Jouannet, P. 1997. Evidence for regional differences of semen quality among fertile French men. *Hum Reprod* 12: 740–45.

Fedorcsak, P., Storeng, R., Dale, P.O., Tanbo, T., and Abyholm, T. 2000. Obesity is a risk factor for early pregnancy loss after IVF or ICSI. *Acta Obstet Gynecol Scand* 29: 43–48.

Feldman, H.A., Goldstein, I., Hatzichristou, D.G., Krane, R.J., and McKinlay, J.B. 1994. Impotence and its medical and psychosocial correlates: Results of the Massachusetts male aging study. *J Urol* 151: 54–61.

Ferrari, R.M., Cooney, M.A., Vexler, A., Liu, A., and Buck Louis, G.M. 2007. Time to pregnancy and multiple births. *Hum Reprod* 22: 407–13.

Gesink Law, D.C., Klebanoff, M.A., Brock, J.W., Dunson, D.B., and Longnecker, M.P. 2005. Maternal serum levels of polychlorinated biphenyls and 1,1-dichloro-2,2-bis (p- chlorophenyl)ethylene (DDE) and time to pregnancy. *Am J Epidemiol* 162: 523–32.

Gini, C. 1924. Premieres Recherches sur la fecondabilite de la femme. *Proceedings of the International Mathematical Congress* 2: 889–92.

Gini, C. 1926. Decline in the birth rate and "fecundability" of women. *Eugen Rev* 17: 258–74.

Gnoth, C., Godehardt, D., Godehardt, E., Frank-Herrmann, P., and Freundl, G. 2003. Time of pregnancy: Results of the German prospective study and impact on the management of infertility. *Hum Reprod* 18: 1959–66.

Hall, JG. 2003. Twinning. *Lancet* 362: 735–43.

Hamilton, B.E. and Ventura, S.J. 2006. Fertility and abortion rates in the United States, 1960–2002. *Int J Androl* 29: 34–45.

Handyside, A.H., Kontogianni, E.H., Hardy, K., and Winston, R.M. 1990. Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature* 344: 768–70.

Harley, K.G., Marks, A.R., Chevrier, J., Bradman, A., Sjodin, A., and Eskenazi, B. 2010. PBDE concentrations in women's serum and fecundability. *Environ Health Perspect* 18: 699–704.

Hassan, M.A. and Killick, S.R. 2004. Negative lifestyle is associated with a significant reduction in fecundity. *Fertil Steril* 81: 384–92.

Hauser, R. 2006. The environment and male fertility: Recent research on emerging chemicals and semen quality. *Sem Reprod Med* 24: 156–67.

Hemon, D., Berger, C., and Lazar, P. 1981. Twinning following oral contraceptive discontinuation. *Int J Epidemiol* 10: 319–28.

Henderson, K.D., Bernstein, L., Henderson, B., Kolonel, L., and Pike, M.C. 2008. Predictors of the timing of natural menopause in the multiethnic cohort study. *Am J Epidemiol* 167: 1287–94.

Herman-Giddens, M.E., Slora, E.J., Wasserman, R.C., Bourdony, C.J., Bhapkar, M.V., Koch, G.G., et al. 1997. Secondary sexual characteristics and menses in young girls seen in office practice: A study from the Pediatric Research in Office Settings network. *Pediatrics* 99: 505–12.

Herman-Giddens, M.E., Wang, L., and Koch, G. 2001. Secondary sexual characteristics in boys: Estimates from the National Health and Nutrition Examination Survey III, 1988–1994. *Arch Pediatr Adoles Med* 155: 1022–28.

Howards, P.P., Hertz-Picciotto, I., and Poole, C. 2006. Conditions for bias from differential left truncation. *Am J Epidemiol* 165: 444–52.

Jackson, R.A., Gibson, K.A., and Croughan, M.S. 2004. Perinatal outcomes in singletons following in vitro fertilization: A meta analysis. *Obstet Gynecol* 103: 551–63.

James, W.H. 1987. The human sex ratio. Part 1: A review of the literature. *Hum Biol* 59: 721–52.

James, W.H. 2004. Further evidence that mammalian sex ratios at birth are partially controlled by parental hormone levels around the time of conception. *Hum Reprod* 19: 1250–56.

Jensen, T.K., Toppari, K., Keiding, N., and Skakkebaek, N.E. 1995. Do environmental estrogens contribute to the decline in male reproductive health? *Clin Chem* 41: 1896–901.

Jensen, T.K., Carlsen, E., Jorgensen, N., Berthelsen, J.G., Keiding, N., Christensen, K., et al. 2002. Poor semen quality may contribute to recent decline in fertility rates. *Hum Reprod* 17: 1437–40.

Jensen, T.K., Jorgensen, N., Punab, M., Haugen, T.B., Suominen, J., Zilaitiene, B., et al. 2004. Association of *in utero* exposure to maternal smoking with reduced semen quality and testis size in adulthood: A cross-sectional study of 1,770 young men from the general population in five European countries. *Am J Epidemiol* 159: 49–58.

Jensen, T.K., Jacobsen, R., Christensen, K., Nielsen, N.C., and Bostofte, E. 2009. Good semen quality and life expectancy: A cohort study of 43,277 men. *Am J Epidemiol* 170: 559–65.

Joffe, M., Villard, L., Li, Z., Plowman, R., and Vessey, M. 1993. Long-term recall of time-to-pregnancy. *Fertil Steril* 60: 99–104.

Joffe, M., Key, J., Keiding, N., Scheike, N., and Jensen, T.K. 2005. Studying time to pregnancy by use of a retrospective design. *Am J Epidemiol* 162: 115–24.

Joint SOGC-CFAS. 2008. Guidelines for the number of embryos to transfer following in vitro fertilization No. 182, September 2006. *Int J Gynaecol Obstet* 102: 203–16.

Jorgensen, N., Carlsen, E., Nermoen, I., Punab, M., Suominen, J., Andersen, A.G., et al. 2002. East-West gradient in semen quality in the Nordic-Baltic area: A study of men from the general population in Denmark, Norway, Estonia and Finland. *Hum Reprod* 17: 2199–208.

Juul, S., Karamus, W., Olsen, J.,. 1999. Regional differences in waiting time to pregnancy: Pregnancy-based surveys from Denmark, France, Germany, Italy and Sweden: The European Infertility and Subfertility Study Group. *Hum Reprod* 14: 1250–54.

Karpati, A.M., Rubin, C.H., Kieszak, S.M., Marcus, M., and Troiano, R.P. 2002. Stature and pubertal stage assessment in American boys: The 1988-1994 Third National Health and Nutrition Examination Survey. *J Adolesc Health* 30: 205–12.

Kaufmann, R.B., Morris, L., and Spitz, A.M. 1997. Comparison of two question sequences for assessing pregnancy intentions. *Am J Epidemiol* 145: 810–16.

Keiding, N., Kvist, K., Hartvig, H., and Tvede, M. 2002. Estimating time to pregnancy from current durations in a cross-sectional sample. *Biostatistics* 3: 565–78.

Keith, L.G., Oleszczuk, J.I., and Keith, D.M. 2000. Multiple gestations: Reflections on epidemiology, causes, and consequences. *Int J Fertil* 45: 206–14.

Kim, S., Sundaram, R., and Louis, G.M.B. 2010. Joint modeling of intercourse behavior and human fecundability using structural equation models. *Biostatistics* 1: 59–71.

Kinsey, A.C., Pomeroy, W.B., Martin, C.E., and Gebhard, P.H. 1953. *Sexual behavior in the human female*. Philadelphia, PA: WB Saunders Co.

Knuth, U.A., Kuhne, J., Bals-Pratsch, M., and Nieschlag, E. 1988. Intraindividual variation of sperm velocity, linearity, lateral head displacement and beat frequency in healthy volunteers. *Andrologia* 20: 243–48.

Krackow, S., Schmidt, T.A., and Elepfandt, A. 2003. Sexual growth dimorphism affects birth sex ratio in house mice. *Proc Biol Sci* 270: 943–47.

Laumann, E.O., Paik, A., and Rosen, R.C. 1999. Sexual dysfunction in the United States: Prevalence and predictors. *JAMA* 281: 537–44.

Leridon, H. and Spira, A. 1984. Problems in measuring the effectiveness of infertility therapy. *Fertil Steril* 41: 580–86.

Leridon, H. 2004. Can assisted reproduction technologies compensate for the natural decline in fertility with age? A model assessment. *Hum Reprod* 19: 1548–53.

Leridon, H. 2007. Studies of fertility and fecundity: Comparative approaches from demography and epidemiology. *C R Biol* 330: 339–46.

Lewis, S.E. and Aitken, R.J. 2005. DNA damage to spermatozoa has impacts on fertilization and pregnancy. *Cell Tissue Res* 322: 33–41.

Louis, G.M. and Cooney, M.A. 2007, Effects of environmental contaminants on ovarian function and fertility. In *Novel concepts in ovarian endocrinology*, pp. 249–268. A. Gonzalez-Bulnes, ed. Kerala, India: Research Signpost.

Luke, B., Brown, M.B., Alexandre, P.K., Kinoshi, T., O'Sullivan, M.J., Martin, D., et al. 2005. The cost of twin pregnancy: Maternal and neonatal factors. *Am J Obstet Gynecol* 192: 909–15.

Luke, B., Brown, M.B., Grainger, D.A., Baker, V.L., Ginsburg, E., and Stern, J.E. 2009. The sex ratio of singleton offspring in assisted-conception pregnancies. *Fertil Steril* 92: 1579–85.

Lutjen, P., Trounson, A., Leeton, J., Findlay, J., Wood, C., and Renou, P. 1984. The establishment and maintenance of pregnancy using in vitro fertilization and embryo donation in a patient with primary ovarian failure. *Nature* 307: 174–75.

Lynch, C.D., Jackson, L.W., and Buck Louis, G.M. 2006. Estimation of the day-specific probabilities of conception: Current state of the knowledge and the relevance for epidemiological research. *Paediatr Perinat Epidemiol* 20(Suppl.1): 3–12.

Mackenzie, C.A., Lockridge, A., and Keith, M. 2005. Declining sex ratio in a first nation community. *Environ Health Perspect* 113: 1295–98.

Martin, J.A., Hamilton, B.E., Sutton, P.D., Ventura, S.J., Menacker, F., Kirmeyer, S., and Mathews, T.J. 2009. Births: Final data for 2006. *National vital statistics reports*; vol 57. no 7. Hyattsville, MD: National Center for Health Statistics. 1–104.

Mathews, T.J. and Hamilton, B.E. 2005. Trend analysis of the sex ratio at birth in the United States. *Natl Vital Stat Rep* 53: 1–17.

Menken, J., Trussell, J., and Larsen, U. 1986. Age and infertility. Science 233: 1389–94.

Mikolajczyk, R., Louis, G.B., Cooney, M., Lynch, C., and Sundaram, R. 2010. Heterogeneity of vaginal bleeding and its implications for measuring time in time-to-pregnancy. *Paediatr Perinat Epidemiol* 24: 24–30.

Missmer, S.A., Hankinson, S.E., Spiegelman, D., Barbieri, R.L., Marshall, L.M., and Hunter, D.J. 2004. Reproductive history and endometriosis among premenopausal women. *Fertil Steril* 82: 1501–08.

Mocarelli, P., Gerthoux, P.M., Ferrari, E., Patterson, D.G., Jr., Kieszak, S.M., Brambilla, P., et al. 2000. Paternal concentrations of dioxin and sex ratio of offspring. *Lancet* 355: 1858–63.

Mohr, L.R. and Trouson, A.O. 1985. Cryopreservation of human embryos. *Ann NY Acad Sci* 442: 536–43.

Moller, H. 1996. Change in male-female ratio among newborn infants in Denmark. *Lancet* 348: 828–29.

Nichols, H.B., Trentham-Dietz, A., Hampton, J.M., Titus-Ernstoff, L., Egan, K.M., Willett, W.C., and Newcomb, P.A. 2006. From menarche to menopause: Trends among US women born from 1912–1969. *Am J Epidemiol* 164: 1003–11.

Nguyen, R.H.N. and Baird, D.D. 2005. Accuracy of men's recall of their partner's time to pregnancy. *Epidemiology* 16: 694–98.

Niswander, K.R. and Gordon, M. 1972. *The women and their pregnancies*. Philadelphia, PA: W.B. Saunders.

Northam, S. and Knapp, T.R. 2006. The reliability and validity of birth certificates. *J Obstet Gynecol Neonatal Nurs* 35: 3–12.

North American Menopause Society. 2002. *Menopause core curriculum study guide*. Cleveland, OH: North American Menopause Society.

Nguyen, R.H.N. and Baird, D.D. 2005. Accuracy of men's recall of their partner's time to pregnancy. *Epidemiology* 16: 694–98.

Olsen, J. and Andersen, P.K. 1999. We should monitor human fecundity, but how? A suggestion for a new method that may also be used to identify determinant of low fecundity. *Epidemiology* 10: 419–21.

Olsen, J., Bonde, J.P., Hjollund, N.H., Basso, O., and Ernst, E. 2005. Using infertile patients in epidemiologic studies. *Hum Reprod* 11: 607–11.

O'Rahilly, R. and Muller, F. 1992. *Human embryology and teratology*. New York: Wiley-Liss.

Ostermeier, G.C., Miller, D., Huntriss, J.D., Diamong, M.P., and Krawetz, S.A. 2004. Reproductive biology: Delivering spermatozoan RNA to the oocyte. *Nature* 429: 154.

Palmero, G., Joris, H., Devroey, P., and Van Steirteghem, A.C. 1992. Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. *Lancet* 340: 17–18.

Parazzini, F., Villa, A., Moroni, S., Tozzi, L., and Restelli, S. 1994. The epidemiology of multiple pregnancies. *Acta Genet Med Gemellol (Roma)* 43: 17–23.

Pasqualotto, F.F., Junior, E.B., and Pasqualotto, E.B. 2008. The male biological clock is ticking: A review of the literature. *Sao Paulo Med J* 126: 197–201.

Peterson, R. and Moos, M. 1997. Defining and measuring unintended pregnancy: Issues and concerns. *Womens Health Issues* 7: 234–40.

Pyeritz, R.E. 1998. Sex: What we make of it. *JAMA* 279: 269.

Ramlau-Hansen, C.H., Thulstrup, A.M., Nohr, E.A., Bonde, J.P., Sorensen, T.I.A., and Olsen, J. 2007. Subfecundity in overweight and obese couples. *Hum Reprod* 22: 1634–37.

Reynolds, M.A., Schieve, L.A., Martin, J.A., Jeng, G., and Macaluso, M. 2003. Trends in multiple births conceived using assisted reproductive technology, United States, 1997–2000. *Pediatrics* 111: 1159–62.

Ridout, M.S. and Morgan, B.J.T. 1991. Modelling digit preferences in fecundability studies. *Biometrics* 47: 1423–33.

Rockett, J.C., Lynch, C.D., and Buck Louis, G.M. 2004. Biomarkers for assessing reproductive development and health: Part 1: Pubertal development. *Environ Health Perspect* 112: 105–12.

Rosenfeld, C.S. and Roberts, R.M. 2004. Maternal diet and other factors affecting offspring sex ratio: A review. *Biol Reprod* 71: 1063–70.

Russell, R.B., Petrini, J.R., Damus, K., Mattison, D.R., and Schwarz, R.H. 2003. The changing epidemiology of multiple births in the United States. *Obstet Gynecol* 101: 129–35.

Rylander, L., Wetterstrand, B., Haugen, T.B., Malm, G., Malm, J., Bjorsvik, C., et al. 2009. Single semen analysis as a predictor of semen quality: Clinical and epidemiological implications. *Asian J Androl* 11: 723–30.

Salma, R., Ducot, B., Carstensen, L., Corente, C., de La Rochebrochard, E., Leridon, H., et al. 2006. Feasibility of the current-duration approach to studying human fecundity. *Epidemiology* 17: 440–49.

Salumets, A., Suikkari, A-m., Möls, T., Söderström-Anttila, V., and Tuuri, T. 2002. Influence of oocytes and spermatozoa on early embryonic development. *Fertil Steril* 78: 1082–87.

Samaras, T.T. and Storms, L.H. 2002. Secular growth and its harmful ramifications. *Med Hypotheses* 58: 93–112.

Scheike, T.H. and Jensen, T.K. 1997. A discrete survival model with random effects: An application to time to pregnancy. *Biometrics* 53: 318–29.

Schieve, L.A., Cohen, B., Nannini, A., Ferre, C., Reynolds, M.A., Zhang, Z., et al. 2007. A population-based study of maternal and perinatal outcomes associated with assisted reproductive technology in Massachusetts. *Matern Child Health J* 11: 517–25.

Schieve, L.A., Devine, O., Boyle, C.A., Petrini, J.R., and Warner, L. 2009. Estimation of the contribution of non-assisted reproductive technology ovulation stimulation fertility treatments to US singleton and multiple births. *Am J Epidemiol* 170: 1396–407.

Schwartz, D., MacDonald, P.D.M., and Heuchel, V. 1980. Fecundability, coital frequency and the viability of ova. *Popul Stud* 34: 397–400.

Schwartz, D., Mayaux, M.J., Spira, A., Moscato, M.L., Jouannet, P., Czyglik, F., and David, G. 1983. Semen characteristics as a function of age in 833 fertile men. *Fertil Steril* 39: 530–35.

Sharpe, R.M. 1994. Regulation of spermatogenesis. In *The physiology of reproduction*, 2nd ed. E. Knobil and J.D. Neale, eds. New York: Raven Press.

Shape, R.M. and Franks, S. 2002. Environment, lifestyle and infertility: An intergenerational issue. *Nat Cell Biol* 4(Suppl): 33–40.

Skakkebaek, N.E., Rajpert-De Meyts, E., and Main, K.M. 2001. Testicular dysgenesis syndrome: An increasingly common developmental disorder with environmental aspects. *Hum Reprod* 16: 972–78.

Slama, R., Ducot, B., Carstensen, L., Lorente, C., de La Rochebrochard, E., Leridon, H., et al. 2006. Feasibility of the current-duration approach to studying human fecundity. *Epidemiology* 17: 440–49.

Stanford, J.B. and Dunson, D.B. 2007. Effects of sexual intercourse patterns in time to pregnancy studies. *Am J Epidemiol* 165: 1088–95.

Steptoe, P.C. and Edwards, R.G. 1978. Birth after the reimplantation of a human embryo. *Lancet* 2: 366.

Stevens, V.C. 1997. Some reproductive studies in the baboon. *Hum Reprod Update* 3: 533–40.

Stokes-Riner, A., Thurston, S.W., Brazil, C., Guzick, D., Liu, F., Overstreet, J.W., et al. 2007. One semen sample or two: Insights from a study of fertile men. *J Androl* 28: 638–43.

Taylor, K.C., Jackson, L.W., Lynch, C.D., Kostyniak, P.J., and Buck Louis, G.M. 2006. Preconception maternal polychlorinated biphenyl concentrations and the secondary sex ratio. *Environ Res* 103: 99–105.

Tietze, C., Guttmacher, A.F., and Rubin, S. 1950. Time required for conception in 1727 planned pregnancies. *Fertil Steril* 1: 338–46.

Uhlenhaut, N.H., Jakob, S., Anlag, K., Eisenberger, T., Sekido, R., Kress, J., et al. 2009. Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell* 139: 1130–42.

Viudes-de-Castro, M.P. and Vicente, J.S. 1997. Effect of sperm count on the fertility and prolificity rates of meat rabbits. *Anim Reprod Sci* 46: 313–19.

Wang, X., Chen, C., Wang, L., Chen, D., Guang, W., French, J., and The Reproductive Health Study Group. 2003. Conception, early pregnancy loss, and time to clinical pregnancy: A population-based prospective study. *Fertil Steril* 79: 577–84.

Weinberg, C.R., Baird, D.D., and Wilcox, A.J. 1994. Sources of bias in studies of time to pregnancy. *Stat Med* 13: 671–81.

Weinberg, C.R., Gladen, B.C., and Wilcox, A.J. 1994. Models relating the timing of intercourse to the probability of conception and the sex of the baby. Biometrics 50: 358–36.

Weisskopf, M.G., Anderson, H.A., and Hanrahan, L.P. 2003. Great Lakes Consortium. *Environ Health* 2(1): 2.

Williams, F.L., Ogston, S.A., and Lloyd, O.L. 1995. Sex ratios of births, mortality, and air pollution: Can measuring the sex ratios of births help to identify health hazards from air pollution in industrial environments? *Occup Environ Med* 52: 164–69.

Wilcox, A.J., Weinberg, C.R., O'Connor, J.F., Baird, D.D., Schlatterer, J.P., Canfield, R.E., et al. 1988. Incidence of early loss of pregnancy. *N Engl J Med* 319: 189–94.

World Health Organization. 2010. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 5th edition. Geneva: World Health Organization.

Wright, V.C., Chang, J., Jeng, G., and Macaluso, M. 2008. Assisted reproductive technology surveillance- United States, 2005. *MMWR Surveill Summ* 57: 1–23.

Wyatt, K.M., Dimmock, P.W., Walker, T.J., and O'Brien, P.M.S. 2001. Determination of total menstrual blood loss. *Fertil Steril* 76: 125–31.

Zelen, M. 2004. Forward and backward recurrence times and length biased sampling: age specific models. *Lifetime Data Anal* 10: 325–34.

Zieluis, G.A., Hulscher, M.E.J.L., and Florack, E.I.M. 1992. Validity and reliability of a questionnaire on fecundability. *Int J Epidemiol* 21: 1151–56.

Zou, K., Yuan, Z., Yang, Z., Luo, H., Sun, K., Zhou, L., et al. 2009. Production of offspring from a germline stem cell line derived from neonatal ovaries. *Nat Cell Biol* 11: 631–36.

4

Fecundity Impairments

GERMAINE M. BUCK LOUIS

Fecundity impairments are an important component of reproductive epidemiology, given the considerable heterogeneity in time-to-pregnancy (TTP) among couples attempting to become pregnant and given that some couples are incapable of conceiving even with treatment. The inefficiency of human reproduction, as described in Chapter 3, coupled with considerable pregnancy loss further underscores the importance of fecundity impairments in relation to reproductive health. For example, Roberts and Lowe (1975) estimated pregnancy loss for married women aged 20–29 years residing in England and Wales in 1971, and concluded that only 22% of menstrual cycles at risk for pregnancy resulted in a live birth, suggesting that many conceptions or early pregnancy losses go unobserved. Identifying the etiologic determinants of (un)successful pregnancy and delivery should help identify interventions aimed at reducing fecundity impairments.

Fecundity impairments are distinct but somewhat interrelated outcomes and include difficulty conceiving, difficulty carrying a pregnancy to term, or the inability to conceive. Translated to endpoints, they become conception delay, pregnancy loss, and infecundity or the more colloquial term *infertility*. In the United States, the National Survey of Family Growth (NSFG) is an important source of information for reproductive health for a cross-sectional representative sample of the U.S. household population aged 15–44 years. Data from the NSFG suggest that the prevalence of impaired fecundity increased from 8% in 1982 and 1988 to 10% in 1995 among all women aged 15–44 years (Chandra and Stephen 1998). Similar increases were observed when restricted to married women (i.e., 11% in 1982 and 1988, and 13% in 1995). It should be noted that impaired fecundity has an unusual definition in the NSFG in that it is defined as difficulty conceiving, difficulty carrying a pregnancy to term, or the inability to conceive after 3 years of unprotected intercourse. The prevalence of infertility during this time interval, as measured in the NSFG, reflected a somewhat different pattern—namely a purported decline in 12-month infertility between 1982 and 2002 (Stephen and Chandra 2006). However, the later analyses were restricted to married women, despite dramatic increases in cohabitation rates among U.S. women (Bumpass and Lu 2000). To some extent, characterizing fertility impairments needs to be done with respect to two populations—the true population at risk and the population at risk that seeks medical care. Only a few prospective

studies with preconception enrollment of couples/women have longitudinally measured fecundity for at least 12 months to capture conception delay and infecundity, and even fewer have followed women through their "trying" period and ensuing pregnancies to capture pregnancy losses across the spectrum of gestation, as noted in Chapter 3. In fact, the only study known to date that followed couples up for 12 months of trying through pregnancy and delivery is the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development's recently completed Longitudinal Study of Fertility and the Environment (LIFE Study; www.lifestudy.us). One aim of the LIFE Study is to more carefully characterize fecundity and related impairments in relation to parental exposures, consistent with the couple-dependent nature of reproductive outcomes.

CONCEPTION DELAYS

Conception delay is defined in a number of ways, depending upon the length of time a couple has been at risk for pregnancy or believed to have been trying. Most often, researchers define conception delay as a TTP between 7–12 months. For purposes of this text, we define conception delay as a TTP of more than 6 but of 12 months or less; infecundity is defined as a TTP more than 12 months, as described later in this chapter. Subfecundity is also called subfertility by some authors and has been operationalized differently across investigators, disciplines, and countries, thus making it difficult to fully synthesize the weight of evidence with regard to its descriptive and analytic epidemiology. Despite conflicting terminology for subfecundity, the key elements include a trying time of at least 6 months with an eventual pregnancy.

To a certain extent, there is some behavioral-biological basis for defining conception delay as a TTP of more than 6 months, given that prospective pregnancy studies with preconception enrollment of women/couples who are followed for up to 12 menstrual cycles report that the majority of women becoming pregnant do so within the first 6 months. In three prospective pregnancy studies in which women were followed for 12 months/cycles, cumulative pregnancies at 6 months ranged from 76% to 90% (Gnoth et al. 2003; Wang et al. 2003; Buck Louis et al. 2009). These findings reflect declining conception probabilities across the trying period as the more fecund couples achieve pregnancy faster and are dropped from the denominator in comparison to the less fecund. Many of the so-called fecundity determinants also are relevant for conception delay and infertility, since each of these outcomes is defined upon some categorization of TTP. Rather than to review individual risk factors for conception delay, the reader is encouraged to consider the fertility determinants listed in Table 3.1 in relation to conception delay and infertility. Suffice to say, age (Fédération CECOS et al. 1982), various biologic factors as recently reviewed by subtype of infertility (Evers 2002), and a host of environmental factors including lifestyle (Fédération CECOS et al. 1982; Evers 2002; Mendola and Buck Louis 2010) have been associated with conception delays. Timing of exposures during the periconception or in utero sensitive windows need careful consideration when assessing determinants or risk factors

for conception delays, particularly for those exposures with endocrine-disrupting potential or other inducers of epigenetic changes, consistent with the highly timed and interrelated processes underlying human reproduction (Sharpe and Skakkebaek 1993).

Secular patterns for conception delay are not well described in that most representative surveys query women about infertility, per se, or the inability to conceive after 12 months of trying. Some authors have articulated the need to monitor TTP as a measure of female or couple fecundity either through prospective (Olsen et al. 2003) or retrospective (Joffe et al. 2003) approaches to track fecundability and, thereby, conception delay and infertility. Other authors argue that monitoring fecundity is not feasible, given the population- and individual-level factors that influence TTP (Sallmen et al. 2005). Concerns about prospective monitoring include potential selection bias and logistical considerations about feasibility and cost. Retrospective monitoring also gives rise to notable concerns, including reporting errors stemming from recall or digit preferences for TTP (Ridout et al. 1991), and validity considerations, especially after long periods of recall (Cooney et al. 2009).

PREGNANCY LOSS

No uniform typology exists for defining pregnancy loss, although there is agreement that the probability of loss varies across sensitive windows of human development. Specifically, an inverse relation has long been recognized between days postconception or gestation and risk of loss. Pregnancy loss may arise at varying stages of human development, including failure of the blastocyst to implant as a result of various mechanisms including poor uterine receptivity or oocyte quality, or delayed implantation (Spandorfer and Rosenwaks 1999; Wilcox et al. 1999). Implantation and placentation are essential for the establishment and support of pregnancy.

Figure 4.1 illustrates the continuum of pregnancy loss and its varying terminology that coincides with the timing of the event. In addition, risk of pregnancy loss varies by study design and methods utilized for selecting the study cohort and ascertaining pregnancy. For example, pregnancy loss is reported to range from 18% to 62% of human chorionic gonadotropin (hCG) pregnancies, from 12% to 33% of pregnancies using life table approaches, and from 8% to 17% using women's reproductive histories (Modvig et al. 1990). The exact timing of most pregnancy losses is unknown, thus making it difficult to come up with precise estimates across gestation. For simplicity purposes, pregnancy loss can be conceptualized as an early or clinical loss reflecting the manner in which pregnancy is affirmed. It is important, however, to note that irrespective of definition used for loss, there may be no biologic basis for any categorization based on gestational age. As such, the fetuses-at-risk paradigm may be appropriate for conceptualizing all pregnancy loss, as discussed in Chapters 13 and 14. *Early pregnancy loss* is measured by the detection of hCG, whereas a *clinical pregnancy loss* is defined as any loss following "clinical" confirmation of pregnancy. Numerous approaches are available for

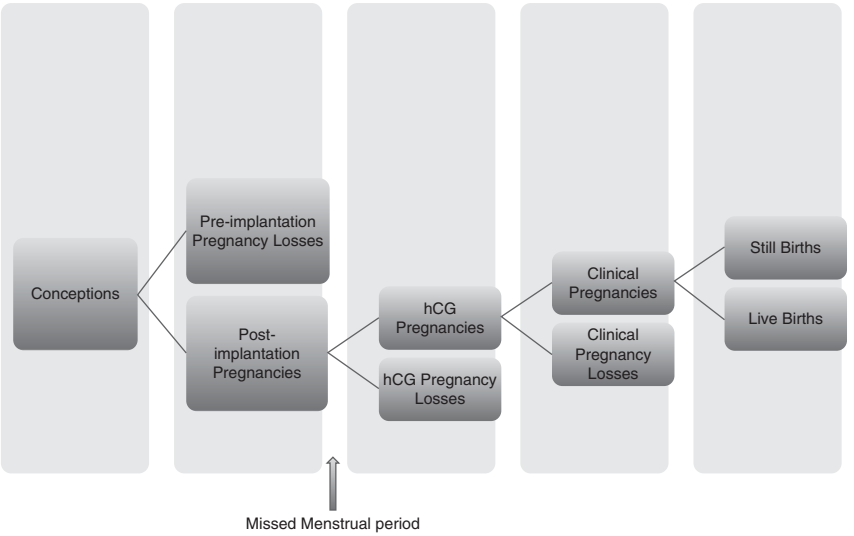


Figure 4.1 Estimated incidence of pregnancy loss from conception through birth.

defining pregnancy within each of these two categories. For example, early loss can be defined on the basis of home pregnancy test kits or upon single/serial quantification of hCG in urine or blood. Clinical pregnancy losses arise at any point following entry into clinical care.

Although beyond the scope of this chapter, it is important for the reader to note that approximately 5% of women experience recurrent pregnancy loss (RPL), which is defined as two or more failed pregnancies; or three or more failures, which affects approximately 1% of women (Rai and Regan 2006; ASRM 2008). A number of causes have been suggested for RPL, including genetic or chromosomal disorders (e.g., translocation), maternal age of 40 years or older, low progesterone concentrations, metabolic and uterine abnormalities, anticardiolipin syndrome, thrombophilias, and male factors (e.g., abnormal sperm integrity). However, no cause is determined in approximately half of affected women (ASRM 2008). The absence of universally accepted definitions for RPL, inconsistent clinical approaches for when and how to evaluate couples, and various sources of bias (e.g., ascertainment, misclassification) make it challenging to conduct epidemiologic research (recently summarized by Christiansen et al. 2006).

Early pregnancy loss

Early pregnancy loss (EPL) is defined as the loss of an hCG-detected pregnancy irrespective of method and before a woman seeks medical care. As such, only the woman observes her pregnancy or sometimes the researcher if active monitoring is in place. A relatively large percentage of conceptions are unrecognized, given our inability to measure conception at the population level and the estimated high

incidence of losses before detection. Although uncertain, up to 78% of all fertilized ova may be lost, thus underscoring the inefficiency of human reproduction (Roberts and Lowe 1975; Regan and Rai 2000; Macklon et al. 2002). As such, a fundamental methodologic consideration in the study of pregnancy loss is our, as yet, lack of a standard definition of pregnancy that is applicable for epidemiologic and clinical research. With no biomarker of conception, particularly one amenable for population-based research, estimates of pregnancy loss primarily capture the interval between implantation of the blastocyst at approximately 5–6 days postovulation and delivery. Risk of pregnancy loss has long been known to vary inversely with length of gestation (French and Bierman 1962; Erhardt 1963; Harlap et al. 1980). As such, studies that comprise already pregnant women will have a lower incidence of pregnancy loss than will studies with preconception enrollment of women.

Human chorionic gonadotropin is the glycoprotein hormone of pregnancy that is detectable in blood and urine in a variety of isoforms. It is responsible for controlling aspects of reproduction, working in concert with other closely related glycoprotein hormones, such as luteinizing hormone (LH) and follicle-stimulating hormone (FSH) (Birken et al. 2001). Human chorionic gonadotropin is secreted mainly by the syncytial cytotrophoblasts of the placenta at approximately 5–8 days postfertilization, and is essential for pregnancy maintenance and progression; the pituitary secretes a small amount of hCG that may contribute to a very small number of false-positive pregnancies or the so-called phantom hCG (Cole 2000). The many roles of hCG include the continued production of steroid hormones and growth factors in the corpus luteum (Rull and Laan 2005), blastocyst implantation (Srisuparp et al. 2001), uterine vascularization and angiogenesis (Zygmunt et al. 2002), and immunological adaptation during pregnancy (Rao 2001). Synthesis of hCG begins shortly after fertilization at the two-cell stage (Jurisicova et al. 1999), and peaks in blood at approximately 9–10 weeks of gestation, followed by a decrease through week 16 and an eventual rise at week 22 through delivery (Hay 1988). Since hCG is not exclusively produced by trophoblasts, pregnancy determination based upon hCG requires detection above a threshold or detection limit for a period of time. Even with such safeguards, considerable laboratory variability exists for hCG quantification, especially at lower limits.

Concentrations of hCG are informative about pregnancy outcome, with low first-trimester concentrations being associated with adverse outcomes such as pregnancy loss, ectopic pregnancy, Down syndrome, unsuccessful outcomes of assisted reproductive technologies (ART), and trophoblastic disease (Gaspard et al. 1980; Gerhard and Runnebaum 1984; Cuckle 2000; Letterie and Hibbert 2000; Poikkeus et al. 2002). A recent study described the urinary pattern of hCG secretion in the 7 days following natural implantation among 142 women with spontaneously conceived clinical pregnancies and noted that the daily mean hCG concentrations rose rapidly following implantation, ranging from a threefold to a 1.6-fold rise by days 6–7 (Nepomnaschy et al. 2008). Overall, hCG concentrations were variable and followed a log-quadratic trajectory during the first week. In contrast to previous work largely based upon women with clinically recognized pregnancy, hCG patterns were unrelated to plurality, spontaneous abortion, or

infant sex. Lohstroh and colleagues (2005) longitudinally collected blood and urine samples for 63 women undergoing donor insemination and measured hCG, LH, and FSH in blood via immunoassay. Approximately 13% of cycles resulted in an EPL or loss before 22 days, and no significant differences were observed in the occurrence or concentration of hCG by pregnancy-loss status. These findings need to be cautiously interpreted, given the absence of longitudinal analytic techniques capable of handling the correlatedness of cycles.

At the population level, EPL also can be defined as a change in home pregnancy test kits from a positive to negative reading during a specified period of time or a decline in hCG concentrations as measured in blood or urine. The absence of an embryonic sac or fetal heart beat at approximately 6–7 weeks of gestation following hCG confirmation may be indicative of EPL, particularly for subgroups of the population under clinical observation. Recognizably, this may introduce confusion as to whether such a loss is an EPL or a clinical loss. To this end, clinicians often refer to EPL as biochemical or subclinical pregnancies, particularly for couples undergoing ART. Estimating the incidence of EPL among hCG-confirmed pregnancies requires prospective study designs with preconception enrollment and longitudinal measurement. Although not always possible, failure to use such a design introduces the possibility of truncation bias, given that women would be recruited at varying gestational ages whereas others would not be recruited if they had already experienced a loss (Howards et al. 2006).

Morris and Udry (1967) published one of the earliest reports of hCG-detected pregnancy loss, which generated considerable interest and research in relation to estimating pregnancy and the incidence of loss. Subsequently, Miller and colleagues (1980) conducted a prospective study focusing on EPL involving 197 volunteer women who stopped hormonal contraception for purposes of becoming pregnant. Women collected urine on day 21 of each cycle, then on alternating days until the onset of menstruation or upon becoming pregnant. The loss rate was 43%, including 50 hCG and 14 clinical pregnancy losses. Subsequent authors estimated the incidence of EPL based upon measurements of hCG in daily urine samples, with a range from 8% (Whittaker et al. 1983) to 57% (Edmonds et al. 1982). It is important to note that the wide range may reflect varying EPL definitions and assays for quantifying hCG (Wilcox et al. 1985). Using the reportedly most sensitive immunoradiometric assay for hCG (Canfield et al. 1987), Wilcox and colleagues (1988) estimated the incidence of EPL to be 22%. Among infecund couples seeking ART, the incidence of EPL when hCG concentrations are measured on days 14–17 following oocyte retrieval with ensuing ultrasonography at about 6–7 weeks of gestation is reported to range from 12% to 48% (Acosta et al. 1990; Simon et al. 1999).

Table 4.1 presents the incidence of EPL among women with hCG-confirmed pregnancies who participated in prospective cohort studies that utilized preconception recruitment sampling frameworks. Incidence ranges from 40% to 12%. Of note, the lowest incidence may reflect use of home pregnancy test kits in an era before the availability of digital kits that replaced subjective symbols (e.g., +/–) with actual digital words (pregnant/nonpregnant) to remove subjectivity in interpreting test results. One of the key findings, as summarized in Table 4.1, is the

Table 4.1 Incidence of early and clinical pregnancy loss in prospective pregnancy studies with preconception enrollment of women

Authors	*Measurement hCG*	*% EPL*	*% EPL & clinical losses*
Miller et al. 1980	Urinary [hCG] ≥5 μg/L on one occasion or ≥2 μg/L on two occasions	33	9
Wilcox et al. 1988	Urinary [hCG] 0.025 ng/mL for 3 days	22	31
Sweeney et al. 1988	Serum radioimmunoassay and 5 monoclonal antibody urinary hCG tests	13	19
Hakim et al. 1995	Urinary [hCG] 0.25 ng/mL for 2 consecutive days	37	52
Eskenazi et al. 1995	Urinary [hCG] ≥0.15 ng/mg creatine for 2 of 3 days	40	52
Zinaman et al. 1996	Urinary [hCG] 0.15 ng/mL for 3+ consecutive days	13	31
Ellish et al. 1996	Urinary [hCG] ≥4.00 pmol/L for 3 days or ≥5.33 for 2 days or ≥6.67 on last day	17	—
Bonde et al. 1998	Urinary [hCG] >0.8–1.0 IU followed by a decline	17	29
Wang et al. 2003	Urinary [hCG] 0.01 ng/mL	25	33
Buck Louis et al. 2009	Urinary home pregnancy test kits capable sensitive ≤50 m IU	12	17

Some percentages were estimated based upon data given in papers. Caution is required in comparing percentages, as definitions vary across studies. Percentages rounded.

EPL, early pregnancy loss; [ ] denotes concentration

high percentage of pregnancy losses, particularly EPL, thus underscoring the importance of this endpoint for epidemiologic investigation.

The occurrence of EPL is challenging to identify and only a few such studies have quantified this window. A delay in implantation has been cited as a risk factor for EPL (Baird et al. 1991). The median duration from ovulation to loss ranged from 14 (Nepomnaschy 2005) to 16 days (Vitzhum et al. 2006), whereas the median duration based upon cycles resulting in EPL ranged from 30.5 (Vitzhum et al. 2006) to 32 days (Wilcox et al. 1999). This timing underscores the importance of preconception enrollment when attempting to identify etiologic determinants or in assessing potential reproductive or developmental toxicants. Other promising approaches for measuring pregnancy loss include gonadotropin and ovarian hormone profiles to aid in the detection of endocrine signaling pathways relevant for pregnancy maintenance or demise, particularly since the ovary may respond to signals from the embryo at approximately 6 days following the LH surge (Stewart et al. 1993).

Despite the high incidence of EPL, there is limited epidemiologic study focusing on its natural history. Harville and colleagues (2003) estimated that 9% of women participating in a prospective pregnancy study with longitudinal capture of menstruation and using a sensitive hCG biomarker for pregnancy reported 1 or more days of bleeding during the first 8 weeks of pregnancy. The occurrence of bleeding tended to be light and at the time menstruation was expected, rather than at the time of implantation, which has long been hypothesized to be associated with bleeding in some women (Speert and Guttmacher 1954). Despite the long-standing belief that bleeding may be indicative of an adverse pregnancy outcome

(Ananth and Savitz 1994), 86% of affected women gave birth. Using data from the same cohort study as Harville and colleagues (2003), Promislow and colleagues (2007) evaluated bleeding patterns following an EPL before 6 weeks and reported an increased duration of bleeding of approximately 0.4 days longer than average menstrual bleeding, thus underscoring the difficulty in recognizing EPL on the basis of bleeding alone or in the absence of hCG testing. To date, no studies have reported the natural history of EPL or clinical pregnancy loss, making it difficult to order and quantify the causal chain of events as measured by signs and symptoms. In what may be the only natural history study of the onset of pregnancy symptoms that may be informative for the natural history of EPL, Sayle and colleagues (2002) assessed the timing of pregnancy symptoms (yes/no) in the interval between discontinuing contraception through 8 weeks past the last menstrual period (LMP) among 136 women participating in a prospective pregnancy study with live births. Approximately 94% of women with clinical pregnancies reported symptoms upon becoming pregnant or after 3 consecutive days with urinary hCG exceeding 0.025 ng/mL. Half the women reported symptoms within 20 days following ovulation, and most by the end of 6 weeks post LMP. In contrast, only 21% of women experiencing EPL reported symptoms. Women experiencing clinical pregnancy losses proportionately reported fewer symptoms than did women giving birth. Last, it is important to keep in mind that a number of sociocultural and biologic factors may affect the natural history of pregnancy loss. These include women's ability to recognize pregnancy as influenced by planning intentions, menstrual cycle characteristics, birth control practices, knowledge of the fertile window, reproductive history, and access to health care.

Numerous risk factors for EPL have been investigated, including sociodemographic, medical, and reproductive history; lifestyle; environmental toxicants; and population-level factors, such as season. Morphologically abnormal blastocysts (Hertig et al. 1959), poor embryonic quality (Boué and Boué 1975), and genetic abnormalities (Stein et al. 1975) are reported to be high among EPLs. Aneuploidy, defined as an abnormal number of chromosomes that is not an exact multiple of the haploid number, is a marker of poor oocyte quality and is reported to be associated with loss. Incidence of aneuploidy varies by stage of development (65%–75% of early spontaneous losses, 35% of clinical spontaneous losses, 4% of stillbirths, and 0.3% of newborns (Plachot et al. 1988; Hassold et al. 1996). Perhaps the most consistent risk factors for EPL are advancing maternal age and prior history of a loss. Few studies have compared risk factors for EPL relative to clinical pregnancy loss, partly a result of so few prospective cohort studies from preconception through delivery. Wilcox and colleagues (1990) compared the risk factors for EPL (losses within 42 days of LMP based upon hCG) and reported increased risks for prenatal diethylstilbestrol (DES) exposure, current cigarette smoking, recent caffeine use, history of a previous spontaneous abortion, and gravidity. Reduced risks were observed for recent birth control usage, IUD user, thin stature, and childhood exposure to smoke (Wilcox et al. 1990). Of late, there is greater recognition of the need to consider the role of parentally mediated effects and not just maternal factors, consistent with the couple-dependent nature of successful reproduction.

Some of the strongest maternal risk factors are observed for male partners as well, and include paternal age (Slama et al. 2003). Specifically, the risk (rate ratio) of spontaneous abortion was 2.13-fold higher (95% confidence interval [CI] 1.07–4.26) for women age 25 years whose partner was age 35 years or older, in comparison with women age 25 years whose partner was younger than age 35 years. No such increased risk of spontaneous abortion with male age was estimated when the woman was age 35 years. Male and female alcohol intake during the week of conception also has been associated with risk of EPL, thus underscoring the importance of a couple-based exposure paradigm for the identification of etiologic mechanisms for EPL (Henriksen et al. 2004). Last, conceptions occurring as a part of ART provide an interesting and exciting opportunity for identifying determinants and possible risk factors. Many risk factors for EPL have been observed among couples receiving ART and include obesity, maternal cigarette smoking, and transfer of poor-quality embryos (Fedorcsak et al. 2000; Winter et al. 2002).

As yet, we have limited understanding regarding the mechanisms underlying the maintenance of early pregnancy. Such critical data gaps make it difficult to conceptualize etiologic mechanisms that impact choice of study design and methods for capturing early pregnancy and ensuing losses. For example, adequate progesterone production by the corpus luteum is critical for maintaining the pregnancy until the placenta begins functioning at approximately 7–9 weeks of pregnancy. Still, we do not fully understand progesterone's mode of action or the interplay between maintenance of early pregnancy and placental growth and differentiation, including the differentiation of trophoblasts. Once pregnancy is well established, we still do not fully appreciate the immunologic regulation of the fetal semi-allograft, and why the mother does not reject her fetus.

Clinical pregnancy loss

Clinical pregnancy loss refers to all losses occurring after clinical recognition or confirmation of pregnancy, irrespective of when or how pregnancy was determined. Clinical pregnancy loss can be defined to include or exclude losses occurring late in pregnancy or after approximately 22–28 weeks' gestation, which are typically defined as fetal deaths or stillbirths, as discussed in Chapter 12. The terminology for clinical pregnancy (similar to EPL) is quite varied and sometimes uses gestational age at the timing of observed loss for its definition despite any strong biologic basis for such categorization. In essence, clinical pregnancy loss denotes that loss occurred after the woman has received clinical care and nothing more. Categorizing gestation rather than considering it as a continuum distorts our ability to delineate risk factors for pregnancy loss across the continuum of pregnancy, especially given the inverse relation between gestation and loss. This issue is similar to the fetuses-at-risk paradigm discussed in greater detail in Chapters 13 and 14. Research initiatives that utilize a prospective approach for loss across the continuum of pregnancy may generate a more complete understanding of pregnancy loss.

The incidence of clinical pregnancy loss varies tremendously depending upon study design and methods. Two early landmark studies reported a range from

16.9% to 60% of pregnancies. French and Bierman (1962) conducted one of the first prospective pregnancy studies in Kauai that involved 3,000 pregnant women and estimated loss to be approximately 18.6%. Treloar and colleagues (1967) conducted a prospective study of 2,000 college women who contributed more than 300,000 menstrual cycles of observation and reported that 16.9% of recognized pregnancies were lost.

Irrespective of the timing of pregnancy loss, most investigators continue to look at the same spectrum of possible risk factors. As noted above, we have virtually no information that allows us to assess exposures or risk factors from TTP through pregnancy and delivery. Nor do we have studies that measured both male and female factors across the trying period and through pregnancy and delivery to assess maternal, paternal, or couple-mediated exposures and outcomes. As such, it is difficult to try to synthesize the literature to identify those risk factors specific to conception delay or loss across the spectrum of gestation. Suffice to say, global categories of risk factors for clinical pregnancy loss include sociodemographic, medical and reproductive history, lifestyle, environmental toxicants, and population-level factors.

INFECUNDITY

Infecundity is defined as the inability to conceive, irrespective of pregnancy intentions, as in the case of congenitally absent reproductive organs or genetic syndromes resulting in sterility. It pertains to men, women, and couples. Infecundity is frequently defined in the context of couples attempting pregnancy who fail to become pregnant within 12 months of regular sexual intercourse without contraception. Much of the scientific and lay literature uses the term *infertility*, which technically is a misnomer, if one defines *fecundity* as the biologic capacity for reproduction and *infecundity* as its absence. Despite such recognition, and to be consistent with definitions utilized by the authors cited in this chapter, we use the term *infertility*.

Implicit in the definition of infertility is the assumption that couples want to be and can be treated to achieve pregnancy. In fact, many retrospective TTP studies truncate self-reported TTP at 12 months, consistent with the time frame within which couples were historically encouraged to seek care. The biological basis for a 12-month cutpoint does not exist, per se, particularly in light of growing evidence supporting a sizeable number of spontaneous pregnancies that occur within 12–24 or 12–36 months of trying, as discussed below. Thus, it is important to note that couples are not sterile, per se, if they do not achieve pregnancy at 12 months. Rather, they have lower fecundity and will require more time or, possibly, medical interventions.

Infertility is classified by type and diagnostic subtype. Type includes primary and secondary infertility, with *primary infertility* denoting the inability to conceive among couples who have never been pregnant, and *secondary infertility* as the inability to conceive among couples with a previous pregnancy, regardless of outcome. Infertile couples who achieve pregnancy are sometimes referred to as

having *resolved infertility*, whereas couples without pregnancy are said to have *unresolved infertility*. Clinical subtype(s) of infertility are illustrated in Figure 4.2. However, it is important to keep in mind that there is no universally accepted scheme for classifying couples, and current approaches are dependent upon physician practices, extensiveness of clinical testing, the sensitivity and specificity of testing, the relatively limited discriminatory capability of tests, and the very limited information about what constitutes normal variations in human fecundity. Guzick and colleagues (1994) followed fertile couples through a routine infertility evaluation and reported that one or more infertility factors were observed in two-thirds of women (only 10/32 fertile couples had completely normal tests). This very interesting study raises the question: What is a normal fertility clinical workup? The validity and reliability of self-reported infertility subtypes have received limited study except for a recent study involving 9,164 Dutch women undergoing in vitro fertilization (IVF) (de Boer et al. 2005). Causes of infertility (including multiple causes) as reported by women were tubal (38%), male factor (38%), unexplained (23%), endometriosis (13%), and hormonal factors (17%). When compared to medical reports, validity ranged by subtype: that is, tubal (84%), male factor (78%), unexplained infertility (59%), cervical (40%), and uterine (46%). This study supports the need to validate self-reported diagnostic infertility subtypes.

Several questions remain about the probability of spontaneous conceptions among infertile couples, irrespective of medical care-seeking behaviors for infertility treatment. Dunson and colleagues (2004) estimated that 43% to 63% of infertile couples would conceive within an additional 12 cycles of trying.

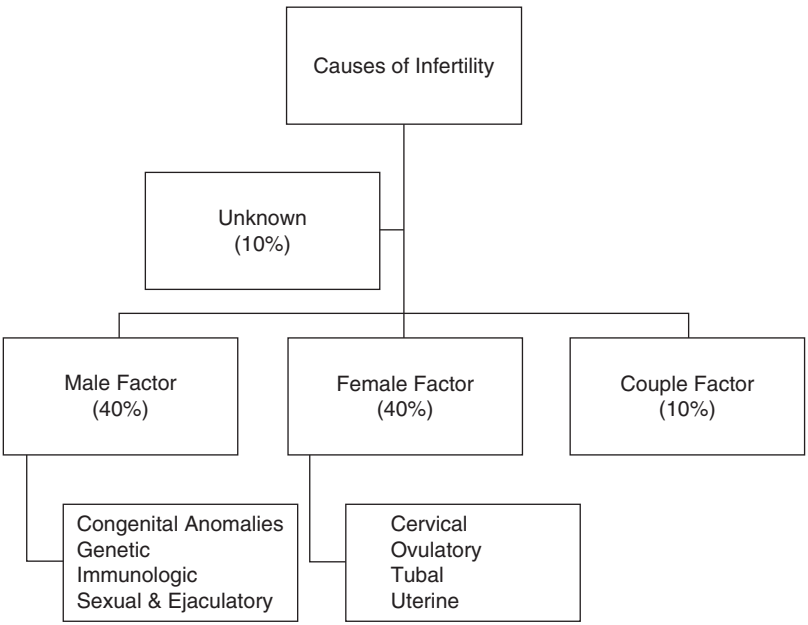


Figure 4.2 Diagnostic subtypes of infertility.

Among infertile women clinically found to be ovulating and with no evidence of tubal obstruction, approximately 60% spontaneously conceived, of which half the pregnancies resulted in a live birth (Kapiteijn et al. 2006). Stanford and colleagues (2010) conducted simulations to model the cumulative probability of pregnancy among subfertile couples without a definitive diagnosis, and estimated that 33% of couples with an infertility history of 2 years' duration would become pregnant within 1 year in the context of frequent intercourse, compared with 20% of couples with less-frequent intercourse. These simulations emphasize the need to consider biology (fecundity) with behavior (timing and frequency of sexual intercourse). Pregnancy outcomes also vary by source of infertility clinical care. For example, approximately 52.5% of pregnant couples initiating infertility treatment from primary-care centers had spontaneous pregnancies resulting in live births at 36 months compared with 25.2% of couples receiving services at tertiary-care centers (Collins et al. 1995; Snick et al. 1997).

The incidence of infertility is largely unknown, although prevalence has been described with respect to person, place, and time characteristics. Some authors argue that infertility is increasing, as evidenced by the increasing number of couples seeking care, the adoption of unhealthy lifestyles detrimental to fertility, and growing environmental contamination. Opponents allege that the so-called increasing prevalence is a simple reflection of changes in population characteristics, such as intentional delays in childbearing, a more socially conducive milieu for seeking care, and other sociodemographic characteristics. To some extent, information on infertility is limited to women seeking care, which is currently estimated to only be 38.5% of U.S. women aged 22–44 years in 2002 (Chandra and Stephen 2010). This raises legitimate concerns at the population level, given that use of services among U.S. women aged 22–44 years is associated with older age, nulliparity, marital status, and higher socioeconomic status (Chandra and Stephen 2010). In addition, there is considerable heterogeneity in care-seeking behavior for fecundity impairments, which include at a minimum help in becoming pregnant or in preventing miscarriage. In 2002, approximately 2.3 million women aged 22–44 years reported seeking medical help for becoming pregnant, and 0.9 million to prevent miscarriage (Chandra and Stephen 2010). The number of women receiving ovulation-induction medications was higher than those seeking ARTs, 1.2 and 0.9 million, respectively. Although approximately half of infertile couples are reported to seek medical service (Olsen et al. 1996), estimates vary by type of infertility, study population, and definitions of care-seeking behavior. For example, the prevalence of medical care-seeking behavior ranged from 32% to 95% for women with primary infertility and from 22% to 79% for women with secondary infertility in three cross-sectional studies (Schmidt and Munster 1995). In both developed and less-developed nations, the percentage of couples seeking medical care was similar and ranged from 42% to 76.3% and from 27% to 74.1%, respectively (Boivin et al. 2007). However, approximately 22.4% of couples undergo treatment.

Reasons for seeking or not seeking infertility-related services have received some study, but most of the research focuses on women. Care-seeking behavior is associated with sociodemographic factors such as older age, married, and higher

socioeconomic status (Stephen and Chandra 2000). Using data from the 2002 NSFG, 7.5% of sexually active men aged 15–44 years reported a visit for help having a child during their lifetime, of which 2.2% reported a visit in the past year (Andersen et al. 2009). Male demographic characteristics associated with care-seeking behavior included age and marital status. In the United States, approximately 59% of women reporting having sought infertility services noted that they became pregnant within 2 years of their last visit, of which 78% had a live birth (Farr et al. 2009). Thirty-two percent of women discontinued services before achieving pregnancy, most within the first month. Socioeconomic differences in care-seeking behavior are reported for countries with national insurance inclusive of infertility services (Moreau et al. 2010).

Couple withdrawal from infertility treatment is sizeable and an important consideration when interpreting results (treatment success or otherwise) restricted to couples receiving and completing care. Approximately 23%–60% of couples withdraw from treatment. For example, between 17% and 70% of couples undergoing ART stop treatment, as recently reviewed by Brandes et al. (2009). Reasons cited include spontaneous pregnancy, emotional distress, financial implications, and medical considerations. In the first study assessing when couples discontinue medical care, Brandes and colleagues (2009) noted that half of couples stop before any treatment, and two-thirds before IVF was initiated. Major reasons cited included rejecting treatment, emotional distress, or poor prognosis.

Both incidence and prevalence of infertility are a function of how both the numerator and denominator are defined for a study. Numerators can include women with current and/or lifetime infertility or be further restricted to include only those who seek care. Denominators range from the population at risk for pregnancy, such as all women aged 15–44 years, or women actually attempting pregnancy. In addition, the denominator should exclude women/couples who are sexually inactive or who have known sterility for either biologic (genetic syndromes) or contraceptive (tubal sterilization/vasectomy) reasons. Prevalence of infertility varies across population and by definition used. In an early study, prevalence of infertility was shown to be a function of definition used and ranged from 6.1% for physician-diagnosed infertility to 32.6% based upon a self-report of 12 months with unprotected intercourse (Marchbanks et al. 1989). Worldwide evidence suggests that the prevalence of infertility during one's lifetime ranges from 6.6% (Rostad et al. 2006) to 33.6% (Schmidt et al. 1995).

One of the earliest estimates of subfertility (or infertility) reported that one in six couples in the United Kingdom was affected (Hull et al. 1985). Varying prevalence has been reported for European countries as well. Schmidt and colleagues (1995) reported that current infertility ranged from 3.6% to 14.3% and lifetime infertility from 12.5% to 33.6%. Juul and colleagues (1999) estimated the prevalence of infertility in various European countries and reported an overall prevalence of 16%, which ranged from 10% in Southern Italy to 24% in East Germany. Karmaus and colleagues (1999) estimated the prevalence of infertility defined as 12 months or more of unprotected intercourse in five population-based samples of women aged 25–44 years. For first pregnancies, 19% of couples required 12 months or more, but geographic variation was noted ranging from 14.8% to 33.3%

when restricting to first attempts in the previous 5 years. In developing countries, prevalence is believed to be high and, again, geographically variable. Prevalence in Asia and Latin America was reported to range from 8% to 12% (WHO 1991). In a 25 population-based survey comprising 172,413 women, prevalence ranged from 3.5% to 16.7% (Boivin and colleagues 2007). Within sub-Saharan Africa, prevalence ranged from 9% in Gambia (Sundby et al. 1998) to 20%–30% in Nigeria (Okonofua 1996; Larsen 2000).

The secular patterns of infertility in the United States vary somewhat by the time period or subgroups being assessed. Prevalence was reported to have decreased from 8.5% in 1982 to 7.4% in 2002 among married women aged 15–44 years, although higher rates were observed for non-Hispanic black and Hispanic women and for women without a college education (Stephen and Chandra 2006). For example, prevalence for black women rose from 7.1% in 1982 to 11.6% in 2002 and from 7.2% to 7.7% among Hispanic women. Based upon NSFG data, prevalence of 12-month infertility or impaired fecundity ranged from 5.1 million women aged 15–44 years in 1982 to 6.7 million in 1995 and 7.9 million in 2002 (Chandra and Stephen 2005). Restricting to married women aged 15–44 years produces estimates of approximately 3.6, 4.3, and 4.8 million, respectively. Combined, these data suggest a temporal trend for increasing prevalence, particularly among unmarried women or specific racial/ethnic subgroups. Given the growing proportion of pregnancies and births to unmarried women, it is important to fully assess infertility irrespective of marital status.

As with any fecundity outcome, various risk factors have been suggested or preliminarily assessed. These include biologic factors such as age (Henry 1953; Menken et al. 1986; LaRochebrochard et al. 2003; Leridon 2004), higher body mass index (Fedorcsak et al. 2004; Aggerholm et al. 2008; Sneed et al. 2008), and a host of environmental factors including lifestyle and behaviors (Joffe and Li 1994). A review of risk factors for female infertility categorized as probable, potential, and suspected for female infertility highlighted the limited information for many diagnostic subtypes of infertility (Buck et al. 1997). A recent paper assessed couples' lifestyle at two points in time (2001 and 2007) in relation to subfertility or a TTP of more than 12 months. Despite modest improvement in lifestyle, the incidence of subfertility was not reduced, thus raising concerns about our ability to mediate risk (Killick et al. 2009). Much of what we know about risk factors is dependent upon medical care-seeking behavior, particularly for clinical study populations or samples. Since approximately only half of infertile couples seek care (Olsen et al. 1996), external validity or generalizability of findings to populations not seeking care is limited. Clinical care-seeking populations are appropriate for etiologic investigation, particularly during the sensitive windows of fertilization through implantation and pregnancy (Olsen et al. 2005). Moreover, couples are unable to self-identify the diagnostic subtype of their infertility.

There remain many critical data gaps underlying infertility. Valid and reliable incidence figures are lacking, making it difficult to empirically delineate its secular pattern in response to concerns about declining human fecundity. In addition, there has been limited study of couples, despite their being the appropriate unit

of analysis for couple-dependent outcomes such as conception. In fact, prior to the 1950s, infertility was largely viewed as a female problem, and this may be a partial explanation for the continual overemphasis on female factors in the absence of a male partner context. Using data from the European Fecundability Study, Dunson and colleagues (2004) reported that prevalence of infertility increased with increasing paternal age among women aged 35 years. In comparison to men aged less than 35 years, infertility prevalence ranged from 18% to 28% for male partners aged 35 and 40 years, respectively. This paper has sparked attention on the need for couple-based research scenarios for assessing fecundity and its impairments.

HEALTH IMPLICATIONS OF IMPAIRED FECUNDITY

To many, human fecundity and fertility denotes pregnancy and nothing more. This overly simplistic conceptualization ignores the many nonpregnant aspects of fecundity, such as gynecologic or urologic health and minimizes the implications of both for health status in general. During the past few decades, two conceptual changes have arisen in our thinking about human fecundity and fertility. These paradigm shifts include the growing recognition that human fecundity and fertility have an early origin, possibly commencing in the periconception sensitive window, and that fecundity and fertility have implications for health across the lifespan. A brief discussion of each conceptual consideration is presented to encourage the reader to think of fecundity and fertility in the context of overall health.

Evidence of a relation between fecundity impairments and adverse perinatal outcomes has existed for several decades. This includes evidence that conception delay is associated with decrements in gestation and birth weight (Williams et al. 1991; Joffe and Li, 1994; Henriksen et al. 1999). More recently, an evolving body of evidence during the past few decades suggests that environmental exposures, possibly those occurring in utero, may alter the hormonal milieu, resulting in a permanent reprogramming of the fetus for diminished fecundity and fertility, and the later appearance of adult-onset diseases. Skakkebaek and colleagues (2001) synthesized the literature for males and posited the testicular dysgenesis hypothesis (TDS), building upon the fetal-origins-of-disease hypothesis (Barker 1992). This hypothesis cites declining semen quality, increasing genital-urinary malformation rates, and increasing rates of testicular cancer as possible evidence of early alterations in the hormonal milieu, with permanent changes throughout the lifespan. An example of such a link between fecundity and later-onset diseases is evidence that subfecund men may be at higher risk of developing testicular cancer than are fecund men (Moller and Skakkebaek 1999; Baker et al. 2005).

Recently, investigators have attempted to synthesize changes in female fecundity and fertility in relation to later-onset disease using the TDS paradigm. To this end, the ovarian dysgenesis syndrome (ODS) hypothesis provides a framework for synthesizing findings linking environmental exposures during critical and sensitive windows to female fecundity and fertility, which in turn impact health

during pregnancy and later in life (Louis et al. 2006; Buck Louis and Cooney 2007). For example, DES exposure has been associated with endometriosis (Missmer et al. 2006), which in turn is associated with a decreased risk of preeclampsia and other gravid diseases (Brosens et al. 2007), but an increased risk of autoimmune disorders (Sinaii et al. 2002) and reproductive site cancers (Bertelsen et al. 2007) later in adulthood. Polycystic ovarian syndrome (PCOS) has been associated with in utero androgen exposure and is associated with a higher risk of gravid diseases, such as preeclampsia and gestational diabetes (Brosens et al. 2007), which in turn is associated with a higher risk of insulin resistance, type 2 diabetes, cardiovascular diseases, and metabolic syndrome (Dunaif 1993; Ovalle and Azziz 2002; Glueck et al. 2003). Another example of an association between gravid and later-onset disease is an 8.12-fold higher risk of death from cardiovascular disease for women with a history of preeclampsia and a preterm delivery in comparison with unaffected women (Irgens et al. 2001).

Fertility outcomes also have been linked to later-onset diseases. Female infants born small-for-gestational age or with evidence of intrauterine growth restriction are reported to have smaller ovaries, more anovulatory cycles as adults, and decreased responsiveness to follicle-stimulating hormone in comparison with unaffected girls (deBruin et al. 1998; Ibanez et al. 2002). Such girls may be predisposed to PCOS and the adult diseases believed associated with PCOS. Male infants of lower birth weight corrected for gestational age are at greater risk for genital-urinary malformations, such as cryptorchidism and hypospadias, than are larger-sized infants (Akre et al. 1999; Weidner et al. 1999). With regard to multiples, there is some evidence that in unlike-sex (male/female) pairs, the presence of a female fetus is associated with a prolonged gestation for the male fetus, leading to higher birth weight for him (Loos et al. 2001).

To date, the exact mechanisms underlying the associations observed between fecundity and fertility outcomes and later-onset diseases have not been identified with certainty, although epigenetic mechanisms or changes in gene expression have been suggested, with DES serving as the prototype for an intrauterine or even transgenerational exposure. Epigenetic research incorporates both genetic and environmental factors that affect phenotype, including those during sensitive windows of human reproduction and development. Hopefully, this avenue of research will delineate causal mechanisms underlying successful reproduction.

In sum, there is evidence in support of an early origin for human fecundity and fertility, which in turn has implications for health and disease across the lifespan. These findings and others support a life course approach for epidemiologic study as advocated, to understand health and disease in the broadest extent possible (Ben-Shlomo and Kuh 2002). Given the longstanding recognition of the racial/ethnic disparity in health, including reproductive health (Butts and Seifer 2010), it is important to keep in mind that such disparities exist for fecundity and fertility as well. Thus, epidemiologists will need to meet the challenges of designing research to include the health implications beyond an exposure- and/or endpoint-specific perspective to include a life-course perspective if we are to more fully understand the origin of reproductive health and its ramifications for health and well-being.

REFERENCES

Acosta, A.A., Oehninger, S., Hammer, J., Hammer, J., Muasher, S.J., Liang, H.M., and Jones, D.L. 1990. Preclinical abortions: Incidence and significance in the Norfolk in vitro fertilization program. *Fertil Steril* 53: 673–76.

Aggerholm, A.S., Thulstrup, A.M., Toft, G., Ramlau-Hansen, C.C.H., and Bonde, J.P. 2008. Is overweight a risk factor for reduced semen quality and altered serum sex hormone profile? *Fertil Steril* 90: 619–626.

Akre, O., Lipworth, L., Cnattingius, S., Sparen, P., and Ekbom, A. 1999. Risk factor patterns for cryptorchidism and hypospadias. *Epidemiology* 10: 364–69.

American Society for Reproductive Medicine. 2008. Patient's Fact Sheet: Recurrent Pregnancy Loss. Birmingham, AL: ASRM.

Ananth, C.V. and Savitz, D.A. 1994. Vaginal bleeding and adverse reproductive outcomes: A meta analysis. *Paediatr Perinatal Epidemiol* 8: 62–78.

Anderson, J.E., Farr, S.L., Jamieson, D.J., Warner, L., and Macaluso, M. 2009. Infertility services reported by men in the United States: National survey data. *Fertil Steril* 91: 2466–70.

Baird, D.D., Weinberg, C.R., Wilcox, A.J., McConnaughey, D.R., Musey, P.I., and Collins, D.C. 1991. Hormonal profiles of normal conceptive cycles ending in early, unrecognized loss. *J Clin Endocrinol Metab* 72: 793–800.

Baker, J.A., Buck, G.M., Vena, J.E., and Moysich, K.B. 2005. Fertility patterns prior to testicular cancer diagnosis. *Cancer Causes and Control* 16: 295–99.

Barker, D.J.P., ed. 1992. Fetal and infant origins of adult disease. London: BMJ Publishing.

Ben-Shlomo, Y. and Kuh, D. 2002. A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol* 31(2): 285–93.

Birken, S., Kovalevskaya, G., and O'Conor, J. 2001. Immunochemical measurement of early pregnancy isoforms of hCG: Potential applications to fertility research, prenatal diagnosis, and cancer. *Arch Med Res* 32: 635–43.

Boivin, J., Bunting, L., Collins, J.A., and Nygren, K.G. 2007. International estimates of infertility prevalence and treatment-seeking: Potential need and demand for infertility medical care. *Hum Reprod* 22: 1506–12.

Bonde, J.P., Hjollund, N.H., Jensen, T.K., Ernst, E., Kolstad, H., Henriksen, T.B., et al. 1998. A follow-up study of environmental and biologic determinants of fertility among 430 Danish first-pregnancy planners: Design and methods. *Reprod Toxicol* 12: 19–27.

Boué, J., Boué, A., and Lazar, P. 1975. Retrospective and prospective epidemiological studies of 1500 karyotyped spontaneous human abortions. *Teratology* 12: 11–26.

Brandes, M., van der Steen, J.O.M., Bokdam, S.B., Hamilton, C.J.C.M., de Bruin, J.P., and Nelen, W.L.D.M. 2009. When and why do subfertile couples discontinue their fertility care? A longitudinal cohort study in a secondary care subfertility population. *Hum Reprod* 24: 3127–35.

Bertelsen, L., Mellemkjaer, L., Frederiksen, K., Kjaer, S.K., Brinton, L.A., Sakoda, L.C., et al. 2007. Risk for breast cancer among women with endometriosis. *Int J Cancer* 120: 1372–75.

Brosens, I.A., De Sutter, P., Hamerlynck T, Imeraj, L., Yao, Z., Cloke, B., Brosens, J.J., and Dhont, M. 2007. Endometriosis is associated with a decreased risk of pre-eclampsia. *Hum Reprod* 22: 1725–29.

Buck, G.M., Sever, L.E., Batt, R.E., and Mendola, P. 1997. Review of lifestyle factors and female infertility. *Epidemiology* 8: 435–41.

Buck Louis, G.M. and Cooney, M.A. 2007. Effects of environmental contaminants on ovarian function and fertility. In *Novel concepts in ovarian endocrinology*, pp. 249–68. A. Gonzalez-Bulnes, ed. Kerala, India: Research Signpost.

Buck Louis, G.M., Dmochowski, J., Lynch, C.D., Kostyniak, P.J., McGuinness, B.M., and Vena, J.E. 2009. Polychlorinated biphenyl concentrations, lifestyle and time-to-pregnancy. *Hum Reprod* 24: 451–58.

Bumpass, L. and Lu, H-H. 2000. Trends in cohabitation and implications for children's context in the United States. *Pop Stud* 54: 29–41.

Butts, S.F. and Seifer, D.B. 2010. Racial and ethnic differences in reproductive potential across the life cycle. *Fertil and Steril* 93: 681–90.

Canfield, R.E., O'Connor, J.F., Birken, S., Krichevsky, A., and Wilcox, A.J. 1987. Development of an assay for a biomarker of pregnancy and early fetal loss. *Envir Health Perspect* 74: 57–66.

Chandra, A., Martinez, G.M., Mosher, W.D., M., Abma, J.C., and Jones, J. 2005. Fertility, family planning, and reproductive health of U.S. women: Data from the 2002 National Survey of Family Growth. *Vital Health Stat* 23: 1–60.

Chandra, A. and Stephen, E.H. 1998. Impaired fecundity in the United States: 1982–1995. *Fam Plan Perspect* 30: 34–42.

Chandra, A. and Stephen, E.H. 2006. Declining estimates of infertility in the United States: 1982–2002. *Fertil Steril* 86: 516–23.

Chandra, A. and Stephen, E.H. 2010. Infertility service use among U.S. women: 1995 and 2002. *Fertil Steril* 93: 725–36.

Christiansen, O.B., Nielsen, H.S., Kolte, A., and Pedersen, A.T. 2006. Research methodology and epidemiology of relevance in recurrent pregnancy loss. *Sem Reprod Med* 24: 5–16.

Cole, L.A. 2000. hCG, its free subunits and metabolites in pregnancy and trophoblastic disease. *J Reprod Med* 43: 3–10.

Collins, J.A., Burrows, E.A., and Willan, A.E. 1995. The prognosis for live birth among untreated infertile couples. *Fertil Steril* 64: 22–28.

Cooney, M.A., Buck Louis, G.M., Sundaram, R., McGuiness, B.M., and Lynch, C.D. 2009. Validity of self-reported time to pregnancy. *Epidemiology* 26: 56–59.

Cuckle, H. 2000. Biochemical screening for Down syndrome. *Eur J Obstet Gynecol Reprod Biol* 92: 97–101.

De Boer, E.J., den Tonkelaar, I., Burger, C.W., and van Leeuwen, F.E. 2005. Validity of self-reported causes of subfertility. *Am J Epidemiol* 161: 978–86.

de Bruin, J.P., Dorland, M., Bruinse, H.W., Spliet, W., Nikkels, P.G., and Te Velde, E.R. 1998. Fetal growth retardation as a cause of impaired ovarian development. *Early Hum Dev* 51: 39–46.

Dunaif, A. 1993. Insulin resistance in polycystic ovarian syndrome. *Ann NY Acad Sci* 687: 60–64.

Dunson, D.B., Baird, D.D., and Colombo, B. 2004. Increased infertility with age in men and women. *Obstet Gynecol* 103: 51–56.

Edmonds, D.K., Lindsay, K.S., Miller, J.F., Williamson, E., and Wood, P.J. 1982. Early embryonic mortality in women. *Fertil Steril* 38: 447–53.

Ellish, N.J., Saboda, K., O'Connor, J., Nasca, P.C., Stanek, E.J., and Boyle C. 1996. A prospective study of early pregnancy loss. *Hum Reprod* 11: 406–12.

Erhardt, C.D. 1963. Pregnancy losses in New York City, 1960. *Am J Pub Health* 53: 1337–52.

Eskenazi, B., Gold, E.B., Lasley, B.L., Samuels, S.J., Hammond, K., Wight, S., et al. 1995. Prospective monitoring of early fetal loss and clinical spontaneous abortion among female semiconductor workers. *Am J Indust Med* 28: 833–46.

Evers, J.L.H. 2002. Female subfertility. *Lancet* 360: 151–59.

Farr, S.L., Anderson, J.E., Jamieson, D.J., Warner, L., and Macaluso, M. 2009. Predictors of pregnancy and discontinuation of infertility services among women who received medical help to become pregnant, National Survey of Family Growth, 2002. *Fertil Steril* 91: 988–97.

Fédération, C.E.C.O.S., Schwartz, D., and Mayaux, M.J. 1982. Female fecundity as a function of age: Results of artificial insemination in 2193 nulliparous women with azoospermic husbands. *N Engl J Med* 306: 404–06.

Fedorcsak, P., Storeng, R., Dale, P.O., Tanbo, T., and Abyholm, T. 2000. Obesity is a risk factor for early pregnancy loss after IVF or ICSI. *Acta Obstet Gynecol Scand* 79: 43–48.

Fedorcsak, P., Dale, P.O., Storeng, R., Ertzeid, G., Bjercke, S., Oldereid, N., et al. 2004. Impact of overweight and underweight on assisted reproduction treatment. *Hum Reprod* 19: 2523–28.

French, F.E. and Bierman, J.M. 1962. Probabilities of fetal mortality. *Pub Health Rep* 77: 835–47.

Gaspard, U.J., Reuter, A.M., Deville, J.L., Vrindts-Gevaert, Y., Bagshawe, K.D., and Franchimont, P. 1980. Serum concentration of human chorionic gonadotropin and its alpha and beta subunits. 2. Trophoblastic tumors. *Clin Endocrinol* 13: 319–29.

Gerhard, I. and Runnebaum, B. 1984. Predictive value of hormone determination in the first half of pregnancy. *Eur J Obstet Gynecol Reprod Biol* 17(1): 1–17.

Glueck, C.J., Papanna, R., Wang, P. Goldenberg, N., and Sieve-Smith, L. 2003. Incidence and treatment of metabolic syndrome in newly referred women with confirmed polycystic ovarian syndrome. *Metabolism* 52(7): 908–15.

Gnoth, C., Godehardt, D., Godehardt, E., Frank-Herrmann, P., and Freundl, G. 2003. Time of pregnancy: Results of the German prospective study and impact on the management of infertility. *Hum Reprod* 18: 1959–66.

Guzick, D.S., Grefenstette, I., Baffone, K., Berga, S.L., Krasnow, J.S., et al. 1994. Infertility evaluation in fertile women: A model for assessing the efficacy of infertility testing. *Hum Reprod* 9: 2306–10.

Hakim, R.B., Gray, R.H., and Zacur, H. 1995. Infertility and early pregnancy loss. *Am J Obstet Gynecol* 172: 1510–17.

Harlap, S. and Shiono, P.H. 1980. Alcohol, smoking, and incidence of spontaneous abortion in the first and second trimester. *Lancet* 26: 173–76.

Harville, E.W., Wilcox, A.J., Baird, D.D., and Weinberg, C.R. 2003. Vaginal bleeding in very early pregnancy. *Hum Reprod* 18: 1944–47.

Hassold, T., Abruzzo, M., Adkins, K., Griffin, D., Merrill, M., Millie, E., et al. 1996. Human aneuploidy: Incidence, origin, and etiology. *Environ Mol Mutagen* 28: 167–75.

Hay, D.L. 1988. Placental histology and the production of human choriogonadotropin and its subunits in pregnancy. *Br J Obstet Gynaecol* 95: 1268–75.

Henriksen, T.B., Baird, D.D., Olsen, J., Hedegaard, M., Secher, N.J., and Wilcox, A.J. 1999. Time to pregnancy and preterm delivery. *Obstet Gynecol* 89(4): 595–99.

Henriksen, T.B., Hjollund, N.H., Jensen, T.K., Bonde, J.P., Andersson, A.M., Kolstad, H., et al. 2004. Alcohol consumption at the time of conception and spontaneous abortion. *Am J Epidemiol* 160: 661–67.

Henry, L. 1953. Fondements théoriques des méasures de la fécondité naturelle. *Rev Inst Internat Statist* 21: 135–17.

Hertig, A.T., Rock, J., Adams, E.C., and Menkin, M.C. 1959. Thirty-four fertilized human ova, good, bad and indifferent, recovered from 210 women of known fertility. *Pediatrics* 23: 202–11.

Howards, P.P., Hertz-Picciotto, I., and Poole, C. 2006. Conditions for bias from differential left truncation. *Am J Epidemoi* 165: 444–52.

Hull, M.G., Glazener, C.M., Kelly, N.J., Conway, D.I., Foster, P.A., Hinton, R.A., et al. 1985. Population study of causes, treatment, and outcome of infertility. *Br Med J* 291: 1693–97.

Ibanez, L., Valls, C., Cols, M., Ferrer, A., Marcos, M.V., and De Zegher, F. 2002. Hypersecretion of FSH in infant boys and girls born small for gestational age. *J Clin Endocrinol Metab* 87: 1986–88.

Irgens, H.U., Reisaeter, L., Irgens, L.M., and Lie, R.T. 2001. Long term mortality of mothers and fathers after pre-eclampsia: Population based cohort study. *BMJ* 323: 1213–16.

Joffe, M. and Li, Z. 1994. Association of time to pregnancy and the outcome of pregnancy. *Fertil Steril* 62(1): 71–75.

Joffe, M. 2003. The potential for monitoring of fecundity and the remaining challenges. (Invited Commentary). *Am J Epidemiol* 157: 89–93.

Joffe, M. and Li, Z. 1994. Male and female factors in fertility. *Am J Epidemiol* 140: 921–29.

Jurisicova, A., Antenos, M., Kapasi, K., Meriano, J., and Casper, R.F. 1999. Variability in the expression of trophectodermal markers β-human chorionic gonadotrophin, human leukocyte antigen-G and pregnancy specific β-1 glycoprotein by the human blastocyst. *Hum Reprod* 14: 1852–58.

Juul, S., Karmaus, W., and Olsen, J. 1999. Regional differences in waiting time to pregnancy: Pregnancy-based surveys from Denmark, France, Germany, Italy and Sweden. *Hum Reprod* 14: 1250–54.

Kapiteijn, K., de Bruijn, C.S., de Boer, E., de Craen, A.J.M., Burger, C.W., and van Leeuwen, F.F. 2006. Does subfertility explain the risk of poor perinatal outcome after IVF and ovarian hyperstimulation? *Hum Reprod* 21: 3228–34.

Karmaus, W. and Juul, S. 1999. Infertility and subfecundity in population-based samples from Denmark, Germany, Italy, Poland and Spain. *Eur J Pub Health* 9: 229–35.

Killick, S., Trussell, J., Cleland, K., and Moreau, C. 2009. Factors associated with subfertility among women attending an antenatal clinic in Hull. *Hum Fertil* 12: 191–97.

LaRochebrochard, E. and Thonneau, P. 2003. Paternal age ≥50 years: An important risk factor for infertility. *Am J Obstet Gynecol* 189: 901–05.

Larsen, U. 2000. Primary and secondary infertility in Tanzania. *Int J Epidem* 29: 285–91.

Letterie, H. and Hibbert, M. 2000. Serial serum human chorionic gonadotropin (hCG) levels in ectopic pregnancy and first trimester miscarriage. *Arch Gynecol Obstet* 263: 168–69.

Leridon, H. 2004. Can assisted reproduction technology compensate for the natural decline in fertility with age? A model assessment. *Hum Reprod* 19: 1548–53.

Lohstroh, P.N., Overstreet, J.W., Stewart, D.R., Nakajima, S.T., Cragun, J.R., Boyers, S.P., and Lasley, B.L. 2005. Secretion and excretion of human chorionic gonadotropin during early pregnancy. *Fertil Steril* 83: 1000–11.

Loos, R.J.F., Derom, C., Eeckels, R., Derom, R., and Vlietinck, R. 2001. Length of gestation and birthweight in dizygotic twins. *Lancet* 358: 560–61.

Louis, G.M.B., Lynch, C.D., and Cooney, M.A. 2006. Environmental influences on female fecundity and fertility. *Semin Reprod Med* 24(3): 178–90.

Macklon, N.S., Geraedts, J.P.M., and Fauser, B.C.J.M. 2002. Conception to ongoing pregnancy: The "black box" of early pregnancy loss. *Hum Reprod Update* 8: 333–43.

Marchbanks, P.A., Peterson, H.B., Rubin, G.L., Wingo, P., Cancer and Steroid Hormone Study Group. 1989. Research on infertility: Definition makes a difference. *Am J Epidemiol* 130: 259–67.

Mendola, P. and Buck Louis, G. 2010. Environmental contaminants, female reproductive health and fertility. In *Environmental impacts on reproductive health and fertility*, pp. 161–72. T.J. Woodruff, S. Janssen, L. Guillette, and L. Giudice, eds. Cambridge, UK: Cambridge Press.

Menken, K., Trussell, J., and Larsen, U. 1986. Age and infertility. *Science* 233: 1389–94.

Miller, J.F., Williamson, E., Glue, J., Gordon, Y.B., Grudzinskas, J.G., and Sykes, A. 1980. Fetal loss following implantation: A prospective study. *Lancet* 1: 554–56.

Missmer, S.A., Hankinson, S.E., Spiegleman, D., Barbieri, R.L., Michels, K.B., and Hunter, D.J. 2004. Reproductive history and endometriosis among premenopausal women. *Fertil Steril* 82: 1501–08.

Modvig, J., Schmidt, L., and Damsgaard, M.T. 1990. Measurement of total risk of spontaneous abortion: The virtue of conditional risk estimation. *Am J Epidemiol* 132: 1021–38.

Moller, H. and Skakkebaek, N.E. 1999. Risk of testicular cancer in subfertile men? Case-control study. *Br Med J* 318: 559–62.

Moreau, C., Bouyer, J., Ducot, B., Spira, A., and Slama, R. 2010. When do involuntary infertile couples choose to seek medical help? *Fertil Steril* 93: 737–44.

Morris, N.M. and Udry, J.R. 1967. Daily immunologic pregnancy testing of initially non-pregnant women. *Am J Obstet Gynecol* 15(98): 1148–50.

Nepomnaschy, P.A., Weinberg, C.R., Wilcox, A.J., and Baird, D.D. 2008. Urinary hCG patterns during the week following implantation. *Hum Reprod* 23: 271–77.

Okonofua, F.E. 1996. The case against new reproductive technologies in developing countries. *Br J Obstet Gynaecol* 103: 957–62.

Olsen, J. and Rachootin, P. 2003. Monitoring fecundity over time: If we do it, then let's do it right. (Invited Commentary). *Am J Epidemiol* 157: 94–97.

Olsen, J., Kuppers-Chinnow, M., and Spinelli, A. 1996. Seeking medical help for subfecundity: A study based upon surveys in five European countries. *Fertil Steril* 66: 95–100.

Olsen, J., Bonde, J.P., Hjollund, N.H., Basso, O., and Ernst, E. 2005. Using infertile patients in epidemiologic studies on subfecundity and embryonal loss. *Hum Reprod* 11: 607–11.

Ovalle, F. and Azziz, R. 2002. Insulin resistance, polycystic ovary syndrome, and type 2 diabetes mellitus. *Fertil Steril* 77: 1095–105.

Plachot, M., Grouchy de, J., Junca, A.M., Mandelbaum, J., Turleau, C., Couilin, P., et al. 1988. From oocyte to embryo: A model, deduced from in vitro fertilization, for natural selection against chromosome abnormalities. *Ann Genet* 30: 22–32.

Poikkeus, P., Hiilesmaa, V., and Tiitinen, A. 2002. Serum HCG 12 days after embryo transfer in predicting pregnancy outcome. *Hum Reprod* 17: 1901–05.

Promislow, J.H.E., Baird, D.D., Wilcox, A.J., and Weinberg, C.R. 2007. Bleeding following pregnancy loss before 6 weeks' gestation. *Hum Reprod* 22: 853–57.

Rai, R. and Regan, L. 2006. Recurrent miscarriage. *Lancet* 368: 601–11.

Rao, C.V. 2001. Tropic effects of LH and hCG on early pregnancy events in women's reproductive tract. *Early Pregnancy* 5: 18–19.

Regan, L. and Rai, R. 2000. Epidemiology and the medical causes of miscarriage. *Bailliere's Clin Obstet Gynaecol* 14: 839–54.

Ridout, M.S. and Morgan, B.J.T. 1991. Modelling digit preferences in fecundability studies. *Biometrics* 47: 1423–33.

Roberts, C.J. and Lowe, C.R. 1975. Where have all the conceptions gone? *Lancet* 305: 498–99.

Rostad, B., Schei, B., and Sundby, J. 2006. Fertility in Norwegian women: Results from a population-based health survey. *Scand J Pubic Health* 34: 5–10.

Rull, K. and Laan, M. 2005. Expression of β-subunit of human chorionic gonadotropin genes during the normal and failed pregnancy. *Hum Reprod* 20: 3360–68.

Sallmen, M., Weinberg, C.R., Baird, D.D., Lindbohm, M-L., and Wilcox, A.J. 2005. Has human fertility declined over time? Why we may never know. *Epidemiology* 16: 494–99.

Sayle, A.E., Wilcox, A.J.H., Weinberg, C.R., and Baird, D.D. 2002. A prospective study of the onset of symptoms of pregnancy. *J Clin Epidemiol* 55: 676–80.

Schmidt, L. and Munster, K. 1995. Infertility, involuntary infecundity, and the seeking of medical advice in industrialized countries 1970–1992: A review of concepts, measurements and results. *Hum Reprod* 10: 1407–18.

Sharpe, R.M. and Skakkebaek, N.E. 1993. Are oestrogens involved in falling sperm counts and disorders in the male reproductive tract? *Lancet* 341: 1392–95.

Simon, C., Landeras, J., Zuzuarregui, J.L., Martin, J.C., Remohi, J., and Pellicer, A. 1999. Early pregnancy losses in in vitro fertilization and oocyte donation. *Fertil Steril* 72: 1061–65.

Sinaii, N., Cleary, S.D., Ballweg, M.L., Nieman, L.K., and Stratton, P. 2002. High rates of autoimmune and endocrine disorders, fibromyalgia, chronic fatigue syndrome and atopic diseases among women with endometriosis: A survey analysis. *Hum Reprod* 17: 2715–24.

Skakkebaek, N.E., Rajpert-De Meyts, E., and Main, K.M. 2001. Testicular dysgenesis syndrome: An increasingly common developmental disorder with environmental aspects. *Hum Reprod* 16: 972–78.

Slama, R., Werwatz, A., Boutou, O., Ducot, B., Spira, A., and Härdle, W. 2003. Does male age affect the risk of spontaneous abortion? An approach using semiparametric regression. *Am J Epidemiol* 157: 815–24.

Sneed, M.L., Uhler, M.L., Grotjan, H.E., Rapisarda, J.J., Lederer, K.J., and Beltsos, A.N. 2008. Body mass index: Impact of IVF success appears age-related. *Hum Reprod* 23: 835–39.

Snick, H.K., Snick, T.S., Evers, J.L., and Collins, J.A. 1997. The spontaneous pregnancy prognosis in untreated subfertile couples: The Walcheren primary care study. *Hum Reprod* 12: 1582–88.

Spandorfer, S. and Rosenwaks, Z. 1999. The impact of maternal age and ovarian age on implantation efficacy. In *Embryo implantation: Molecular, cellular and clinical aspects*, pp. 12–19. DD Carson, ed. New York: Springer-Veerlag.

Speert, H. and Guttmacher, A. 1954. Frequency and significance of bleeding in early pregnancy. *JAMA* 155: 712–15.

Srisuparp, S., Strakova, Z., and Fazleabas, A.T. 2001. The role of chorionic gonadotropin (CG) in blastocyst implantation. *Arch Med Res* 32: 627–34.

Stanford, J.B., Mikolajczyk, R.T., Lynch, C.D., and Simonsen, S.E.2010. Cumulative pregnancy probabilities among couples with subfertility: Effects of varying treatments. *Fertil Steril* 93;2175–2181.

Stein, Z., Susser, M., Warburton, D., Wittes, S., and Kline, J. 1975. Spontaneous abortion as a screening device: The effect of fetal survival on the incidence of birth defects. *Am J Epidemiol* 102: 275–90.

Stephen, E.H. and Chandra, A. 2000. Use of infertility services in the United States: 1995. *Fam Plan Perspect* 32: 132–37.

Stephen, E.H. and Chandra, A. 2006. Declining estimates of infertility in the United States: 1982–2002. *Fertil Steril* 86: 516–23.

Steward, D.R., Overstreet, J.W., Nakajima, S.T., and Lasley, B.L. 1993. Enhanced ovarian steroid secretion prior to implantation in early human pregnancy. *J Clin Endocrinol Metab* 76: 1470–76.

Sundby, J., Mboge, R., and Sonko, S. 1998. Infertility in Gambia: Frequency and health care seeking. *Soc Sci Med* 46: 891–99.

Sweeney, A.M., Meyer, M.R., Aarons, J.H., Mills, J.L., and LaPorte, R.E. 1988. Evaluation of methods for the prospective identification of early fetal losses in environmental epidemiology studies. *Am J Epidemiol* 127: 843–50.

Treloar, A.E., Boynton, R.E., Behn, B.G., and Brown, B.W. 1967. Variation of the human menstrual cycle through reproductive life. *Int J Fertil* 12: 77–126.

Vitzhum, V.J., Spielvogel, H., Thornburg, J., and West, B. 2006. A prospective study of early pregnancy loss in humans. *Fertil Steril* 86: 373–79.

Wang, X., Chen, C., Wang, L., Chen, D., Guang, W., and French, J. 2003. Conception, early pregnancy loss, and time to clinical pregnancy: A population-based prospective study. *Fertil Steril* 79: 577–84.

Weidner, I.S., Moller, H., Jensen, T.K., and Skakkebaek, N.E. 1999. Risk factors for cryptorchidism and Hypospadias. *J Urol* 161: 1606–09.

Whittaker, P.G., Taylor, A., and Lind, T. 1983. Unsuspected pregnancy loss in healthy women. *Lancet* 1: 1126–27.

WHO. 1991. Infertility: A tabulation of available data on prevalence of primary and secondary infertility. Programme on Maternal and Child Health and Family Planning, Division of Family Health, World Health Organization. Geneva: Switzerland.

Wilcox, A.J., Weinberg, C.R., Wehmann, R.E., Armstrong, E.G., Canfield, R.E., and Nisula, B.C. 1985. Measuring early pregnancy loss: Laboratory and field methods. *Fertil Steril* 44: 366–74.

Wilcox, A.F., Weinberg, C.R., O'Connor, J., Baird, D., Schlatterer, J., Canfield, R., et al. 1988. Incidence of early pregnancy loss. *New Engl J Med* 319: 189–94.

Wilcox, A.J., Weinberg, C.R., and Baird, D.D. 1990. Risk factors for early pregnancy loss. *Epidemiology* 1: 382–85.

Wilcox, A.J., Baird, D.D., and Weinberg, C.R. 1999. Time of implantation of the conceptus and loss of pregnancy. *New Engl J Med* 340: 1796–99.

Williams, M., Goldman, M.B., Mittendorf, R., and Monson, R. 1991. Subfertility and the risk of low birthweight. *Fertil Steril* 56(4): 668–71.

Winter, E., Wang, J., Davies, M.J., and Norman, R. 2002. Early pregnancy loss following assisted reproductive technology treatment. *Hum Reprod* 17: 3220–23.

Zygmunt, M., Herr, F., Keller-Schoenwetter, S., Kunzi-Rapp, K., Munstedt, K., Rao, C.V., et al. 2002. Characterization of human chorionic gonadotropin as a novel angiogenic factor. *J Clin Endocrinol Metab* 87: 5290–96.

5

Pregnancy, Prenatal Care, Weight, and Maternal Age

MARY L. HEDIGER

Once pregnant, nearly all women enter into prenatal care, so that the course of the pregnancy can be monitored and complications managed. At the time of entry there is a baseline examination and completion or updating of a medical history, including sociodemographic characteristics (age, race/ethnicity), a family and individual medical history, and a reproductive history. In addition to examination for pathology and tests for preexisting infection, this baseline examination almost always entails measurement of *vital signs*, that is, body temperature, pulse rate (or heart rate), blood pressure, respiratory rate, a series of laboratory tests for maternal iron status, and measurement of maternal height and weight, as global indicators of nutritional status. These baseline measurements all have recognized normal courses during pregnancy and need to be evaluated in relation to the timing of assessment in terms of gestation. This chapter addresses a number of areas of concern, such as pregnancy at the extremes of maternal age, weight status, and weight gain, which have long been the subject of epidemiologic investigation and surveillance. Other vital signs and symptoms that are affected by pregnancy (e.g., blood pressure and blood glucose levels) are discussed in Chapter 6.

PRENATAL CARE AND ITS UTILIZATION

Although the efficacy of many of its components are still debated, organized prenatal (or antenatal) care and its adequacy are important factors in pregnancy outcome (Fiscella 1995; Phelan 2008). Inadequate or no prenatal care, however defined, is viewed as a significant risk factor for a poor pregnancy outcome, whereas adequate prenatal care is associated with a minimal risk of poor outcomes. The schedule of prenatal visits suggested by the American College of Obstetricians and Gynecologists (ACOG) for women carrying a single fetus is monthly through the seventh month (28 weeks), every 2 to 3 weeks until the ninth month (36 weeks), and weekly thereafter until delivery (American Academy of Pediatrics and the American College of Obstetricians and Gynecologists 2007).

From entry to prenatal care and through pregnancy, not only is the fetus monitored for adequacy of growth and screened for genetic and congenital anomalies, but gravid women are given a physical examination and routinely checked for changes in their blood pressure, weight gain, signs of genitourinary tract infections, anemia, and blood or urine sugar. Any of these, when abnormal, may affect pregnancy outcome or be signs of an impending complication, and early intervention may prevent a poor outcome. Information on substance use during pregnancy (i.e., cigarette smoking, alcohol intake, prescription and over-the-counter drugs) is also collected, and women are counseled to stop smoking, avoid drinking alcoholic beverages, and restrict drug use during pregnancy unless the drug has been specifically found safe for using during pregnancy and prescribed. At entry to care, gravid women are almost always prescribed specially formulated prenatal vitamins or some combination of vitamins and minerals, including iron and folate.

Several global indices of prenatal care adequacy, based on utilization—that is, the timing of entry to care with respect to the trimester of pregnancy (first, from the date of the last normal menstrual period to 13 completed weeks; second, 14 to 26 completed weeks; third, 27 weeks to delivery) and the number of subsequent prenatal obstetric visits—have been derived, since this information has long been available on the U.S. Standard Certificates of Live Birth or can be readily collected from prenatal care charts and registries (Alexander and Kotelchuck 1996). The earliest of these, the Kessner Index, defined three levels of prenatal care: *adequate, intermediate*, and *inadequate* (Kessner et al. 1973). At the extremes, adequate care was defined as entry during the first or early second trimester, with most subsequently scheduled prenatal visits kept, whereas inadequate care was based on a late entry to care (i.e., third trimester) or few visits following an earlier entry. However, the Kessner Index is somewhat flawed, and in application one of the initial criteria for adequacy—that the pregnancy be delivered by a private obstetric practice—has been largely ignored (Alexander and Kotelchuck 1996). Birth certificates at the time the Kessner Index was proposed limited the number of reportable prenatal visits to 9 (a single data entry column) and thus did not capture the excessive numbers of visits that would be associated with a complicated pregnancy or multiple birth. Consequently, the relation between prenatal care utilization and pregnancy outcome was attenuated by the fact that many complicated pregnancies that required multiple visits in addition to those routinely scheduled and with poor outcomes were grouped into the adequate category.

Subsequent schemes have defined an additional level of care utilization—*intensive* (Alexander and Cornely 1987) or *adequate plus* (Kotelchuck 1994; Alexander and Kotelchuck, 1996), that appears to discriminate between uncomplicated pregnancies with adequate care at low risk for poor outcomes and complicated pregnancies that have been followed frequently and intensively because of a higher risk or complications. Although all the indices have comparable associations with fetal growth and small-for-gestational age (SGA) outcomes, they differ with respect to their associations with preterm birth or infant mortality and should not be used interchangeably (VanderWeele et al. 2009).

MATERNAL WEIGHT AND WEIGHT GAIN DURING PREGNANCY

The measurements of maternal weight and weight gain during pregnancy are routine components of prenatal care, and weight and weight gain are frequently used as proxies for maternal nutritional status and as an indicator of fetal well-being. Despite serious and persistent concerns about the benefit of weight gain as a clinical indicator (Hytten 1990; Dawes and Grudzinskas 1991; Dawes et al. 1992; Farrar and Duley 2007), the weight of a pregnant women and her gain between prenatal visits are one of the first indicators of maternal nutritional status assessed, evaluated, and discussed at every prenatal visit. Clinicians use maternal weight gain in the assessment of maternal nutritional status, to estimate risk of pregnancy complications, and as an indicator of fetal growth (Institute of Medicine 1990, 2009).

Historical perspectives

The measurement of maternal weight gain attained importance in prenatal care early in the last century for reasons more related to the health of the mother than to the growth of the fetus. Maternal gestational weight gain was monitored carefully throughout prenatal care because excessive weight gain (defined as more than 1 lb/week) in the last two trimesters was viewed as a clinical sign of impending preeclampsia and because excessive gains were related to larger infant size, more difficult delivery, and an increased risk for operative delivery (forceps or cesarean). From the 1950s through the 1960s, a tight restriction on gestational weight gain was imposed, with the targeted upper limit being set at around 20 lb (9.1 kg) for mature women. However, the National Academy of Sciences concluded in their 1970 study that perinatal mortality rates were actually lowest at weight gains of 24–27 lb (11.0–12.3 kg), thereby beginning to refocus weight gain considerations away from the mother and on to the growing fetus (National Research Council 1970).

In the 1980s, new emphasis was placed on low or inadequate weight gains and dietary deficiencies, and their consequences for poor fetal growth and pregnancy outcome, and in this context the Committee on Nutritional Status During Pregnancy and Lactation of the Institute of Medicine (IOM) undertook to revise recommendations for weight gain during pregnancy. The first IOM report in 1990 was the benchmark for subsequent recommendations and research until the release of the new report reexamining the guidelines in 2009.

The 1990 Institute of Medicine findings and recommendations

The central recommendation of the IOM with regard to weight gain is that appropriate weight gain during pregnancy (Table 5.1) should be encouraged (by dietary intervention and counseling), especially among high-risk groups, such as young adolescents (within 2 years of menarche), black women, women with low pregravid weight or short stature (<157 cm), and mothers of twins and higher-order multiples, to enhance fetal growth and diminish the risk of low birth weight (LBW, <2,500 g) and perinatal morbidity (Institute of Medicine 1990).

Table 5.1 Comparisons of Institute of Medicine 1990 and 2009 recommendations for weight gain during pregnancy

	1990 Recommendations				*2009 Recommendations*		
	BMI range (kg/m²)	*Target gain (kg) at term*	*Weight gain trimester 1 (kg)*	*Rates of weight gain trimesters 2 and 3 (kg/wk)*	*BMI range (kg/m²)*	*Target gain (kg) at term*	*Rates* of weight gain trimesters 2 and 3 (kg/wk)*
Underweight	<19.8	12.5–18.0	2.3	0.49	<18.5	12.5–18.0	0.51
Normal weight	19.8–26.0	11.5–16.0	1.6	0.44	18.5–24.9	11.5–16.0	0.42
Overweight	26.1–29.0	7.0–11.5	0.9	0.30	25.0–29.9	7.0–11.5	0.28
Obese	>29.0	≥6.8	—	—	≥30.0	5.0–9.0	0.22

*Assumes a 0.5–2.0 kg weight gain during the first trimester.

Sources: Institute of Medicine, Committee on Nutritional Status During Pregnancy and Lactation, Subcommittee on Nutritional Status and Weight Gain During Pregnancy. 1990. Nutrition during pregnancy. Washington, DC: National Academy Press.

Institute of Medicine, Committee to Reexamine IOM Pregnancy Weight Guidelines, Rasmussen KM, Yaktine AL (eds.). 2009. Weight gain during pregnancy: reexamining the guidelines. Washington, DC: National Academy Press.

A unique feature of the IOM recommendations was that the recommended target weight gain ranges were specific to maternal pregravid weight status, using the body mass index (BMI, kg/m^2) as the preferred index of maternal pregravid nutritional status and provisional cutoffs to define underweight, overweight, and obesity. Body mass index is a better indicator of nutritional status than weight alone. In an exhaustive review of the literature to that point, the IOM established that, in both developed and developing countries and for all racial/ethnic groups: (a) a positive relation exists between weight gain and birth weight, with correlations between birth weight and total weight gain ranging between 0.20 and 0.30; (b) maternal pregravid weight or BMI (kg/m^2) and weight gain have independent and additive effects on birth weight outcome; (c) the average magnitude of the effect on birth weight (in women with a normal BMI) is, assuming a base birth weight of about 3,000 g, approximately 20 g of birth weight for every 1 kg of total gain (Abrams and Laros 1986); and (d) pregravid BMI is a strong effect modifier. That is, the effect of gestational weight gain is modified by maternal pregravid BMI; the impact of a given weight gain is greatest in thin women and least in the overweight and obese.

The IOM also concluded that low gestational weight gain is strongly associated with an increased risk for LBW and fetal growth restriction (FGR). The IOM committee relied heavily in their deliberations on original analyses of data from the 1980 National Natality Survey (Kleinman 1990), which demonstrated a greater than twofold risk of term LBW with a low total weight gain (≤10 kg) for both underweight and normal weight women, whereas the relation was attenuated at this level of weight gain for overweight women.

Despite the strong conclusions about the relation between total weight gain and low birth weight, there were other areas where the literature was undeveloped, and the IOM called for additional research (Institute of Medicine 1990; Lederman

1993). These included whether weight gain, particularly low weight gain, was causally related to preterm delivery, the effects of differing patterns of gain on birth weight, maternal body composition changes and their effects on birth weight, and appropriate recommendations for nutritionally vulnerable groups (underweight and obese women, black women, adolescents). Although the recommendations did not address the issues of excessive weight gain, maternal obesity, and their relation with large-for-gestational age births (LGA, ≥90th percentile), cesarean delivery, and maternal postpartum weight retention, these concerns were raised soon after (Johnson et al. 1992; Johnson and Yancey 1996).

Evaluations of the Institute of Medicine recommendations and reevaluation

Since the IOM report was issued, a number of studies have directly evaluated various aspects of the weight gain recommendations and used the originally suggested BMI categories as part of their evaluative criteria (Siega-Riz et al. 2009). The primary outcomes of interest have been small-for-gestational age births (SGA, <10th percentile birth weight-for-gestational age) or LBW (<2,500g), which are increased with weight gain below the IOM recommended ranges; and LGA births, cesarean delivery, and maternal postpartum weight retention, which are increased with weight gain above or in excess of the IOM range. These results appeared to validate the IOM recommendations as reasonable targets for total weight gain or at least show that pregnancy weight gain within the recommended ranges is associated with the best outcomes (Abrams et al. 2000).

However, weight gain in most pregnant women is not within the IOM recommended ranges (Chu et al. 2009). A series of studies (Caulfield et al. 1996, 1998) on 3,870 singleton pregnancies to white and black women from Baltimore examined total gain above and below recommended ranges by BMI categories in relation to SGA and LGA outcomes. Only 28% of black and 33% of white women were found to have weight gains within the recommended ranges (Caulfield et al. 1996). Among underweight women, each 50 g/week gained was associated with a 18% reduction in risk of SGA, whereas among overweight and obese women each 50 g/week gained was associated with only a 5% reduction (Caulfield et al. 1998). On the other hand, the risk of LGA was increased by 22% for every additional 50 g/week gained for underweight women, but only by 12% among the overweight and obese. Although black women were at increased risk for SGA and decreased risk for LGA overall, there were no differences in the effect of weight gain by race.

Based on this and other evidence (Siega-Riz et al. 2009), the IOM recommended that the guidelines for weight gain and weight gain rates be adjusted to the definitions of BMI status (Table 5.1) currently used by the World Health Organization and the National Heart, Lung, and Blood Institute (1998). The IOM committee also suggested strongly that women be counseled to avoid excessive weight gain to minimize postpartum weight retention, and it set guidelines for weight gain in obese gravidas (5.0–9.0 kg). Further, the report concluded that there was insufficient evidence to support modifications of the recommendations for women of short stature (<157 cm), adolescents (<20 years old), and different

racial/ethnic groups (Institute of Medicine 2009). For the first time, the IOM also endorsed provisional weight gain guidelines for women carrying twins (reestimated from Luke et al. 2003) based on the newer BMI cutoffs and interquartile ranges for a series of twin pregnancies that delivered between 37–42 weeks' gestation and where the twins weighed 2,500 g or more on average: normal weight, 17–25 kg (37–54 lb); overweight, 14–23 kg (31–50 lb); and obese, 11–19 kg (25–42 lb).

Weight gain and preterm delivery

Although more controversial, most studies on adults conducted subsequent to the 1990 IOM report in response to the need for additional research have shown that, although total weight gain is an important predictor of fetal size and birth weight, low rates of gain and to some extent the pattern of gain appear significant in predicting preterm delivery (Carmichael and Abrams 1997; Schieve et al. 2000; Dietz et al. 2006; Stotland et al. 2006). Low rates of weight gain in the last half of pregnancy were also found to be particularly predictive of preterm delivery in adolescent gravidas, for whom weight gain is generally higher and compensates for a lower pregravid BMI (Hediger et al. 1990). One of the earliest studies relating weight gain late in pregnancy to preterm delivery was of 1,790 adolescent pregnancies from Camden County, New Jersey (Hediger et al. 1989). Although a low early weight gain (<4.3 kg by 24 weeks' gestation) was associated with an increased risk of having an SGA infant, preterm delivery at <37 completed weeks' gestation was associated instead with low weight gain rates (<0.40 kg/week) after 24 weeks' gestation. There was more than a 50% increased risk of preterm delivery when weight gain rates were less than 0.40 kg/week after 24 weeks, and the risk was more than 2.5-fold when weight gain was low both before 24 weeks (<4.3 kg) and the rates of gain were low after 24 weeks. Further, the low rates after 24 weeks were associated with a 70% increased risk of preterm delivery even when the total pregnancy weight gain was within targets set in clinical standards (Hediger et al. 1989).

Studies of adults have confirmed that low rates of weight gain, either in the latter half of pregnancy, the second, and/or third trimesters, are associated with preterm delivery, although the identified low rates of gain that are associated with preterm delivery are much lower than those for adolescents: <0.24 kg/week after 20 weeks' gestation (Wen et al. 1990); <0.38 kg/week with a pregravid BMI of <19.8 kg/m^2, <0.37 kg/week with a pregravid BMI of 19.8–26.0 kg/m^2 (Hickey et al. 1995); <0.34 kg/week for underweight, <0.35 kg/week for normal, and <0.30 kg/week for overweight and obese women in the third trimester (Siega-Riz et al. 1996); and <0.23 kg/week between 14 and 28 weeks' gestation. Very low rates of weight gain based on total weight gain and the average rate across an entire pregnancy are also associated with preterm delivery in epidemiologic studies (Dietz et al. 2006; Stotland et al. 2006), although these findings fail to be useful for clinical purposes because the cutoffs established are not specific to pregnancy trimesters.

That low rates of weight gain late in pregnancy appear now firmly associated with preterm delivery does not mean that these low rates are causal. They may reflect

indirectly maternal nutrition (Carmichael and Abrams 1997). Poor nutrition in the first and early second trimester may affect placental development and vascularization, leading to fetal growth restriction (FGR). Early poor nutrition, whether characterized or not by low weight gain, might also affect the integrity of the chorioamnionic membranes, leading to preterm premature rupture of membranes (PPROM) (Allen 1991) or increased susceptibility to vaginal or urinary tract infection. For example, studies of zinc levels have found that low zinc intakes are related to an increased risk of preterm delivery (Scholl et al. 1993), and the risk of preterm delivery with low dietary zinc intake was nearly tripled for those with PPROM. Zinc has well-known antiseptic effects, so that low levels of intake might indicate an increased susceptibility to infection. On the other hand, the low rates of gestational gain may also reflect a slowed fetal growth associated with pregnancy complications and preterm delivery and be a marker for impending preterm birth.

BODY COMPOSITION, PHYSIOLOGICAL CHANGES, AND PREGNANCY OUTCOME

Drawing inferences from weight gain during pregnancy is complicated because weight gain is a composite measure (Pitkin 1976; Lederman 1993; Scholl and Hediger 1995). Maternal weight gain represents both the products of conception, including the fetus, placenta, and amniotic fluid, and maternal tissue and fluid accretion. Over half of the total gain at term is accounted for by expansion of maternal tissue, including increases in uterine and breast tissue, an expansion of maternal blood volume (Hytten 1985), extracellular fluid, and maternal fat stores (Pitkin 1976). By the beginning the third trimester, expansions in maternal tissue account for nearly 70% of total gain, with a significant proportion attributable to the enlargement of the uterus and breasts and the expansion of maternal blood volume.

In the third trimester, from about 30 weeks' gestation to term, not much additional gain is attributed to the maternal compartment; gains in the third trimester reflect primarily the growth of the fetus and increases in amniotic fluid volume. At term, the products of conception account for about half of the total weight gain; the fetus alone increases from less than 20% to nearly 30% of the total at term. In general, most of the weight gained during the first and second trimesters is related to expansion of maternal tissue, whereas gains in the third trimester involve the fetal compartment.

There is every reason to believe that the early expansion of maternal tissue, particularly plasma volume expansion (Hytten 1985; Duvekot et al. 1995), and changes in maternal body composition during pregnancy are critical for fetal growth. This may be why early weight gain is important. For example, in the Camden Study, although low weight gains (<4.3 kg by 24 weeks) both early and later in pregnancy were independently associated with an increased risk of SGA births in all adolescents, when the sample was restricted to those adolescents whose total weight gain at delivery fell within the optimal ranges defined in clinical references, the association of SGA births with early weight gain was only slightly attenuated (Hediger et al. 1989). Thus, even when later weight gains are compensatory and bring

the total gain to within recommended ranges, low early weight gains were independently associated with SGA. The reason for this association is probably found in the pattern of maternal body composition changes and volume expansion that accompany pregnancy.

Only a few smaller clinical studies have used techniques of body composition analysis (deuterium dilution, hydrodensitometry, dual energy x-ray absorptiometry) to examine maternal body composition and composition changes in relation to fetal growth and infant birth weight (Lederman et al. 1997; Mardones-Santander et al. 1998; Lederman et al. 1999; Butte et al. 2003) expanding on earlier studies that have relied on less precise field techniques (skinfolds, anthropometry, bioelectric impedance analysis) (Taggart et al. 1967; Langhoff-Roos et al. 1987; Villar et al. 1992; Hediger et al. 1994). Using a deuterium dilution techniques, Mardones-Santander et al. (1998) estimated maternal fat-free mass and fat mass for 224 low-income, adult women from Santiago, Chile, near term (34–40 weeks' gestation) and found that maternal fat-free mass was the most important variable predicting birth weight. Although maternal fat mass near term was significant in predicting birth weight, it accounted for only 5% of the variance.

Using a similar technique, but longitudinally, Lederman et al. (1997, 1999) looked at changes in body composition at 14 weeks' gestation and 37 weeks in predicting infant birth weight for a sample of 200 well-nourished, adult women in New York City. Total body water, representing nonfat tissue, was measured directly at both points in time by deuterium dilution, and body fat was estimated using a multicompartment model that included the measures of total body water and body density determined by hydrodensitometry. Although estimated net maternal body fat (subtracting out fetal size) was not associated with birth weight, each additional liter of net maternal body water at term was associated with a 23 g increase in birth weight. Among well-nourished women who began pregnancy with adequate fat reserves, expansion of nonfat tissue as indexed by total body water was the most important component of maternal weight gain predicting birth weight.

Likewise, Butte et al. (2003) studied the body composition of 63 women (low BMI, 17 women; normal BMI, 34 women; and high BMI, 12 women) using measurements of total body nitrogen by prompt-γ activation analysis, total body potassium by whole body counting, and a multicompartment based on total body water by deuterium dilution, body volume by densitometry, and bone mineral content by dual energy x-ray absorptiometry (DXA) before pregnancy, and at 9, 22, and 36 weeks' gestation. Gestational weight gain was found to correlate significantly with gains in total body water, total body potassium, protein, fat-free mass, and fat mass. And, although gains in total body water, total body potassium, protein, and fat-free mass did not differ among BMI groups, fat mass gain was higher in the high BMI group (Butte et al. 2003). Confirming the findings of Lederman et al. (1997, 1999), birth weight was correlated positively with gain in total body water, total body potassium, and fat-free mass, but not fat mass.

These changes in body composition are consistent with patterns of weight gain and how the patterns are related to birth weight. The studies that have looked at patterns of weight gain by trimester have found that weight gain rates are highest in the second trimester (14–26 weeks), when maximal gains made in the maternal

compartments are most predictive of birth weight (Abrams and Selvin 1995; Carmichael et al. 1997).

PREGNANCY AT THE EXTREMES OF MATERNAL AGE

Pregnancy at the extremes of maternal age presents special circumstances, whether chronological age is considered as an independent risk factor, risk indicator, effect modifier, or confounder.

Adolescent pregnancy

It is well recognized that pregnancy in adolescence is a particular concern, especially in the United States, where rates are the highest of all developed countries. Most studies have focused on the fact that teenage pregnancy is more common among disadvantaged and/or minority adolescents (African-Americans, Puerto Ricans, Mexican-Americans), and the pregnancies are at risk because of factors, such as lack of prenatal care and inadequate nutrition, that are more common among the socially disadvantaged. For 337 pregnant adolescents in Galveston, Texas, the risk of low weight gain (<20 lb) and, subsequently, lower birth weight was increased more than twofold when the adolescent had a sexually transmitted disease, more than fivefold when she was battered during the pregnancy, or more than eightfold if the pregnancy was unplanned (Berenson et al. 1997). On the other hand, especially in developed countries, even minority adolescents are at a gynecologic and obstetric advantage because they are generally healthier, with less chronic disease and less likely to be obese (Perry et al. 1996).

The other unique aspect of adolescent pregnancy—the fact that teenagers may be biologically immature—has been less well studied. This is because there are no reliable markers for maternal growth and development during adolescent pregnancy that do not involve specialized techniques of assessment (Scholl and Hediger 1993; Scholl et al. 1994) that have been difficult to replicate (Jones et al. 2010). Most, although not all, growth in stature has been accomplished by the time of menarche (Roche and Davila 1972), but continued normal growth and body composition changes (development) occur in late adolescence that are important for pregnancy and may be exaggerated with pregnancy, with adverse long-term consequences. That is, girls post-menarche have normal gains in lean body or fat-free mass, a tendency to accrue subcutaneous fat centrally and to accrue visceral fat (Guo et al. 1990; Hediger et al. 1995), and there is continued growth and development of the reproductive organs, particularly the uterus (Da Costa et al. 2004). With weight gain during pregnancy, and especially the excessive weight gain common for pregnant adolescents, the long-term cardiovascular and metabolic disease risk profile for teenagers will be increased. A longitudinal study of girls from 10–11 through 18–19 years showed that both primiparous and multiparous adolescents at 18–19 years of age had disproportionately increased waist-to-hip ratios, evidence of central adiposity (Gunderson et al. 2009). It also means that pregnancy during adolescence may be complicated by the immature

uterus having less expansile capacity and, thereby, being prone to contractions with distension (Lockwood and Kuczynski 1999).

In fact, the major hallmark of adolescent pregnancy is its nearly universal association with preterm labor and delivery, an association which is strongest in those teenagers who are less gynecologically mature. Gynecologic maturity is conventionally indexed as gynecologic age, defined as number of years since menarche. A low gynecologic age is defined as being within 2 completed years of menarche and is primarily studied in relation to adolescent or teenage pregnancy. A study of 366 minority adolescents (56% African-American, 35% Puerto Rican, 9% white) who were less than 16 years of age at the beginning of pregnancy, compared with 239 women who were 18–29 years, showed an overall twofold increase in preterm delivery with preterm labor with young age, whereas there was a modest decreased risk of preterm delivery with other causes (premature rupture of the chorionic membranes, medical indications) with young age (Hediger et al. 1997). Among the 36% of the young teenagers with low gynecologic age (within 2 completed years of menarche), the risk of preterm delivery with preterm labor was highest, at nearly threefold.

National, cohort, and clinical studies from the United States (Chen et al. 2007), elsewhere in the Western hemisphere (Canada, Jamaica, Brazil), Western Europe, Africa (Nigeria, Ethiopia), the Middle East (Turkey, Jordan, Saudi Arabia), and Asia (India, Nepal, Hong Kong, Singapore, Taiwan, Thailand), to name just some, have all reported an association of teenage pregnancy with preterm delivery. That the association may be biological, and not just because of social disadvantage, is further evidenced by the fact that among Utah teenagers 17 years of age and younger in better social circumstances (white, married, adequate prenatal care, appropriate education for age), the relative risk of preterm delivery was nearly doubled compared with similar women 20–24 years of age (Fraser et al. 1995). The relative risk of LBW 2,500 g) was also significantly increased, but to a lesser extent. As further evidence that the effect of young age may be to shorten gestation as opposed to reducing fetal growth, Salihu and colleagues (2006) reported for the United States that the effect of young maternal age (<20 years) was actually significantly protective for stillbirth, once they accounted for preterm delivery.

Advanced maternal age

At the other end of the reproductive age spectrum are women of *advanced* maternal age, generally defined as 35 years of age or older, among whom the rates of pregnancy have been increasing steadily in the last several decades. Healthy women in their late 30s and early 40s have no particular increased risks in the delivery of singletons, but the risk of adverse pregnancy outcomes at older maternal age is higher because older women are more likely to be already obese and/or have chronic disease, including type 2 diabetes mellitus, chronic hypertension, renal and thyroid disease, autoimmune disorders, and neurological and psychiatric disorders, all of which are associated with poor pregnancy outcomes (Cleary-Goldman et al. 2005; Newburn-Cook and Onyskiw 2005). In addition, older

women are at higher risk for obstetric complications, including pregnancy-induced hypertension, gestational diabetes mellitus, and cesarean deliveries.

This cumulative increase in obesity and chronic diseases among women as they age has been called *weathering* (Geronimus 1992) and is related to racial disparities in health. Minority and disadvantaged women, especially African-Americans, have an earlier onset of obesity and chronic disease and carry a disproportionate burden through adulthood (Geronimus et al. 2006). In terms of adverse pregnancy outcomes (e.g., preterm delivery), the racial/ethnic disparity widens with maternal age (Holzman et al. 2009). This has led to the still-controversial hypothesis that, despite the social implications, it is biologically adaptive for U.S. minority women to bear children at younger ages, when the pregnancies should be less complicated and the outcomes optimal.

The other major obstetric concern that increases with maternal age is twinning. Since the 1980s, the rates of twinning have nearly doubled. It has been estimated that 70%–75% of the increase is attributable to infertility treatment and the use of assisted reproductive technologies, but somewhere between 25% and 30% is attributable to an increase in the age at which women are conceiving. Although there is some increase in monozygotic twinning with age, by far most of the increase is in dizygotic, dichorionic twins, where the course and outcomes are better even for women of advanced maternal age (Fox et al. 2009). Given this observation, the most likely pathway accounting for an increase in twinning with maternal age involves increases in follicle-stimulating hormone (FSH) with aging, in response to diminishing ovarian reserve. In susceptible women, the ovarian stimulation will result in multiple ovulation.

PREGNANCY AND OBESITY

Obese women are known to be at increased risk for a number of pregnancy complications, including infection, gestational diabetes, pregnancy-induced hypertension, cesarean delivery, and postpartum morbidity (Gunderson 2009), and they make more frequent use of health care services during pregnancy (Chu et al. 2008). Obese women have infants at increased risk for birth defects, open neural tube defects and congenital heart disease—in particular (Stothard et al. 2009; Biggio et al. 2010), with larger birth weights and both LGA births and macrosomia. Overly large infants are, in turn, at risk for delivery complications (such as shoulder dystocia) and postpartum morbidity. Long-term, infants born to obese women are more likely themselves to be overweight, obese, and subject to the complications of excess weight (Oken 2009).

Whether these complications and outcomes are also independently related to or exacerbated by gestational weight gain is still unclear, and controversies still surround whether obese gravidas should be counseled to lose weight exclusively before pregnancy to improve reproductive functioning and reduce risk of complications, or continue weight loss during pregnancy. It is also still not known whether the same weight gain recommendation of 5.0–9.0 kg at term (Table 5.1) is appropriate for all classes of obese women, especially the morbidly obese (Institute of Medicine 2009;

Artel et al. 2010). A growing number of morbidly obese women are undergoing bariatric surgery for weight loss before trying to conceive, and the long-term consequences of this procedure are largely unknown (Wax 2009).

CONCLUSION

There has been a demonstrable shift over the past several decades in the characteristics of women giving birth in the United States, as well as an increase in pregnancy complications, such as preeclampsia and gestational diabetes, that are likely be associated with increased rates of adverse infant outcomes, including preterm delivery and FGR. The birth rates for teenagers, which had for nearly 15 years been declining, started to rise again in 2007, especially among older teenagers (Hamilton et al. 2009). More women over 35–40 years of age ("advanced" maternal age) and women who are less optimally healthy (i.e., with preexisting obesity, hypertension, diabetes, or metabolic syndrome) are giving birth for the first time (Cleary-Goldman et al. 2005; Newburn-Cook and Onyskiw 2005). There are more obese women of reproductive age, and more of them are giving birth. As the characteristics of reproductively active women continue to change, research must move forward in identifying optimal maternal weight and nutritional status for ensuring good pregnancy outcomes and long-term maternal health. Not only the optimal total amount of weight that should be gained, but also the optimal pattern of weight gain, rates of gain, and composition need to be identified.

REFERENCES

Abrams, B., Altman, S.L., and Pickett, K.E. 2000. Pregnancy weight gain: still controversial. *Am J Clin Nutr* 71(5 Suppl):1233S–41S.

Abrams, B.F. and Laros, R.K. 1986. Prepregnancy weight, weight gain, and birth weight. *Am J Obstet Gynecol* 154:503–509.

Abrams, B. and Selvin, S. 1995. Maternal weight gain pattern and birth weight. *Obstet Gynecol* 86:163–69.

Alexander, G.R. and Cornely, D.A. 1987. Prenatal care utilization: its measurement and relationship to pregnancy outcome. *Am J Prev Med* 3:243–53.

Alexander, G.R. and Kotelchuck, M. 1996. Quantifying the adequacy of prenatal care: a comparison of indices. *Public Health Rep* 111:408–18.

Allen, S.R. 1991. Epidemiology of premature rupture of the fetal membranes. *Clin Obstet Gynecol* 34:685–93.

American Academy of Pediatrics and the American College of Obstetricians and Gynecologists. 2007. *Guidelines for perinatal care*, 6th ed. Elk Grove Village, IL: American Academy of Pediatrics, American College of Obstetricians and Gynecologists.

Artal, R., Lockwood, C.J., and Brown H.L. 2010. Weight gain recommendations in pregnancy and the obesity epidemic. *Obstet Gynecol* 115:152–55.

Biggio, J.R. Jr., Chapman, V., Neely, C., Cliver, S.P., and Rouse, D.J. 2010. Fetal anomalies in obese women: the contribution of diabetes. *Obstet Gynecol* 115(2 pt 1):290–96.

Berenson, A.B., Wiemann, C.M., Rowe, T.F., and Rickert, V.I. 1997. Inadequate weight gain among pregnant adolescents: risk factors and relationship to infant birth weight. *Am J Obstet Gynecol* 178:1220–27.

Butte, N.F., Ellis, K.J., Wong, W.W., Hopkinson, J.M., and Smith, E.O. 2003. Composition of gestational weight gain impacts maternal fat retention and infant birth weight. *Am J Obstet Gynecol* 189:1423–32.

Carmichael, S.L. and Abrams, B. 1997. A critical review of the relationship between gestational weight gain and preterm delivery. *Obstet Gynecol* 89:865–73.

Carmichael, S., Abrams, B., and Selvin, S. 1997. The pattern of maternal weight gain in women with good pregnancy outcomes. *Am J Public Health* 87:1984–88.

Caulfield, L.E., Witter, F.R., and Stoltzfus, R.J. 1996. Determinants of gestational weight gain outside the recommended ranges among black and white women. *Obstet Gynecol* 87:760–66.

Caulfield, L.E., Stoltzfus, R.J., and Witter, F.R. 1998. Implications of the Institute of Medicine weight gain recommendations for preventing adverse pregnancy outcomes in black and white women. *Am J Pub Health* 88:1168–74.

Chen, X.K., Wen, S.W., Fleming, N., Demissie, K., Rhoads, G.G., and Walker, M. 2007. Teenage pregnancy and adverse birth outcomes: a large population based retrospective cohort study. *Int J Epidemiol* 36:368–73.

Chu, S.Y., Bachman, D.J., Callaghan, W.M., Whitlock, E.P., Dietz, P.M., Berg, C.J., et al. 2008. Association between obesity during pregnancy and increased use of health care. *N Engl J Med* 358:1444–53.

Chu, S.Y., Callaghan, W.M., Bish, C.L., and D'Angelo, D. 2009. Gestational weight gain by body mass index among US women delivering live births, 2004–2005: fueling future obesity. *Am J Obstet Gynecol* 200:271.e1–7.

Cleary-Goldman, J., Malone, F.D., Vidaver, J., Ball, R.H., Nyberg, D.A., Comstock, C.H., et al., for the FASTER Consortium. 2005. Impact of maternal age on obstetric outcome. *Obstet Gynecol* 105:983–90.

Da Costa, A.G., Filho, F.M., Ferreira, A.C., Spara, P., and Mauad, F.M. 2004. Uterine volume in adolescents. *Ultrasound Med Biol* 30:7–10.

Dawes, M.G., Green, J., and Ashurst, H. 1992. Routine weighing in pregnancy. *Br Med J* 304:487–89.

Dawes, M.G. and Grudzinskas, J.G. 1991. Repeated measurement of maternal weight during pregnancy. Is this a useful practice? *Br J Obstet Gynaecol* 98:189–94.

Dietz, P.M., Callaghan, W.M., Cogswell, M.E., Morrow, B., Ferre, C., and Schieve, L.A. 2006. Combined effects of prepregnancy body mass index and weight gain during pregnancy on the risk of preterm delivery. *Epidemiology* 17:170–77.

Duvekot, J.J., Cheriex, E.C., Pieters, F.A.A., Menheere, P.P.C.A., Schouten, H.J.A., and Peeters, L.L.H. 1995. Maternal volume homeostasis in early pregnancy in relation to fetal growth restriction. *Obstet Gynecol* 85:361–67.

Farrar, D. and Duley, L. 2007. Commentary: But why should women be weighed routinely during pregnancy? *Int J Epidemiol* 36:1283–84.

Fiscella, K. 1995. Does prenatal care improve birth outcomes? A critical review. *Obstet Gynecol* 85:468–79.

Fox, N.S., Rebarber, A., Dunham, S.M., and Saltzman, D.H. 2009. Outcomes of multiple gestations with advanced maternal age. *J Matern Fetal Neonatal Med* 22:593–96.

Fraser, A.M., Brockert, J.E., and Ward, R.H. 1995. Association of young maternal age with adverse reproductive outcomes. *N Engl J Med* 332:1113–17.

Geronimus, A.T. 1992. The weathering hypothesis and the health of African-American women and infants: evidence and speculations. *Ethn Dis* 2:207–21.

Geronimus, A.T., Hicken, M., Keene, D., and Bound, J. 2006. "Weathering" and age patterns of allostatic load scores among blacks and whites in the United States. *Am J Public Health* 96:826–33.

Guo, S.S., Chumlea, W.C., Roche, A.F., and Siervogel, R.M. 1990. Age- and maturity-related changes in body composition during adolescence into adulthood: The Fels Longitudinal Study. *Int J Obesity* 21:1167–75.

Gunderson, E.P. 2009. Childbearing and obesity in women: weight before, during, and after pregnancy. *Obstet Gynecol Clin N Am* 36:317–32.

Gunderson, E.P., Striegel-Moore, R., Schreiber, G., Hudes, M., Biro, F., Daniels, S., and Crawford, P.B. 2009. Longitudinal study of growth and adiposity in parous compared with nulligravid adolescents. *Arch Pediatr Adolesc Med* 163:349–56.

Hamilton, B.E., Martin, J.A., and Ventura, S.J. 2009. *Births: preliminary data for 2007. National vital statistics reports*, web release; vol. 57 no 12. Hyattsville, MD: National Center for Health Statistics. Released March 18.

Hediger, M.L., Scholl, T.O., Ances, I.G., Belskym D.H., and Salmon, R.W. 1990. Rate and amount of weight gain during adolescent pregnancy: associations with maternal weight-for-height and birth weight. *Am J Clin Nutr* 52:793–99.

Hediger, M.L., Scholl, T.O., Belsky, D.H., Ances, I.G., and Salmon, R.W. 1989. Patterns of weight gain in adolescent pregnancy: effects on birth weight and preterm delivery. *Obstet Gynecol* 74:6–12.

Hediger, M.L., Scholl, T.O., Schall, J.I., and Cronk, C.E. 1995. One-year changes in weight and fatness in girls during late adolescence. *Pediatrics* 96:253–58.

Hediger, M.L., Scholl, T.O., Schall, J.I., Healey, M.F., and Fischer, R.L. 1994. Changes in maternal upper arm fat stores are predictors of variation in infant birth weight. *J Nutr* 124:24–30.

Hediger, M.L., Scholl, T.O., Schall, J.I., and Krueger, P.M. 1997. Young maternal age and preterm labor. *Ann Epidemiol* 7:400–406.

Hickey, C.A., Cliver, S.P., McNeal, S.F., Hoffman, H.J., and Goldenberg, R.L. 1995. Prenatal weight gain patterns and spontaneous preterm birth among nonobese black and white women. *Obstet Gynecol* 85:909–14.

Holzman, C., Eyster, J., Kleyn, M., Messer, L.C., Kaufman, J.S., Laraia, B.A., et al. 2009. Maternal weathering and risk of preterm delivery. *Am J Public Health* 99:1864–71.

Hytten, F. 1985. Blood volume changes in normal pregnancy. *Clin Haematol* 14:601–12.

Hytten, F. 1990. Is it important or even useful to measure weight gain in pregnancy? *Midwifery* 6:28–32.

Institute of Medicine, Committee on Nutritional Status During Pregnancy and Lactation, Subcommittee on Nutritional Status and Weight Gain During Pregnancy. 1990. *Nutrition during pregnancy*. Washington, DC: National Academy Press.

Institute of Medicine, Committee to Reexamine IOM Pregnancy Weight Guidelines, Rasmussen KM, Yaktine AL (eds.). 2009. *Weight gain during pregnancy: Reexamining the guidelines*. Washington, DC: National Academy Press.

Johnson, J.W.C., Longmate, J.A., and Frentzen, B. 1992. Excessive maternal weight and pregnancy outcome. *Am J Obstet Gynecol* 167:353–72.

Johnson, J.W.C. and Yancey, M.K. 1996. A critique of the new recommendations for weight gain in pregnancy. *Am J Obstet Gynecol* 174:254–58.

Jones, R.L., Cederberg, H.M.S., Wheeler, S.J., Poston, L., Hutchinson, C.J., Seed, P.T., et al. 2010. Relationship between maternal growth, infant birthweight and nutrient partitioning in teenage pregnancies. *BJOG* 117:200–11.

Kessner, D.M., Singer, J., Kalk, C.E., and Schlesinger, E.R. 1973. *Infant death: An analysis by maternal risk and health care*. Washington, DC, Institute of Medicine, National Academy of Sciences.

Kleinman, J.C. 1990. *Maternal weight gain during pregnancy: Determinants and consequences*. NCHS Working Paper Series No. 33. Hyattsville, MD: National Center

for Health Statistics, Public Health Service, U.S. Department of Health and Human Services.

Kotelchuck, M. 1994. An evaluation of the Kessner Adequacy of Prenatal Care Index and a proposed Adequacy of Prenatal Care Utilization Index. *Am J Public Health* 84:1414–20.

Langhoff-Roos, J., Lindmark, G., and Gebre-Medhin, M. 1987. Maternal fat stores and fat accretion during pregnancy in relation to infant birthweight. *Br J Obstet Gynaecol* 94:1170–77.

Lederman, S.A. 1993. Recent issues related to nutrition during pregnancy. *J Am Coll Nutr* 12:91–100.

Lederman, S.A., Paxton, A., Heymsfield, S.B., Wang, J., Thornton, J., and Pierson, R.N. 1997. Body fat and water changes during pregnancy in women with different body weight and weight gain. *Obstet Gynecol* 90:483–88.

Lederman, S.A., Paxton, A., Heymsield, S.B., Wang, J., Thornton, J., and Pierson, R.N. 1999. Maternal body fat and water during pregnancy: Do they raise infant birth weight? *Am J Obstet Gynecol* 180:235–40.

Lockwood, C.J. and Kuczynski, E. 1999. Markers of risk for preterm delivery. *J Perinat Med* 27:5–20.

Luke, B., Hediger, M.L., Nugent, C., Newman, R.B., Mauldin, J.G., Witter, F.R., and O'Sullivan, M.J. 2003. Body mass index-specific weight gains associated with optimal birth weight in twin pregnancies. *J Reprod Med* 48:217–24.

Mardones-Santander, F., Salazar, G., Rosso, P., and Villarroel, L. 1998. Maternal body composition near term and birth weight. *Obstet Gynecol* 91:873–77.

National Heart, Lung, Blood Institute, Expert Panel on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. 1998. Executive summary of the clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults. *Arch Intern Med* 158:1855–67.

National Research Council, National Academy of Sciences. 1970. *Maternal nutrition and the course of pregnancy*. Washington, DC: National Academy of Sciences.

Newburn-Cook, C.V. and Onyskiw, J.E. 2005. Is older maternal age a risk factor for preterm birth and fetal growth restriction? A systematic review. *Health Care Women Int* 26:852–75.

Oken, E. 2009. Maternal and child obesity: the causal link. *Obstet Gynecol Clin N Am* 36:362–77.

Perry, R.L., Mannino, B., Hediger, M.L., and Scholl, T.O. 1996. Pregnancy in early adolescence: are there obstetric risks? *J Matern Fetal Med* 5:333–39.

Phelan, S.T. 2008. Components and timing of prenatal care. *Obstet Gynecol Clin N Am* 35:339–53.

Pitkin, R.M. 1976. Nutritional support in obstetrics and gynecology. *Clin Obstet Gynecol* 19:489–513.

Roche, A.R. and Davila, G.H. 1972. Late adolescent growth in stature. *Pediatrics* 50:874–80.

Salihu, H.M., Sharma, P.P., Ekundayo, O.J., Kristensen, S., Badewa, A.P., Kirby, R.S., and Alexander, G.R. 2006. Childhood pregnancy (10-14 years old) and risk of stillbirth in singletons and twins. *J Pediatr* 148:522–26.

Schieve, L.A., Cogswell, M.E., Scanlon, K.S., Perry, G., Ferre, C., Blackmore-Prince, C., et al., for the NMIHS Collaborative Working Group. 2000. Prepregnancy body mass index and pregnancy weight gain: associations with preterm delivery. *Obstet Gynecol* 96:194–200.

Scholl, T.O. and Hediger, M.L. 1993. A review of the epidemiology of nutrition and adolescent pregnancy: maternal growth during pregnancy and its effect on the fetus. *J Am Coll Nutr* 12:101–107.

Scholl, T.O. and Hediger, M.L. 1995. Weight gain, nutrition, and pregnancy outcome: findings from the Camden Study of teenage and minority gravidas. *Sem Perinatol* 19:171–81.

Scholl, T.O., Hediger, M.L., Schall, J.I., Fischer, R.L., and Khoo, C.S. 1993. Low zinc intake during pregnancy: its association with preterm and very preterm delivery. *Am J Epidemiol* 137:1115–24.

Scholl, T.O., Hediger, M.L., Schall, J.I., Khoo, C.S., and Fischer, R.L. 1994. Maternal growth during pregnancy and the competition for nutrients. *Am J Clin Nutr* 60: 183–88.

Siega-Riz, A.M., Adair, L.S., and Hobel, C.J. 1996. Maternal underweight status and inadequate rate of weight gain during the third trimester of pregnancy increase the risk of preterm delivery. *J Nutr* 126:146–53.

Siega-Riz, A.M., Viswanathan, M., Moos, M.-K., Deierlein, A., Mumford, S., Knaack, J., et al. 2009. A systematic review of outcomes of maternal weight gain according to the Institute of Medicine recommendations: birthweight, fetal growth, and postpartum weight retention. *Am J Obstet Gynecol* 201:339.e1–14.

Stothard, K.J., Tennant, P.W., Bell, R., and Rankin, J. 2009. Maternal overweight and obesity and the risk of congenital anomalies: a systematic review and meta-analysis. *JAMA* 301:636–50.

Stotland, N.E., Caughey, A.B., and Lahiff, M. 2006. Weight gain and spontaneous preterm birth: the role of race or ethnicity and previous preterm birth. *Obstet Gynecol* 108: 1448–55.

Taggart, N.R., Holliday, R.M., Billewicz, W.Z., Hytten, F.E., and Thomson, A.M. 1967. Changes in skinfolds during pregnancy. *Br J Nutr* 21:439–51.

VanderWeele, T.J., Lantos, J.D., Siddique, J., and Lauderdale, D.S. 2009. A comparison of four prenatal care indices in birth outcome models: comparable results for predicting small-for-gestational-age outcome but different results for preterm birth or infant mortality. *J Clin Epidemiol* 62:438–45.

Villar, J., Cogswell, M., Kestler, E., Castillo, P., Menendez, R., and Repke, J.T. 1992. Effect of fat and fat-free mass deposition during pregnancy on birth weight. *Am J Obstet Gynecol* 167:1344–52.

Wax, J.R. 2009. Risks and management of obesity in pregnancy: current controversies. *Curr Opin Obstet Gynecol* 21:117–23.

Wen, S.W., Goldenberg, R.L., Cutter, G.R., Hoffman, H.J., and Cliver, S.P. 1990. Intrauterine growth retardation and preterm delivery: prenatal risk factors in an indigent population. *Am J Obstet Gynecol* 162:213–18.

6

Pregnancy Complications

MICHELLE A. WILLIAMS

METABOLIC AND ANATOMICAL CHANGES IN PREGNANCY

Pregnancy is characterized by profound metabolic and other physiological alterations involving virtually every organ system (Williams and Mittendorf 2000). Alterations include changes in circulating concentrations of steroid and adrenocorticotropic hormones, circulating plasma lipids, and hemodynamic changes such as plasma volume, red cell mass, and blood viscosity. These alterations, when overlaid upon chronic medical illnesses predating pregnancy, may contribute to a worsening, improvement, or no change in maternal health; and they may also contribute to adverse neonatal outcomes. Among apparently healthy women, the metabolic challenges, or "stress test" of pregnancy may also unmask susceptibility for chronic disorders. Available clinical and epidemiologic data are supportive of the hypothesis that common medical complications of pregnancy including gestational diabetes mellitus (GDM), and hypertensive disorders of pregnancy may represent, at least for some women, their underlying risk for developing related chronic medical conditions as they age.

This chapter is not intended to be a comprehensive review of maternal medical complications during pregnancy. Rather, it first provides a brief overview of maternal metabolic, physiological, and anatomical changes during pregnancy, with a particular focus on changes that contribute to significant alterations in glucose and lipid metabolism, as well as to blood pressure control. These are followed by a summary of the clinical and epidemiologic characteristics of asthma, a relatively prevalent medical condition among reproductive-age women, which may be exacerbated as a result of the physiological demands of pregnancy. Next, a review is presented of maternal medical conditions (e.g., GDM and preeclampsia) arising in apparently healthy women, along with a discussion of the possible association of these conditions with subsequent chronic metabolic disorders. Last, there is consideration of the influence of other factors, including intimate partner violence, a relatively common stressor and determinant of maternal physical and mental well-being.

Changes in lipid and carbohydrate metabolism

Every aspect of lipid metabolism is influenced by pregnancy (Rosso 1990). Maternal serum or plasma cholesterol and triglyceride concentrations increase (1.5- and 2- to 3-fold, respectively) during pregnancy, with the major increase occurring in the second and third trimesters of pregnancy (Potter and Nestel 1979). In addition to absolute changes in plasma lipid levels, changes in lipoprotein composition (Potter and Nestel 1979; Knopp et al. 1982) are also noted in uncomplicated pregnancy. Some lipoproteins, such as high-density lipoproteins (HDL) and low-density lipoproteins (LDL), become more triglyceride-enriched; and very low-density lipoproteins (VLDL) increase their cholesterol and triglyceride concentrations even more sharply, reaching up to fivefold the levels seen in nonpregnant women (Montes et al. 1984; Knopp et al. 1985).

Many of the observed alterations in maternal lipid metabolism during pregnancy are attributed to changes in hormonal status, although few investigators have examined the specific effects of hormones on lipids and lipoprotein changes during pregnancy (Al et al. 1995). Estrogen and progesterone concentrations increase to values 16- and 7-fold their prepregnancy levels by 30 weeks' gestation (Desoye et al. 1987). In general, studies assessing the effect of exogenous hormones suggest that the effects of estrogens on lipids and lipoproteins are opposite in direction to those of progestogens. The continuous elevation of triglyceride levels in all lipoprotein fractions is thought to be a result of increasing estrogen concentrations (Knopp et al. 1985). Elevations in HDL and associated lipoprotein subfractions throughout the first two trimesters of pregnancy are thought to be related to rises in estrogen levels (Fahraeus et al. 1985), whereas the decline in HDL-cholesterol that occurs after 24 weeks' gestation is thought to be associated with increasing insulin concentrations (Desoye et al. 1987). Collectively, this combination of hormonal and carbohydrate metabolic changes in uncomplicated pregnancies makes pregnancy a "diabetogenic state." Thus, it is important to recognize that uncomplicated pregnancy is characterized by progressive insulin resistance, hyperinsulinemia, and a deterioration of glucose tolerance in the third trimester.

Why are there such profound changes in maternal lipid and carbohydrate metabolism in normal pregnancy? Knopp et al. (1982) have provided a comprehensive summary for the apparent physiologic hyperlipidemia and hyperinsulinemia of pregnancy. First, the authors argue that increases in plasma triglycerides likely enhance the availability of essential and nonessential fatty acids for placental transfer to the fetus, and as such, pregnancy-associated hypertriglyceridemia represents a physiologic adaptation that would favor fetal nutrition and growth. Second, pregnancy-associated hypercholesterolemia may provide the substrate needed for placental progesterone synthesis and transplacental cholesterol transfer to the fetus. Third, elevations in maternal plasma triglyceride concentrations may be the barometer of a general metabolic adaptation by the mother to augment nutrient flow to the fetus. Fourth, the hyperlipidemia may stress maternal lipid homeostasis to an extent that subclinical or mild hyperlipidemia becomes clinically detectable, analogous to the prediabetes condition recognized in women

who develop gestational diabetes (Potter and Nestel 1979). And fifth, the hyperlipidemia, and particularly the triad of increased triglycerides and LDL, and decreased HDL, resembles some dyslipidemic syndromes in nonpregnant individuals, such as metabolic syndrome, and could thus function as an arteriosclerosis risk factor.

Hemodynamic changes

A large number of hemodynamic changes are noted throughout the 40 weeks of a normal uncomplicated pregnancy. These changes represent the maternal adaptive responses necessary for meeting the mother's own circulatory needs and those of the developing fetoplacental unit. The changes include an increase in plasma volume from the time of conception to approximately 32 weeks, peaking at 40%–50% above the nonpregnant baseline (Metcalf et al. 1981). This occurs despite the increase in red blood cell mass on the order of 20%–35% above baseline. The greater increase in plasma volume accounts for pregnancy-associated anemia and decreased blood viscosity. The modest decrease in systolic blood pressure and a more pronounced decrease in diastolic blood pressure are secondary to a decrease in systemic vascular resistance that occurs in uncomplicated pregnancies. Heart rate is increased by 10%–20% in pregnancy (Katz et al. 1978). Cardiac output increases late in the first trimester, peaking at 40% above baseline in the early second trimester, and may increase further during labor, depending upon pain control and anesthesia (Danzell 1998). These changes—increased cardiac output, increased blood volume, and decreased total systemic vascular resistance—are considered the most important alterations in maternal circulation during pregnancy (Duvekot and Peeters 1998). The largest change in diastolic blood pressure is believed to occur before the eighth week of gestation (Robson et al. 1989; Duvekot and Peeters 1998). By the eighth week of gestation, cardiac output is increased 13% above prepregnancy levels, peaking at 40% above prepregnancy levels in the early second trimester (Danzell 1998). Plasma volume expansion, detectable as early as 7 weeks' gestation, continues to increase until about 32 weeks' gestation and peaks at approximately 40%–50% above prepregnant levels. Together, these early changes lead to a decrease in systemic peripheral vascular resistance, often regarded as the earliest hemodynamic adaptation to pregnancy (Duvekot and Peeters 1998). Pregnancies complicated by intrauterine growth restriction, gestational hypertension, and preeclampsia are associated with lower maternal cardiac output and relatively higher total peripheral vascular resistance than are uncomplicated pregnancies.

Other metabolic, immunological, and anatomical changes

Alterations in renal hemodynamics during pregnancy include increased glomerular filtration rates and renal plasma flow. Consequently, creatinine clearance is increased, and serum creatinine and urea nitrogen decreases. Alterations in tubular function and dilation of the renal collecting duct systems are also noted in uncomplicated pregnancies (Mason et al. 1998). Additionally, all aspects of maternal

hemostasis are altered during normal pregnancy (Brenner 2004). On balance, these changes, including increased concentrations of most clotting factors, decreasing concentrations of some of the natural anticoagulants, and diminishing fibrinolytic activity, are important for maintaining placental function during pregnancy and for meeting the hemostatic challenges associated with labor and deliver (Brenner 2004). Importantly, these changes may also predispose pregnant women to thrombosis and placental vascular complications, including preeclampsia and abruptio placentae.

Several alterations in immunological status occur during pregnancy. These changes include a shift away from cell-mediated immunity to humoral immunity (Sargent et al. 1987; Wegmann et al. 1993). The parallel increase in the predominance of type 1 or proinflammatory (helper T cells) over type 2, or anti-inflammatory (helper T cells) is thought to play an important role in maternal tolerance of the fetus. An inversion of this shift during the latter half of the third trimester may contribute to the initiation of labor. Although details of the nature and direction of the influence of altered immunological status in pregnancy are still highly debated, there is a growing consensus that pregnancy-associated immunosuppressive cytokines may account for modifications in the behavior of autoimmune disorders, including antiphospholipid syndrome (APS), multiple sclerosis, systemic lupus erythematosus (SLE), and autoimmune thyroid diseases during pregnancy (Borchers et al. 2009).

Uncomplicated pregnancy is associated with a 20% increase in oxygen consumption and a 15% increase in maternal metabolic rate (Nelson-Piercy and Moore 1996). This demand is met by impressive anatomical and physiological changes, including a flaring of the lower ribs; an increase in the subcostal angle (approximately 2 cm) as the transverse diameter of the chest increases; and an elevation of the diaphragm to compensate for the enlarged uterus, particularly in the third trimester. Placentally derived hormones are also known to stimulate alterations in respiration during pregnancy, resulting in compensatory respiratory alkalosis. Briefly, progesterone, which peaks during the third trimester, stimulates the respiratory centers of the brain to produce hyperventilation. Consequently, resting minute ventilation increases by as much as 40%–50% over the pregestational state. Hyperventilation may result in a decrease in partial pressure of arterial carbon dioxide, resulting in respiratory alkalosis, decreased bicarbonate levels, and changes in arterial pH.

Summary

In summary, pregnancy is associated with significant changes in organ function and physiological levels from the nonpregnant state. An understanding of the natural history and determinants of these alterations (i.e., variations in magnitude and timing), as well as how these changes differ for women with medically complicated and uncomplicated pregnancies, are important for better defining, categorizing, surveilling, and ultimately managing clinically heterogeneous medical complications of pregnancy (such as preeclampsia or hemolysis, elevated liver enzymes, and low platelet count [HELLP] syndrome). Additionally, an understanding of specific

pathophysiological characteristics of these medical complications may allow for more precise assessments of how underlying genetic and nongenetic determinants of plasma concentrations of steroid hormones, lipoproteins, adipocytokines, cytokines, and other inflammatory markers, for example, contribute to observed variations noted in uncomplicated and complicated pregnancies. Finally, for phenotypically diverse pregnancy-associated disorders, such as preeclampsia and preterm delivery (PTD), classification of disorders according to metabolic disturbances associated with the clinical diagnosis may assist in defining homogeneous case populations. Such improvements may enhance the resolution of etiological studies of the disorder and accelerate the processes for identifying appropriate and efficient prevention strategies.

PREGNANCY COMPLICATIONS

This chapter continues with a summary of the clinical and epidemiologic characteristics of asthma, a relatively prevalent medical condition among reproductive age women, and one that may be exacerbated as a result the physiological demands of pregnancy. Following that, is a review of the clinical and epidemiologic characteristics of GDM and preeclampsia, two of the most common medical complications of pregnancy. This section also includes a summary of emerging evidence documenting associations of hypertensive and psychiatric disorders with relatively novel risk factors, including maternal migraine history and exposure to intimate partner violence. The chapter concludes with a section on the influence of sleep disorders on maternal physical and mental well-being.

Asthma

Nearly 20% of reproductive age women suffer from some form of allergic disorder that affects pulmonary function (Schatz and Dombrowski 2009), with asthma accounting for most cases. Asthma, a heterogeneous chronic lung disease, is characterized by recurrent bouts of wheezing and dyspnea resulting in airway obstruction, and it is the most common respiratory crisis encountered in pregnancy. Approximately 4% of all pregnancies are complicated by an asthma crisis (Witlin 1997). The airways of asthmatics are hyper-responsive to stimuli such as allergens, viral infections, air pollutants, and cold air. This hypersensitivity is manifested by bronchospasm, mucosal edema, and mucus plugging that results in air trapping and hyperinflation of the lungs. The causes of asthma are unknown, although certain pathophysiologic mechanisms have been implicated.

The effect of pregnancy on the course of this disease is variable. Overall, an equal number of women have asthma symptoms that improve, worsen, or are unchanged through pregnancy (Gluck and Gluck 2006). Asthma symptoms can worsen during pregnancy because of identifiable factors, such as infection, gastroesophageal reflux disease, reduction of appropriate medications by physician or patient, and smoking. Importantly, undertreatment, which remains a problem during pregnancy, can lead to continued difficulty with asthma. Experiences

during a previous pregnancy may be somewhat predictive of experience in subsequent pregnancies; and there is some evidence to support an association between the severity of asthma prior to pregnancy and likelihood of deterioration during pregnancy (Gluck and Gluck 2006; Beecroft et al. 1998). For instance, Beecroft et al. (1998) noted that the course of asthma during pregnancy was influenced by fetal gender: women carrying female fetuses were more likely to experience moderate to severe deterioration, as compared with those carrying male fetuses.

Overall methodological limitations related to the subjective nature of measuring breathlessness and heightened maternal awareness and concern about disease control during pregnancy—in addition to undertreatment secondary to maternal concerns about possible fetal toxicity—are important limitations in many of these studies. Recent findings, however, from studies that employ objective measures of maternal airway hyper-responsiveness (e.g., using a methacholine challenge test protocol) will substantially enhance causal inferences from pregnancy–asthma studies (Siddiqui et al. 2008). Mechanisms involved in changes of the course of asthma during pregnancy have not been defined. The peaking of plasma progesterone concentrations during the third trimester, coincident with improvements in the course of asthma in the third trimester, is considered a hint for a causal model. Associations of adverse pregnancy outcomes, such as PTD, preeclampsia, and fetal growth restriction, with maternal asthma have been inconsistently reported (Murphy et al. 2006). Recent reviews of the literature suggest that the severity of asthma, whether the mother experiences exacerbations during pregnancy, and the variations in pharmacological agents used to manage asthma symptoms contribute to noted inconsistencies across studies (Dombrowski 2006; Murphy et al. 2006). Women with severe asthma and those with poor asthma control in pregnancy have also been noted to be at increased risk of delivering preterm.

Gestational diabetes mellitus

Gestational diabetes mellitus complicates approximately 4% of all pregnancies in the United States, resulting in approximately 135,000 cases annually. As noted by the American Diabetes Association (ADA), the prevalence of GDM may range from 1% to 14% of pregnancies, depending on the population studied (ADA 2005). Differences in screening programs and diagnostic criteria make it difficult to compare frequencies of GDM among and across various populations. Despite this, the ADA estimates a national GDM cumulative incidence of 4%. Gestational diabetes mellitus is considered to represent nearly 90% of all pregnancies complicated by diabetes (ADA 2005). Available evidence suggests that the cumulative incidence of GDM may be on the increase. In their study of 267,051 pregnancies, Ferrara et al. (2004) noted that the age- and ethnicity-adjusted yearly cumulative incidence of GDM increased from 5.1% in 1991 to 7.4% in 1997 and leveled off to 6.9% among members of the northern California Kaiser Permanente Medical Care Program. The findings from northern California are consistent with increasing prevalence of type 2 diabetes and obesity in the United States and elsewhere (Gabbe and Graves 2003). Gestational diabetes mellitus is of public health importance, in large part because of its association with adverse short- and long-term sequelae in both

mothers and newborns; these are summarized below (Buchanan and Kjos 1999; Ben-Haroush et al. 2004).

Women who develop GDM are thought to have a compromised physiological capacity to adapt to the metabolic challenges of late pregnancy. As discussed above, the third trimester is characterized by profound metabolic stresses on maternal lipid and glucose homeostasis favoring the transfer of nutrients to the fetus (Knopp et al. 1982). Maternal metabolic changes occurring in the third trimester include marked insulin resistance, hyperinsulinemia, a progressive worsening of glucose intolerance, and hyperlipidemia (particularly hypertriglyceridemia). The endocrinological challenges of pregnancy and the variation in maternal physiological adaptation to these challenges have led several investigators (Knopp et al. 1982) to speculate that pregnancy serves to unmask a predisposition to glucose metabolic disorders in some women. Women with a history of GDM have an increased risk of developing type 2 diabetes and impaired glucose tolerance later in life (Ben-Haroush et al. 2004; Lauenborg et al. 2004). In a study of Danish women, Lauenborg et al. (2004) reported that 40% of women with a history of diet-treated GDM developed diabetes at a median of 10 years after pregnancy. This incidence was greater than ten times higher than the incidence noted in the general 30- to 60-year-old female population.

Gestational diabetes mellitus is known to have diverse effects on fetal and infant outcomes. The risk of late intrauterine fetal death has been reported to be three fold higher in gestational diabetics than in the general obstetric population (O'Sullivan 1984). Higher risks of macrosomia, birth trauma, hyperbilirubinemia, hypoglycemia, polycythemia, respiratory distress syndrome, erythrocytosis, postpartum obesity, and others have been reported for the offspring of GDM mothers (O'Sullivan 1984; ADA 1999). A study of maternal and infant cord blood leptin levels in diabetic mother–infant pairs and normoglycemic controls suggests that maternal hyperglycemia has a measurable impact on fetal leptin and insulin status (Maffei et al. 1998). Dorner at al. (2000) reported evidence suggestive of maternal GDM status as an important predisposing factors for increased risk of type 1 childhood diabetes in offspring. Hedderson et al. (2003) reported that the risk of spontaneous PTD increased with increasing levels of glycemia during pregnancy. Women with GDM were 40% more likely to deliver preterm infants than were unaffected mothers. Anderson et al. (2005) reported that GDM is associated with a 2.9-fold (adjusted odds ratio [OR] = 2.9; 95% confidence interval [CI] 1.0–8.4) increased risk of holoprosencephaly (i.e., a spectrum of defects involving the brain and face). The authors also noted evidence of statistical interaction (on a multiplicative scale) of maternal prepregnancy obesity and GDM in relation to risk of central nervous system birth defects. On balance, a fairly large and diverse literature serves to underscore the importance of the maternal metabolic environment on fetal growth and development, gestational length, and postnatal metabolism.

As noted above, women with a history of GDM have a considerably elevated risk of developing type 2 diabetes or impaired glucose tolerance in the years following pregnancy (O'Sullivan 1984; Ben-Haroush et al. 2004; Lauenborg et al. 2004). Among women with a prior history of GDM, the cumulative incidence of developing type 2 diabetes later in life has been reported to range from

22% to 60% (Damm 1998; Foster-Powell and Cheung 1998). In a study of 615 women in Boston, O'Sullivan (1984) noted that 36% of women with a GDM-affected index pregnancy developed overt diabetes within 22–28 years of follow-up. Several other investigators (Damm 1998; Foster-Powell and Cheung 1998) have reported this finding of an increased risk of developing overt diabetes. Variations in the cumulative incidence of developing type 2 diabetes after an GDM-affected pregnancy may be attributed to differences in diagnostic criteria, length of follow-up, analytical approach (life-table versus other methods), and racial/ethnic distribution of study populations.

Several teams of investigators have identified maternal characteristics that are associated with an increased risk of developing type 2 diabetes after a GDM-affected pregnancy (Metzer et al. 1993; Kjos et al. 1995; Peters et al. 1996). Catalano et al. (1993) reported that high fasting glucose and low gestational age at GDM diagnosis were associated with abnormal glucose tolerance at 6 weeks postpartum. These observations underscore the importance of regularly assessing glucose tolerance in women with previous GDM and further delineating the etiology and pathophysiology of the disorder.

The high recurrence risk of GDM also contributes to its prominent position as a clinically significant diagnosis among reproductive-age women. The recurrence risk of GDM has been reported to range from 30% to 91%. Coelingh Bennink et al. (1977) noted in an early study that 30% of women with an index pregnancy complicated by GDM experienced GDM in a subsequent pregnancy. For many of the recently reported studies of North American women, the recurrence risks of GDM have been noted to range from 52% to 91%. Variation in study design, duration of postpartum follow-up, diagnostic tests and criteria, and racial/ethnic distributions of study populations contribute to the variation in reported rates. On balance, however, available information suggests a very high recurrence of GDM in subsequent pregnancies. The literature also suggests that the recurrence of GDM is most common among multiparous women (i.e., women who were multiparas at the index pregnancy), women with a high prepregnancy body mass index (BMI), women requiring insulin therapy for the index pregnancy, and women whose GDM diagnosis was made early in the index pregnancy (Foster-Powell and Cheung 1998). Foster-Powell and Cheung (1998) reported that weight gain between the index pregnancy and the subsequent pregnancy is also an important risk factor for the recurrence of GDM.

Relatively few risk factors have been identified for GDM to date. Several of the consistently reported GDM risk factors are common to type 2 diabetes. Advanced maternal age, a family history of type 2 diabetes, and a prior history of GDM are three well-recognized risk factors of GDM (Berkowitz et al. 1992; Solomon et al. 1997; Williams and Mittendorf 2000;Williams et al. 2003). Women with a positive family history of type 2 diabetes mellitus (particularly maternal history), as compared with women without such a history, have been noted to experience a 1.5- to 2-fold increase in risk for GDM. Women with parental history of both essential hypertension and type 2 diabetes had particularly elevated risks of developing GDM (Williams et al. 2003). Cigarette smoking has not consistently been identified as a risk factor of GDM (Berkowitz et al. 1992; Solomon et al. 1997).

Available data suggest that the magnitude of any possible association between maternal smoking during pregnancy and GDM may be modest. Asian, Hispanic, and Native American women, as compared with non-Hispanic white women have an increased risk of GDM (Jovanovic-Peterson et al. 1989). African-American women have been reported to have an increased risk of GDM, as compared with non-Hispanic whites, by some (Solomon et al. 1997), although not all (Berkowitz et al. 1992) investigators. As is the case for type 2 diabetes, women who were underdeveloped at birth appear to be at increased risk of GDM (Williams et al. 1999). An emerging literature suggests that vitamin C intake, ascorbic acid plasma concentrations, dietary fiber, and fruit and vegetable intake (Zhang et al. 2004a,b, 2006a), as well as dietary fat intake (Saldana et al. 2004), intake of sugar-sweetened beverages (Chen et al. 2009), and vitamin D intake status (Zhang et al. 2008) may be associated with GDM risk.

Women with a high prepregnancy BMI are noted to have an increased risk of GDM. There is a progressive increase in risk of GDM above a prepregnancy BMI of 22 kg/m^2 (Solomon et al. 1997). A study of Mexican women living along the Mexican–U.S. border in Juarez, Mexico, also provides evidence suggestive of an increased risk of GDM associated with short maternal stature (Meza et al. 1995).

Increasingly, investigators have been documenting associations between varying concentrations of adipose tissue–derived cytokines (e.g., leptin and adiponectin) with the occurrence of GDM (Qiu et al. 2004; Williams et al. 2004; Lain et al. 2005). In a prospective, nested case-cohort study of women without pregestational diabetes or chronic hypertension, low maternal plasma adiponectin concentrations, measured in early pregnancy, were associated with a 4.6-fold increased risk of GDM. Additionally, the risk of GDM was found to increase by 20% for each 1 μg/mL decrease in maternal plasma adiponectin concentration. This association appeared to be independent of maternal age, family history of type 2 diabetes, and adiposity in early pregnancy (Williams et al. 2004). Findings from cross-sectional studies have also supported associations between hypoadiponectinemia and GDM (Ranheim et al. 2004; Retnakaran et al. 2004). In a preliminary study of the 1,000-member Omega cohort, Qiu et al. (2004) noted that elevated leptin concentrations in maternal plasma collected at 13 weeks' gestational age, on average, was associated with an increased risk of subsequently developing GDM. On balance, these data suggest that maternal prepregnancy adiposity and alterations in adipose tissue-derived cytokines may be important risk factors and possibly biological mediators of GDM risk.

Several observational studies have examined the association between physical activity before and/or during pregnancy and the risk of developing GDM (Dye et al. 1997; Dempsey et al. 2004a,b, 2005; Harizopoulou et al. 2009). Dye et al. (1997) noted that exercise during pregnancy was associated with a 47% reduction in GDM among obese women. Solomon and colleagues (1997) reported that women who engaged in vigorous activity or brisk walking prior to pregnancy were less likely to develop GDM, although these associations were not statistically significant. In a case-control study of 155 women with GDM and 386 controls in the state of Washington, Dempsey et al. (2004b) noted that women who engaged in any physical activity during pregnancy or during the first 20 weeks

of pregnancy had reductions in risk of 55% and 48%, respectively. Characteristics suggestive of an active lifestyle, such as daily stair climbing, were also associated with statistically significant reductions in risk of GDM, irrespective of participation in recreational physical activity during pregnancy (Dempsey et al. 2004b).

Similar reductions in GDM risk were associated with physical activity before and during pregnancy in a prospective cohort study (Dempsey et al. 2004a). Women who spent 4.2 hours/week or more engaged in recreational physical activity experienced a 76% decrease in GDM risk, and those who expended 21.1 MET-hours/week or more (the equivalent of 5.3 hours/week of moderate-intensity exercise, such as brisk walking) experienced a 74% reduction, compared with inactive women. These observations were corroborated by Zhang et al. (2006b), who reported a reduction in GDM risk associated with vigorous physical activity before pregnancy. Collectively, these data suggest that physical activity before and during pregnancy may be an important risk-protective factor for GDM. Inferences from these few studies, however, are limited due to small sample sizes and other methodological considerations.

Evidence from epidemiologic, controlled, clinical metabolic and animal studies suggests that physical activity may impact the occurrence of GDM through a number of biological pathways. Potential intermediate effects include improved insulin sensitivity, decreased concentrations of proinflammatory cytokines in peripheral circulation, reduced oxidative stress, and improved plasma lipid and lipoprotein concentrations (Durstine et al. 2001; Butler et al. 2004).

As discussed above, during pregnancy, the increased fetal demand for glucose, the major substrate for growth and development, produces a shift in maternal metabolism toward greater fat production and storage and decreased glucose utilization. Moderate- and vigorous-intensity physical activity is associated with improved insulin sensitivity (Vitoratos et al. 2002) and reductions in fat mass (McMurray et al. 1993). Aerobic dance and walking, the most popular forms of exercise among pregnant women (Ning et al. 2003), have been shown to result in a reduction in plasma insulin during pregnancy (Clapp and Capeless 1991). Importantly, carefully designed and conducted metabolic studies suggest that pregnancy may be characterized by a set of compensatory mechanisms that attenuate the adverse influences exercise may have in glucose handling. Clapp and Capeless (1991) reported that the typical exercise-induced hyperglycemia observed in nonpregnant subjects appears to be reversed in healthy pregnant women who regularly engage in recreational physical activities. The authors attribute the variation in exercise-induced response to a pregnancy-associated decrease in hepatic glucose synthesis and an increase in fractional glucose use by maternal muscle during exercise. These data provide evidence to support the biological plausibility of using recreational physical activity as a disease prevention modality in pregnant women.

Recreational physical activity is also associated with improvements in lipid concentrations in men and nonpregnant women, specifically, reduced plasma triglycerides and increased HDL (Durstine et al. 2001). Every aspect of lipid metabolism is dramatically altered during pregnancy. Maternal serum or plasma cholesterol and triglyceride concentrations increase 1.5- and 3-fold, respectively,

above nonpregnant levels by the mid third trimester (Potter and Nestel 1979). Pregnancy-associated hyperlipidemia is further exaggerated in women with gestational diabetes. Butler et al. (2004) reported that mean triglyceride concentrations were lower (–23.6 mg/dL) among women in the highest tertiles (>12 hrs per week) of time performing physical activity, as compared with inactive women. Reductions in mean total cholesterol were also observed among women in the highest levels of time performing physical activity, energy expenditure, and peak intensity. Linear relations were observed across levels of physical activity measures for triglyceride and total cholesterol. There was no association between physical activity and HDL cholesterol. These data suggest that habitual physical activity performed during pregnancy may mitigate the pregnancy-associated dyslipidemia commonly noted in hypertensive and diabetic pregnancies.

Moderate-intensity physical activity results in decreased concentrations of proinflammatory cytokines and C-reactive protein in peripheral circulation. Clapp and Kiess (2000) reported recently that regular weight-bearing exercise during pregnancy influences alterations in plasma tumor necrosis factor-α (TNF-α) during pregnancy. Women randomized to the physical activity group, as compared with inactive women, experienced a greater attenuation of the proinflammatory cytokine concentrations during pregnancy. As late pregnancy is characterized by both an increase in insulin resistance and elevated circulating TNF-α levels, regular weight-bearing exercise during pregnancy may moderate insulin resistance.

In summary, physical activity may reduce the risk of GDM through several pathways. Emerging results from small clinical and epidemiologic studies in pregnant women offer compelling evidence of the short- and long-term benefits of physical activity in pregnancy.

Preeclampsia

Hypertensive disorders, including pregnancy-induced hypertension (with and without proteinuria), and pregnancy-induced hypertension superimposed on chronic hypertension, complicate approximately 12% of pregnancies in the United States. The majority of women with hypertensive disorders of pregnancy have preeclampsia (defined as hypertension diagnosed after 20 weeks' gestation with proteinuria), making it the most common medical disease in pregnancy, as well as a leading cause of maternal mortality and PTD worldwide. In the United States, preeclampsia is the third leading cause of maternal mortality (NHBPEP 2000) and is an important cause of PTD, fetal growth retardation, and perinatal mortality. In the United Kingdom, it accounts for 40% of nonspontaneous preterm deliveries (ACOG 2000). At present, no proven preventive intervention for preeclampsia exists. Hence, the disorder is managed by screening asymptomatic women and inducing delivery when necessary to mitigate harm to mother or fetus.

The cardinal features of preeclampsia are hypertension, proteinuria, and edema. When seizures or coma further complicate these pregnancies, the condition is generally referred to as eclampsia. Elevated hypertension without concomitant proteinuria is generally referred to as gestational hypertension. Hereinafter, preeclampsia will be used to refer to proteinuric pregnancy-induced hypertension.

Pregnancy-induced hypertension without proteinuria will be referred to as gestational hypertension. Pathologic changes associated with preeclampsia include maladaptation of spiral arteries of the placental bed, hypertriglyceridemia, hypercholesterolemia, excessive lipid peroxidation, endothelial cell dysfunction, sympathetic nervous system over-reactivity, plasma elevations of proinflammatory cytokines and adipocytokines, an imbalance in thromboxane and prostacyclin in favor of vasoconstriction, glucose intolerance or hyperinsulinemia, hyperuricemia, hypovolemia, decreased angiotensin II, and alterations in angiogenic factors (Williams and Mittendorf 2000; ACOG 2000; NHBPEP 2000). Oxidative damage of maternal and placental vascular endothelium, probably secondary to early events linked to placental hypoxemia, is thought to be of central importance in the pathophysiology of preeclampsia. These physiological alterations offer important clues as to possible intrinsic and extrinsic causes of the disorder. The etiology, early predictive markers, and means of effective prevention of the disorder remain elusive.

Epidemiologic research supports the hypothesis that the etiology of preeclampsia is multifactorial. Several relatively consistent clinico-obstetric risk factors have been identified. These include nulliparity, primigravidity, young and advanced maternal age, previous preeclamptic pregnancy, family history of preeclampsia, second-trimester placental abnormalities, and uterine artery Doppler profile (ACOG 2000). Hypertriglyceridemia, antiphospholipid antibodies, coagulation disorders, and increased placental mass associated with multifetal gestation are associated with increased risk (Zhang et al. 1997; Williams and Mittendorf 2000). Several lifestyle characteristics have been reported to be associated with preeclampsia. These include changed paternity, high BMI, and sedentary lifestyle. Paradoxically, cigarette smoking is associated with a reduced risk (Zhang et al. 1997; England and Zhang 2007). It is worth noting that many of the aforementioned risk factors of preeclampsia are also predictive of coronary heart disease (CHD) in nonpregnant individuals. Additionally, many of the protective factors for preeclampsia, except for smoking, are also protective of CHD (Williams and Mittendorf 2000; Dempsey et al. 2005; Garovic and Hayman 2007). Much of the overlap in risk factors is concentrated among behavioral and metabolic characteristics that are known to be related to lipid and carbohydrate metabolism.

The recurrence risk of preeclampsia is very high. Women with a previous history of preeclampsia are approximately seven times more likely to have a subsequent pregnancy complicated by the disorder, as compared with women with no history of preeclampsia (Dildy et al. 2007). This suggests that a subgroup of women may be predisposed to developing this very dangerous complication of pregnancy. Indeed, mounting evidence suggests that preeclampsia risk is shared by family members. In their Norwegian birth registry study, investigators noted that the daughters of preeclamptic women had more than twice the risk of having preeclampsia (OR = 2.2, 95% CI 2.0–2.4), and that men born from a pregnancy complicated by preeclampsia had a moderately increased risk of fathering a preeclamptic pregnancy (OR = 1.5; 95% CI 1.3–1.7) (Skjaerven et al. 2005). Collectively, these and other similar data concerning the high recurrence risk and familial aggregation of preeclampsia suggest that maternal and fetal genetic factors are important in the causation of the disorder.

The number of candidate gene association studies of preeclampsia has exponentially increased in the last decade. To date, studies have largely focused on approximately 50 candidate genes largely selected on account of their potential functional role in regulating features of established preeclampsia. Essentially, investigators have focused on genes encoding elements of the renin–angiotensin system, which regulates blood pressure (angiotensinogen [*AGT*], angiotensin-converting enzyme [*ACE*], and angiotensin receptors [*AGTR1*, *AGTR2*]); inherited thrombophilias (coagulation factor V Leiden variant [*F5*], prothrombin [*F2*] and methylene tetrahydrofolate reductase [*MTHFR*]); the NOS3 gene regulating the synthesis of the vasorelaxant eNOS (endothelial nitric oxide synthase [*NOS3*]); and the gene encoding the cytokine TNF-α [*TNF*] (Chappell and Morgan 2006; GOPEC Consortium 2005; Ward 2008). This literature, however, is characterized by the absence of consistency and reproducibility of findings. Possible reasons for inconsistent findings across studies include population stratification, different definition of the preeclampsia phenotype, gene–environment interactions, small sample size, limited statistical power, and failure to address concerns about linkage disequilibrium. The absence of adjustments for multiple comparisons all have compromised causal inferences from these studies. Larger and more rigorously designed studies, including those that include genome-wide evaluations of variants in both the maternal and fetal genome (Ward 2008), as well as whole-genome gene expression profiling studies (Enquobahrie et al. 2008) in the context of environmental influences will likely contribute to our increased understanding of the genetic etiology of preeclampsia.

Migraine

Migraine, a common chronic-intermittent neurovascular headache disorder, is ranked among the world's 20 most disabling medical conditions by the World Health Organization (Leonardi and Mathers 2003). A migraine is characterized by episodic severe headache accompanied by autonomic nervous system dysfunction. Some patients with migraines have headaches that are accompanied by transient neurological symptoms and are thus classified as having migraines with aura (Russell et al. 2002). Women are more commonly affected than men, with reported lifetime prevalence estimates of 16%–32% for women and 6%–9% for men (Stewart et al. 1996). Migraine risk varies considerably across the life course. The prevalence of migraine in women rises after the average age of menarche and peaks before the average age of menopause (Lipton et al. 2001). Thus, migraines are most prevalent among women in their childbearing years.

Associations between migraine and vascular disease have long been hypothesized and considered over the last century, but results reported in the epidemiologic literature are contradictory. Available evidence suggests a consistent relation between migraine and vasospastic disorders, such as variant angina and Raynaud's phenomenon, as well as ischemic stroke in young women (Rosamond 2004). A recent population-based cross-sectional study of Dutch adults found that female migraineurs were at increased risk of subclinical brain infarcts and white matter lesions (OR = 2.1; 95% CI 1.0–4.7), which may increase the risk of future stroke and dementia (Kruit et al. 2004).

Studies have shown that preexisting migraine diminishes or disappears in the majority of pregnant women, but for some the disorder may remain unchanged or worsen, and for others migraine may appear for the first time during pregnancy (Adeney and Williams 2006). Spontaneous improvement tends to occur in the first trimester of pregnancy, when dramatic changes in estradiol levels occur and pain thresholds increase. Women with headaches persisting into the second trimester are less likely to improve. The likelihood of improvement may also depend on a history of menstrual migraine, parity, and headache changes during prior pregnancy (Marcus 2003). Despite the results reported by Callaghan (1968), in which 17% of women experienced migraine onset during pregnancy, this phenomenon is not common and may indicate a more serious problem. The most recent study found that, among women without headache attacks before pregnancy, only 1.9% developed migrainous symptoms de novo in pregnancy (Ertresvag et al. 2005).

Some investigators, on the basis of their review of the literature, reported that migraine has no adverse effect on the outcome of pregnancy (Aube 1999; Matharu et al. 2002). However, this conclusion is based on relatively few studies and is likely to be hindered by residual confounding. Moreover, until very recently, most studies of migraine and pregnancy outcome were focused only on the potential teratogenicity of medications used to treat migraines, particularly sumatriptan, during pregnancy. Briefly, these studies have generally found no difference in the incidence of infant malformation between patients who did and did not use sumatriptan after conception (Fox et al. 2002; Kallen and Lygner 2001). Olesen et al. (2000) linked birth records to a prescription database and found an increased risk of PTD with sumatriptan use (adjusted OR = 3.3; 95% CI 1.3–8.5), and an increased risk of low birth weight (LBW) among all migraineurs who delivered at term, compared with healthy controls (OR = 3.0; 95% CI 1.3–7.0).

Mounting evidence (Adeney et al. 2005; Scher et al. 2005; Sanchez et al. 2008; Facchinetti et al. 2009) suggests consistent associations of migraine with preeclampsia. Using data from the Genetic Epidemiology of Migraine (GEMS) Study, investigators noted that migraineurs were significantly more likely to report a history of gestational hypertension after adjusting for age, socioeconomic status, smoking, and alcohol use (OR = 1.63; 95% CI 1.2–2.1) (Scher et al. 2005). Similarly, in a secondary analysis of a case-control study of 244 preeclamptic women and 470 controls, Adeney et al. (2005) reported a significantly increased risk of preeclampsia among women who reported a history of physician-diagnosed migraine (adjusted OR = 1.8; 95% CI 1.1–2.7). These observations from cross-sectional case-control studies were recently corroborated by Facchinetti et al. (2009), who completed a prospective cohort study of 702 Italian pregnant women. The authors reported that maternal history of migraines, defined using the International Headache Society diagnostic criteria, was associated with a 2.85-fold increased risk of incident hypertensive disorders of pregnancy (gestational hypertension or preeclampsia combined) (95% CI 1.40–5.81). In addition to these documented associations, available data indicate that migraine and preeclampsia share many common epidemiologic and pathophysiological characteristics, including overlapping cardiovascular risk factors and increased risks of adverse long-term outcomes including ischemic stroke (Wilson et al. 2003; Kruit et al. 2004).

Furthermore, endothelial dysfunction, platelet activation, hypercoagulation, and inflammation are common to both disorders (Hayashi et al. 2002; Welch 2003).

In summary, mounting evidence suggests that migraine is positively associated with preeclampsia risk. Increased knowledge concerning the epidemiology and pathophysiology of preeclampsia and migraines, particularly as they relate to proinflammatory platelet adhesion to leukocytes during pregnancy, will aid understanding the underlying pathophysiologic mechanisms and lead to the identification of possible prevention and therapeutic strategies.

MATERNAL PSYCHIATRIC DISORDERS AND EXPOSURE TO INTIMATE PARTNER VIOLENCE

Psychiatric disorders, such as depression and anxiety disorders, are more prevalent in reproductive-aged women, and are linked to hormonal and reproductive events (O'Hara 1995). For example, blunted memory and diminished anxiety during pregnancy have been associated with progesterone and glucocorticoids (Brett and Baxendale 2001). Posttraumatic stress disorder (PTSD), a well-defined mental disorder and one of the most severe stress-related illnesses, is twice as common among women as men (Brunello et al. 2001). Posttraumatic stress disorder and depression commonly co-occur, and comorbid PTSD and depression is of particular concern because this condition may increase symptom severity (Shalev et al. 1998), contribute to PTSD chronicity (Freedman et al. 1999), and increase the risk of adverse health outcomes (Kimerling 2004). In addition, preexisting major depression can increase the risk of exposure to traumatic events and PTSD, and vice versa (Breslau et al. 2000).

Perinatal depression encompasses major and minor depressive episodes that occur either during pregnancy or within a year after delivery. The prevalence of depression during pregnancy has been found to be close to 25%, based on self-report measures, and 12% based on diagnostic measures (Flynn 2005). Depression and PTSD around the time of pregnancy is of concern because such pathologies have been linked to numerous negative health-related behaviors and outcomes, including poor nutrition, increased substance use (including alcohol, illicit drugs, and tobacco), inadequate prenatal care, decreased fetal growth, preeclampsia, premature infant delivery, delivery of LBW infants, postnatal depression, comorbidity, lower recall of memory-related PTSD symptoms, and suicide (Barrio and Burt 2000; Hoffman and Hatch 2000; Horrigan et al. 2000; Kurki et al. 2000; Najman et al. 2000).

Intimate partner violence (IPV) is a common problem, affecting large numbers of women and girls globally. Intimate partner violence takes on many forms, including psychological, emotional, financial, physical, and sexual abuse, and its effects on the physical and mental health of victims are varied. Among women, the prevalence of physical abuse by an intimate partner ranges from 10% to 52%; whereas the prevalence of sexual abuse among women and girls ranges from 10% to 27% (Krug et al. 2002; Garcia-Moreno et al. 2006). Physical IPV is often accompanied by psychological abuse, and between 33% and 50% of victims of IPV experience both physical and sexual abuse (Watts and Zimmerman 2002;

Krug et al. 2002). Intimate partner violence is associated with depression (Campbell 2002), anxiety (Hedin et al. 1999), suicide, suicide ideation (Kernic et al. 2000), and PTSD (Woods 2000). Results from a meta-analysis of 18 studies indicated that 47.6% of abused women are depressed (Golding 1999). Yet, other studies have shown that depression and depressive symptoms are most prevalent among women abused both as children and as adults (McCauley et al. 1997). Prevalence estimates of PTSD among abused women have ranged from 16% to 85% (Golding 1999; Coker et al. 2000), compared with 5%–10% observed among women in the general population (Kessler et al. 1995). Jones et al. (2001) reported positive trends of increasing PTSD symptomatology and severity with severity of IPV. Intimate partner violence may impact maternal mental health directly through increasing psychological stress (Cohen et al. 1999). In particular, depression is known to occur among abused women because of the social, emotional, and physical isolation; separation; loss; and the unpredictability exerted by the abuser upon the abused woman (Campbell et al. 1997).

Psychiatric disorders and preterm delivery

Preterm births comprise approximately 12.5% of pregnancies in the United States. Infants born preterm (<37 weeks' gestation) are at greater risk for mortality and a wide range of medical and developmental complications when compared with infants born at term as discussed in Chapter 12. The principal pathways leading to preterm birth are spontaneous preterm labor (sPTL), preterm premature rupture of membranes (PPROM), and medical induction (MEDPTD). Associations of maternal psychosocial stress with pregnancy outcomes, including PTD, are provided in several reviews and a recent Institute of Medicine report (IOM 2006a). Maternal stress (of any severity), distress, PTSD, generalized anxiety, and depression have each been shown to be positively associated with PTD in some (Copper et al. 1996; Hedegaard et al. 1996; Hobel et al. 1999), but not all studies (Yost et al. 2005). Hedegaard et al. (1996), in their population-based prospective study, identified a dose–response relation between psychological distress during the third trimester and PTD risk. These findings of increased PTD risk with increasing severity of maternal distress and depression (Rondo et al. 2003) suggest the importance of evaluating PTD risk in relation to measures of severity and persistence of mental health disorders.

Intimate partner violence and preterm delivery

Most, although not all studies have documented positive results of IPV and adverse pregnancy outcomes, including PTD. Women who reported experiencing any IPV during pregnancy are more likely to deliver preterm than are their nonabused counterparts (Coker et al. 2004). In a study that examined the severity of violence during pregnancy and PTD, severe violence (including hitting, kicking, beating, injuring with a weapon, or injuring the abdomen) was significantly associated with PTD (RR = 2.7; 95% CI 1.7–4.4) and very PTD (RR = 4.6; 95% CI 1.6–13.6) (Covington et al. 2001). Physical abuse during pregnancy, in addition to being independently

associated with PTD, is also associated with PTD risk factors including alcohol consumption, illicit drug use, underutilization of prenatal care services, and maternal psychosocial stress (Heaman 2005). Emerging evidence suggests that IPV influences maternal physiological status via multiple chronic and acute pathophysiological pathways. For instance, some have speculated that the chronic effects of IPV may be mediated through a pathway of episodic and persistent psychological stress and distress. This thesis is supported by evidence of chronic hypothalamic-pituitary-adrenal (HPA) axis dysregulation among IPV survivors (Pico-Alfonso et al. 2004). Acute injury to the abdomen may be one pathway that contributes to the increased risk of PTD among IPV victims.

Pathophysiology of psychiatric disorders and preterm delivery

A wealth of information derived from basic, clinical, and epidemiologic studies has increased our knowledge about the multiple physiological disturbances associated with mood and anxiety disorders. Studies have repeatedly and consistently demonstrated, for example, alterations in hormonal and immune function associated with various characteristics of mood and anxiety disorders, including severity and duration of such illness in pregnant and nonpregnant patients (Wadhwa et al. 2001). Here, we provide a brief overview of these neuroendocrine and immune alterations.

Psychiatric disorders and immune/inflammatory status

Substantial evidence suggests associations between mood-anxiety disorders and altered immune function (Agarwal and Marshall 1998; Miller and Raison 2006). Depressed patients, as compared with their nondepressed counterparts, are known to be in a persistent inflammatory state, as evidenced by their higher plasma/serum concentrations of proinflammatory cytokines and acute-phase proteins. Concomitant with this proinflammatory state, depressed patients are also more likely to have higher concentrations of chemokines and cellular adhesion molecules (consistent with diffuse vascular endothelia activation) than are nondepressed patients (Maymon et al. 2000). These clinical observations are consistent with evidence from basic and clinical research that demonstrates an activation of the inflammatory response system in the contexts of acute states of anxiety, major depression (Maes 1995; Maymon et al. 2000; Raison et al. 2006), and, particularly, PTSD (Maes et al. 1999). Animal and human clinical studies have shown that proinflammatory cytokines interact with many of the pathophysiological domains that characterize depression, including neurotransmitter metabolism, neuroendocrine function, synaptic plasticity, and behavior (Maymon et al. 2000; Miller and Raison 2006). Available evidence also suggests that proinflammatory cytokines may promote HPA axis hyperactivity (frequently observed in depressive disorders). Proinflammatory cytokines are thought to cause HPA axis hyperactivity by disturbing the negative feedback inhibition of circulating corticosteroids on the HPA axis (Schiepers et al. 2005). On balance, available data from diverse sources suggest intrinsic and complex interactions between brain functioning and immune status and support the "cytokine hypothesis of depression" (Maes 1995).

Few investigators, however, have evaluated maternal antepartum immune/inflammatory status and mental health status in the same cohort.

Psychiatric disorders and neuroendocrine function

One of the most robust pathophysiological findings in the study of major depression is the increase in HPA activity. Patients with major depression exhibit decreased HPA axis sensitivity to negative feedback and higher basal cortisol levels, whereas those with PTSD, including comorbid PTSD, tend to exhibit increased HPA axis sensitivity to negative feedback and lower basal cortisol concentrations (Yehuda 2002a,b). Catecholamine and serotonin also are promising potential biomarkers of mood and anxiety disorders (Yehuda et al. 1998). Southwick et al. (1999) reviewed literature documenting the dysregulation of catecholamine and serotonin in patients with PTSD and documented symptoms such as hypervigilance, exaggerated startle, irritability, impulsivity, aggression, intrusive memories, depressed mood, and suicidality with these biomarkers.

Substantial evidence suggests that dysregulation of corticotropin-releasing hormone (CRH) may be of pathophysiological importance in PTD (Erickson et al. 2001). Serum concentrations of placental CRH are known to be elevated even at 15–20 weeks of pregnancy in women who deliver preterm (Hobel et al. 1999; Leung et al. 1999). Furthermore, CRH has been found to be a mediator between maternal prenatal anxiety and length of gestation (Mancuso et al. 2004). Maternal psychosocial stress levels at midgestation was also found to significantly predict the magnitude of increase in maternal CRH levels between mid- and later gestation (Hobel et al. 1999). A small literature suggests that CRH–PTD associations may be dependent on the chronicity of the stressor (Lockwood and Kuczynski 1999).

In summary, numerous neuroimmunologic biomarkers of PTD have been identified, and some of these biomarkers are also associated with mood and anxiety disorders. Increased understanding of the biological mechanisms through which depressive and anxiety disorders can lead to PTD may be gained from further research. At present, however, little is known about how IPV-induced mood and anxiety disorders are associated with PTD.

SLEEP DISORDER AND PREGNANCY OUTCOMES

Insufficient sleep and poor sleep quality, considered endemic in modern society, are associated with obesity, impaired glucose tolerance, and diabetes (IOM 2006b). Despite awareness of pregnancy-associated metabolic and morphological changes that contribute to poor and fragmented sleep during pregnancy (Hedman et al. 2002; Pien and Schwab 2004), relatively little, however, is known about the consequences of insufficient sleep and poor sleep quality during pregnancy on maternal cardiometabolic and psychiatric status.

Insomnia, sleep-disordered breathing (snoring), and restless legs syndrome are noted common complaints of pregnant women (Franklin et al. 2000; Hedman et al.

2002; Pien and Schwab 2004). Although prior studies of sleep and pregnancy have provided useful descriptive information concerning the prevalence of these sleep disorders, few have evaluated the contributions of short sleep duration to the incidence of obstetric complications. A number of studies have documented increased risks of gestational hypertension, preeclampsia, and fetal growth restriction with maternal self-reported snoring during pregnancy (Franklin et al. 2000; Izci et al. 2005). Recently, Koken et al. (2007) reported that some, but not all, biomarkers of oxidative stress were elevated in the pregnant women who snored during pregnancy compared with those who did not. Further, in their pilot study of 19 healthy pregnant women, Okun et al. (2007) noted that maternal self-reported short sleep duration and poor sleep efficiency in both mid and late pregnancy were associated with higher concentrations of the inflammatory cytokine, IL-6, which is predictive of PTD and other adverse perinatal outcomes. Additionally, a number of studies suggest that women's sleep patterns are greatly affected during pregnancy and the postpartum period (IOM 2006b).

Although results from these and other prior pregnancy sleep studies have provided compelling information, inferences are limited in part because of cross-sectional study design, failure to exclude women with preexisting medical conditions, small sample size, measurement errors attributable to reliance on subjective recall of sleep parameters, and failure to control for confounding. In addition to these limitations, to date, there are no studies of maternal sleep duration and sleep quality in relation to incident preeclampsia or gestational diabetes. Longitudinal cohort studies that integrate repeated objective measures of sleep duration and quality with biological markers of maternal cardiometabolic status during pregnancy are needed to move this area of perinatal research forward and to fill gaps that have been identified by the IOM (IOM 2006b).

CONCLUSION

Studies primarily conducted and published during the past two decades have confirmed that dyslipidemia, hypoleptinemia, hyperhomocyst(e)inemia, systematic inflammation, antioxidant deficiency, and alternations in endothelial cell function are evident 8–20 weeks before the clinical manifestations of common maternal metabolic complications, including preeclampsia and GDM. Investigators have increasingly turned toward assessing a broader spectrum of medical, psychosocial, and behavioral factors (e.g., sleep disorders, migraine, and exposure to violence), and they are documenting associations of these factors with maternal cardiometabolic and psychiatric health status. Collectively, this available literature motivates the next set of research aims, which ideally will involve characterization of the maternal preconceptional cardiometabolic and psychiatric profiles (i.e., establishing a true metabolic baseline) and quantification of interindividual changes from the preconceptional period throughout pregnancy and on to the postpartum period. Such studies that describe and characterize interindividual variation in maternal adaptation to pregnancy (e.g., change in lipid profile, change in inflammation status from baseline throughout pregnancy and the postpartum

period) and that evaluate how variation in adaptive responses to pregnancy are related to pregnancy outcomes (e.g., preeclampsia, GDM, migraines, and postpartum depression) will provide important new physiological, clinical, and epidemiologic information useful for developing evidence-based strategies for health promotion and disease prevention activities that specifically target reproductive-aged women and their newborns.

REFERENCES

Adeney, K.L. and Williams, M.A. 2006. Migraine headaches and preeclampsia: An epidemiologic review. *Headache* 46: 794–803.

Adeney, K.L., Williams, M.A., Miller, R.S., Frederick, I.O., Sorensen, T.K., and Luthy, D.A. 2005. Risk of preeclampsia in relation to maternal history of migraine headaches. *J Matern Fetal Neonat Med* 18: 167–72.

Agarwal, S.K. and Marshall, G.D., Jr. 1998. Glucocorticoid-induced type 1/type 2 cytokine alterations in humans: A model for stress-related immune dysfunction. *J Interferon Cytokine Res* 18: 1059–68.

Al, M., Houwelingen, A.C., Badart-Smook, A., and Honstra, G. 1998. Some aspects of neonatal essential fatty acid status are altered by linoleic supplementation of women during pregnancy. *J Nutr* 125: 2822–30.

American College of Obstetricians and Gynecologists. 2000. Report of the National High Blood Pressure Education Program working group on blood pressure in pregnancy. *Am J Obstet Gynecol* 183: S1–S2000.

American Diabetes Association. 1999. Gestational diabetes mellitus. *Diabetes Care* 22(Suppl): S74–S76.

American Diabetes Association. 2005. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 28: S37–S42.

Anderson, J.L., Waller, D.K., Canfield, M.A., Shaw, G.M., Watkins, M.L., and Werler, M.M. 2005. Maternal obesity, gestational diabetes and central nervous system birth defects. *Epidemiology* 16: 87–92.

Aube, M. 1999. Migraine in pregnancy. *Neurology* 53: S26–S28.

Barrio, L. and Burt, V. 2000. Depression in pregnancy: Strategies for primary care management. *Dealing with an underdiagnosed undertreated problem. Women's Health in Primary Care* 3: 490–98.

Beecroft, N., Cochrane, G.M., and Milburn, H.J. 1998. Effects of sex on fetus on asthma during pregnancy: Blind prospective study. *BMJ* 317: 856–57.

Ben-Haroush, A., Yogev, Y., and Hod, M. 2004. Epidemiology of gestational diabetes and its association with type 2 diabetes. *Diabet Med* 21: 103–13.

Berkowitz, G.S., Lapinski, R.H., Wein, R., and Lee, D. 1992. Race/ethnicity and other risk factors for gestational diabetes. *Am J Epidemiol* 135: 965–73.

Borchers, A.T., Naguwa, S.M., Keen, C.L., and Gershwin, M.E. 2010. The implications of autoimmunity and pregnancy. *J Autoimmun*. 34: J287–99.

Brenner, B. 2004. Haemostatic changes in pregnancy. *Thrombosis Res* 114: 409–14.

Breslau, N., Davis, G.C., Peterson, E.L., and Schultz, L.R. 2000. A second look at comorbidity in victims of trauma: The posttraumatic stress disorder-major depression connection. *Biol Psychiatry* 48: 902–909.

Brett, M. and Baxendale, S. 2001. Motherhood and memory: A review. *Psychoneuroendocrinology* 26(4): 339–62; correction, 2002; 27: 299–302.

Brunello, N., Davidson, J.R., Deahl, M., Kessler, R.C., Mendlewicz, J., Racagni, G., Shalev, A.Y., and Zohar, J. 2001. Posttraumatic stress disorder: Diagnosis and epidemiology, comorbidity and social consequences, biology and treatment. *Neuropsychobiology* 43: 150–62.

Buchanan, T.A. and Kjos, S.L. 1999. Gestational diabetes: Risk or myth? *J Clin Endocrinol Metab* 84: 1854–57.

Butler, C.L., Williams, M.A., Sorensen, T.K., Frederick, I.O., and Leisenring, W.M. 2004. Relation between maternal recreational physical activity and plasma lipids during early pregnancy. *Am J Epidemiol* 160: 350–59.

Callaghan, N. 1968. The migraine syndrome in pregnancy. *Neurology* 18: 197–99.

Campbell, J. 2002. Health consequences of intimate partner violence. *Lancet* 359: 1331–36.

Campbell, J.C., and Lewandowski, L.A. 1997. Mental and physical health effects of intimate partner violence on women and children. *Psychiatr Clin North Am* 20: 353–74.

Catalano, P.M., Tyzbir, E.D., Wolfe, R.R., Calles, J., Roman, N.M., Amini, S.B., and Sims, E.A. 1993. Carbohydrate metabolism during pregnancy in control subjects and women with gestational diabetes. *Am J Physiol*. 264(1 Pt 1): E60–E67.

Chappell, S. and Morgan, L. 2006. Searching genetic clues to the causes of pre-eclampsia. *Clin Sci* 110: 443–58.

Chen, L., Hu, F.B., Yeung, E., Willett, W., and Zhang, C. 2009. Prospective study of pre-gravid sugar-sweetened beverage consumption and the risk of gestational diabetes mellitus. *Diabetes Care* 32: 2236–2241.

Clapp, J.F., III and Capeless, E.L. 1991. The changing glycemic response to exercise during pregnancy. *Am J Obstet Gynecol* 165: 1678–83.

Clapp, J.F., III and Kiess, W. 2000. Effects of pregnancy and exercise on concentrations of the metabolic markers tumor necrosis factor α and leptin. *Am J Obstet Gynecol* 182: 300–306.

Coelingh Bennink, H.J.T. 1977. Recurrence of gestational diabetes mellitus. *Eur J Obstet Gynecol Reprod Biol* 7: 359–63.

Cohen, S., Doyle, W.J., and Skoner, D.P. 1999. Psychological stress, cytokine production, and severity of upper respiratory illness. *Psychosom Med* 61: 175–80.

Coker, A.L., Sanderson, M., Dong, B. 2004. Partner violence during pregnancy and risk of adverse pregnancy outcomes. *Paediatr Perinat Epidemiol* 18: 260–69.

Coker, A.L., Smith, P.H., McKeown, R.E., and King, M.J. 2000. Frequency and correlates of intimate partner violence by type: Physical, sexual, and psychological battering. *Am J Public Health* 90: 553–59.

Copper, R.L., Goldenberg, R.L., Das, A., Elder, N., Swain, M., Norman, G., et al. 1996. The preterm prediction study: Maternal stress is associated with spontaneous preterm birth at less than thirty-five weeks' gestation. *National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. Am J Obstet Gynecol* 175: 1286–92.

Covington, D.L., Hage, M., Hall, T., and Mathis, M. 2001. Preterm delivery and the severity of violence during pregnancy. *J Reprod Med* 46: 1031–39.

Damm, P. 1998. Gestational diabetes mellitus and subsequent development of overt diabetes mellitus. *Danish Med Bull* 45: 495–509.

Danzell, J.D. 1998. Pregnancy and pre-existing heart disease. *J Louisiana State Med Soc* 150: 97–102.

Dempsey, J.C., Butler, C.L., and Williams, M.A. 2005. No need for a pregnant pause: Physical activity may reduce the occurrence of gestational diabetes mellitus and preeclampsia. *Exerc Sport Sci Rev* 33: 141–49.

Dempsey, J.C., Sorensen, T.K., Lee, I.-M., Thompson, M.L., Miller, R.S., Frederick, I.O., and Williams, M.A. 2004b. A case-control study of maternal recreational physical activity and risk of gestational diabetes mellitus. *Diabetes Res Clin Pract* 66: 203–15.

Dempsey, J.C., Sorensen, T.K., Williams, M.A., Lee, I.-M., Miller, R.S., Dashow, E.E., and Luthy, D.A. 2004a. Prospective study of gestational diabetes mellitus in relation to maternal recreational physical activity before and during pregnancy. *Am J Epidemiol* 159: 663–70.

Desoye, G., Schweditsch, M., Pfieffer, K.P., Zechner, R., and Kostner, G.M. 1987. Correlation of hormones with lipid and lipoprotein levels during normal pregnancy and postpartum. *J Clin Endocrinol Metab* 64: 704–12.

Dildy, G.A. 3rd, Belfort, M.A., and Smulian, J.C. 2007. Preeclampsia recurrence and prevention. *Semin Perinatol* 31: 135–41.

Dombrowski, M.P. 2006. Outcomes of pregnancy in asthmatic women. *Immunol Allergy Clin North Am*. 26: 81–92.

Dorner, G., Plagemann, A., Neu, A., and Rosenbauer, J. 2000. Gestational diabetes as possible risk factor for type 1 childhood-onset diabetes in the offspring. *Neuroendocrinol Lett* 21: 355–59.

Durstine, J.L., Grandjean, P.W., Davis, P.G., Ferguson, M.A., Alderson, N.L., and DuBose, K.D. 2001. Blood lipid and lipoprotein adaptations to exercise: A quantitative analysis. *Sports Med* 31: 1033–62.

Duvekot, J.J. and Peeters, L.L.H. 1998. Very early changes in cardiovascular physiology. In: G. Chamberlain and F.B. Pipkin, eds., *Clinical physiology in obstetrics*, 3rd ed., pp 3–32. Malden, MA: Blackwell Science.

Dye, T.D., Knox, A.L., Artal, R., Aubry, R.H., and Wojtowycz, M.A. 1997. Physical activity, obesity, and diabetes in pregnancy. *Am J Epidemiol* 146: 961–65.

England, L. and Zhang, J. 2007. Smoking and risk of preeclampsia: A systematic review. *Front Biosci* 12: 2471–83.

Enquobahrie, D.A., Meller, M., Rice, K., Psaty, B.M., Siscovick, D.S., and Williams, M.A. 2008. Differential placental gene expression in preeclampsia. *Am J Obstet Gynecol* 199: 566.e1–11.

Erickson, K., Thorsen, P., Chrousos, G., Grigoriadis, D.E., Khongsaly, O., McGregor, J., and Schulkin, J. 2001. Preterm birth: Associated neuroendocrine, medical, and behavioral risk factors. *J Clin Endocrinol Metab* 86: 2544–52.

Ertresvag, J.M., Swart, J.-A., Helde, G., Johnsen, H.-J., and Bovim, G. 2005. Headache and transient focal neurological symptoms during pregnancy, a prospective cohort. *Acta Neurol Scand* 111: 233–37.

Facchinetti, F., Allais, G., Nappi, R.E., D'Amico, R., Marozio, L., Bertozzi, L., et al. 2009. Migraine is a risk factor for hypertensive disorders in pregnancy: A prospective cohort study. *Cephalalgia* 29: 286–92.

Fahraeus, L., Larsson-Cohn, U., and Wallentin, L. 1985. Plasma lipoproteins including high density lipoprotein subfractions during normal pregnancy. *Obstet Gynecol* 66: 468–72.

Ferrara, A., Kahn, H.S., Quesenberry, C.P., Riley, C., and Hedderson, M.M. 2004. An increase in the incidence of gestational diabetes mellitus: Northern California, 1991–2000. *Obstet Gynecol* 103: 526–33.

Flynn, H.A. 2005. Epidemiology and phenomenology of postpartum mood disorders. *Psychiatr Ann* 35: 544–51.

Foster-Powell, K.A. and Cheung, N. 1998. Recurrence of gestational diabetes. *Aust N Zealand J Obstet Gynaecol* 38: 384–87.

Fox, A.W., Chambers, C.D., Anderson, P.O., Diamond, M.L., and Spierings, E.L.H. 2002. Evidence-based assessment of pregnancy outcome after sumatriptan exposure. *Headache* 42: 8–15.

Franklin, K.A., Holmgren, P.A., Jonsson, F., Poromaa, N., Stenlund, H., and Svanborg, E. 2000. Snoring, pregnancy-induced hypertension, and growth retardation of the fetus. *Chest* 117: 137–41.

Freedman, S.A., Brandes, D., Peri, T., and Shalev, A. 1999. Predictors of chronic post-traumatic stress disorder. *A prospective study. Br J Psychiatry* 174: 353–59.

Gabbe, S.G. and Graves, C.R. 2003. Management of diabetes mellitus complicating pregnancy. *Obstet Gynecol* 102: 857–68.

Garcia-Moreno, C., Jansen, H.A., Ellsberg, M., Heise, L., and Watts, C.H. 2006. Prevalence of intimate partner violence: Findings from the WHO multi-country study on women's health and domestic violence. *Lancet* 368: 1260–69.

Garovic, V.D. and Hayman, S.R. 2007. Hypertension in pregnancy: An emerging risk factor for cardiovascular disease. *Nat Clin Pract Nephrol* 3: 613–22.

Gluck, J.C. and Gluck, P.A. 2006. The effect of pregnancy on the course of asthma. *Immunol Allergy Clin N Am* 26: 63–80.

Golding, J. 1999. Intimate partner violence as a risk factor for mental disorders: A meta-analysis. *J Fam Violence* 14: 99–132.

GOPEC Consortium. 2005. Disentangling fetal and maternal susceptibility for pre-eclampsia: A British multicenter candidate-gene study. *Am J Hum Genet* 77: 127–31.

Harizopoulou, V.C., Kritikos, A., Papanikolaou, Z., Saranti, E., Vavilis, D., Klonos, E., et al. 2009. Maternal physical activity before and during early pregnancy as a risk factor for gestational diabetes mellitus. *Acta* Diabetol DOI: 10.1007/s00592-009-0136-1 [Epub ahead of print].

Hayashi, M., Inoue, T., Hoshimoto, K., Negishi, H., Ohkura, T., and Inaba, N. 2002. Characterization of five marker levels of the hemostatic system and endothelial status in normotensive pregnancy and pre-eclampsia. *Eur J Haematol* 69: 297–302.

Heaman, M.I. 2005. Relationships between physical abuse during pregnancy and risk factors for preterm birth among women in Manitoba. *J Obstet Gynecol Neonatal Nurs* 34: 721–31.

Hedderson, M.M., Ferrara, A., and Sacks, D.A. 2003. Gestational diabetes mellitus and lesser degrees of pregnancy hyperglycemia: Association with increased risk of spontaneous preterm birth. *Obstet Gynecol* 102: 850–56.

Hedegaard, M., Henriksen, T.B., Secher, N.J., Hatch, M.C., and Sabroe, S. 1996. Do stressful life events affect duration of gestation and risk of preterm delivery? *Epidemiology* 7: 339–45.

Hedin, L.W., Grimstad, H., Moller, A., Schei, B., and Janson, P.O. 1999. Prevalence of physical and sexual abuse before and during pregnancy among Swedish couples. *Acta Obstet Gynecol Scand* 78: 310–15.

Hedman, C., Pohjasvaara, T., Tolonen, U., Suhonen-Malm, A.S., and Myllyla, V.V. 2002. Effects of pregnancy on mothers' sleep. *Sleep Med* 3: 37–42.

Hobel, C.J., Dunkel-Schetter, C., Roesch, S.C., Castro, L.C., and Arora, C.P. 1999. Maternal plasma corticotropin-releasing hormone associated with stress at 20 weeks' gestation in pregnancies ending in preterm delivery. *Am J Obstet Gynecol* 180(1 Pt 3): S257–S263.

Hoffman, S. and Hatch, M.C. 2000. Depressive symptomatology during pregnancy: Evidence for an association with decreased fetal growth in pregnancies of lower social class women. *Health Psychol* 19: 535–43.

Horrigan, T.J., Schroeder, A.V., and Schaffer, R.M. 2000. The triad of substance abuse, violence, and depression are interrelated in pregnancy. *J Subst Abuse Treat* 18: 55–58.

Institute of Medicine. 2006a. *Preterm birth: Causes, consequences, and prevention*. Washington, DC: National Academy of Sciences Press.

Institute of Medicine, Committee on Sleep Medicine and Research. 2006b. *Sleep disorders and sleep deprivation: An unmet public health problem*. Washington, DC: National Academy of Sciences Press.

Izci, B., Martin, S.E., Dundas, K.C., Liston, W.A., Calder, A.A., and Douglas, N.J. 2005. Sleep complaints: Snoring and daytime sleepiness in pregnant and pre-eclamptic women. *Sleep Med* 6: 163–69.

Jones, L., Hughes, M., and Unterstaller, U. 2001. Post-traumatic stress disorder (PTSD) in victims of domestic violence: A review of the research. *Trauma Violence Abuse* 2: 99–119.

Jovanovic-Peterson, L., Durak, E.P., and Peterson, C.M. 1989. Randomized trial of diet versus diet plus cardiovascular conditioning on glucose levels in gestational diabetes. *Am J Obstet Gynecol* 161: 415–19.

Kallen, B. and Lygner, P.E. 2001. Delivery outcome in women who used drugs for migraine during pregnancy with special reference to sumatriptan. *Headache* 41: 351–56.

Katz, R., Karliner, J.S., and Resnik, R. 1978. Effects of a natural volume overload state (pregnancy) on left ventricular performance in normal human subjects. *Circulation* 58: 434–41.

Kernic, M.A., Wolf, M.E., and Holt, V.L. 2000. Rates and relative risk of hospital admission among women in violent intimate partner relationships. *Am J Public Health* 90: 1416–20.

Kessler, R.C., Sonnega, A., Bromet, E., Hughes, M., and Nelson, C.B. 1995. Posttraumatic stress disorder in the National Comorbidity Survey. *Arch Gen Psychiatry* 52: 1048–60.

Kimerling, R. 2004. An investigation of sex differences in nonpsychiatric morbidity associated with posttraumatic stress disorder. *J Am Med Womens Assoc* 59: 43–47.

Kjos, S.L., Peters, R.K., Xiang, A., Henry, O.A., Montoro, M., and Buchanan, T.A. 1995. Predicting future diabetes in Latino women with gestational diabetes. *Utility of early postpartum glucose tolerance testing. Diabetes* 44: 586–95.

Knopp, R.H., Bergelin, R.O., Wahl, P.W., and Walden, C.E. 1985. Effects of pregnancy, postpartum lactation and oral contraceptive use on the lipoprotein cholesterol/triglyceride ratio. *Metabolism* 34: 893–99.

Knopp, R.H., Bergelin, R.O., Wahl, P.W., Walden, C.E., Chapman, M., and Irvine, S. 1982. Population-based lipoprotein lipid reference values for pregnant women compared to nonpregnant women classified by sex hormone status. *Am J Obstet Gynecol* 143: 626–37.

Koken, G., Sahin, F.K., Cosar, E., Saylan, F., Yilmaz, N., Altuntas, I., et al. 2007. Oxidative stress markers in pregnant women who snore and fetal outcome: A case control study. *Acta Obstet Gynecol Scand* 86: 1317–21.

Krug, E.G., Mercy, J.A., Dahlberg, L.L., and Zwi, A.B. 2002. The world report on violence and health. *Lancet* 360: 1083–88.

Kruit, M.C., van Buchem, M.A., Hofman, P.A.M., Bakkers, J.T.N., Terwindt, G.M., Ferrari, M.D., and Launer, L.J. 2004. Migraine as a risk factor for subclinical brain lesions. *JAMA* 291: 427–34.

Kurki, T., Hiilesmaa, V., Raitasalo, R., Mattila, H., and Ylikorkala, O. 2000. Depression and anxiety in early pregnancy and risk for preeclampsia. *Obstet Gynecol* 95: 487–90.

Lain, K., Speer, P., McGonigal, S., and Roberts, J. 2005. Early pregnancy adiponectin concentrations are associated with gestational diabetes. *[Abstract] Am J Obstet Gynecol* 191(Suppl): S52.

Lauenborg, J., Hansen, T., Jensen, D.M., Vestergaard, H., Molsted-Pedersen, L., Hornnes, P., et al. 2004. Increasing incidence of diabetes after gestational diabetes. *Diabetes Care* 27: 1194–99.

Leonardi, M. and Mathers, C.D. 2003. Global burden of migraine in the year 2000: Summary of methods and data sources. GBD 2000 Working Paper. Geneva: World Health Organization. Online at: http:www.who.int/evidence/bod.

Leung, T.N., Chung, T.K., Madsen, G., McLean, M., Chang, A.M., and Smith, R. 1999. Elevated mid-trimester maternal corticotrophin-releasing hormone levels in pregnancies that delivered before 34 weeks. *Br J Obstet Gynaecol* 106: 1041–46.

Lipton, R.B., Stewart, W.F., Diamond, S., Diamond, M.L., and Reed, M. 2001. Prevalence and burden of migraine in the United States: Data from the American Migraine Study II. *Headache* 41: 646–57.

Lockwood, C.J. and Kuczynski, E. 1999. Markers of risk for preterm delivery. *J Perinat Med* 27: 5–20.

Maes, M. 1995. Evidence for an immune response in major depression: A review and hypothesis. *Prog Neuropsychopharmacol Biol Psychiatry* 19: 11–38.

Maes, M., Lin, A.H., Delmeire, L., Van Gastel, A., Kenis, G., De Jongh, R., and Bosmans, E. 1999. Elevated serum interleukin-6 (IL-6) and IL-6 receptor concentrations in posttraumatic stress disorder following accidental man-made traumatic events. *Biol Psychiatry* 45: 833–39.

Maffei, M., Volpe, L., Di Cianni, G., Bertacca, A., Ferdeghini, M., Murru, S., et al. 1998. Plasma leptin levels in newborns from normal and diabetic mother. *Horm Metab Res* 30: 575–80.

Mancuso, R.A., Schetter, C.D., Rini, C.M., Roesch, S.C., and Hobel, C.J. 2004. Maternal prenatal anxiety and corticotropin-releasing hormone associated with timing of delivery. *Psychosom Med* 66: 762–69.

Marcus, D.A. 2003. Headache in pregnancy. *Curr Pain Headache Rep* 7: 288–96.

Mason, E., Rosene-Montella, K., and Powrie, R. 1998. Medical problems during pregnancy. *Med Clin N Am* 82: 249–69.

Matharu, M.S., Levy, M.J., and Goadsby, P.J. 2002. Understanding migraine in women. *Practitioner* 246: 272–78.

Maymon, E., Romero, R., Pacora, P., Gomez, R., Athayde, N., Edwin, S., and Yoon, B.H. 2000. Human neutrophil collagenase (matrix metalloproteinase 8) in parturition, premature rupture of the membranes, and intrauterine infection. *Am J Obstet Gynecol* 183: 94–99.

McCauley, J., Kern, D.E., Kolodner, K., Dill, L., Schroeder, A.F., DeChant, H.K., et al. 1997. Clinical characteristics of women with a history of childhood abuse: Unhealed wounds. *JAMA* 277(17): 1362–68.

McMurray, R.G., Mottola, M.F., Wolfe, L.A., Artal, R., Millar, L., and Pivarnik, J.M. 1993. Recent advances in understanding maternal and fetal responses to exercise. *Med Sci Sports Exerc* 25: 1305–21.

Mertzer, B.E., Cho, N.H., Roston, S.M., and Radvany, R. 1993. Pre-pregnancy weight and antepartum insulin secretion predicts glucose tolerance five years after gestational diabetes mellitus. *Diabetes Care* 16: 1598–1605.

Metcalf, J., McAnulty, J.H., and Ueland, K. 1981. Cardiovascular physiology. *Clin Obstet Gynecol* 24: 693–710.

Meza, E., Barraza, L., Martínez, G., Fernández, V., Ramos-Jáquez, E., Cano-Vargas, C., et al. 1995. Gestational diabetes in a Mexican-U.S. *Border population: Prevalence and epidemiology. Rev Invest Clin* 47: 433–38.

Miller, A.H. and Raison, C.L. 2006. Cytokines, p38 MAP kinase and the pathophysiology of depression. *Neuropsychopharmacology* 31: 2089–90.

Montes, A., Walden, C.E., Knopp, R.H., Cheung, M., Chapman, M.B., and Albers, J.J. 1984. Physiologic and supraphysiologic increases in lipoprotein lipids and apoproteins in late pregnancy and postpartum: Possible markers for the diagnosis of "prelipemia." *Arteriosclerosis* 4: 407–417.

Murphy, V.E., Clifton, V.L., and Gibson, P.G. 2006. Asthma exacerbations during pregnancy: Incidence and association with adverse pregnancy outcomes. *Thorax* 61: 169–76.

Najman, J.M., Andersen, M.J., Bor, W., O'Callaghan, M.J., and Williams, G.M. 2000. Postnatal depression-myth and reality: Maternal depression before and after the birth of a child. *Soc Psychiatry Psychiatr Epidemiol* 35: 19–27.

National High Blood Pressure Education Program. 2000. *Working group report on high blood pressure in pregnancy*. NIH Publication No. 00-30, July 29.

Nelson-Piercy, C. and Moore, G.J. 1996. Asthma in pregnancy. *Br J Hospital Med* 55: 115–17.

Newcomer, J.W., Craft, S., Hershey, T., Askins, K., and Bardgett, M.E. 1994. Glucocorticoid-induced impairment in declarative memory performance in adult humans. *J Neurosci* 14: 2047–53.

Ning, Y., Williams, M.A., Dempsey, J.C., Sorensen, T.K., Frederick, I.O., and Luthy, D.A. 2003. Correlates of recreational physical activity in early pregnancy. *J Matern Fetal Neonat Med* 13: 385–93.

O'Hara, M.W. 1995. Postpartum depression. In: L.B. Alloy, ed. *Series in psychopathology*, pp. 1–27. New York: Springer-Verlag.

Okun, M.L., Hall, M., and Coussons-Read, M.E. 2007. Sleep disturbances increase interleukin-6 production during pregnancy: Implications for pregnancy complications. *Reprod Sci* 14: 560–67.

Olesen, C., Steffensen, F.H., Sorensen, H.T., Nielsen, G.L., and Olsen, J. 2000. Pregnancy outcome following prescription for sumatriptan. *Headache* 40: 20–24.

O'Sullivan, J.B. 1984. Long term follow-up of gestational diabetes. In: R.A. Camerini Davalos and H.S. Cole, eds. *Early diabetes*, pp. 1009–1027. New York: Academic Press.

Peters, R.K., Kjos, S.L., Xiang, A., and Buchanan, T.A. 1996. Long-term diabetogenic effect of single pregnancy in women with previous gestational diabetes mellitus. *Lancet* 347: 227–30.

Pico-Alfonso, M.A., Garcia-Linares, M.I., Celda-Navarro, N., Herbert, J., and Martinez, M. 2004. Changes in cortisol and dehydroepiandrosterone in women victims of physical and psychological intimate partner violence. *Biol Psychiatry* 56: 233–40.

Pien, G.W. and Schwab, R.J. 2004. Sleep disorders during pregnancy. *Sleep* 27: 1405–17.

Potter, J.M. and Nestel, P.J. 1979. The hyperlipidemia of pregnancy in normal and complicated pregnancies. *Am J Obstet Gynecol* 15: 165–70.

Qiu, C., Williams, M.A., Vadachkoria, S., Frederick, I.O., and Luthy, D.A. 2004. Increased plasma leptin in early pregnancy is predictive of gestational diabetes mellitus. *Obstet Gynecol* 103: 519–25.

Raison, C.L., Capuron, L., and Miller, A.H. 2006. Cytokines sing the blues: Inflammation and the pathogenesis of depression. *Trends Immunol* 27: 24–31.

Ranheim, T., Haugen, F., Staff, A.C., Braekke, K., Harsem, N.K., and Drevon, C.A. 2004. Adiponectin is reduced in gestational diabetes mellitus in normal weight women. *Acta Obstet Gynecol Scand* 83: 341–47.

Retnakaran, R., Hanley, A.J., Raif, N., Connelly, P.W., Sermer, M., and Zinman, B. 2004. Hypoadiponectinaemia in South Asian women during pregnancy: Evidence of ethnic variation in adiponectin concentration. *Diabet Med* 21: 388–92.

Robson, S.C., Hunter, S., Boys, R.J., and Dunlop, W. 1989. Serial study of factors influencing changes in cardiac output during human pregnancy. *Am J Physiol* 256(4 Pt 2): H1060–H1065.

Rondo, P.H., Ferreira, R.F., Nogueira, F., Ribeiro, M.C., Lobert, H., and Artes, R. 2003. Maternal psychological stress and distress as predictors of low birth weight, prematurity and intrauterine growth retardation. *Eur J Clin Nutr* 57: 266–72.

Rosamond, W. 2004. Are migraine and coronary heart disease associated? An epidemiologic review. *Headache* 44(Suppl 1): S5–S12.

Rosso, P. 1990. *Nutrition and metabolism in pregnancy: Mother and fetus* pp. 41–65. New York: Oxford University Press.

Russell, M.B., Ulrich, V., Gervil, M., and Olesen, J. 2002. Migraine without aura and migraine with aura are distinct disorders. *A population-based twin survey. Headache* 42: 332–36.

Saldana, T.M., Siega-Riz, A.M., and Adair, L.S. 2004. Effect of macronutrient intake on the development of glucose intolerance during pregnancy. *Am J Clin Nutr* 79: 479–86.

Sanchez, S.E., Qiu, C., Williams, M.A., Lam, N., and Sorensen, T.K. 2008. Headaches and migraines are associated with an increased risk of preeclampsia in Peruvian women. *Am J Hypertens* 21: 360–64.

Sargent, I.L., Arenas, J., and Redman, C.W. 1987. Maternal cell-mediated sensitization to paternal HLA may occur, but is not a regular event in normal human pregnancy. *J Reprod Immunol* 10: 111–20.

Schatz, M. and Dombrowski, M.P. 2009. Clinical practice. *Asthma in pregnancy. N Engl J Med* 360: 1862–69.

Scher, A.I., Terwindt, G.M., Picavet, H.S.J., Verschuren, W.M.M., Ferrari, M.D., and Launer, L.J. 2005. Cardiovascular risk factors and migraine: The GEM population-based study. *Neurology* 64: 614–20.

Schiepers, O.J., Wichers, M.C., and Maes, M. 2005. Cytokines and major depression. *Prog Neuropsychopharmacol Biol Psychiatry* 29: 201–17.

Shalev, A.Y., Freedman, S., Peri, T., Brandes, D., Sahar, T., Orr, S.P., and Pitman, R.K. 1998. Prospective study of posttraumatic stress disorder and depression following trauma. *Am J Psychiatry* 155: 630–37.

Siddiqui, S., Goodman, N., McKenna, S., Goldie, M., Waugh, J., and Brightling, C.E. 2008. Pre-eclampsia is associated with airway hyperresponsiveness. *BJOG* 115: 520–22.

Skjaerven, R., Vatten, L.J., Wilcox, A.J., Rønning, T., Irgens, L.M., and Lie, R.T. 2005. Recurrence of pre-eclampsia across generations: Exploring fetal and maternal genetic components in a population based cohort. *BMJ* 331: 877.

Solomon, C.G., Willett, W.C., Carey, V.J., Rich-Edwards, J., Hunter, D.J., Colditz, G.A., et al. 1997. A prospective study of pregravid determinants of gestational diabetes mellitus. *JAMA* 278: 1078–83.

Southwick, S.M., Paige, S., Morgan, C.A., 3rd, Bremner, J.D., Krystal, J.H., and Charney, D.S. 1999. Neurotransmitter alterations in PTSD: Catecholamines and serotonin. *Semin Clin Neuropsychiatry* 4: 242–48.

Stewart, W.F., Lipton, R.B., and Liberman, J. 1996. Variation in migraine prevalence by race. *Neurology* 47: 52–59.

Vitoratos, N., Kassanos, D., Salamalekis, E., Sirisratidis, C., Baimacou, D., and Creatsas, G. 2002. Maternal homocysteine levels and plasma lipids in gestational diabetes: Is there any relationship? *J Obstet Gynecol* 22: 366–69.

Wadhwa, P.D., Culhane, J.F., Rauh, V., Barve, S.S., Hogan, V., Sandman, C.A., et al. 2001. Stress, infection and preterm birth: A biobehavioural perspective. *Paediatr Perinat Epidemiol* 15(Suppl 2): 17–29.

Ward, K. 2008. Searching for genetic factors underlying pre-eclampsia: Recent progress and persistent challenges. *Minerva Ginecol* 60: 399–419.

Watts, C. and Zimmerman, C. 2002. Violence against women: Global scope and magnitude. *Lancet* 359: 1232–37.

Wegmann, T.G., Lin, H., Guilbert, L., and Mosmann, T.R. 1993. Bidirectional cytokine interactions in the maternal-fetal relationship: Is successful pregnancy a TH2 phenomenon? *Immunol Today* 14: 353–56.

Welch, K.M. 2003. Contemporary concepts of migraine pathogenesis. *Neurology* 61: S2–S8.

Williams, M.A., Emanuel, I., Kimpo, C., Leisenring, W.M., and Hale, C.B. 1999. A population-based cohort study of the relation between maternal birth weight and risk of gestational diabetes mellitus in four race/ethnic groups. *Paediatr Perinat Epidemiol* 13: 452–65.

Williams, M.A. and Mittendorf, R. 2000. Maternal Morbidity. In: M.B. Goldman and M. Hatch, eds., *Women and health*, pp. 172–81. New York: Academic Press.

Williams, M.A., Qiu, C., Dempsey, J.C., and Luthy, D.A. 2003. Familial aggregation of type 2 diabetes and chronic hypertension in women with gestational diabetes. *J Reprod Med* 48: 955–62.

Williams, M.A., Qiu, C.F., Muy-Rivera, M., Vadachkoria, S., Song, T., and Luthy, D.A. 2004. Plasma adiponectin concentrations in early pregnancy and subsequent risk of gestational diabetes mellitus. *J Clin Endocrinol Metab* 17: 574–81.

Wilson, B.J., Watson, M.S., Prescott, G.J., Sunderland, S., Campbell, D.M., Hannaford, P., and Smith, W.C. 2003. Hypertensive diseases of pregnancy and risk of hypertension and stroke in later life: Results from cohort study. *BMJ* 326: 845–52.

Witlin, A.G. 1997. Asthma in pregnancy. *Semin Perinatol* 21: 284–97.

Woods, S.J. 2000. Prevalence and patterns of posttraumatic stress disorder in abused and post-abused women. *Issues Ment Health Nurs* 21: 309–24.

Yehuda, R. 2002a. Post-traumatic stress disorder. *N Engl J Med* 346: 108–14.

Yehuda, R. 2002b. Current status of cortisol findings in post-traumatic stress disorder. *Psychiatr Clin N Am* 25: 341–68, vii.

Yehuda, R., Siever, L.J., Teicher, M.H., Levengood, R.A., Gerber, D.K., Schmeidler, J., and Yang, R.K. 1998. Plasma norepinephrine and 3-methoxy-4-hydroxyphenylglycol concentrations and severity of depression in combat posttraumatic stress disorder and major depressive disorder. *Biol Psychiatry* 44: 56–63.

Yost, N.P., Bloom, S.L., McIntire, D.D., and Leveno, K.J. 2005. A prospective observational study of domestic violence during pregnancy. *Obstet Gynecol* 106: 61–65.

Zhang, C., Liu, S., Solomon, C.G., Hu, F.B. 2006a. Dietary fiber intake, dietary glycemic load, and the risk for gestational diabetes mellitus. *Diabetes Care* 29: 2223–30.

Zhang, C., Solomon, C.G., Manson, J.E., and Hu, F.B. 2006b. A prospective study of pregravid physical activity and sedentary behaviors in relation to the risk of gestational diabetes mellitus. *Arch Intern Med* 166: 543–48.

Zhang, C., Qiu, C., Hu, F.B., David, R.M., van Dam, R.M., Bralley, A., and Williams, M.A. 2008. Maternal plasma 25-hydroxyvitamin D concentrations and the risk for gestational diabetes mellitus. *PLoS One* 3: e3753.

Zhang, C., Williams, M.A., Frederick, I.O., King, I.B., Sorensen, T.K., Kestin, M.M., et al. 2004b. A case-control study of vitamin C and risk of gestational diabetes mellitus. *J Reprod Med* 49: 257–66.

Zhang, C., Williams, M.A., Sorensen, T.K., King, I.B., Kestin, M.M., Thompson, M.L., et al. 2004a. Maternal plasma ascorbic acid (Vitamin C) concentrations and risk of gestational diabetes mellitus. *Epidemiology* 15: 597–604.

Zhang, J., Zeisler, J., Hatch, M.C., and Berkowitz, G. 1997. Epidemiology of pregnancy-induced hypertension. *Epidemiol Rev* 19: 218–32.

7

Obstetric Intervention

JUN ZHANG, K.S. JOSEPH, AND WILLIAM D. FRASER

Obstetric intervention refers to a range of therapeutic measures directed at safeguarding or improving the health of the pregnant woman and the fetus. The gamut of interventions available in contemporary obstetric practice include Rhesus (Rh) immune globulin prophylaxis (to prevent Rh hemolytic disease of the newborn), prenatal diagnosis of congenital anomalies (with pregnancy termination or fetal therapy being the principal therapeutic options), and iatrogenic early delivery given fetal compromise or maternal indication (effected through labor induction or cesarean delivery). These and other interventions have had a substantial impact on maternal, fetal, and infant health in recent decades. The introduction of Rh immune globulin prophylaxis has been referred to as the greatest obstetric achievement of the previous quarter century (Gravenhorst 1989), whereas prenatal diagnosis and pregnancy termination for serious congenital anomalies has reduced the birth prevalence of congenital anomalies (Roberts 1995; Wortelboer 2000; Collins 2008; Crider 2008), impacted overall infant mortality (Liu 2002; Davidson 2005) and, most recently, displaced congenital anomalies as the leading cause of infant death in some countries (Public Health Agency of Canada 2008). On the other hand, recent dramatic increases in labor induction and cesarean delivery have been received with some ambivalence both by the public and by the medical community (Rasmussen 1996; Kitzinger 2006; Klein 2006; Joseph 2007; Malloy 2008). This chapter provides a brief overview of prenatal screening for congenital anomalies, discusses the role of early delivery in modern obstetrics, and provides details regarding trends in labor induction and cesarean delivery and their impact on perinatal outcomes.

PRENATAL SCREENING

The history of prenatal screening

Organized programs of prenatal screening for the early diagnosis of Down syndrome were introduced in industrialized countries in the 1970s, when the safety of amniocentesis was considered sufficient to warrant its use among older women. The program involved identifying women at high risk for Down syndrome infants (i.e., older women) and confirming the diagnosis of Down syndrome by amniocentesis.

A maternal age cutoff of 35 years was chosen to identify women at high risk for Down syndrome infants because it yielded a screen positive rate of 5% (5% of pregnant women at that time were ≥35 years). However, the population impact of this strategy was limited. Although women aged 35 years or older have a relatively high risk of bearing a Down syndrome infant, 95% of cases occur in women under 35 years of age (Bell 1986; Rappaport 2008). A reluctance among older women to partake in such screening also contributed to the limited population impact.

Screening for open neural tube defects using maternal serum α-fetoprotein was introduced in the 1980s. α-Fetoprotein typically enters the maternal circulation after 12 weeks' gestation, and high concentrations are observed in maternal serum if fetal body wall defects allow excess protein to leak into the amniotic fluid. Levels of serum α-fetoprotein vary by laboratory, maternal race, weight, diabetic status, number of fetuses, and gestational age. The test is typically performed between 14 and 22 weeks' gestation and, given laboratory variation, is reported in multiples of the median (MoM) value observed for the particular population (Cunningham 2005). Maternal serum α-fetoprotein levels over 3.5 MoM indicate an increased fetal risk for an open neural tube defect, and values between 2.0 and 3.5 MoM require repeat testing.

Contemporary prenatal screening

The widespread availability of various biochemical markers for congenital abnormalities and routine ultrasonography has revolutionized prenatal diagnosis. Current prenatal screening programs are multifaceted and include preconception testing, maternal serum screening using several different biochemical markers, and fetal evaluation by targeted ultrasonography, with antenatal screening administered in the first and second trimesters of pregnancy.

Preconception testing typically serves to identify genetic diseases among high-risk groups, such as those with a family history of specific malformations and ethnic groups that have a genetic predisposition for particular diseases. For instance, the American Congress of Obstetricians and Gynecologists (ACOG) recommends that screening for Tay-Sachs disease, Canavan disease, cystic fibrosis, and familial dysautonomia be offered to couples of Eastern European Jewish descent (ACOG 2004).

Maternal serum screening for Down syndrome typically includes testing for biochemical markers, such as pregnancy-associated plasma protein A (PAPP-A) and the free β subunit of human chorionic gonadotropin (fβ-hCG) during the first trimester, and screening for serum α-fetoprotein, total human chorionic gonadotropin (hCG), unconjugated estriol, and inhibin A during the second trimester. Assessment of risk is based on the quantitative values of these markers, along with maternal age. Ultrasonographic screening for Down syndrome involves measurement of nuchal translucency (thickness of the echolucent area in the posterior aspect of the neck of the fetus, reported in multiples of the median) during the first trimester of pregnancy. Ultrasonography is also used to identify various structural abnormalities in the early second trimester, including those of the central

nervous system, heart, soft tissue, gastrointestinal tract, abdominal wall, and genitourinary tract.

Large studies have attempted to identify the optimal algorithm that incorporates available screening tests for Down syndrome and other aneuploidy (abnormal number of chromosomes, such as trisomy 18). The First Trimester Maternal Serum Biochemistry and Fetal Nuchal Translucency Screening Study (Wapner 2003) used maternal age, serum levels of fβ-hCG and PAPP-A, and nuchal translucency among women with gestational ages between 74 and 98 days to identify the risk for Down syndrome and trisomy 18. The final diagnosis of Down syndrome or trisomy 18 was based on fetal karyotyping or an evaluation of the phenotype at birth. This study reported a 5% false-positive rate, and a 78.7% (95% confidence interval [CI] 66.3%–88.1%) detection rate for Down syndrome. The detection rate for trisomy 18 was 90.9% (95% CI 58.7%–99.8%) with a false-positive rate of 2%.

The First- and Second-Trimester Evaluation of Risk Trial (Malone 2005) contrasted several screening algorithms to determine the optimal screening regimen for Down syndrome. Tests included measurement of nuchal translucency, PAPP-A, and fβ-hCG between 10 weeks and 3 days through 13 weeks and 6 days of gestation; and α-fetoprotein, total hCG, unconjugated estriol, and inhibin A between 15 and 18 weeks. The highest detection rate (96%, 95% CI 92%–97% for a false-positive rate of 5%) was obtained with the fully integrated screening (i.e., PAPP-A and nuchal translucency measured at 11 weeks' gestation, the quadruple markers (mentioned above) measured between 15 to 17 weeks, and each women provided with a single risk estimate based on all test results combined). The detection and false-positive rates were not substantially different following stepwise sequential screening (i.e., PAPP-A, fβ-hCG, and nuchal translucency in the first trimester; quadruple screening in the second trimester, with results of first trimester tests provided in the first trimester and combined first and second trimester test results provided after quadruple screening): detection rate 95% (95% CI 91%–97%) at a false-positive rate of 4.9%. The choice between these approaches to prenatal screening depends on cost considerations, patient preference, and the availability of newer biomarkers.

Confirmatory (diagnostic) tests

In routine prenatal care, the results of prenatal screening for aneuploidy are typically confirmed through fetal karyotyping, with fetal cells obtained by amniocentesis (amniotic fluid sampling by ultrasound-guided, transabdominal needle aspiration of the amniotic sac) or chorionic villus sampling (CVS, transabdominal or transcervical sampling of placental chorionic villi). Amniocentesis is preferably performed in the second trimester, whereas CVS is usually carried out between 10 and 13 weeks' gestation. Mid-trimester amniocentesis (between 14 and 20 weeks) is safer than early amniocentesis (at 11–14 weeks) and CVS, although CVS has the advantage of providing results earlier in pregnancy and thereby reducing parental anxiety and enabling earlier/safer termination of pregnancy, if necessary (Cunningham 2005). Other methods for obtaining fetal cells include

cordocentesis (sampling of fetal blood from the umbilical cord) and fetal tissue biopsy. Currently, techniques are being developed to harvest fetal cells from the maternal circulation for diagnosing single-gene disorders and for aneuploidy screening (South 2008).

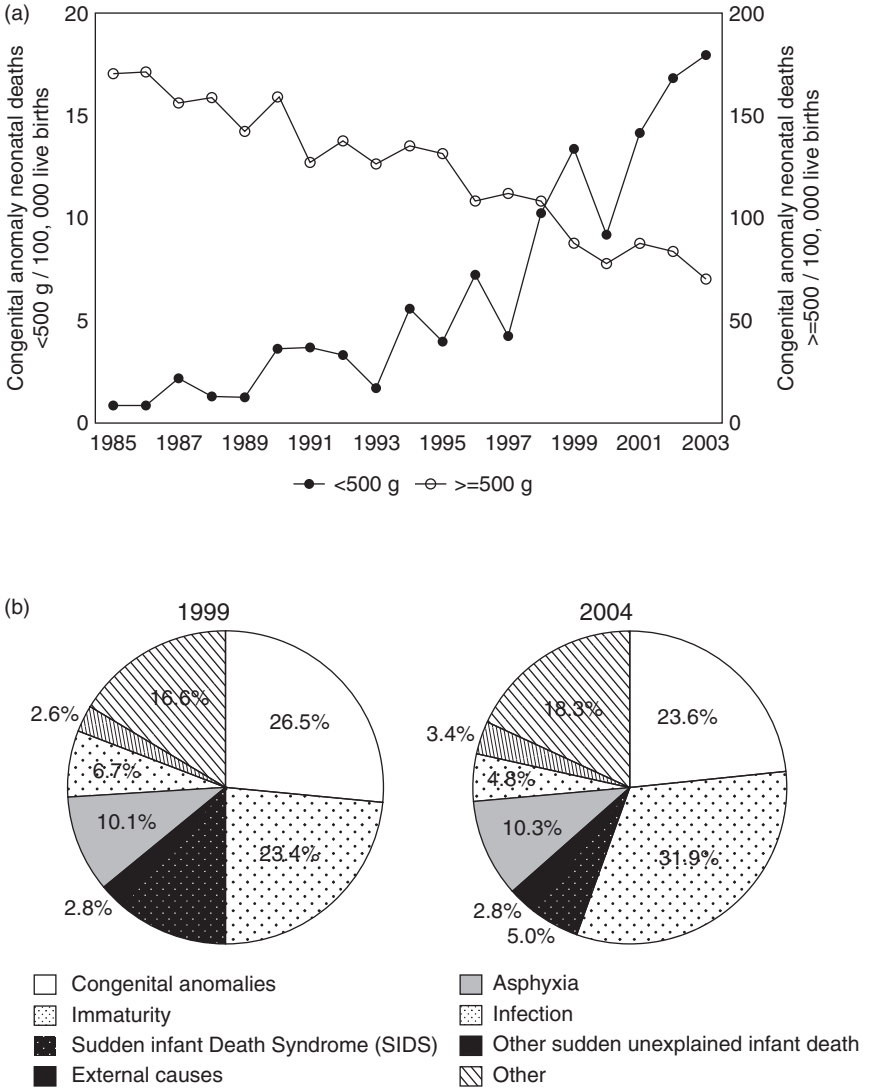
Therapeutic options

Although fetal blood transfusions and fetal surgery are infrequently used, and therapies such as stem cell transplantation and gene transfer remain in the experimental realm, prenatal diagnosis does enable the scheduling of elective delivery for fetuses who require surgery immediately after birth. The more common benefit following prenatal diagnosis of serious congenital anomalies is the option it provides parents who wish to terminate the pregnancy at an early gestation.

Population impact

Prenatal diagnosis and pregnancy termination have had an impact on the birth prevalence of specific congenital anomalies. For instance, the birth prevalence of neural tube defects decreased significantly well before the introduction of folic acid fortification of food in 1998 in Canada and the United States (Roberts 1995; Public Health Agency of Canada 2008). Changes in the birth prevalence of Down syndrome have been variable, given the increases in older maternal age in most industrialized countries. The birth prevalence of some serious cardiac conditions, such as hypoplastic left heart syndrome, has also declined substantially (Allan 1991; Bull 1999). The rate at which cardiac malformations are being diagnosed prenatally has increased, whereas the gestational age at diagnosis has decreased. In Paris, France, for instance, the rate of prenatal diagnosis for congenital heart disease increased twofold between 1983–1988 and 1995–2000, with a median gestational age at diagnosis declining from 24 to 22 weeks (Khoshnood 2005). The percentage of cases with hypoplastic left heart syndrome that was diagnosed prenatally increased from 31.8% in 1983–1988 to 88.9% in 1995–2000; 13.6% and 63.0% of such pregnancies were prenatally terminated in the two periods.

Prenatal diagnosis and pregnancy termination has had a dual impact on overall perinatal and infant mortality (Davidson 2005; Public Health Agency of Canada 2008). In countries such as Canada and Australia, which have definition-based birth registration of live births and stillbirths, prenatal diagnosis and pregnancy termination has led to increases in stillbirths and neonatal deaths due to congenital anomalies at the borderline of viability (i.e., at 20–24 weeks' gestation and <500 g birth weight; Figure 7.1A). On the other hand, as Figure 7.1B shows, stillbirth and neonatal and infant mortality due to congenital anomalies at birth weights of 500 grams or more has declined substantially (Public Health Agency of Canada 2008). Some of this decrease is attributable to factors other than prenatal diagnosis, including clinical interventions (such as surgical treatment following birth), folic acid supplementation on the part of women, and folic acid fortification of food at the population level. Nevertheless, prenatal diagnosis and pregnancy termination has had an unprecedented effect on late fetal and infant mortality



Figures 7.1A and 7.1.B Increases in congenital anomaly–related neonatal death among live births with a birth weight <500 g (primary y-axis), declines in congenital anomaly-related neonatal death among live births with a birth weight ≥500 g (secondary y-axis), Canada, 1985–2003 (A) and changes in the leading causes of infant death, Canada, 1999 versus 2004. Adapted with permission (Public Health Agency of Canada 2008; Health Canada 2003).

(Public Health Agency of Canada 2008), with congenital anomalies recently falling into second place as a cause of infant death (behind immaturity-related conditions) in countries such as Canada (Figure 7.1B).

EARLY DELIVERY IN OBSTETRICS

Early (iatrogenic) delivery for fetal and/or maternal indications

Selective, carefully timed early delivery for fetal and/or maternal indication is the cornerstone of modern obstetrics (Joseph 2007). The decision to effect early delivery, either by labor induction or cesarean delivery, is made when the balance of risks and benefits indicate that birth and supportive neonatal care are preferable to an intrauterine environment that is adversely affecting fetal well-being. This decision is highly influenced by the gestational age of the fetus, the degree of fetal compromise, and the technologic package available for effecting early delivery and clinical management of the newborn.

Induction of labor is an intervention for achieving early delivery that was first used in the mid-18th century as a clinical management option for contracted pelvis (O'Dowd 1994). Early delivery after 35 weeks' gestation was routinely used to prevent stillbirth in severe cases of Rh hemolytic disease in the 1950s (Mollison 1993). More recently, with advances in the diagnosis of fetal compromise (e.g., biophysical profile, umbilical artery Doppler velocimetry) and in neonatal care (e.g., antenatal corticosteroids, surfactant, assisted ventilation), rates of medically indicated labor induction and cesarean delivery have increased substantially in industrialized countries at preterm and term gestation (Millar 1992; Breart 1995; Joseph 1998; Sue-A-Quan 1999; Foix-L'Helias 2000; Joseph 2002; Ananth 2005a; Ananth 2005b). These changes have been accompanied by declines in perinatal mortality in several industrialized countries (Joseph 1998; L'Helias 2000; Joseph 2002; Ananth 2005a,b).

The concept of selective and carefully timed early delivery and some of the complexities involved in its use may be illustrated by considering specific scenarios encountered in obstetric practice (Joseph 2007). The first scenario involves an obstetric emergency (e.g., placental abruption with fetal bradycardia), in which a rapid absolute increase in the (incidence density) rate of perinatal death is anticipated over a short time span (minutes). It is expected that early delivery (i.e., an emergency cesarean delivery carried out within 15–30 minutes) will prevent perinatal death in more than half the fetuses (Kayani 2003).

A second scenario involves a serious pregnancy complication and fetal compromise (e.g., severe preeclampsia with fetal growth restriction), in which early delivery through labor induction and/or cesarean delivery would decrease the risk of mortality and morbidity for both the fetus and the mother. Choosing the option of early delivery in this situation can be challenging and is dependent on the gestational age of the fetus and the level of neonatal care services that are available. The need to extend the duration of pregnancy to achieve fetal maturity has to be balanced against the serious risks to both mother and infant that could be

averted by early delivery (Churchill 2002). The gestational age at which this decision is most challenging varies widely between and within industrialized and less industrialized countries, given differences in the availability of supportive neonatal care.

The final scenario involves routine early delivery at 41 weeks' gestation for an uncomplicated singleton pregnancy. Early delivery under this scenario is premised on the assumption that routine delivery at 41 weeks for singletons serves to prevent perinatal mortality and serious neonatal morbidity (Gülmezoglu 2006).

One controversial issue of interest to perinatal epidemiologists is the specific risk(s) upon which to predicate clinical decision-making for intervention. It has been suggested that the entity of clinical interest is the "prospective risk of stillbirth" (Cotzias 1999) at any gestation; that is, the cumulative incidence of stillbirth at any gestation, with the duration over which incidence is measured left open-ended (similar to the lifetime cumulative incidence of breast cancer at any given age). For instance, the prospective risk of stillbirth among women with diabetes mellitus at 32 weeks' gestation is estimated by documenting the experience of such a cohort of women from 32 weeks until delivery. All stillbirths between 32 weeks and delivery would contribute to the numerator of such an open-ended cumulative incidence rate.

An alternative measure is the cumulative incidence of stillbirth (or other outcome) in the next week (Boulvain 2000; Hilder 2000; Yudkin 2000). The 1-week cumulative incidence of stillbirth among women with diabetes mellitus at 32 weeks' gestation is estimated by documenting the experience of such a cohort of women from the beginning to the end of the 32nd week of gestation. All stillbirths that occur within these 7 days would contribute to the numerator of this 1-week cumulative incidence rate, as noted in chapter 13. Medically indicated early delivery is based on the repeated assessment of short-term risks (e.g., ≤1 week) of serious events such as stillbirth, perinatal death, and serious neonatal morbidity, since these can change substantially over a relatively short time span (Joseph 2008).

Early (iatrogenic) delivery at preterm gestation

Early delivery is carried out at preterm gestation only if the overall risks to the fetus (or mother) of a continuing pregnancy are judged to exceed those of early delivery and supportive neonatal care. As mentioned, recent advances in neonatal care have led to increasing medically indicated early delivery at preterm gestation, especially at 34 to 36 weeks' gestation (Millar 1992; Breart 1995; Joseph 1998; Sue-A-Quan 1999; Foix-L'Helias 2000; Joseph 2002; Ananth 2005a,b). The consequent increase in iatrogenic preterm birth (PTB) has been associated with simultaneous declines in perinatal mortality in several countries including Canada, France, and the United States (Joseph 1998; Foix-L'Helias 2000; Joseph 2002; Ananth 2005a,b). Concerns regarding the long-term effects of such intervention remain, however, given insufficient information on long-term health status of infants delivered at late preterm gestation (Lee 2007). Future studies on this issue may affect decisions regarding delivery at 34 to 36 weeks by showing that the

long-term risks are greater than anticipated. What makes such decisions particularly complex is the need to formally or informally consider therapeutic indices such as the number needed to treat (NNT). How many early deliveries at 36 weeks are justifiable to prevent one stillbirth at 37 weeks' gestation? For instance, in a somewhat different context, the Term Breech Trial (which changed obstetric practice with regard to the preferred mode of delivery for breech presentation at term), showed that the NNT for planned cesarean delivery (compared with planned vaginal delivery) was 29 (Hannah 2000). In other words, 29 women with breech presentation at term have to undergo planned cesarean delivery to prevent one case of perinatal mortality or serious neonatal morbidity. As always, patient values and preferences are critical to such decision-making, and such situations enhance the need for clear communication between physician and patient.

Early (iatrogenic) delivery at term gestation

Early delivery at term gestation is widely practiced for several reasons, including the evidence that shows that routine delivery at 41 weeks' gestation reduces perinatal death rates compared with expectant management (Gülmezoglu 2006). There is also an expectation that such intervention averts potential complications, such as shoulder dystocia, by preventing macrosomia, although some studies suggest that such practices are ineffective (Gherman 2006). Whatever the motivation for attempting early delivery at term, the practice has led to some dramatic results. After increasing for several decades, mean birth weight at term in the United States has begun to fall in recent years, as have rates of macrosomia and appropriate-for-gestational age term live births (Zhang et al. 2010).

INDUCTION OF LABOR

Trends

Rates of labor induction have increased significantly in many developed countries, including a doubling of the U.S. rate between 1990 and 2000 (i.e., 9.5% in 1990 to 20.2% in 2000) (Martin 2007). Figure 7.2 shows a doubling of rates for singleton pregnancies for all gestational age categories, and particularly for term pregnancies. The induction rate appears to be leveling off for PTBs since 2000. Nonetheless, approximately one in seven preterm and one in four term births follow labor inductions (Martin 2007). The rate of induction in Canada and the United Kingdom is similar to that in the United States (21.8% and 19.8%, respectively, in 2005) (National 2008; Public 2008), in comparison to 22.8% in the United States (Martin 2007).

The labor induction rate varies between and within physicians, hospitals, and region. For example, the induction rate varied from 11% to more than 40% within the United States in 1998 (Zhang 2002). Similar variations were observed across the Canadian provinces and territories (Public 2008). Glantz (2003) reported the induction rate ranged from 10% to 39% among 16 upstate New York hospitals in

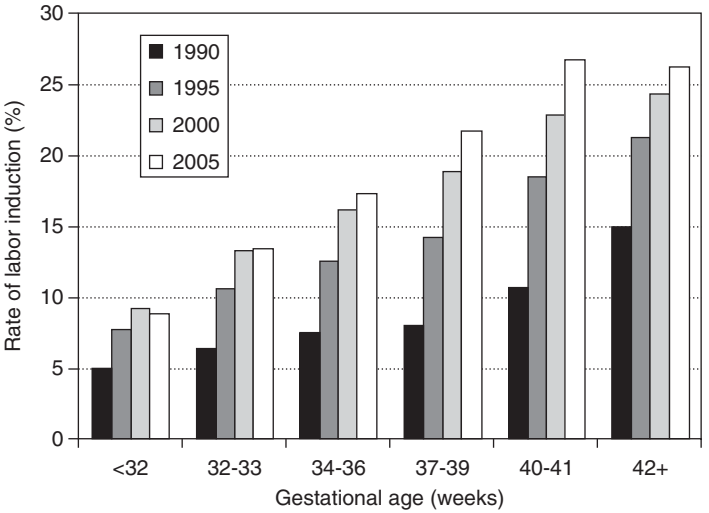


Figure 7.2 Rates of induction of labor by gestational age, singleton births: United States, 1990, 1995, 2000, 2005 (Martin 2007).

1998–1999, with only 12.6% of the variation in induction rates being explained by patient characteristics (Glantz 2004). Furthermore, labor induction tended to occur earlier in gestation in 1998 than in 1989 (MacDorman 2002), which may have contributed, in part, to the shifting of the timing of delivery toward earlier gestational age in recent years (Martin 2007).

Indications

The reasons for the substantial increase in labor induction are multifactorial. Labor induction is clinically indicated when the benefits to the mother and/or fetus outweigh those associated with continuing the pregnancy. The common clinical indications include hypertensive disorders of pregnancy (e.g., preeclampsia), diabetes, postterm pregnancy (>42 weeks), chorioamnionitis, premature rupture of the membranes, and intrauterine growth restriction (Wing 2006). Possibly as a result of delayed childbearing or increased body mass indices of parturient women, more women may be affected with gravid health conditions, such as gestational diabetes or chronic hypertension. Better detection has also identified more at-risk pregnancies (Joseph 2002). However, Zhang et al. (2002) showed that clinically indicated inductions increased approximately 60%–70% during the 1990s, and could not explain the doubling in the overall rate of induction during the same period. Other factors, including nonmedical factors, also have contributed to the increase.

Rayburn et al. (2002) postulated that an ability to plan the timing of birth by the physician, woman, and her family is the most common reason for an induction, since it accommodates everyone's schedules and time frames. Women's physical discomfort in late pregnancy, concern for rapidly progressing labor precluding timely arrival at the hospital or epidural placement, and ongoing concerns for

maternal or fetal complications may also encourage women to request a scheduled birth.

Physicians' concern for the pregnant women's physical comfort, distance from hospital, or ongoing risk profile may also make elective inductions more acceptable. In addition, scheduled induction provides advantages for the obstetricians and hospitals by allowing better planning of resources and staffing requirements (Rayburn 2002). Overall, patient preference, physician's fear of medicolegal action, work-related issues, and the availability of newer and more effective methods of preinduction cervical ripening are all factors encouraging labor induction even in the absence of clinical indications (Rayburn 2002).

Methods of induction

During pregnancy, the cervix is closed and rigid to support the enlarging uterus. The cervix is predominantly a connective-tissue and collagen structure that "ripens" as a part of spontaneous labor progression. This process allows the cervix to become soft and pliable, with ensuing dilation. Cervical status is a critical factor in determining the degree of difficulty and success of labor induction (Crane 2006). If the cervical status is favorable (i.e., soft and pliable) and the fetal head is engaged, a uterotonic agent such as intravenous oxytocin causes the uterus to contract and induces labor (Bowes 2004). If the cervix is not ripe, preinduction cervical ripening is required. The methods of cervical ripening can be chemical (e.g., prostaglandin agents or analogues) or mechanical (e.g., Foley balloon or osmotic dilators). The cervical dilation process can be a lengthy (over 24 hours) process.

Even when the cervix is ready for labor, normal labor itself is a long process. In nulliparous women, the average labor lasts 8–12 hours. Parous women usually have a labor that is one-third to one-half shorter than nulliparous women. The labor is traditionally divided into three stages. The first stage is from onset of labor contractions to full dilation of the cervix (10 cm). The second stage is from the full dilation of the cervix to delivery of the fetus. And the third stage is from the delivery of the fetus to expulsion of the placenta. The first stage can further be divided into two phases: latent phase and active phase. Cervical dilation and effacement (i.e., shortening and thinning of the cervix) are the most important measures in the first stage. In the latent phase, the cervix effaces and dilates gradually, whereas the active phase is characterized by rapid cervical dilation. The division between the latent and active phases is often difficult to pinpoint.

Based on the incidence of cesarean delivery for dystocia, some studies suggest that artificial ripening may not provide the same favorable cervical condition as spontaneous ripening, particularly in nulliparous women (Johnson 2003). The majority of women who have cesarean delivery following labor induction are still in the latent phase when the procedure is performed (Vahratian 2005).

Impact on maternal and perinatal health

For medically indicated induction of labor, the benefits for the mother and/or the fetus are indisputable. Induction can resolve a worsening condition for women

with preeclampsia, reduce the risk of intrauterine infection due to premature rupture of the membranes, and reduce the risk of macrosomia in women with poorly controlled diabetes. However, controversies arise when labor is induced for women with marginally clinically indicated reasons or in the absence of clinical indication.

Fetal maturity is a major concern in labor induction. Despite substantial advances in neonatal care, late PTB (e.g., 34–36 weeks) or early-term birth (e.g., 37 and 38 weeks) are still associated with an increased risk of neonatal and child morbidity (Petrini 2009; Tita 2009). Tita et al. (2009) examined neonatal morbidity among repeated cesarean deliveries using a composite of neonatal death and any of several adverse events, including respiratory complications, treated hypoglycemia, newborn sepsis, and admission to the neonatal intensive care unit (NICU). In comparison with births at 39 weeks' gestations, births occurring at 37 and 38 weeks were associated with an increased risk of the outcome (adjusted odds ratio [AOR] 2.1; 95% CI, 1.7–2.5; and AOR 1.5; 95% CI, 1.3–1.7, respectively). These findings support ACOG's guidelines that elective induction should not be considered prior to 39 weeks of gestation (ACOG 1999). However, evidence indicates that this guideline is not rigorously followed. For instance, elective induction prior to 39 weeks is not uncommon in some hospitals (Vahratian 2005). Furthermore, estimation of gestational age can also be erroneous, particularly for women with uncertain dates of conception. Induction, therefore, can result in a delivery prior to optimal maturity of the fetus.

Whether the elective induction of labor increases the risk of cesarean delivery has drawn substantial attention. Observational studies consistently show that nulliparous women with an unfavorable cervix have a much higher risk of cesarean delivery than do women with a favorable cervix or women with spontaneous labor (Grobman 2007). On the other hand, elective induction is not associated with an increased risk of cesarean delivery in multiparous women with a favorable cervix. Evidence is less clear in nulliparous women with a favorable cervix or multiparous women with an unfavorable cervix. Some studies also suggested that induction of labor increases the risk of postpartum hemorrhage and blood transfusion (Grobman 2007).

Thus, elective induction of labor may bear a stiff price when considering increases in costs related to maternal morbidity, cesarean delivery, and the likelihood of repeat cesarean deliveries in subsequent pregnancies. Costs for the longer hospital stays among women with labor induction in comparison with women with spontaneous labor are substantial. Another associated cost is the more frequent request for epidural analgesia.

Common pitfalls in research on labor induction

Research on labor induction, including efficacy of the treatment and its effects on perinatal outcomes, is characterized by several methodological considerations. Since labor progression and method of delivery (e.g., incidence of cesarean section) differ tremendously between nulliparous and multiparous women, and between induction with and without preinduction cervical ripening, parity and

cervical ripening are important effect modifiers. For example, in women who had a trial of labor, the rate of cesarean delivery was 15%–20% in nulliparas, but only 3%–5% in multiparas (Vahratian 2005; Hoffman 2006). In nulliparas, induction of labor with cervical ripening increases the risk of cesarean delivery by 2.5-fold, compared with spontaneous onset of labor. But induction without cervical ripening dose not increase the risk of cesarean delivery (Vahratian 2005). Therefore, it may not be appropriate to use a multivariable regression model, such as logistic regression, to simply control for these variables. Instead, stratified analysis, or testing and quantifying an interaction term in the model should be performed.

In most retrospective observational studies, the authors often compared women with induced labor versus women with spontaneous onset of labor at the same gestational age. This comparison has deficiencies (Grobman 2007). As a pregnancy continues after term, the fetus continues to grow and the placenta begins to age. The risk of cesarean delivery and pregnancy complications, therefore, increases. In practice, clinicians have to make a decision on whether to induce now or wait until labor starts spontaneously (expectant management). As time passes, the risk changes in the expectant management group of women. Therefore, comparisons between induced and spontaneous labor at the same gestational week on perinatal outcomes may not be appropriate. The induced labor would have to be compared with spontaneous labor at a later gestational week. Randomized controlled trials are the only approach for definitively answering this question.

Finally, mechanical and chemical methods for cervical ripening often work through different mechanisms (Gelber 2006; Keirse 2006). For example, the mechanical method has a direct effect on cervical dilation but less of an effect on cervical softening than does the chemical method. After the cervix is dilated by a mechanical method, a uterotonic agent (e.g., oxytocin) is given to cause the uterus to contract. One can assume that the labor starts at the time when oxytocin is first given. However, some chemical methods have the dual effects of cervical ripening and uterine contractions (i.e., induction). This results in a challenge in determining when the induction starts because the cervical ripening process can be quite long. Selecting a starting time as the onset of labor substantially affects the calculation of the duration of labor as either an exposure or outcome.

CESAREAN DELIVERY

Estimating cesarean section rates

Cesarean rates can be estimated from several sources: birth certificates, hospital discharge summary data, regional or national hospital surveys (Thomas 2001) or, where vital statistics are not collected, by direct questioning of mothers. All estimates have errors. For the purpose of program evaluation, errors in rate estimates that are as small as 1%–2% resulting from poor data quality can seriously impact on recommendations regarding the adequacy of services (Stanton 2005). Particularly in the developing world setting, small differences in cesarean rates may distinguish between adequate and inadequate access to cesarean delivery.

Although overall cesarean rates are useful in describing secular and geographic variations, comparisons between health care facilities are often confounded by case mix and practice patterns. For example, Robson (2001) proposed that comparisons be based on selected patient groups defined by parity, fetal presentation (e.g., vertex, breech), type of labor onset, and the presence or absence of a uterine scar. The ACOG recommends that between-hospital comparisons should focus on nulliparous, term, singleton vertex (NTSV) women, as cesarean rates, under optimal conditions, would be expected to vary minimally within this homogeneous group (ACOG 2000). Variations in primary cesarean rates within one U.S. health network, as measured by the ratio of standard deviation to the mean, were two- to threefold greater than variations in treatments given for conditions such as acute myocardial infarction, congestive heart failure, and pneumonias (Clark 2007). It has been proposed that when major variations do occur, they are largely influenced by health care system–related factors and may reflect differences in quality of care (Bailit 2007).

Trends in cesarean rates

The prevalence of cesarean births in the United States achieved a record high of 31.1% in 2006 and, by 2008, likely reached one-third of all births (MacDorman 2008). This represents an increase of 50% in the past decade alone. Rates increased across all age and ethnic groups. In Canada, a similar trend occurred over a 10-year period, rising from 17.6% in 1996–1997 to 25.6% in 2004–2005. Dystocia and elective repeat cesarean delivery were the indications most contributing to this increase (Public 2008). Not all countries have experienced this trend. Belizan et al. (1999) noted wide variations in cesarean rates across 18 Latin American countries, ranging from 4.9% in Bolivia to 40% in Chile. National rates were strongly correlated with the level of economic development, with higher rates in private than in public hospitals. Ronseman et al. (2006), reporting on cesarean rates in 42 developing countries, noted wide international variations (0.38% in Chad to 36.5% in Brazil). Huge disparities in cesarean rates were noted across income quartiles in many countries, with rich to poor ratios ranging from 1.0 in Uganda to 122 in Bangladesh in 1994–2002. A recent report of cesarean rates representing approximately 90% of births in the developing world found approximately 2.9% cesarean births in the region of sub-Saharan Africa, with marked intercountry variations (Stanton 2006). This compared to a rate of 25.8% in Latin America and 26.8% in East Asia. In sub-Saharan Africa, the annual rate of change between 1991 and 1998 stagnated or was negative, whereas, during the 1990s, rates in several Latin American countries increased at 4%–8% per year and at an annual rate of 16.7% in India.

Factors influencing trends in cesarean rates

Changes in women's characteristics over time may influence cesarean rates. In Nova Scotia, Joseph et al. (2003) found that changes over a 10-year period in maternal characteristics (e.g., maternal prepregnancy weight, age, parity)

accounted for the majority of the observed increase in primary cesarean section rate. Similarly, among Scottish births from 1980 to 2005, increases in maternal age accounted for more than one-third of the increase in cesarean rate (Smith 2008). Although racial and ethnic variations in cesarean rates are frequently reported, the inclusion of race and ethnicity did not modify the performance of a multivariable model to predict primary cesarean risk (Bailit 2008). In the United States, the primary cesarean birth rate increased between 1991 and 2003 for both nulliparous (Hamilton 2005) and parous women (Ford 2008). Increases in rates for parous women were greater at earlier gestational ages, and could not be explained by changes in age and ethnic distribution of the population.

Several institutional and health care provider–related characteristics may account for increases in cesarean rates. Main et al. (2006) reported a correlation ($r = 0.73$; $p < 0.001$) between the institutional cesarean birth rate and the combined proportion of labor inductions and early admissions in labor among women with NTSV. In contrast, increases in national cesarean rates in the Netherlands among term singleton nulliparous women were not accompanied by changes in induction rates (Kwee 2007). Cesarean rates varied from 10.3% to 34.2% in women with NTSV among 40 hospitals in Arizona, in the United States. Paradoxically, lower cesarean rates were found in centers with higher levels of neonatal care, higher proportions of publicly insured patients, and in public rather than private hospitals. Interestingly, the presence of an in-house obstetrician/gynecology or maternal-fetal medicine specialist was correlated with increased rates. Abenhaim (2007) reported that the presence of the woman's obstetrician rather than an on-call obstetrician conferred a protective effect on cesarean birth rates. Similarly, a study of obstetric complication rates in U.S. hospitals found that delivery on weekends was associated with an increased probability of both cesarean (36% increase in risk) and of birth trauma, including hypoxia and palsies (Bendavid 2007). Higher cesarean rates were also associated with increasing premiums for malpractice insurance. Multivariable analyses demonstrated that for each annual $10,000 insurance premium increase, the primary cesarean delivery rate increased by 15.7 per 1,000 for nulliparous women. This association also was evident for multiparous women, who had an increase in cesarean deliveries of 4.7 per 1,000 for every $10,000 increase (Murthy 2007).

The results of obstetrical randomized clinical trials have had a recent impact on cesarean rates. For example, the Term Breech Trial (Hannah 2000) showed improvements in perinatal health indicators associated with a policy of elective cesarean section as opposed to planned vaginal delivery for breech presentation. The dissemination of these findings led to marked increases in cesarean delivery for breech presentation in several regions. In the Netherlands (Kwee 2007), the cesarean rate among women with breech presentation increased from 50% to 80% within 2 months of the Trial's publication. Similar results were noted in Sweden (Alexandersson 2005). A large observational study conducted in France and Belgium failed to confirm the benefits of elective cesarean section on perinatal outcome for breech presentation (Goffinet 2006). However, cesarean section among women with breech presentation has also increased in France (Carayol 2008). In the United States, the impact of the Term Breech Trial on cesarean rates

appears to have been more modest than in Europe, but with wide variation across states (Lee 2008). The failure to demonstrate longer-term health benefits in the follow-up of those infants born in the Term Breech Trial (Whyte 2004) has raised questions as to whether vaginal delivery of carefully selected term breech may not still be a valid option.

A marked reduction in vaginal birth after cesarean (VBACs) has contributed to the increased cesarean rate in the United States, due to a reduction in the proportion of women with a previous cesarean who undergo a trial of labor (MacDorman 2008). Evidence from observational studies suggests that a policy of labor trial results in a small but significant increase in the absolute risk of adverse perinatal (Landon 2004) and maternal (MacMahon 1996) outcomes. With a risk difference of 0.25%, approximately 400 cesarean deliveries would need to be performed to prevent one perinatal adverse event (Landon 2004).

In the United States, primary cesarean delivery prior to the onset of labor as a proportion of overall cesarean sections increased 43.6% from 19.7% in 1991 to 28.3% in 2001 (Meikle 2005). Also, approximately a 33% increase in the proportion of unspecified indications for the cesarean was noted. Elective cesarean without medical indication has gained a degree of acceptance in obstetrical practice in many developed countries (Habiba 2006). With the widening debate concerning the practice of cesarean on request, a National Institutes of Health State-of-the-Science Conference was convened to explore the potential risks and benefits of this practice (NIH 2006). At present, the proportion of women undergoing elective primary cesarean without medical indication is uncertain, as there is no code for this procedure in most perinatal databases. Cases must be imputed from the absence of identified risk factors in the source file. Using a state-wide perinatal database, 87% of cesarean births with no risk factors identified on birth certificates had at least one risk factor coded on the hospital discharge summary database, indicating the importance of the method of ascertainment (Kahn 2009). Based on this latter method, approximately 4% of women in the state underwent cesarean without documented medical risk factors (Kahn 2009). However, this rate is likely to underestimate the true rate of cesarean delivery on maternal request, because most medical risk factors are not strong indications for cesarean delivery. The presence of a medical risk factor does not automatically lend an indication for cesarean delivery. The subjectivity in diagnosing some medical conditions and openness in reporting an elective cesarean delivery make it even more difficult to distinguish between cesarean-on-request and indicated deliveries.

McCourt et al. (2007) conducted a systematic literature review to assess the evidence concerning the proportion of women who actually prefer elective cesarean rather than attempted vaginal delivery. She found limited data to support the view that cesarean delivery on demand is primarily motivated by women's request. However, cultural factors likely play a role. Analyzing data from 21 surveillance cities and counties in southeastern China with a one-child policy in effect, Zhang et al. (2008) reported that cesarean-on-request increased from 0.8% in 1994 to 22% in 2003, moderating to 20% in 2006. Overall cesarean rates in these three periods were 22%, 60%, and 56%, respectively.

Association between the increasing cesarean rate and maternal health indicators

In areas of the developing world where access to emergency obstetrical services is limited, an inverse relation between cesarean rates and maternal mortality has been reported (Ronsmans 2003). In contrast, a concern has been raised that the high cesarean rates observed in some regions may increase maternal mortality and morbidity (Victora 2006; Belizan 2007). A recent review assessed the evidence concerning the effect(s) of cesarean delivery on maternal mortality and concluded that methodologically stronger studies do not support an increased risk for cesarean delivery (Vadnais 2006). However, with respect to serious maternal morbidity, Liu et al. (2007) conducted a retrospective population-based cohort study of women in Canada who underwent a primary elective cesarean for breech presentation—a surrogate for a low-risk elective cesarean delivery—in 1991–2005 and compared them with women with planned vaginal deliveries. The elective cesarean group had increased postpartum risks for cardiac arrest (AOR 5.1; 95% CI 4.1–6.3), wound hematoma (OR 5.1; 95% CI 4.6–5.5), hysterectomy (OR 3.2; 95% CI 2.2–4.8), major puerperal infection (OR 3.0; 95% CI 2.7–3.4), anesthetic complications (OR 2.3; 95% CI 2.0–2.6), venous thromboembolism (OR 2.2; 95% CI 1.5–3.2), and hemorrhage requiring hysterectomy (OR 2.1; 95% CI 1.2–3.8). The absolute increases in risk for severe maternal morbidity rates were low (e.g., for postpartum cardiac arrest, the increase with planned cesarean delivery was 1.6 per 1,000 deliveries, 95% CI 1.2–2.1). Similar increases in serious maternal morbidity associated with elective cesarean delivery have been reported in Finland (Pallasmaa 2008). Last, a large prospective cohort study in Latin America reported that maternal mortality and a range of serious morbidity indicators were significantly increased with elective, as well as with intrapartum, cesarean delivery (Villa 2008).

Association between cesarean rates and perinatal health indicators

Concerns have also been raised regarding potential adverse perinatal effects associated with increasing cesarean rates, particularly in the absence of true medical indication. The increase in the U.S. PTB rate between 1996 and 2004 occurred primarily among women delivered by cesarean delivery (Bettegowa 2008), but rates increased across all gestational age groups. The extent to which cesarean-on-request is contributing to iatrogenic late PTB is, at present, unknown. The contribution of late and moderate PTBs to infant mortality has been well documented (Kramer 2000). As well, the question of the risks and benefits of lower cesarean thresholds for infants with relative indications for preterm cesarean delivery is in urgent need of investigation.

In an attempt to assess the competing risks of elective cesarean versus planned vaginal delivery, Signore et al. (2006) conducted a decision analysis modeling the probability of perinatal death in a hypothetical cohort of 2 million births. The results of analysis suggest that overall perinatal death was increased among babies

born to women with planned vaginal delivery, although neonatal mortality was increased among cesarean births, with 1,441 cesareans needed to prevent one perinatal death.

Strategies to reduce cesarean rates

In 1985, a WHO Consensus conference on cesarean section suggested, on the basis of expert opinion, that rates in excess of 10%–15% reflect inappropriate levels of intervention and may lead to iatrogenic morbidity (WHO 1985). There have been numerous published reports of strategies to reduce cesarean rates. The simple dissemination of guidelines regarding standards of care appears to be singularly ineffective (Oppenheimer 2006). Chaillet et al. (2007) conducted a systematic review to assess the evidence concerning the effectiveness of strategies to reduce cesarean rates. Strategies that involve care providers in the identification of barriers to change and those that use multifaceted approaches including education, feedback, audit, and peer review appear to be the most effective. A multicentre cluster-randomized trial of an education, quality improvement, audit, and feedback intervention is currently in progress in Canada.

REFERENCES

Abenhaim, H., Benjamin, A., Koby, R.D., Kinch, R.A., and Kramer, M. 2007. Comparison of obstetric outcomes between on-call and patients' own obstetricians. *CMAJ* 177: 352–56.

ACOG. 1999. *Induction of labor*. In ACOG Practice Bulletin Number 10. Washington, DC: American College of Obstetricians and Gynecologists.

ACOG. 2000. *Evaluation of cesarean delivery*. Washington, DC: American College of Obstetricians and Gynecologists.

ACOG committee opinion. 2004. Number 298, August 2004. Prenatal and preconceptional carrier screening for genetic diseases in individuals of Eastern European Jewish descent. ACOG Committee on Genetics. *Obstet Gynecol* 104: 425–28.

Alexandersson, O., Bixo, M., and Högberg, U. 2005. Evidence-based changes in term breech delivery practice in Sweden. *Acta Obstet Gynecol Scand* 84: 584–87.

Allan, L.D., Cook, A., Sullivan, I., and Sharland, G.K. 1991. Hypoplastic left heart syndrome: Effects of fetal echocardiography on birth prevalence. *Lancet* 337: 959–61.

Ananth, C.V., Joseph, K.S., Demissie, K., and Vintzileos, A.M. 2005b. Trends in twin preterm birth subtypes in the United States, 1989 through 2000: Impact on perinatal mortality. *Am J Obstet Gynecol* 193: 1076–82.

Ananth, C.V., Joseph, K.S., Oyelese, Y., Demissie, K., and Vintzileos, A.M. 2005a. Trends in preterm birth and perinatal mortality among singletons: United States, 1989 through 2000. *Obstet Gynecol* 105: 1084–91.

Bailit, J. 2007. Measuring the quality of inpatient obstetrical care. *Obstet Gynecol Surv* 62: 207–13.

Bailit, J. and Love, T. 2008. The role of race in cesarean delivery rate case mix adjustment. *Am J Obstet Gynecol* 198: 69.e1–69.e5.

Belizan, J., Althabe, F., Barros, C., and Alexandre, S. 1999. Rates and implications of cesarean sections in Latin America: Ecological study. *BMJ* 319: 1397–402.

Belizan, J., Althabe, F., and Cafferata, M.L. 2007. Health consequences of the increasing cesarean section rates. *Epidemiology* 18: 485–86.

Bell, J., Hilden, J., Bowling, F., Pearn, J., Brownlea, A., and Martin, N. 1986. The impact of prenatal diagnosis on the occurrence of chromosome abnormalities. *Prenat Diagn* 6: 1–11.

Bendavid, E., Kaganova, Y., Needleman, J., Gruenberg, L., and Weissman, J. 2007. Complication rates on weekends and weekdays in US hospitals. *Am J Med* 120: 422–28.

Bettegowa, V.R., Dias, T., Davidoff, M.J., Damus, K., Callaghan, W.M., and Petrini, J.R. 2008. The relationship between cesarean delivery and gestational age among US singleton births. *Clin Perinatol* 35: 309–23.

Boulvain, M., Faltin, D., Ibecheole, V., and Irion, O. 2000. Randomised trials of earlier induction of labour are needed. *BMJ* 320: 445.

Bowes, W.A., Jr., and Thorp, J.M. 2004. Clinical aspects of normal and abnormal labor. In *Maternal-fetal medicine: Principles and practice*, 5th ed., pp. 671–705. R.K. Creasy, R. Resnik, J.D. Iams, eds. Philadelphia, PA: Saunders.

Breart, G., Blondel, B., Tuppin, P., Grandjean, H., and Kaminski, M. 1995. Did preterm deliveries continue to decrease in France in the 1980s? *Paediatr Perinat Epidemiol* 9: 296–306.

Bull, C. 1999. Current and potential impact of fetal diagnosis on prevalence and spectrum of serious congenital heart disease at term in the UK. British Paediatric Cardiac Association. *Lancet* 354: 1242–47.

Carayol, M., Blondel, B., Zeitlin, J., Breart, G., and Goffinet, F. 2007. Changes in the rates of caesarean delivery before labour for breech presentation at term in France: 1972–2003. *Eur J Obstet Gynecol Reprod Biol* 132: 20–26.

Chaillet, N. and Dumont, A. 2007. Evidence-based strategies for reducing cesarean section rates: A meta-analysis. *Birth* 34: 53–64.

Churchill, D. and Duley, L. 2002. Interventionist versus expectant care for severe pre-eclampsia before term. *Cochrane Database Syst Rev* 3; CD003106. DOI:10.1002/14651858.CD003106.

Clark, S., Belfort, M., Hankins, G., Meyers, J., and Houser, F. 2007. Variation in the rates operative delivery in the United States. *Am J Obstet Gynecol* 196: 526.e1–526.e5.

Collins, V.R., Muggli, E.E., Riley, M., Palma, S., and Halliday, J.L. 2008. Is Down syndrome a disappearing birth defect? *J Pediatr* 152: 20–4, 24.e1.

Cotzias, C.S., Paterson-Brown, S., and Fisk, N.M. 1999. Prospective risk of unexplained stillbirth in singleton pregnancies at term: Population-based analysis. *BMJ* 319: 287–88.

Crane, J.G. 2006. Factors predicting labor induction success: A critical analysis. *Clin Obstet Gynecol* 49: 573–84.

Crider, K.S., Olney, R.S., and Cragan, J.D. 2008. Trisomies 13 and 18: Population prevalences, characteristics, and prenatal diagnosis, metropolitan Atlanta, 1994–2003. *Am J Med Genet A* 146: 820–26.

Cunningham, F.G., Leveno, K.J., Bloom, S.L., Hauth, J.C., Gilstrap, L.C., III, and Wenstrom, K.D., eds. 2005. *Williams obstetrics*, 22nd ed. Toronto: McGraw-Hill.

Davidson, N., Halliday, J., Riley, M., and King, J. 2005. Influence of prenatal diagnosis and pregnancy termination of fetuses with birth defects on the perinatal mortality rate in Victoria, Australia. *Paediatr Perinat Epidemiol* 19: 50–55.

Foix-L'Helias, L. and Blondel, B. 2000. Changes in risk factors of preterm delivery in France between 1981 and 1995. *Paediatr Perinat Epidemiol* 14: 314–23.

Ford, J., Grewal, J., Mikolajczyk, R., Seikle, S., and Zhang, J. 2008. Primary cesarean delivery among parous women in the United States 1990–2003. *Obstet Gynecol* 112: 1235–41.

Gelber, S. and Sciscione, A. 2006. Mechanical methods of cervical ripening and labor induction. *Clin Obstet Gynecol* 49: 642–57.

Gherman, R.B., Chauhan, S., Ouzounian, J.G., Lerner, H., Gonik, B., and Goodwin, T.M. 2006. Shoulder dystocia: The unpreventable obstetric emergency with empiric management guidelines. *Am J Obstet Gynecol* 195: 657–72.

Glantz, J.C. 2003. Labor induction rate variation in Upstate New York: What is the difference? *Birth* 30: 168–74.

Glantz, J.C. and Guzick, D.S. 2004. Can differences in labor induction rates be explained by case mix? *J Reprod Med* 49: 175–81.

Goffinet, F., Carayol, M., Foidart, J.M., Alexander, S., Uzan, S., Subtil, D., et al. 2006. Is planned vaginal delivery for breech presentation at term still an option? *Am J Obstet Gynecol* 194: 1002–11.

Gravenhorst, J.B. 1989. Rhesus isoimmunization. In *Effective care in pregnancy and childbirth*, vol. 1: Pregnancy, pp. 656–77. I. Chalmers, M. Enkin, M.J.N.C. Kierse, eds. Toronto: Oxford University Press.

Grobman, W.A. 2007. Elective induction: When? Ever? *Clin Obstet Gynecol* 50: 537–46.

Gülmezoglu, A.M., Crowther, C.A., and Middleton, P. 2006. Induction of labour for improving birth outcomes for women at or beyond term. *Cochrane Database Syst Rev* 4;CD004945. DOI: 10.1002/14651858.CD004945.pub2.

Habiba, M., Kaminski, M., Da Fré, M., Marsal, K., Bleler, O., Librero, J., et al. 2006. Cesarean section of request: A comparison of obstetricians' attitudes in eight European countries. *BJOG* 113: 647–56.

Hamilton, B.E., Martin, J.A., and Ventura, S.J. 2007. *Births: Preliminary data for 2006. National vital statistics reports*. Hyattsville, MD: National Center for Health Statistics.

Hamilton, B.E., Martin, J.A., Ventura, S.J., Menacker, F., Kirmeyer, S. et al. 2007. Births: Final data for 2005. *Natl Vital Stat Rep* 54: 1–17.

Hannah, M.E., Hannah, W.J., Hewson, S.A., Hodnett, E.D., Saigal, S., and Willan, A.R. 2000. Planned caesarean section versus planned vaginal birth for breech presentation at term: A randomised multicentre trial. Term Breech Trial Collaborative Group. *Lancet* 356: 1375–83.

Health Canada. Canadian Perinatal Health Report, 2003. Ottawa: Minister of Public Works and Government Services Canada, 2003. Online at: http://www.phac-aspc.gc.ca/publicat/cphr-rspc03/index-eng.php.

Hilder, L., Costeloe, K., and Thilaganathan, B. 2000. Prospective risk of stillbirth. *BMJ* 320: 444.

Hoffman, M.K., Vahratian, A., Sciscione, A.C., Troendle, J.F., and Zhang, J. 2006. Comparison of labor progression between induced and noninduced multiparous women. *Obstet Gynecol* 107: 1029–234.

Johnson, D.P., Davis, N.R., and Brown, A.J. 2003. Risk of cesarean delivery after induction at term in nulliparous women with an unfavorable cervix. *Am J Obstet Gynecol* 188: 1565–69.

Joseph, K.S. 2007. Theory of obstetrics: An epidemiologic framework for justifying medically indicated early delivery. *BMC Pregnancy Childbirth* 7: 4.

Joseph, K.S. 2008. The fetuses-at-risk approach: Semantic and conceptual misapprehension. *BMC Pregnancy Childbirth* 8: 11.

Joseph, K.S., Demissie, K., and Kramer, M.S. 2002. Obstetric intervention, stillbirth, and preterm birth. *Semin Perinatol* 26: 250–59.

Joseph, K.S., Kramer, M.S., Marcoux, S., Ohlsson, A., Wen, S.W., Allen, A., and Platt, R. 1998. Determinants of preterm birth rates in Canada from 1981 through 1983 and from 1992 through 1994. *N Engl J Med* 339: 1434–39.

Joseph, K.S., Young, D.C., Dodds, L., O'Connell, C.M., Allen, V.M., Chandra, S., and Allen, A.C. 2003. Changes in maternal characteristics and obstetric practice and recent increases in primary cesarean delivery. *Obstet Gynecol* 102: 791–800.

Kahn, E.B., Berg, C.J., and Callaghan, W. 2009. Cesarean delivery among women with low-risk pregnancies. *Obstet Gynecol* 113: 33–40.

Kayani, S.I., Walkinshaw, S.A., and Preston, C. 2003. Pregnancy outcome in severe placental abruption. *BJOG* 110: 679–83.

Keirse, M.J.N.C. 2006. Natural prostaglandins for induction of labor and preinduction cervical ripening. *Clin Obstet Gynecol* 49: 609–26.

Khoshnood, B., De Vigan, C., Vodovar, V., Goujard, J., Lhomme, A., Bonnet, D., and Goffinet, F. 2005. Trends in prenatal diagnosis, pregnancy termination, and perinatal mortality of newborns with congenital heart disease in France, 1983–2000: A population-based evaluation. *Pediatrics* 115: 95–101.

Kitzinger, S., Green, J.M., Chalmers, B., Keirse, M., Lindstrom, K., and Hemminki, E. 2006. Why do women go along with this stuff? *Birth* 33: 154–58.

Klein, M.C., Sakala, C., Simkin, P., Davis-Floyd, R., Rooks, J.P., and Pincus, J. 2006. Why do women go along with this stuff? *Birth* 33: 245–50.

Kramer, M.S., Demissie, K., Yang, H., Platt, R.W., Sauvé, R., Liston, R., and the Fetal and Infant Health Study Group of the Canadian Perinatal Surveillance System. 2000. The contribution of mild and moderate preterm birth to infant mortality. *JAMA* 284: 843–49.

Kwee, A., Elferink-Stinkens, P.M., Reuwer, P.J., and Bruinse, H.W. 2007. Trends in obstetric interventions in the Dutch obstetrical care system in the period 1993–2002. *Eur J Obstet Gynecol Reprod Biol* 132: 70–75.

Landon, M.B., Hauth, J.C., Leveno, K.J., Spong, Y., Leindecker, S., et al. 2004. Maternal and perinatal outcomes associated with a trial of labor after prior cesarean delivery. *N Engl J Med* 351: 2581–89.

Lee, H.C., Yasser, Y., and Gould, J.B. 2008. Population trends in cesareans delivery for breech presentation in the United States, 1997–2003. *Am J Obstet Gynecol* 199: 9.e1–59.e8.

Lee, S. and Armson, A. 2007. Consensus statement on healthy mothers-healthy babies: How to prevent low birth weight. *Int J Technol Assess Health Care* 23: 505–14.

Liu, S., Joseph, K.S., Kramer, M.S., Allen, A.C., Sauvé, R., Rusen, I.D., and Wen, S.W. 2002. Relationship of prenatal diagnosis and pregnancy termination to overall infant mortality in Canada. *JAMA* 287: 1561–67.

Liu, S., Liston, R.M., Joseph, K.S., Heaman, M., Sauvé, R., Kramer, M.S., and the Maternal Health Study Group of the Canadian Perinatal Surveillance System. 2007. Maternal mortality and severe morbidity associated with low-risk planned cesarean delivery versus planned vaginal delivery at term. *CMAJ* 176: 455–60.

MacDorman, M.F., Mathews, T.J., Martin, J.A., and Malloy, M.H. 2002. Trends and characteristics of induced labour in the United States, 1989–1998. *Paediatr Perinat Epidemiol* 16: 263–73.

MacDorman, M.F., Menacker, F., and Declercq, E. 2008. Cesarean birth in the United States: Epidemiology, trends, and outcomes. *Clin Perinatol* 35: 293–307.

Main, E.K., Moore, D., Farrell, B., Schimmel, L.D., Altman, R.J., et al. 2006. Is there a useful caesarean birth measure? Assessment of the nulliparous term singleton vertex caesarean birth rate as a tool for obstetric quality improvement. *Am J Obstet Gynecol* 194: 1644–52.

Malloy, M.H. 2008. Impact of cesarean section on neonatal mortality rates among very preterm infants in the United States, 2000–2003. *Pediatrics* 122: 285–92.

Malone, F.D., Canick, J.A., Ball, R.H., Nyberg, D.A., Comstock, C.H., et al. 2005. First- and Second-Trimester Evaluation of Risk (FASTER) Research Consortium. *N Engl J Med* 353: 2001–11.

Martin, J.A., Hamilton, B.E., Sutton, P.D., Ventura, S.J., Menacker, F., Kirmeyer, S., and Munson, M.L. 2007. Births: Final data for 2005. *National vital statistics reports*, vol. 56, no. 6. Hyattsville, MD: National Center for Health Statistics.

McCourt, C., Weaver, J., Statham, H., Beake, S., Gamble, J., and Creedy, D.K. 2007. Elective cesarean section and decision making: A critical review of the literature. *Birth* 34: 65–79.

McMahon, M.J., Luther, E.R., Bowes, W.A., Jr., and Olshan, A.F. 1996. Comparison of a trial of labor with an elective second cesarean section. *N Engl J Med* 335: 689–95.

Meikle, S.F., Steiner, C.A., Zhang, J., and Lawrence, W.L. 2005. A national estimate of the elective primary caesarean delivery rate. *Obstet Gynecol* 105: 751–56.

Millar, W.J., Wadhera, S., and Nimrod, C. 1992. Multiple births: Trends and patterns in Canada, 1974–1990. *Health Rep* 4: 223–50.

Mollison, P.L., Engelfriet, C.P., and Contreras, M. 1993. *Blood transfusion in clinical medicine*, 9th ed., pp. 543–91. Boston, MA: Blackwell Scientific Publications.

Murthy, K., Grobman, W.A., Lee, T.A., and Holl, J.L. 2007. Association between rising professional liability insurance premiums and primary cesarean delivery rates. *Obstet Gynecol* 110: 1264–69.

National Collaborating Center for Women's and Children's Health. 2008. *Induction of labor*. London: RCOG Press.

NIH State-of-the Science. 2006. Conference Statement on Cesarean Delivery on Maternal Request. *NIH Consens Sci Statements* 23: 1–29.

O'Dowd, M.J. and Philipp, E.E. 1994. *The history of obstetrics and gynecology*. New York: Parthenon Publishing Group.

Oppenheimer, L.W., Holmes, P., Yang, Q., Yang, T., Walker, M., and Wu Wen, S. 2007. Adherence to guidelines on the management of dystocia and cesarean section rates. *Am J Perinatol* 24: 271–75.

Pallasmaa, N., Ekblad, U., and Gissler, M. 2008. Severe maternal morbidity and the mode of delivery. *Acta Obstet Gynecol* 87: 662–68.

Petrini, J.R., Dias, T., McCormick, M.C., Massolo, M.L., Green, N.S., and Escobar, G.J. 2009. Increased risk of adverse neurological development for late preterm infants. *J Pediatr* 154: 169–76.

Public Health Agency of Canada. Canadian Perinatal Health Report, 2008 Edition. Ottawa: Minister of Public Works and Government Services Canada, 2008. Online at: http://www.phac-aspc.gc.ca/publicat/2008/cphr-rspc/index-eng.php.

Rappaport, V.J. 2008. Prenatal diagnosis and genetic screening: Integration into prenatal care. *Obstet Gynecol Clin N Am* 35: 435–58.

Rasmussen, S., Irgens, L.M., Bergsjo, P., and Dalaker, K. 1996. Perinatal mortality and case fatality after placental abruption in Norway 1967–1991. *Acta Obstet Gynecol Scand* 75: 229–34.

Roberts, H.E., Moore, C.A., Cragan, J.D., Fernhoff, P.M., and Khoury, M.J. 1995. Impact of prenatal diagnosis on the birth prevalence of neural tube defects, Atlanta, 1990–1991. *Pediatrics* 96: 880–83.

Robson, S.M. 2001. Can we reduce the cesarean section rate? *Best Pract Res Clin Obstet Gynaecol* 15: 179–94.

Ronsmans, C., Etard, J.F., Walraven, G., Hoj, L., Dumont, A., de Bernis, L., and Dodio, B. 2003. Maternal mortality and access to obstetric services in West Africa. *Trop Med Int Health* 8: 940–48.

Ronsmans, C., Holtz, S., and Stanton, C. 2006. Socioeconomic differentials in caesarean rates in developing countries: A retrospective analysis. *Lancet* 368: 1516–23.

Signore, C., Hemachandra, A., and Klebanoff, M.A. 2006. Neonatal mortality and morbidity after elective cesarean delivery versus expectant management: A decision analysis. *Semin Perinat* 30: 288–95.

Smith, G.C.S., Cordeaux, Y., White, I.A., Pasupathy, D., Missfelder-Lobos, H., Pell, J.P., et al. 2008. The effect of delaying childbirth on primary caesarean section rates. *PLoS Med* 5: e144.

South, S.T., Chen, Z., and Brothman, A.R. 2008. Genetic medicine in prenatal diagnosis. *Clin Obstet Gynecol* 51: 62–73.

Stanton, C.K., Dubourg, D., De Brouwere, V., Pujades, M., and Ronsmans, C. 2005. Reliability of data on cesarean sections in developing countries. *Bull World Health Organ* 83: 449–55.

Stanton, C. and Holtz, S. 2006. Levels and trends in cesarean birth in the developing world. *Stud Fam Plann* 37: 41–48.

Sue-A-Quan, A.K., Hannah, M.E., Cohen, M.M., Foster, G.A., and Liston, R.M. 1999. Effect of labour induction on rates of stillbirth and cesarean section in post-term pregnancies. *CMAJ* 160: 1145–49.

Thomas, J. 2001. *Paranjothy Royal College of Obstetricians and Gynaecologists Clinical Effectiveness Support Unit. National Sentinel Caesarean Section Audit Report*. London: RCOG Press.

Tita, A.T.N., Landon, M.B., Spong, C.Y., Lai, Y., Leveno, K.J., et al. 2009. Timing of elective repeat cesarean delivery at term and neonatal outcomes. *N Engl J Med* 360: 111–20.

Vadnais, M. and Sach, B. 2006. Maternal mortality with cesarean delivery: A literature review. *Semin Perinatol* 30: 242–46.

Vahratian, A., Zhang, J., Troendle, J.F., Sciscione, A., and Hoffman, M. 2005. Labor progression and risk of cesarean delivery in electively induced nulliparas. *Obstet Gynecol* 105: 698–704.

Wing, D.A. 2006. Symposium on cervical ripening and labor induction. *Clin Obstet Gynecol* 49: 549–50.

Victora, C.G. and Barros, F.C. 2006. Beware: Unnecessary cesarean sections may be hazardous. *Lancet* 367: 1796–97.

Villar, J., Carroli, G., Zavaleta, N., Donner, A., Wojdyla, D., et al. 2007. Maternal and neonatal individual risks and benefits associated with caesarean delivery: Multicentre prospective study. *BMJ* 335(7628): 1025.

Wapner, R., Thom, E., Simpson, J.L., Pergament, E., Silver, R. et al. 2003. First Trimester Maternal Serum Biochemistry and Fetal Nuchal Translucency Screening (BUN) Study Group. First-trimester screening for trisomies 21 and 18. *N Engl J Med* 349: 1405–13.

WHO. 1985. PAHO statement: Appropriate technology for birth. *Lancet* 24;2(8452): 436–37.

Whyte, H., Hannah, M.E., Saigal, S., Hannah, W.J., Hewson, S., Amankwah, K., et al. 2004. Outcomes of children at 2 years after planned caesarean birth versus planned vaginal birth for breech presentation at term: The International Randomized Term Breech Trial. *Am J Obstet Gynecol* 191: 864–71.

Wortelboer, M.J., De Wolf, B.T., Verschuuren-Bemelmans, C.C., Reefhuis, J., Mantingh, A., Beekhuis, J.R., and Cornel, M.C. 2000. Trends in live birth prevalence of Down syndrome in the Northern Netherlands 1987–1996: The impact of screening and prenatal diagnosis. *Prenat Diagn* 20: 709–13.

Yudkin, P. and Redman, C. 2000. Impending fetal death must be identified and pre-empted. *BMJ* 320: 445.

Zhang, J., Liu, Y., Meikle, S., Zheng, J., Sun, W., and Li, Z. 2008. Cesarean delivery on maternal request in southeast China. *Obstet Gynecol* 111: 1077–82.

Zhang, J., Yancey, M.K., and Henderson, C.E. 2002. U.S. national trends in labor induction, 1989–1998. *J Reprod Med* 47: 120–24.

Zhang, X., Joseph, K.S., and Kramer, M.S. 2010. Decreased term and postterm birthweight in the United States: impact of labor induction. *Am J Obstet Gynecol*. 203(2): 124.e1–7.

8

Duration of Gestation and Timing of Birth

JUN ZHANG AND DAVID A. SAVITZ

Human pregnancy lasts 9 months on average. However, an accurate estimate of gestational age in individual pregnancies is not easy to obtain. Several methods are commonly used, each of which has its advantages and disadvantages. This chapter reviews these methods and discusses the implications of errors derived from these methods.

Despite its deficiencies, gestation dating is clinically critical and usually established as early in pregnancy as possible. Many clinical decisions are gestational age–dependent because fetal maturity is directly related to the duration of gestation. When a fetus is delivered too early (i.e., preterm birth [PTB]) or too late (i.e., post-term birth), perinatal and infant long-term morbidity and mortality increase. Thanks to widespread use of ultrasonography and induction of labor, post-term birth has been mostly eliminated in many countries. However, progress in understanding the etiology of PTB, and in identifying effective prevention and treatment has been frustratingly slow. This chapter provides a comprehensive but succinct review of current knowledge in PTB. Of note, we use the terms "preterm" and "post-term" births to reflect the timing of delivery, rather than "premature" and "postmature" births, which have complex pathological meanings.

DURATION OF GESTATION

Duration of pregnancy is commonly estimated by three dating methods: last menstrual period (LMP), ultrasonography, and neonatal assessment.

Dating based on last menstrual period

Theoretically, duration of gestation should be calculated as the day of conception to the day of delivery. Although methods of ovulation monitoring, such as charting of basal body temperature, monitoring of cervical mucus, and home-based urine test kits to detect luteinizing hormone surge (Eichner 2004), are available tools to detect conception, the majority of women are not aware of the exact timing

of conception. Thus, the first day of the LMP as a landmark for calculating duration of gestation has long been used and is still the most commonly used approach, particularly in regions where an early ultrasound exam is unavailable.

It is customary to estimate the expected date of delivery by adding 1 year and 7 days to the first day of the LMP and then counting back 3 months (Naegele's rule). Gestational age is usually expressed as "completed weeks" rather than "rounded weeks." For example, if gestational age is 37 weeks plus 5 days from the LMP, it is often recorded as 37 instead of 38 weeks. Occasionally, the term *menstrual age* or *menstrual week* is used to distinguish from *embryonic age, ovulatory age*, or *fertilization age*, which are calculated from conception (Cunningham 2001).

Problems with the LMP method are well recognized. A substantial proportion of women do not remember their LMP accurately; this is especially the case among women who do not initiate prenatal care early during pregnancy (Hall 1985; Geirsson 1991). Mid-cycle bleeding and bleeding in early pregnancy may also lead to erroneous recall of LMP. Although physicians often try to confirm the LMP date by a bimanual exam of the uterine size in the first prenatal visit, this method is crude, with variability of 2 weeks or more (Resnik 2004). In addition, using LMP as the starting point of gestation has two assumptions: duration of a menstrual cycle is 28 days, and ovulation occurs on day 14 of the cycle. Such assumptions are often not met due to longer menstrual cycles, irregular menstrual cycles, and delayed ovulation (Lynch 2007). For example, only about half of women have a regular menstrual cycle of 28 ± 2 days (Geirsson 1991). Approximately 30% of women have an average cycle length of 30 days or longer (Berg 1991). Baird et al. (1995) found that only 12% of women ovulated on exactly day 14.

The errors in LMP result in inaccurate estimates of gestational duration. Mid-cycle bleeding and bleeding in early pregnancy, both occurring after the LMP, can make the duration of gestation look shorter than it actually is. On the other hand, long menstrual cycles and delayed ovulation artificially prolong the apparent duration. Erroneous recalls of LMP and irregular cycles can lead to errors in either direction.

Dating based on ultrasound measures

Using ultrasonography, researchers carefully measured fetal biometry among healthy women with regular menses and certain LMP (Hadlock 1982). Mean values of the fetal measurements at given gestational weeks were generated as the standard. Ultrasound estimation of duration of pregnancy is a reserve process. It measures fetal size to estimate gestational age. The earlier an ultrasound measurement is obtained, the more accurate the estimate of gestational age is because of the limited variability in embryonic size in early gestation. Crown–rump length is a widely accepted measure in the first trimester (Filly 2000). It can be reliably obtained from 7 to 13 weeks. Accuracy of this method in predicting gestational age is 3 to 5 days.

If an ultrasound examination is not done in the first trimester, an early second trimester ultrasound measure can still estimate gestational age with reasonable

accuracy. The basic fetal measurements used for dating are biparietal diameter, head circumference, abdominal circumference, and femoral length (Degani 2001). Although any single measure may yield an estimate of gestational age, an estimate derived from multiple measures has been found to be the most accurate (Hadlock 1987). In general, the variability (± 2 standard deviations) of ultrasound estimates between 14 and 20 weeks of gestation is within 1 week. The variability increases to 2 weeks in late second trimester and to 3 weeks in the third trimester.

To test whether the ultrasound method is more accurate than the LMP method, a number of studies compared the ability to predict timing of spontaneous onset of labor by "certain" LMP and by ultrasound estimate of gestational age (Nguyen 1999). The ultrasound dating was consistently superior to the LMP method (Lynch 2007). Thus, some authors have concluded that use of LMP even for "certain" dates offers no advantage over ultrasound estimates if they are available (Mongelli 1996).

However, it is worth noting that the ultrasound method uses size to approximate duration (time). It implicitly assumes that, early in pregnancy, all fetuses have the same size at the same gestational age. Variation in fetal size is interpreted as variation in gestational age. For instance, suppose two pregnancies are conceived on the same day. As the gestations advance, the fetal sizes may diverge. The ultrasound method assigns the larger fetus a higher gestational age and the smaller fetus a lower gestational age, introducing a systematic bias. As fetal growth abnormality can now be demonstrated as early as in the first trimester (Smith 1998; Bukowski 2007; Pedersen 2008), the systematic bias introduced by the ultrasound method can potentially distort the results (Lynch 2007). One example is the association between fetal sex and risk of preterm and post-term births (Henriksen 1995). Using reliable LMP methods, male and female fetuses had similar risks for preterm and post-term births. But using the second-trimester ultrasound method, the female fetus had a 13% higher risk of preterm delivery and a 19% lower risk of post-term delivery than did the male fetus.

The American Congress of Obstetricians and Gynecologists recommends a revision in a woman's LMP-based due date only if the LMP and ultrasound-based estimates differ by more than: ± 7 days up to 20 weeks' gestation, ± 14 days from 20 to 30 weeks' gestation, and ± 21 days at 30 weeks' gestation or beyond (ACOG 2004). By keeping the original LMP when it appears reasonably accurate, this approach may reduce the degree of the systematic bias of the ultrasound method (Lynch 2007). On the other hand, a large study showed that the combination approach still did not predict timing of spontaneous onset of labor as well as the ultrasound dating alone did (Mongelli 1996), thus suggesting that the error in the LMP dating may still be larger than the systematic bias in the ultrasound dating (Savitz 2002).

Dating based on neonatal assessment

When an antenatal estimate of gestational age is not available, physicians sometimes estimate the gestational age based on physical and neuromuscular maturity of the newborn. The Dubowitz (1970) and Ballard (1979, 1991) scoring systems are the most common methods. Studies have shown that postnatal dating is the least

accurate method, since it tends to overestimate gestations for infants born before 40 weeks, while underestimating the gestation of infants born at or after 40 weeks (Alexander 1992). With wide accessibility of ultrasonography, however, postnatal dating is now performed infrequently.

With these dating methods in mind, we may now be able to answer a timeless question: how long does an average human pregnancy last? In the early 19th century, Naegele postulated that a human pregnancy lasted 10 menstrual cycles or 280 days from the first day of the LMP (Naegele 1836). Over the years, studies based on "reliable" LMP data produced a wide range of estimated mean durations from 272 to 283 days (Kieler 1995). Smith (2001) studied 1,514 healthy pregnant women in whom the discrepancy between the LMP-based gestational age and the gestational age based on the first trimester crown–rump length was within –1 to +1 day. To overcome the problem that pregnancy may be shortened (or censored) by clinical intervention, such as induction of labor or cesarean delivery, the median duration of gestation using survival analysis was estimated to be 283 days for both fetal sexes, but varied by parity (i.e., 284 and 282 days for nulliparous and multiparous women, respectively).

PRETERM BIRTH

Descriptive epidemiology of preterm birth

Preterm birth comprises all deliveries of less than 37 weeks completed gestation, regardless of the reason. This is not an uncommon outcome of pregnancy, affecting 11.0% of singleton live births and 12.7% of all births based on vital records in the United States in 2005 (Martin 2007), with markedly lower rates in Canada (6%–7%) and France (5%–6%), for example (Blondel 2002). Using a more stringent criterion of less than 34 weeks' completed gestation, 2.9% of singleton births and 3.6% of all births are classified as preterm. The frequency of PTB depends in part upon the method for defining gestational age, differing for last menstrual period versus ultrasound. Temporal trends show a small but steady increase over time, from 9.7% of singletons in 1990 to 10.1% in 2000, and 11.0% in 2005 (Martin 2007). The increase is driven exclusively by births in the 32- to 36-week range, with no discernible increase in births of less than 32 weeks' gestation (Behrman 2007). The rate of PTB in the United States is believed to be higher than in most other developed countries for reasons that are poorly understood, but probably includes a combination of artifacts (e.g., the method for assessing gestational age in vital records, and who is included in the vital records registration systems) as well as true differences in the presence of risk factors and clinical practice with regard to interventions leading to early delivery.

The most notable, recalcitrant health disparity in the risk of PTB pertains to the marked disadvantage experienced by African-Americans, who have notably higher risk compared to non-Hispanic whites. In 2005, 18.4% of live births to African-Americans were preterm, compared with 11.7% among non-Hispanic white women and 12.2% among Hispanic women (Martin 2007). In the most extreme

low end of the gestational age spectrum, the disadvantage for blacks is even more apparent: 1.9% of black infants versus 0.6% of white non-Hispanic infants were born prior to 28 weeks' completed gestation in 2005. The excess risk among African-Americans is not readily explained by typical indicators of socioeconomic status; nor do the few known risk factors, such as smoking, low prepregnancy weight, or bacterial vaginosis, account for the disparity. The lack of excess risk for many Hispanic groups, particularly Mexican-Americans, who also may be economically deprived, poses somewhat of a paradox, pointing to some distinctive aspect of the African-American population that confers this increased risk. However, there is increasing appreciation of ethnic variation within broad (and heterogeneous) groups of Hispanics (with higher risk among Puerto Ricans) and Asians (with higher risk among Filipinos) (Singh 1996), and recognition that risk status changes with time since immigration (Singh 1996) supporting the need for a more comprehensive consideration of the underlying reasons for such patterns.

Classification of preterm birth

Preterm birth is defined solely by the gestational age at which delivery occurs, reflecting a final common pathway for many different clinical and biological processes, resulting from diverse causal mechanisms and a wide range of candidate risk factors. Furthermore, although defined as a dichotomy divided arbitrarily at 37 weeks completed gestation, the consequences of PTB for infant health and survival vary across the continuum of gestational age within the preterm range. The entity of PTB deserves closer consideration and possible subgrouping to study both etiology and prognosis.

Preterm birth may arise from medical interventions, reflecting the judgment of the medical care provider that inducing a preterm delivery is preferable to continuing the pregnancy at that point in time. The most commonly applied basis for division of PTB has been into "spontaneous," in which preterm labor or membrane rupture begins a process culminating in preterm delivery, versus "medically indicated," in which concerns with the health of the pregnancy lead to a clinical judgment that delivery should be induced prior to term. Recent analyses suggest that the proportion of medically indicated PTBs in the United States is rising markedly over time (MacDorman 2002).

Although it is often assumed that medically indicated PTBs are limited to pregnancies that are nearing term, in fact the proportion of all PTBs due to medical indications is fairly constant across the gestational age spectrum (Savitz 2005). Both indicated and spontaneous PTBs are much rarer earlier as opposed to later in gestation, but severe complications are identified throughout the range of gestational ages of PTB that provide justification for early delivery. These indications, such as preeclampsia, fetal distress, or other serious threats to the mother's or fetus's health, lead the clinician to make the judgment that a preterm delivery is the preferred outcome, intended to reduce the risk of more severe outcomes for the mother or fetus, including stillbirth or neonatal death. In fact, interpretation of the slowly rising rate of preterm delivery must take into account the potential for a reduced rate of other more adverse outcomes, and thus this does not necessarily constitute a negative trend.

Within the category of spontaneous PTB, distinctions are sometimes made between those in which the initiating event is labor onset and those in which the initiating event is rupture of the chorioamnionic membranes (Savitz 1991; Moutquin 2003). Arguing in favor of splitting these categories is the potential for distinctive biologic mechanisms, for example, influences on uterine contractility, as distinct from influences on chorioamnionic membrane integrity. On the other hand, there is often a close relationship in time between membrane rupture and labor onset, and even when it seems that membrane rupture has initiated the process leading to early delivery, subclinical contractions may have begun. Despite efforts to identify distinctive etiologic processes using this axis for division, there are questions about the validity of the distinction (Klebanoff 1995), and the empirical support for distinctive risk factor profiles is limited (Meis 1995, 1998; Berkowitz 1998).

Other approaches to dividing births for assessing etiology have been suggested, building on our concepts of etiologic pathways. Klebanoff (1998) proposed subdividing births into those precipitated by inflammation, vascular compromise, and neuroendocrine processes. Complicating the effort to isolate these subsets based on risk factors or mechanistic pathways is the fact that predisposing attributes tend to increase both spontaneous and medically indicated PTBs. That is, the same processes that lead a clinician to intervene tend to result in spontaneous PTB if no intervention is undertaken. In addition, accurate measurement and assignment of the underlying etiology is problematic based on routinely collected data.

Another important dimension of the heterogeneity of PTB pertains to severity, with marked differences in infant health as a function of gestational age at delivery. The frequency of adverse infant outcomes, both mortality and morbidity, is inversely proportional to gestational age, with 70% of all PTBs occurring in the 34- to 36-week range and a diminishing proportion occurring earlier in gestation (Kramer 2000). On the other hand, the severity of consequences follows the opposite pattern, with profoundly different implications for survival for infants born earlier as compared with later within the period of less than 37 weeks. Compared to term births, the relative risk of infant mortality (death within the first year) is 2.9 for a 34- to 36-week birth, 6.6 for a 32- to 33-week birth, 16.2 for a 28- to 31-week birth, and 127 for a birth of less than 28 weeks. In clinical care, the focus tends to be on very early PTBs, often defined as of less than 32 weeks' completed gestation. However, from a public health perspective, the later PTBs are so much more common that even with their lower absolute risk of death, they make a substantial contribution to infant mortality (Kramer 2000; Petrini 2009).

Biologic mechanisms of preterm birth

Multiple biologic pathways are thought to contribute to the etiology of PTB, and these are considered in detail in the Institute of Medicine's report on PTB (Behrman 2007). Beyond the differing clinical presentations that manifest as preterm labor, membrane rupture, or complications leading to medical interventions, underlying biological processes need to be considered that have varying degrees of experimental and clinical support.

The most extensive support is probably for the role of ascending bacterial infection to produce inflammation that culminates in preterm delivery. This mechanism is thought to contribute to nearly half of PTBs, particularly the earliest deliveries, through infection of the chorioamnionic membranes. The proinflammatory cytokine–prostaglandin pathway is thought to play a critical role in the process culminating in PTB. Empirical evidence strongly links intrauterine infection to early delivery, and substantial evidence implicates bacterial vaginosis with PTB (Leitich 2007). However, randomized trials to screen and treat bacterial vaginosis have yielded only mixed results regarding whether intervention reduces risk of preterm delivery. In addition to the role of reproductive tract infection, there is growing evidence from observational studies that systemic infection and periodontal infection may also contribute to an increased risk of PTB, but the results from randomized trials that treat periodontal disease do not suggest benefit in reducing PTB (Macones 2008; Offenbacher 2008).

Vascular compromise, which affects placental nutrient transfer, maternal blood pressure, and other critical processes in fetal development, is an independent pathway by which risk of preterm delivery may be affected. In a sizable fraction of PTBs, particularly those associated with preterm premature rupture of the membranes, there are indications of vascular lesions in the placenta. The pathway is thought to involve thrombosis leading to ischemia. The elevated risk of PTB associated with preeclampsia (Mouldenhaur 2003) is thought to result in part through this mechanism, increasing both spontaneous and medically indicated preterm delivery.

Stress in its various forms is thought to be a contributor to PTB, mediated through neuroendocrine, immune, and/or inflammatory processes. The neuroendocrine pathway involves an increase in release of corticotrophin-releasing hormone from the placenta, which in turn is a signal to the "placental clock" for delivery (Smith 2007). When this mechanism is stimulated prior to term, preterm delivery may result. Corticotropin-releasing hormone, cortisol, and other biomarkers of stress response have been examined in relation to PTB, with mixed results.

Uterine overdistension is thought to be the mechanism by which multiple gestations result in a markedly elevated risk of PTB. There are a number of candidate mechanistic pathways by which increased volume of uterine contents could stimulate pathways that culminate in PTB.

Each of these mechanistic hypotheses has resulted in important, contributory laboratory and clinical research, identifying candidate modifiable risk factors and suggesting possible interventions, some of which have been rigorously evaluated. For example, the potential role of inflammation has led to the study of infectious processes and genes controlling inflammatory response in relation to PTB (Engel 2005), yet trials of antibiotics have not consistently provided the anticipated benefit of such treatment (Varma 2006). Analogously, the neuroendocrine mechanism has generated abundant research on candidate biomarkers, stress, social support, and other psychological pathways, but with little clarity or consistency in the empirical findings.

One of the important distinctions to be made in relating biological mechanisms to epidemiologic findings is the challenge of distinguishing *markers* of risk from *causal determinants* of risk. In the course of pregnancy that is predisposed

to culminate in early delivery, biologic changes can be reflected in biomarkers that are strong, empirical predictors of early delivery. These include fetal fibronectin (an indicator of placental dysfunction), elevated cortisol, increased uterine activity, and others. Examined in the usual way, these markers appear to be strong risk factors for early delivery, yet the evidence that they are on the causal pathway and thus amenable to intervention is lacking. That is, they reflect an etiologic process that is in progress, which allows for preparation for impending PTB (including administration of antenatal steroids to promote lung maturation) but do not necessarily provide an opportunity to prevent the PTB from occurring.

Risk factors for preterm birth

Epidemiologic research on risk factors for PTB is quite substantial in volume and increasingly of high quality, yet with very limited success in identifying causal influences that are strong and amenable to intervention. The clearest predictors of PTB, not directly modifiable, are multiple gestation, with over half of such infants delivering preterm; prior preterm birth, associated with a two- to threefold increased risk; and African-American ethnicity, associated with around a twofold increased risk. The leading indicators of risk that are well-established and offer more potential for preventive measures are bacterial vaginosis, associated with around a twofold increased risk; cigarette smoking, associated with a 1.3- to 2-fold increased risk; and low prepregnancy body mass index (BMI), associated with around a 1.5-fold increased risk (Behrman 2007).

The extensive research concerning other well-studied candidate risk factors warrants mention, even though the work has not yet culminated in preventive measures or, in most instances, in trials of prevention. In the realm of health behaviors (Savitz and Murnane, 2010), a large body of research has considered the potential adverse effects of alcohol, caffeine, and illicit drugs, with only cocaine showing reasonably consistent evidence of an association with increased risk. Low levels of vitamin C, folate, and lack of multivitamin use have been associated with increased risk, with iron providing mixed results depending on the indicator that is used. Leisure time physical activity has been associated with decreased risk. Sexual activity during pregnancy has been linked to both increased and decreased risks.

Psychosocial factors have generated considerable interest and attention, in large part because of the compelling mechanistic pathways discussed earlier. Various indicators of stress and responses to stress (both psychosocial and biological) have been investigated repeatedly, with at best mixed empirical evidence for an increased risk of PTB with higher stress indicators. Depression has also been evaluated, as well as anxiety, lack of social support, neighborhood crime, racial discrimination, and other related constructs, all with mixed empirical support for an association. These constructs are so challenging to measure accurately that the potential remains for an important effect that has not yet been captured in the level of detail that would be required to generate compelling empirical support.

A potentially important but incompletely understood influence on PTB is the rapidly increasing use of assisted reproductive technologies (ART). Beyond the well-known increased risk of multiple gestation and resulting increase in PTB,

there is growing evidence that singleton births conceived through medical interventions may carry an increased risk of PTB. A meta-analysis published in 2004 (Helmerhorst 2004) reported a pooled relative risk estimate of 2.0 (95% confidence interval [CI] 1.8, 2.3) for PTB among singletons conceived through ART, even more pronounced for very preterm births of less than 32 weeks completed gestation (relative risk = 3.3, 95% CI = 2.0, 5.3). Whether the underlying infertility that resulted in assisted conception or the medical treatment to achieve conception is responsible is not clear.

The limited progress in identifying causes of PTB, despite extensive epidemiologic research, calls for thoughtful examination of the reasons. It is, of course, possible that the important determinants of PTB are not strongly affected by the "usual suspects" in the behavioral, social, and medical arenas, yet the persistent racial disparity for African-Americans and the plausible biological mechanisms for a number of these risk factors (infection and stress, for example) argues otherwise. Heterogeneity in the causes of PTB may be an important limitation in studying etiology, analogous to searching for the causes of adult morbidity or mortality in the aggregate—a multiplicity of pathways and risk factors that need to be subdivided in order to identify strong influences. Combined with the rather crude tools used to assess behavior, psychosocial processes, and diet, the mixed results should not be overinterpreted as demonstrating that these factors are *not* contributors to the etiology of PTB.

Prevention strategies

Multiple preventive strategies have been considered, often with a strong rationale and great optimism, yet with one exception—all have failed. The rationale for prevention strategies has been to focus on early identification of preterm labor or other evolving indications that the biological process that may lead to PTB has begun. These efforts include primary prevention, reducing risk factors such as stress (through social support), bacterial vaginosis (through antibiotic treatment), and nutritional deficiency (through supplementation and counseling). A more medically based model for prevention begins with biologic signs of preterm labor based on monitoring of uterine contractions (with pharmacologic efforts to postpone labor onset), identification of cervical insufficiency (treated with cervical cerclage), or biomarkers of incipient early delivery such as detection of fetal fibronectin or a rise in cortisol levels.

During the 1980s, it was thought that if subclinical premature contractions indicative of impending preterm delivery could be detected early enough, and antitocolytic treatment initiated, PTB could be averted. However, despite some early promise, large randomized trials based on early detection through patient education or electronic monitoring failed to result in a decreased risk of PTB. No matter how effective the ability to identify women at high risk of premature onset of labor, the pharmacologic tools available for postponing delivery for days rather than hours are quite limited at present.

Prenatal care, social support, and reduction of physical activity have been considered as more general approaches to reducing risk of PTB by screening and

treating for medical conditions, providing counseling regarding healthful behaviors such as smoking cessation, and reducing stress. One of the common recommendations for women who are identified as being at elevated risk of PTB is bed rest, which has not been demonstrated to be of clinical benefit in preventing PTB as intended. Although there may very well be multiple medical and psychological benefits to such care and attention, there is no indication that PTB rates are reduced as a result.

Most promising and most extensively investigated is the strategy of screening to identify bacterial vaginosis and treating it with antibiotics to eliminate the condition. In the past, a number of reproductive tract infections, including group B streptococcus, *Chlamydia trachomatis*, and *Trichomonas vaginalis* were suspected of causing PTB, but for some time, the main focus has been on an imbalance in the vaginal microflora, labeled as *bacterial vaginosis*. Whatever the origins may be, identification of the presence of bacterial vaginosis during pregnancy has been shown in an extensive array of observational studies to be empirically predictive of an increased risk of PTB (Leitich 2007). Furthermore, medications are available that alter the vaginal flora to assume a normal profile, even if there is some tendency for bacterial vaginosis to recur once the antibiotics are no longer in use. Given this promising evidence, a series of large, randomized trials to screen asymptomatic women in pregnancy and intervene to eliminate bacterial vaginosis have not demonstrated a consistent reduction in the risk of PTB. It seems that bacterial vaginosis is a marker of increased risk and reflective of contributions of infection and inflammation to the etiology of PTB, but the relation is not so simple as cause and effect.

The one strategy that has been proven effective is weekly progesterone injections for women who have a history of prior PTB. In a large, well-designed, randomized controlled trial, women who received progesterone experienced a 40%–55% reduction in the risk of PTB (Meis 2003). The short-term neonatal sequelae associated with PTB were correspondingly reduced, and there is every reason to believe that the results reflect a real and clinically meaningful benefit. The treatment is only proven effective for women at high risk and involves a rather demanding intervention protocol, but nonetheless offers some of the first concrete evidence that PTB can be prevented.

Limited epidemiologic evidence suggests that vitamin C may help to reduce risk, based on mechanistic support for its potential benefit based on its role in collagen cross-linkages and thus in maintaining the integrity of the chorioamnionic membranes. A recent trial addressed the role of antioxidants in the prevention of preeclampsia, but also examined impact on PTB (Poston 2006). However, once again, a promising intervention strategy failed to show benefit.

Short- and long-term consequences of preterm birth

Preterm birth is closely associated with an increased risk of infant mortality, both neonatal (first 28 days) and postneonatal (28 days to 1 year). Figure 8.1 shows the pattern of neonatal mortality by gestational age for black and white infants in the United States for two time periods, 1989 and 2005, across the PTB range.

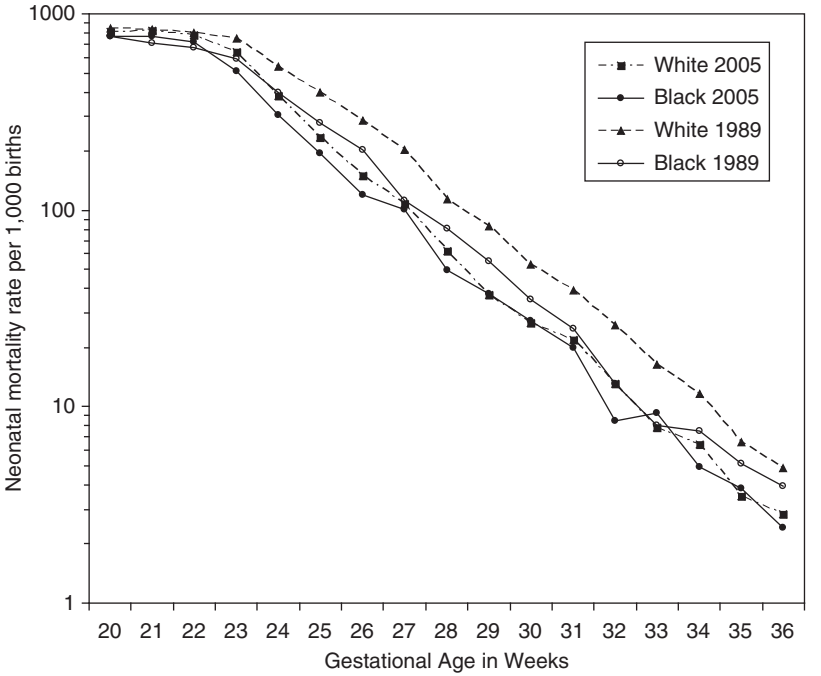


Figure 8.1 Neonatal mortality rates in 1989 and 2005 by gestational age and race.

Noting the log scale, the mortality rate increases exponentially as gestational age at birth declines, lower for the more recent calendar time period.

Beyond the fatal conditions with which PTB is strongly linked, many serious medical conditions are increased among those preterm infants who survive. These include pulmonary complications (e.g., respiratory distress syndrome, bronchopulmonary dysplasia, chronic lung disease), gastrointestinal disease (e.g., necrotizing enterocolitis), immune deficiencies that leave the preterm infant more vulnerable to infection, cardiovascular abnormalities (e.g., patent ductus arteriosis), auditory and visual deficits, and an extensive array of neurodevelopmental abnormalities, many of which are irreversible as more fully described in Chapter 12. Major neurodevelopmental impairments such as cerebral palsy and mental retardation are associated with PTB, as well as more subtle deficits, as reflected in diminished cognitive performance in the form of lower intelligence quotient (IQ) scores throughout life. Aggregate measures of newborn well-being, such as length of hospital stay, are reflective of the increased morbidity of preterm infants as well. As has been found for mortality, the risks of sequelae for the late preterm (born at 34–36 weeks' completed gestation) are elevated relative to term infants, but not as dramatically as for more severely preterm infants. However, the increased frequency of marginally compared to severely PTBs leaves a substantial public health burden in terms of health status, medical services, and/or quality of life.

POST-TERM BIRTH

Prevalence

Post-term gestation is defined as any pregnancy that lasts 294 days (42 completed weeks) or more. The incidence of post-term gestation varies greatly, depending on which dating method is used. Recent studies showed that the incidence of post-term gestation was 10%–12% based on the LMP, but only about 3% based on the ultrasound dating (Taipale 2001; Savitz 2002). The incidence was further reduced from 3.7%, based on the early second-trimester ultrasound dating, to 2.7%, based on the first-trimester ultrasound dating in a recent study (Caughey 2008). It should be noted, however, that the above incidence figures may have been affected by an increasing temporal prevalence for induction of labor in the United States, especially in 41 weeks of gestation.

Etiology and risk factors

The mechanism of onset of parturition in humans has yet to be fully elucidated. However, several risk factors have been identified for post-term gestation. The most important factor "causing" post-term gestation by far is an error in estimating gestational age. Studies show that the mean duration of the follicular phase is 16–18 days, with a standard deviation representing 6–8 days (Lynch 2007). Women are more likely to have delayed ovulation (or later than day 14 of the cycle) than early ovulation in any given cycle (Baird 1995). Thus, even in women who have an accurate LMP, post-term pregnancy may be erroneously diagnosed due to delayed ovulation arising from menstrual cycle lengths greater than the so-called norm of 28 days.

In most cases of true post-term pregnancy, the cause is not known. Nulliparity, maternal overweight and obesity, and history of previous post-term birth are the only established risk factors. Lynch et al. (2007) showed that nulliparous women had a 1.4 to 2.0-fold increased risk to have a post-term birth in comparison to multiparous women. The relative risk for overweight and obese women was 1.3 to 1.4, using normal weight women as the reference group. One population-based study showed that, after one prolonged pregnancy, the risk of a second such pregnancy in the subsequent birth increased from 10% to 27% (Mogren 1999). If there have been two successive prolonged pregnancies, the risk in the subsequent birth rises to 39%.

More recent evidence suggests that genetics also plays a role. For example, if the mother was born after a prolonged pregnancy, her daughter is also at increased risk of a prolonged pregnancy (relative risk [RR] = 1.3, 95% CI 1.0–1.7) (Mogren 1999). A Danish twin study showed that the concordance rate for female twin pairs for a post-term gestation was higher for monozygotic than for dizygotic twin pairs (Laursen 2004). Biometric modeling suggested that genetic factors account for 23%–30% of prolonged gestations. On the other hand, paternal genes had no obvious effect. Some rare conditions also predispose women to post-term pregnancy, such as placental sulfatase deficiency (an X-linked recessive disorder characterized

by a low level of estriol), fetal adrenal insufficiency or hypoplasia, and fetal anencephaly (Norwitz 2007).

Effect on perinatal outcomes

Post-term pregnancy is associated with an increased risk of stillbirth and neonatal and infant mortality (Hilder 1998). As the placenta ages with advancing gestation, placental insufficiency, fetal acidemia, and preeclampsia are more common among post-term births. Approximately 1 in 5 post-term infants has "fetal postmaturity syndrome," characterized by decreased subcutaneous fat, skin desquamation, and long fingernails, often with yellow meconium staining of the nails, skin, and vernix. These pregnancies are more likely to have oligohydramnios, nonreassuring fetal status, intrauterine passage of meconium, and neonatal complications (e.g., hypoglycemia, respiratory insufficiency, and seizures). Meconium aspiration syndrome is mainly a complication in post-term infants. Consequently, the risk of perinatal mortality in ongoing pregnancies at 42 weeks of gestation is twice that at 40 weeks and increases fourfold at 43 weeks and five- to sevenfold at 44 weeks (Norwitz 2007).

Post-term fetuses are often larger than term fetuses, which increase the risk of prolonged labor, cephalopelvic disproportion, and shoulder dystocia. These intrapartum complications in turn increase the risks of both neonatal and maternal morbidity. For instance, cesarean delivery rates double for post-term births (Rand 2000). Also raised are the risks of other maternal complications such as perineal laceration, endometritis, hemorrhage, and thromboembolic disease (Norwitz 2007). With the common practice of preventive induction at 41 weeks, perinatal morbidity and mortality related to post-term gestation is decreasing (Yoder 2002).

In summary, ultrasonography has improved the accuracy of gestation dating substantially. Frequent induction of labor, coupled with accurate dating, prevents the majority of post-term births in many regions. However, preterm birth is still a major problem with few effective prevention or treatment strategies. It remains a high priority for perinatal research.

REFERENCES

ACOG Committee on Practice Bulletins–Obstetrics. 2004. Clinical management guidelines for obstetricians-gynecologists. No. 55 Management of Postterm Pregnancy. *Obstet Gynecol* 104: 639–46.

Alexander, G.R., de Caunes, F., Hulsey, T.C., Tompkins, M.E., and Allen, M. 1992. Validity of postnatal assessments of gestational age: A comparison of the method of Ballard et al. and early ultrasonography. *Am J Obstet Gynecol* 166: 891–95.

Baird, D.D., McConnaughey, D.R., Weinberg, C.R., Musey, P.I., Collins, D.C., Kesner, J.S., Knecht, E.A., and Wilcox, A.J. 1995. Application of a method for estimating day of ovulation using urinary estrogen and progesterone metabolites. *Epidemiology* 6: 547–50.

Ballard, J.L., Khoury, J.C., Wedig, K., Wang, L., Eiler-Walsman, B.L., and Lipp, R. 1991. New Ballard Score, expanded to include extremely premature infants. *J Pediatr* 119: 417–23.

Ballard, J.L., Novak, K.K., and Driver, M. 1979. A simplified score for assessment of fetal maturation of newly born infants. *J Pediatr* 95: 769–74.

Behrman, R.E. and Stith-Butler, A. eds. 2007. *Preterm birth. Causes, consequences, and prevention*. Committee on Understanding Premature Birth and Assuring Health Outcomes. Washington, DC: National Academy Press.

Berg, A.T. 1991. Menstrual cycle length and the calculation of gestational age. *Am J Epidemiol* 133: 585–89.

Berkowitz, G.S., Blackmore-Prince, C., Lapinski, R.H., and Savitz, D.A. 1998. Risk factors for preterm birth subtypes. *Epidemiology* 9: 279–83.

Blondel, B., Kogan, M.D., Alexander, G.R., Dattani, N., Kramer, M.S., Macfarlane, A., and Wen, S.W. 2002. The impact of the increasing number of multiple births on the rates of preterm birth and low birthweight: an international study. *Am J Public Health* 92: 1323–30.

Bukowski, R., Smith, G.C.S., Malone, F.D., Ball, R.H., Nyberg, D.A., Comstock, C.H., et al. 2007. Fetal growth in early pregnancy and risk of delivering low birth weight infant: prospective cohort study. *BMJ* 334: 836.

Caughey, A.B., Nicholson, J.M., and Washington, E. 2008. First- vs second-trimester ultrasound: The effect on pregnancy dating and perinatal outcomes. *Am J Obstet Gynecol* 198:703.e1–703.e6.

Cunningham, F.G., Gant, N.F., Leveno, K.J., Gilstrap, L.C., III, Hauth, J.C., and Wenstrom, K.D. 2001. *Williams obstetrics*, 21st ed., pp. 129–34. New York: McGraw-Hill.

Degani, S. 2001. Fetal biometry: Clinical, pathological, and technical considerations. *Obstet Gynecol Surv* 56: 159–67.

Dubowitz, L.M., Dubowitz, V., and Goldberg, C. 1970. Clinical assessment of gestational age in the newborn infant. *J Pediatr* 77: 1–10.

Eichner, S.F. and Timpe, E.M. 2004. Urinary-based ovulation and pregnancy: Point-of-care testing. *Ann Parmacother* 38: 325–31.

Engel, S.A.M., Erichsen, H.C., Savitz, D.A., Thorp, J., Chanock, S.J., and Olshan, A.F. 2005. Risk of spontaneous preterm birth is associated with common proinflammatory cytokine polymorphisms. *Epidemiology* 16: 469–77.

Filly, R.A. and Hadlock, F.P. 2000. Sonographic determination of menstrual age. In *Ultrasonography in obstetrics and gynecology*, 4th ed., pp. 146–70. P.W. Callen, ed. Philadelphia, PA: Saunders.

Geirsson, R.T. and Busby-Earle, R.M. 1991. Certain dates may not provide a reliable estimate of gestational age. *Br J Obstet Gynaecol* 98: 108–09.

Hadlock, F.P., Deter, R.L., Harrist, R.B., and Park, S.K. 1982. Fetal head circumference: Relation to menstrual age. *Am J Radiology* 138: 649–53.

Hadlock, F.P., Harrist, R.B., Shah, Y.P., King, D.E., Park, S.K., and Sharman, R.S. 1987. Estimating fetal age using multiple parameters: A prospective evaluation in a racially mixed population. *Am J Obstet Gynecol* 156: 955–57.

Hall, M.H., Carr-Hill, R.A., Fraser, C., Campbell, D., and Samphier, M.L. 1985. The extent and antecedents of uncertain gestation. *Br J Obstet Gynaecol* 92: 445–51.

Helmerhorst, F.M., Perguin, D.A.M., Donker, D., and Keirse, M.J.N.C. 2004. Perinatal outcome of singletons and twins after assisted conception: A systematic review of controlled studies. *BMJ* 328: 261.

Henriksen, T.B., Wilcox, A.J., Hedegaard, M., and Secher, N.J. 1995. Bias in studies of preterm and postterm delivery due to ultrasound assessment of gestational age. *Epidemiology* 6: 533–37.

Hilder, L., Costeloe, K., and Thilaganathan, B. 1998. Prolonged pregnancy: Evaluating gestation-specific risks of fetal and infant mortality. *Br J Obstet Gynaecol* 105: 169–73.

Kieler, H., Axelsson, O., Nilsson, S., and Waldenstrom, U. 1995. The length of human pregnancy as calculated by ultrasonographic measurement of the fetal biparietal diameter. *Ultrasound Obstet Gynecol* 6: 353–57.

Klebanoff, M.A. 1998. Conceptualizing categories of preterm birth. *Prenatal Neonatal Med* 3: 13–15.

Klebanoff, M.A. and Shiono, P.H. 1995. Top down, bottom up and inside out: Reflections on preterm birth. *Paediatr Perinatal Epidemiol* 9: 125–29.

Kramer, M.S., Demissie, K., Yang, H., Platt, R.W., Sauvé, R., and Liston, R. 2000. The contribution of mild and moderate preterm birth to infant mortality. *JAMA* 284: 843–49.

Leitich, H. and Kiss, H. 2007. Asymptomatic bacterial vaginosis and intermediate flora as risk factors for adverse pregnancy outcome. *Best Pract Res Clin Obstet Gynaecol* 21: 375–90.

Lynch, C.D. and Zhang, J. 2007. The research implications of the selection of a gestational age estimation method. *Paediatric Perinatal Epidemiol* 21(Suppl 2): 86–96.

MacDorman, M.F., Mathews, T.J., Martin, J.A., and Malloy, M.H. 2002. Trends and characteristics of induced labour in the United States, 1989–1998. *Paediatr Perinatal Epidemiol* 16: 263–67.

Macones, G., Jeffcoat, M., Parry, S., et al. 2008. Screening and treating periodontal disease in pregnancy does not reduce the incidence of preterm birth: Results from the PIPS study. *Am J Obstet Gynecol* 199(Suppl 1): S3 (abstract).

Martin, J.A., Hamilton, B.E., Sutton, P.D., Ventura, S.J., Menacker, F., Kirmeyer, S., and Munson, M.L. 2007. Births: Final data for 2005. In *National vital statistics reports*, vol. 56, no. 6. Hyattsville, MD: National Center for Health Statistics.

Meis, P.J., Goldenberg, R.L., Mercer, B.M., Iams, J.D., Moawad, A.H., Miodovnik, M., et al. 1998. The preterm prediction study: Risk factors for indicated preterm birth. *Am J Obstet Gynecol* 178: 562–67.

Meis, P.J., Klebanoff, M., Thom, E., Dombrowski, M.P., Sibai, B., Moawad, A.H., et al. 2003. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. *N Engl J Med* 348: 2379–85.

Meis, P.J., Michielutte, R., Peters, T.J., Wells, H.B., Sands, R.E., Coles, E.C., Johns, K.A. 1995. Factors associated with preterm birth in Cardiff, Wales. II. Indicated and spontaneous preterm birth. *Am J Obstet Gynecol* 173: 597–602.

Mogren, I., Stenlund, H., Hogberg, U. 1999. Recurrence of prolonged pregnancy. *Int J Epidemiol* 28: 253–57.

Moldenhauer, J.S., Stanek, J., Warshak, C., Khoury, J., and Sibai, B. 2003. The frequency and severity of placental findings in women with preeclampsia are gestational age dependent. *Am J Obstet Gynecol* 189: 1173–77.

Mongelli, M., Wilcox, M., and Gardosi, J. 1996. Estimating the date of confinement: Ultrasonographic biometry versus certain menstrual dates. *Am J Obstet Gynecol* 174: 278–81.

Moutquin, J-M. 2003. Classification and heterogeneity of preterm birth. *BJOG* 110: 30–33.

Naegele, F.C. 1836. *Lehrbuch der Geburtshülfe für Hebammen*, 3rd ed., p. 54. Heidelberg Mohr.

Nguyen, T.H., Larsen, T., Engholm, G., and Moller, H. 1999. Evaluation of ultrasound-estimated date of delivery in 17450 spontaneous singleton births: Do we need to modify Naegele's rule? *Ultrasound Obstet Gynecol* 14: 23–28.

Norwitz, E.R., Snegovskikh, V.V., and Caughey, A.B. 2007. Prolonged pregnancy: When should we intervene? *Clin Obstet Gynecol* 50: 547–57.

Offenbacher, S., Beck, J., Jared, H., et al. 2008. Maternal oral therapy to reduce obstetric risk (MOTOR): A report of a multi-centered periodontal therapy randomized-controlled trial on rate of preterm delivery. *Am J Obstet Gynecol* 199(Suppl 1):S2 (abstract).

Pedersen, N.G., Figueras, F., Wojdemann, K.R., Tabor, A., and Gardosi, J. 2008. Early fetal size and growth as predictors of adverse outcome. *Obstet Gynecol* 112: 765–71.

Petrini, J.R., Dias, T., McCormick, M.C., Massolo, M.L., Green, N.S., and Escobar, G.J. 2009. Increased risk of adverse neurological development for late preterm infants. *J Pediatr* 154: 169–76.

Poston, L., Briley, A.L., Seed, P.T., Kelly, F.J., and Shennan, A.H. 2006. Vitamin C and vitamin E in pregnant women at risk for pre-eclampsia (VIP Trial): Randomized placebo-controlled trial. *Lancet* 367: 1145–54.

Rand, L., Robinson, J.N., Economy, K.E., and Norwitz, E. 2000. Post-term induction of labor revisited. *Obstet Gynecol* 96: 779–83.

Resnik, J.L. and Resnik, R. 2004. Post-term pregnancy. In *Maternal-fetal medicine: Principles and practice*, 5th ed., p. 664. R.K. Creasy, R. Resnik, and J.D. Iams, eds. Philadelphia: Saunders.

Savitz, D.A., Blackmore, C.A., and Thorp, J.M. 1991. Epidemiology of preterm delivery: Etiologic heterogeneity. *Am J Obstet Gynecol* 164: 467–71.

Savitz, D.A., Dole, N., Herring, A.H., Kaczor, D., Murphy, J., Siega-Riz, A.M., et al. 2005. Should spontaneous and medically indicated preterm birth be separated for studying aetiology? *Paediatr Perinatal Epidemiol* 19: 97–105.

Savitz, D.A. and Murnane, P. 2010. Behavioral influences on preterm birth: A review. *Epidemiology* 21: 291–99.

Savitz, D.A., Terry, J.W., Dole, N., Thorp, J.M., Jr., Siega-Riz, A.M., and Herring, A.H. 2002. Comparison of pregnancy dating by last menstrual period, ultrasound scanning, and their combination. *Am J Obstet Gynecol* 187: 1660–66.

Singh, G.K. and Yu, S.M. 1996. Adverse pregnancy outcomes: Differences between US- and foreign-born women in major US racial and ethnic groups. *Am J Public Health* 86: 837–43.

Smith, G.C.S., Smith, M.F.S., McNay, M.B., and Fleming, J.E.E. 1998. First-trimester growth and the risk of low birth weight. *N Engl J Med* 339: 1817–22.

Smith, R. 2007. *Parturition*. N Engl J Med 356: 271–83.

Taipale, P. and Hiilesmaa, V. 2001. Predicting delivery date by ultrasound and last menstrual period in early gestation. *Obstet Gynecol* 97: 189–94.

Varma, R. and Gupta, J.K. 2006. Antibiotic treatment of bacterial vaginosis in pregnancy. Multiple meta-analyses and dilemmas in interpretation. *Eur J Obstet Gynecol Repro Biol* 124: 10–14.

Yoder, B.A., Kirsch, E.A., Barth, W.H., and Gordon, M.C. 2002. Changing obstetric practices associated with decreasing incidence of meconium aspiration syndrome. Obstet Gynecol 99: 731–39.

9

Fetal Growth: Measurement and Evaluation

MARY L. HEDIGER AND K.S. JOSEPH

Fetal size and growth reflect critical dimensions of fetal health, and abnormal growth is strongly associated with perinatal death and serious neonatal morbidity. However, defining what constitutes appropriate fetal growth presents a challenge for several reasons: the normal variations in fetal size, proportions, and growth associated with parental and environmental factors; a limited ability to obtain accurate measurements of the fetus in utero; and the choice of reference for defining abnormal size or patterns of growth can influence the interpretation.

Care is needed in addressing distinctions between what is a measure of fetal size or status and what is a measure of growth, as well as among prevailing concepts, such as those embodied in the terms *small-for-gestational age* (SGA), *large-for-gestational age* (LGA), *fetal growth restriction* (FGR), *intrauterine growth restriction* (IUGR), or *macrosomia*. The former two terms refer to the size of the fetus or neonate (at a particular gestational age) in relation to some normative reference and could indicate a constitutional phenomenon or an abnormality of fetal growth. A fetus suffering from growth restriction, on the other hand, does not have to be SGA, since the growth restriction may refer to a pattern of inadequate growth over time, resulting in the fetus not reaching its potential size, while remaining above the cutoff for absolutely small size. Both of the constructs, FGR and IUGR, were formally defined in the literature as "fetal growth retardation" and "intrauterine growth retardation," but over the past several decades, in recognition that a small fetus is actually more likely to be constitutionally small (biologically or physiologically restricted by parental size, for example) than growth retarded—which implies a pathological cause—the more general term "restriction" has gained favor.

At the other end of the fetal growth spectrum, a different but important clinical distinction is made between the LGA fetus and one in whom growth in one or more compartments has been accelerated and is macrosomic (≥4,000 or 4,500 g). Macrosomic growth is of relevance to obstetrics, given the association of macrosomia with gestational diabetes mellitus (GDM) and a higher propensity for birth trauma. For these and related reasons, measurement and evaluation of fetal size and growth represent a complex and evolving area of research that has an important bearing on obstetric practice and population health.

ROLE OF THE PLACENTA

Fetal blood circulation occurs within a closed system that exchanges oxygen, nutrients, and waste products with the maternal circulation through a placental interface. The role of the placenta extends far beyond that of a physical structure for this exchange, however, and includes the production of steroid and protein hormones in a greater amount and diversity than any other endocrine tissue (Cunningham et al. 2005; Murphy et al. 2006; Salafia et al. 2006). The placental hormones that play an important role in fetal growth include human placental lactogen, growth hormone variant, growth hormone-releasing hormone, and leptin. Human placental lactogen is a major hormone of pregnancy and is responsible for mobilizing maternal nutrient stores through maternal lipolysis and an anti-insulin–like action, which facilitates transport of amino acids to the fetus. It also functions to secrete angiogenic hormones and plays a role in the formation of the fetal vasculature (Cunningham et al. 2005).

Fetal size and growth are associated with several gross aspects of placental structure, including its shape, morphology, and size. Placental size is generally correlated positively with fetal size, although notable exceptions occur that may seem counterintuitive. For example, with maternal anemia, the placenta may be enlarged in relation to the size of the fetus, as the placenta hypertrophies in response to a relatively anoxic environment. The normal placenta is disc shaped, and abnormal lobulation is seen with maternal smoking and other conditions typically associated with poor fetal growth. The location of umbilical cord insertion into the placenta (central vs. peripheral); placental diameter, thickness, and weight; gross placental infarcts; and microscopic lesions in the placental villi (which can lead to restricted umbilical arterial blood flow) also correlate closely with fetal growth (Murphy et al. 2006). Abnormal placentation is more commonly seen in the placentas of twins, and vascular connections within the placenta of monochorionic twins that cause abnormal shunting of blood can lead to one fetus becoming growth restricted and anemic, while the other becomes polycythemic and hydropic (twin-to-twin transfusion syndrome) (El Kateb and Ville 2008).

SOURCES OF INFORMATION ABOUT FETAL GROWTH

Two complementary sources of information are available about fetal size and growth. The first is indirect (no measurement of the fetus in utero) and based on the anthropometric measurement of the abortus (Birkbeck et al. 1975; Kaul et al. 1986), stillbirth (nonmacerated), or live-born neonate, with a weight immediately at delivery (i.e., birth weight) being the most common and reliable measurement and the one most commonly reported to vital registries. Although birth weight is the most reliable measurement, it is not measured totally without error or variation, depending upon the timing of the measurement. For example, weight at birth can be affected by meconium or urine passage or the timing of any feeding relative to weighing.

Birth length, head circumference, and chest circumference are also usually measured in the neonatal nursery and recorded as part of the medical record, but are

rarely included in vital registries. The measurement of birth length can be biased by the fact that, at delivery, the neonate is in a flexed state that only relaxes after several days. As the degree of flexion decreases over the first few days after birth, measured length increases (Shinwell and Shlomo 2003), such that mean measured length increases significantly by approximately 0.2 cm between admission and age 1 day and by a further 0.2 cm by age 2 days. Head molding at delivery can bias measurements of head circumference and may not resolve for several days. In addition to weight, length, and head circumference, the same anthropometric techniques and measurements that are used for children and adults (e.g., trunk and limb circumferences, skinfold thicknesses, and the calculation of limb fat and muscle areas) have been used in neonates to determine if pregnancy complications have affected fetal growth in proportions or compartments (Catalano et al. 1995; Stetzer et al. 2002). There are also a score of other specialized cranial and somatic measurements for which normative references exist (Meany and Farrer 1996), but these are primarily of interest to dysmorphologists and geneticists who use them to diagnose birth defects or confirm the presence of genetic syndromes.

Anthropometry at birth can be augmented by various methods that more directly measure body composition or compartments (e.g., lean body mass, fat, water, bone mineral density), such as bioelectric impedance analysis (BIA), dual energy x-ray absorptiometry (DXA) (Beltrand et al. 2008), and air displacement plethysmography (Ma et al. 2004; Ellis et al. 2007; Ellis 2007), although these latter two especially are rarely used outside of research studies because they require specialized laboratory conditions and equipment. Bioelectric impedance analysis is the only one of the methods that might be useful in clinical or field situations. Bioelectric impedance analysis is based on the principle that lean tissues that contain water and free electrolytes conduct an applied electrical signal more readily than fat tissue, and BIA has been proposed as a measure to track the efficacy of hydration therapies in growth-restricted neonates (Gartner et al. 1994). Its widespread use in neonates has been hampered, however, by the lack of adequate reference data.

The other source of information is direct and obtained from fetal imaging. Fetal imaging was originally done using radiographs and was complicated by the limitations and static nature of the x-ray filming. Even well-defined fetal parameters, such as head circumference and bone lengths, could be difficult to measure if the fetus was not fortuitously positioned, and the fetus was generally not even seen radiographically until after the first trimester. Imaging is now done primarily by ultrasonography, which is not only safer (Duck 2008), but more sensitive and flexible and allows for visualization in multiple planes. The gestational and yolk sacs can be seen as early as 4 to 6 weeks' gestation, and fetal crown–rump length (CRL) is measured in the first trimester to establish dates. Thereafter, a number of equations employ various combinations of ultrasound-measured fetal head circumference (HC), abdominal circumference (AC), biparietal diameter (BPD), and femur length (FL) to estimate fetal weight in relation to gestational age (Anderson et al. 2007) as an indicator of fetal growth and well-being.

Ultrasound methods have advanced over the past 20 years to become increasingly more sensitive, such that not only are bony dimensions and ossification centers (Gottlieb and Galan 2008) visualized as they once were on x-rays, but also elements

of fetal body composition, such as cross-sectional subcutaneous fat layers and muscle (Larciprete et al. 2003; Parretti et al. 2003), and organ dimensions (e.g., kidney, liver) can be measured using two-dimensional (2D) ultrasonography and have been for research purposes. Volumes can be obtained using three-dimensional (3D) ultrasound and are beginning to be used for fetal body composition studies (Lee et al. 2009) and to study organ development in relation to growth restriction or macrosomia, although there appears to be, to date, limited improvement of 3D approaches over conventional 2D in terms of clinical care and diagnostics.

FETAL GROWTH AS AN OUTCOME

Despite its limitations, weight at birth is still used as the primary indicator of neonatal status and reflects broadly the outcome of fetal growth processes and complications. A number of cutoffs, indices, and ratios have been devised to relate birth weight to risk of mortality and morbidity (Table 9.1). One of these metrics, low birth weight (LBW), an outcome defined as a birth weight below 2,500 g (5 lb, 8 oz), is by far the most common and is used worldwide in surveillance (UNICEF 2007) and in epidemiologic studies. Although given the advances in perinatal care, the 2,500 g cutoff now seems arbitrary, historically it was the level below which neonatal mortality was seen to be significantly increased. About two-thirds of LBW infants are preterm by dates, and all term LBW infants are considered to be growth restricted. Low-birth-weight infants are at increased risk for complications at birth, including a number of severe conditions, such as respiratory distress syndrome (RDS), intraventricular hemorrhage (IVH), and necrotizing enterocolitis (NEC), especially if they are also preterm. Very-low-birth-weight (VLBW) infants, born below 1,500 g are almost exclusively preterm, and their risks for severe conditions are proportionately higher.

The problem with the LBW construct is that it represents a mixture of both preterm deliveries and growth-restricted infants and is not very useful when the outcome or disease of interest may be differentially related to prematurity or abnormal growth. It also conveys no information about patterns, growth velocity, or the timing of growth-restricting complications. Similarly, using an absolute cutoff for a high birth weight alone (e.g., ≥4,500 g) might underestimate the proportion of neonates whose mothers had gestational diabetes and who were delivered at a late preterm (35–36 weeks) or early-term gestation to avoid delivery complications.

From neonatal measurements, a number of other indices and ratios have been promoted as being more useful in evaluating fetal growth outcomes and more sensitive in indicating growth retardation or excessive growth than birth weight alone (Table 9.1). A fetal growth ratio, which assesses birth weight in proportion to the mean or median (50th percentile) for gestational age, can be constructed from any birth-weight-for-gestational age reference. Where means and standard deviations (SD) of birth weight are available by gestational age, a relative birth weight (z-score or standard deviation unit, SDU) can be calculated. Because these indices and ratios are based on a measure of gestational age, in developing countries where

Table 9.1 Measures and Common Indices of Birth Size Based on Neonatal Anthropometry

Measure	*Definition and comments*
Neonatal Anthropometry	
Birth weight	Weight in pounds or grams at delivery
Birth length	Recumbent length in inches or centimeters measured at or shortly after delivery; normal neonatal flexed condition may bias the measurement shorter in the first 2 days of life
Head circumference	Head circumference in inches or centimeters measured at or shortly after delivery; head molding at delivery may bias measurement
Chest circumference	Chest circumference in inches or centimeters measured at or shortly after delivery, measured at the level of the nipples
Neonatal Indices-Continuous	
Fetal growth ratio	Percent of mean or median (100%) birth weight for gestational age calculated using means or medians by gestational age (and usually by gender)
Relative birth weight	*Z*-score (SDU) birth weight for gestational age calculated using means and standard deviations by gestational age (and usually by gender)
Weight-for-length	2000 CDC reference percentiles, cutoffs usually at <10th percentile (low weight-for-length) and ≥90th percentile (high weight-for-length); weight-for-length at birth by gender determined for infants <1,500 g
Ponderal index	Calculated as (weight (g)/length^3 (cm)) × 100 or (weight (kg)/length^3 (m)), cutoffs at <10th percentile (low ponderosity-for-length) and ≥90th percentile (high ponderosity-for-length)
Neonatal Indices–Categorical	
Birth weight categories	Low birth weight (LBW) <2,500 g (5 lb, 8 oz) Very low birth weight (VLBW) <1,500 g (3 pounds, 5 ounces) Extremely low birth weight (ELBW) <1,000 g (2 lb, 3 oz) or <500 g (1 lb, 2 oz) High birth weight or macrosomia ≥4,000 (8 lb, 13 oz) or 4,500 g (9 lb, 14 oz)
Birth-weight-for-gestational age	Small-for-gestational age (SGA) <10th, 5th, 3rd, or 2nd percentile [same as small-for-dates (SFD)] Large-for-gestational age (LGA) ≥90th, 95th, 97th, or 98th percentile [same as large-for-dates (LFD)] Appropriate-for-gestational age (AGA) as reference
Growth restriction proportionality	Asymmetric = wasted, –2 SD weight-for-length (or ponderal index) Symmetric = stunted, –2 SD length-for-age, normal ponderal index

there is limited access to prenatal care and determination of gestation, categories based on birth weight alone are still used predominantly for surveillance of birth outcome and determination of historical trends. Other indices, such as the ponderal index (PI), rely on neonatal measures of length, and birth length is only sporadically available in birth registries or from vital statistics. Reference percentiles for PI at birth by gestational age have been constructed for local populations (e.g., Landmann et al. 2006), but have not been widely used.

The Centers for Disease Control and Prevention (CDC) growth charts for the United States (Kuczmarski et al. 2000) include percentile and *z*-scores for birth weights, birth lengths, and weight-for-length at birth on the charts specific to birth

to 36 months of age. These percentiles are applicable for birth weights of greater than or equal to 1,500 g. The World Health Organization (WHO) standards have available both percentiles and *z*-scores for weight and length for well-grown term neonates at birth (WHO 2006).

MEASURING AND DATING ISSUES

Ultrasound measurement of fetal size has become a reliable tool not only for evaluation of fetal well-being and pregnancy prognosis, but also for confirming menstrual dates or dating pregnancies when menstrual dates may be unknown or in error (Table 9.2). In the first trimester, fetal CRL correlates closely with gestational age and is often used as the gold standard for dates (Hadlock et al. 1992; Daya 1993). The CRL has the added advantage of not being biased by gender differences or ethnic differences at these early gestational ages (Sahota et al. 2009), and the variation is small. However, CRL measurements are not in perfect agreement with dates. The discrepancy between known menstrual dates (expected size) and CRL (observed size) has been shown to be predictive of later growth, such that a greater observed size than expected for dates is associated with a larger birth weight at term and a smaller observed size for dates with a lower birth weight and preterm delivery (Smith et al. 1998, 2004; Bukowski et al. 2007).

After the first trimester and up to about 24–26 weeks' gestation, fetal size (weight) as estimated using a variety of biometric parameters, singly or in combination

Table 9.2 Ultrasound Biometric Parameters for Determination of Fetal Growth and Size, and to Establish dates

Parameter	*Applicable Weeks*		*Reference*
	Used to Date	Used for Fetal Growth	
Crown-rump length (CRL)	7–13	7–13	Hadlock et al. 1992; Daya 1993
Biparietal diameter (BPD)	13–24	13–40	
Head circumference (HC)	13–24	13–40	Hadlock et al. 1991;
Femur length (FL)	13–24	13–40	Anderson et al. 2007
Abdominal circumference (AC)	13–24	13–40	
Transcerebellar diameter (TCD)*	15–24	15–40	Goldstein et al. 1987
Fetal foot length	10–24	24–40**	Drey et al. 2005
Other nomograms			
Orbital diameters	14–36	14–36	Goldstein et al. 1998
Clavicle	14–41	14–41	Sherer et al. 2006
Scapula	16–41	16–41	Dilmen et al. 1995
Radius, ulna, tibia	13–24	13–40	Chitty & Altman 2002
Kidneys	24–38	24–38	Konje et al. 2002
Thymus		19–38	Cho et al. 2007

*May be best for dates because of brain-sparing and the preferential shunting (blood flow) to cerebellum

**Of value later in gestation to predict nasotracheal tube length in neonates (Embleton et al. 2001)

(Table 9.2), is still closely associated with dates and can be used to confirm or modify estimated gestational age. After that point in gestation, the differences in fetal size associated with gender, ethnicity, and a number of other maternal factors, such as cigarette smoking, so affect fetal growth rates as to increase the variance and significantly bias estimates of gestational age based on size (Henriksen et al. 1995; Dietz et al. 2007). Gender differences in fetal size, particularly in head dimensions (BPD, HC), with male fetuses having larger dimensions than female fetuses, are measurable on the average by 12–14 weeks' gestation and significant by 24 weeks (Parker et al. 1984; Davis et al. 1993). Based on head size alone, then, in the third trimester, female fetuses would tend to be dated as younger than males (Henriksen et al. 1995).

Also as early as the second trimester, the ethnic differences in body proportions that are evident in children and adults are already seen in the fetus, which can further bias estimates of gestational age or complicate evaluation of fetal proportions for birth defects screening, depending on the reference used. Beginning in the second trimester (15–20 weeks), the FL of non-Hispanic black fetuses are already longer than those of non-Hispanic white fetuses of the same gestational age (Davis et al. 1993; Shipp et al. 2001), whereas the FL of Asian fetuses are shorter than those of non-Hispanic whites (Shipp et al. 2001), and these differences are magnified in the third trimester. Likewise for the arm, the humerus lengths (HL) of non-Hispanic black fetuses tend to be longer than those of non-Hispanic whites, which are longer than those of Asians (Mastrobattista et al. 2004). Because a short HL is used as a marker, along with other sonographic findings, for the prediction of Down syndrome (Schluter and Pritchard 2005) care should be taken in the second and third trimester that suitable ethnic-specific references are used for dimensions or proportions that are considered indicators of birth defects (Meany and Farrer 1986).

MEASUREMENT APPROACHES FOR DETERMINING ABNORMAL FETAL GROWTH

Two different approaches are used to create charts or nomograms to evaluate fetal, infant, or child size and growth velocity. The first is construction of a *reference*. Growth references are considered to be descriptive (how a fetus or child *is* growing) and population-specific. They can be based on a representative sample, but still include abnormal cases or those who have exposures or conditions that may affect growth. The second approach is to create a *standard* of growth. In contrast to a reference, a standard is proscriptive (how a fetus or child *should* be growing) and is generally not population-specific. Samples for a standard are chosen because they are growing under optimal conditions, and cases that have exposures or conditions that may affect growth are excluded. Unlike in a reference, the variance in a standard is restricted, and the cutoffs or other distributional parameters may be very different. Care should be taken then, before using either a reference or a standard to evaluate fetal growth, to understand the selection of the source sample or population and the assumptions of construction.

General approach for creating fetal growth references and standards

Clinical growth charts for infants and children, whether references (Kuczmarski et al. 2000) or standards (WHO 2006), are based on measured values, with the 3rd percentile (or the approximately equivalent mean minus 2 standard deviations) and the 97th percentile (or the approximately equivalent mean plus 2 standard deviations) used clinically as cutoffs for growth faltering or excessive growth. Birth-weight-for-gestational age references, on the other hand, have tended to use the 10th percentile and the 90th percentile of birth-weight-for-gestational age as the cutoffs for the identification of SGA and LGA infants, respectively. The higher-percentile cutoff for fetal growth references (typically created using vital statistics data) accounts for the mixed normal and abnormal population (i.e., infants from complicated and uncomplicated pregnancies) used in creating the reference.

Traditional birth weight-for-gestational age versus ultrasound-based references

Traditionally, fetal growth references have been created by measuring the birth weights of a relatively large numbers of live-born babies at each gestational week and identifying a cutoff that signifies poor or excessive growth using distributional assumptions. With the 10th, 50th, and 90th percentile thereby determined for each gestational week, fetal growth "curves" have been developed that chart the expected fetal weight through pregnancy. The drawback with this methodology is the extrapolation implicit in using cross-sectional information (weight at birth) for modelling a longitudinal experience. If preterm delivery is more likely to occur among a subset of fetuses whose growth patterns or velocities are different from the subset that continues in utero, this would bias the information in the fetal growth curve derived from birth weights (Hediger et al. 1995; Bukowski et al. 2001; Hutcheon and Platt 2008).

As mentioned, ultrasonographic techniques have been increasingly used to measure fetal size in recent decades. Estimation of fetal weight requires biometric measurements of the fetus including measurements of head size (head circumference, biparietal diameter), abdominal circumference, and FL (Platz and Newman 2008). These measures are combined using one of any number of available formulae (Anderson et al. 2007), the most commonly used being the one proposed by Hadlock and colleagues (1991). The addition of other biometric measurements does not appear to improve the accuracy of fetal weight estimation. Validation studies show that fetal weight estimated by these techniques yields reasonable results for fetuses weighing between 1,500 and 3,999 g. Outside these bounds, however, there is tendency to overestimate or underestimate fetal weight, with errors as high as 25% between the actual and predicted weights (Platz and Newman 2008).

Both birth-weight-for-gestational age and ultrasound-based references, and particularly the trajectories of the percentile curves of interest (e.g., 10th or 90th percentile), have been used to derive *rates* of growth and make inferences about fetal growth in relation to complications. However, because of substantial interindividual variation in the timing and tempo of growth and measurement error, increments

or interval growth inferred from cross-sectional estimated fetal weights sampled at different gestational ages and smoothed percentiles will underestimate variation and inflate the proportion growing abnormally. Assessment of rates of growth should ideally be made against an actual velocity reference, but only a couple of small series of fetuses have been evaluated by ultrasound longitudinally to estimate growth velocity (Owen et al. 1996; Milani et al. 2005). A current need still exists for a properly constructed longitudinal ultrasound reference or standard for interval growth.

Customized and individualized fetal growth standards

It is generally acknowledged that male fetuses have patterns of growth that are distinct from those of female fetuses; females infants have lower birth weight for gestational age and also lower perinatal mortality rates than do male infants. For this reason, most traditional fetal growth references are customized by sex; that is, they provide different SGA and LGA cutoffs for male and female fetuses (Zhang and Bowes 1995; Kramer et al. 2001), although references are available for the sexes combined (Alexander et al. 1996) as well. Gardosi and coworkers (1992, 1995) extended this concept and proposed fetal growth standards that predict optimal growth and are simultaneously customized not only for fetal sex but also for maternal height, weight, parity, and ethnicity. A previous variation of this theme involved the individualization of the fetal growth curve based on two ultrasound measurements in early gestation (Deter and Rossavik 1987). This latter proposition assumed that fetal growth patterns deviate from the norm only in late gestation.

Other ultrasound-based approaches for assessing fetal growth problems

Numerous other approaches have been proposed to identify growth restriction including combinations of biometric measures, such as an elevated FL-to-AC ratio (FL/AC), an elevated head-to-AC ratio (HC/AC), or elevated transverse cerebellar diameter-to-AC ratio (TCD/AC) (Gottlieb and Galan 2008; Platz and Newman 2008). The HC/AC is particularly useful for detecting asymmetric growth. Another approach for diagnosing fetal growth restriction is based on the amount of amniotic fluid measurable, with decreased amniotic fluid indicative of poor perfusion of the fetal kidneys and associated growth restriction. Currently, the most useful clinical approach for the diagnosis of compromised fetal growth involves umbilical artery Doppler velocimetry among fetuses found to be growth restricted by biometric methods (Cunningham et al. 2005; Zhang, J. et al. 2010).

Controversies and alternative approaches for assessing growth patterns

One intriguing feature of traditional fetal growth standards is the decline in fetal weight that is observed at late gestation. Do fetuses at post-term gestation lose weight, or is this phenomenon a consequence of cross-sectional information being

used to model longitudinal fetal growth? Some fetal growth references have deployed methods that have eliminated this decline in fetal weight at late gestation (Kramer et al. 2001).

Customized fetal growth standards (Gardosi et al. 1992, 1995), although popular, have also elicited some controversy. Proponents claim that such standards perform better than other references in terms of predicting perinatal mortality (Clausson et al. 2001; McCowan et al. 2005; Ego et al. 2006), but others have argued that this is merely an artifact of the method used, which leads to a higher proportions of preterm births within the customized SGA subset (Zhang et al. 2007; Hutcheon et al. 2008).

The relatively simple methodology required for the creation of fetal growth references (determining percentiles using data from a vital registry, or perinatal or ultrasound database) has led to a plethora of fetal growth references in the literature. Should each county have its own reference or standard of fetal growth? This question is controversial, given the recent creation of a standard for infant and child growth (WHO 2006) that was designed for universal application. The more general question regarding the selection of factors that require customization is also thought-provoking because percentile-based methods could be used to produce separate references for infants of mothers who smoke, for singletons and twins, and for a host of other factors that affect birth weight. This issue reiterates the profound difference between a descriptive reference and a normative standard.

Alternative approaches to fetal growth reference creation have been proposed recently and include one advocating epidemiologic modeling for addressing the missing data problem in traditional fetal growth references (Hutcheon and Platt 2008). Since traditional birth-weight-for-gestational age charts are created based on live births at each gestation, they do not reflect the weights of the large proportion of undelivered fetuses. The weight of these "missing" fetuses needs to be incorporated into the fetal growth chart if the reference is to reflect the growth patterns that prevail in utero.

Another proposed approach to the creation of standards attempts to determine the optimal birth weights at each gestational age over which serious neonatal morbidity and mortality rates are lowest and then determine relevant cutoffs (Joseph et al. 2009). Among singleton males at 40 weeks, serious neonatal morbidity/mortality rates were lowest between 3,012 and 3,978 g. The low end of this optimal birth weight range for females was 37 g less (consistent with a priori expectations). The low end of the optimal birth weight range was 152 g less for twins compared with singletons, suggesting that there might be plurality-specific thresholds for obstetric intervention and for neonatal growth monitoring and nutritional supplementation (Joseph et al. 2009).

FETAL GROWTH AND PRETERM DELIVERY

It has long been recognized that poor fetal growth increases the risk of preterm delivery and that the proportion of growth-restricted infants is likely to be substantially higher among preterm live births than that defined by a birth-weight-for-gestational

age reference and SGA categorization (Goldenberg et al. 1985; Ott 1993; Hediger et al. 1995; Smith-Bindman et al. 2002). This is not surprising, given that fetal compromise and fetal growth restriction are indications for iatrogenic preterm delivery. However, studies have shown that, even within spontaneous preterm live births, there is a relative preponderance of infants with a birth-weight-for-gestational age that is substantially different from the mean weight as defined using an ultrasound-based reference (Morken et al. 2006). Interestingly, in preterm gestational age categories of less than 34 weeks, this preponderance is restricted to infants with birth weights that are much lower than the mean birth weight for gestational age, whereas at 34–36 weeks' gestation, there is an excess of infants with birth weights that are significantly smaller and significantly larger than the mean (Morken et al. 2006).

Ultrasonographic studies that have followed the growth patterns of twins have shown that slow growth for one or both twins was associated with higher rates of preterm delivery (Hediger et al. 2005). Another large study of South American singletons showed a more complex relation between growth patterns and preterm delivery. Larger estimated fetal weight prior to 23 weeks and larger estimated fetal weight between 23 and 28 were both associated with significantly higher rates of preterm birth. This pattern was reversed between 28 and 33 weeks, and fetuses with a higher estimated fetal weight at this gestation had significantly lower rates of preterm delivery (Lampl et al. 2009).

FETAL GROWTH IN MULTIFETAL PREGNANCIES

The trajectory of fetal growth in twins throughout gestation has been observed and described, both using ultrasonography and ultrasonography combined with birth weight, with the finding that the trajectory indicates a slowing of growth relative to singletons for almost all twins after 28 weeks' gestation (Min et al. 2000; Smith et al. 2001). Before 28 weeks' gestation, however, all twins should be expected to grow at rates comparable with singletons. Even twins who had been growing at rates similar to singletons before 28 weeks may show slowed rates of growth in the third trimester (Buckler and Green 1994; Ananth et al. 1998; Glinianaia et al. 2000; Gielen et al. 2008; Joseph et al. 2009).

The major determinant of variation in twin fetal growth is not zygosity, but instead chorionicity. The approximate 65% of monozygotic twin pairs that are dichorionic show patterns of fetal growth similar to dizygotic twins. On the other hand, monozygotic, monochorionic twin pairs tend have a growth trajectory that falls below even that of dichorionic twins, not to mention singletons. A couple of birth-weight-for-gestational age references (Naeye et al. 1966; Ananth et al. 1998) have stratified by chorionicity, but this factor is usually not considered in references for twin fetal growth because chorionicity is generally not included in vital statistics and obstetric databases, and because monochorionicity is considered a risk factor for poor fetal growth that should be evaluated by comparison with normal references, either twin or singleton.

It has been hypothesized that the pattern of growth for even well-growing twins in the third trimester is a normal down-regulation in response to a constricted

uterine environment or limited nutrition and may underscore the need for an ultrasound or birth-weight-for-gestational age reference that is specific for twins. The primary rationale for a separate reference is that optimal intrauterine growth and development is achieved earlier in gestation for twins than for singletons and at a lower birth weight (Luke et al. 1993; Dodd et al. 2003; Joseph et al. 2009), but as for singleton births, it appears that such slow-downs may be associated with earlier delivery and a greater risk for morbidity associated with premature birth (Hediger et al. 2005). This would argue against a separate reference for twins, or, at least, for the development of a growth velocity reference for singletons and twins to allow for the accurate assessment of interval growth.

RISK FACTORS FOR ABNORMAL FETAL GROWTH

Recognizing risk factors for abnormal fetal growth allows clinicians to anticipate potential problems and institute more careful screening regimens. From the community medicine standpoint, knowledge of risk factors helps to explain temporal trends in fetal growth, and to develop policy initiatives to optimize health and to anticipate future trends.

Causes of poor fetal growth

The causes of poor fetal growth may be categorized into maternal, fetal, and placental factors (Das and Sysyn 2004; Maulik 2006). Maternal factors include maternal diseases and pregnancy complications, the most common of which are hypertensive disorders in pregnancy. These include chronic hypertension, preeclampsia, and the other hypertensive variants that lead to a decrease in uteroplacental perfusion. Other maternal diseases that can lead to growth restriction include autoimmune disease (such as antiphospholipid syndrome and systemic lupus erythematosus), diabetes, and renal and cardiac diseases, among others.

One important maternal risk factor for poor fetal growth in industrialized countries is cigarette smoking. It is estimated that the average birth weight among infants of women who smoke is 150 g less than the average birth weight of infants born to nonsmoking women (Kramer 1987). In some South Asian and other less industrialized countries, where some rates of LBW are 29% (UNICEF 2007), maternal prepregnancy size and related factors are responsible for high levels of poor fetal growth. Another factor associated with SGA live births is low socioeconomic status (Joseph et al. 2007). Poverty represents a collection of behavioral and other factors (in addition to maternal smoking) that contributes to poor fetal growth, including higher rates of substance abuse, hard physical work, short interpregnancy interval, and poor maternal health. Various medications, including antiepileptic agents and multiple courses of antenatal steroids (Murphy et al. 2009), are also associated with higher rates of poor fetal growth.

Chromosomal anomalies (such as Down syndrome and trisomy 18) and other congenital malformations of the fetus are strongly associated with poor fetal growth. Chromosomal anomalies in particular are associated with symmetrical

growth restriction (unlike other causes in which the head is spared). Multiple malformations, anencephaly, microcephaly, cardiac defects (such as tetralogy of Fallot), and abdominal wall defects are also associated with high rates of poor growth. Perinatal infections due to protozoan parasites (such as malaria) and viruses (such as rubella, cytomegalovirus, and human immunodeficiency virus [HIV]) are important causes of fetal growth restriction worldwide. The many maternal and fetal factors listed here often exert their adverse effects on fetal growth by affecting the placenta and reducing uteroplacental circulation. Other abnormalities of the placenta, such as placenta previa and placental abruption, are also associated with growth restriction (as causes or as consequences). Placental factors also play an important role in the poor fetal growth seen in multifetal pregnancy.

Causes of excessive fetal growth

The main maternal causes of excessive fetal growth are diabetes mellitus and maternal obesity (Das and Sysyn 2004; Maulik 2006). Uncontrolled maternal diabetes mellitus, which can be preexisting type 1 or type 2, or gestational, leads to maternal hyperglycemia and consequently fetal hyperglycemia. The higher output of fetal insulin produced as a response leads to excessive uptake of glucose by fetal tissues and an increase in fetal growth. Maternal obesity is also associated with excessive fetal growth, although the macrosomic infants of diabetic mothers have a different morphology compared with macrosomic infants of mothers who do not have diabetes mellitus. Some genetic and chromosomal disorders, such as Beckwith-Wiedemann syndrome, are also associated with LGA infants (Taeusch et al. 2004).

Temporal trends in fetal growth

Fetal size has increased in industrialized countries in recent years. Much of this change has occurred due to reductions in SGA live births and increases in LGA live births among term births (Kramer et al. 2002). The factors that have acted to increase fetal growth at the population level include increases in maternal weight, body mass index (BMI, weight-for-height^2), increases in gestational weight gain, and reductions in maternal smoking. Factors that have adversely affected fetal growth trends include increases in maternal age, hypertension, and diabetes. However, some of the documented changes could be a consequence of improvements in the accuracy of gestational age ascertainment due to the increasing use of ultrasonography (Kramer et al. 2002). In recent years, rates of macrosomia have decreased because of increases in early delivery (i.e., by labor induction and/or cesarean delivery), which have reduced post-term birth rates and pregnancy duration even among term pregnancies. Mean birth weight among term births has also begun to decline after increasing steadily for decades (Zhang X. et al. 2010). Figure 9.1 shows temporal trends in singleton SGA and LGA live births in Canada, from 1995 to 2004 (Public Health Agency of Canada 2008).

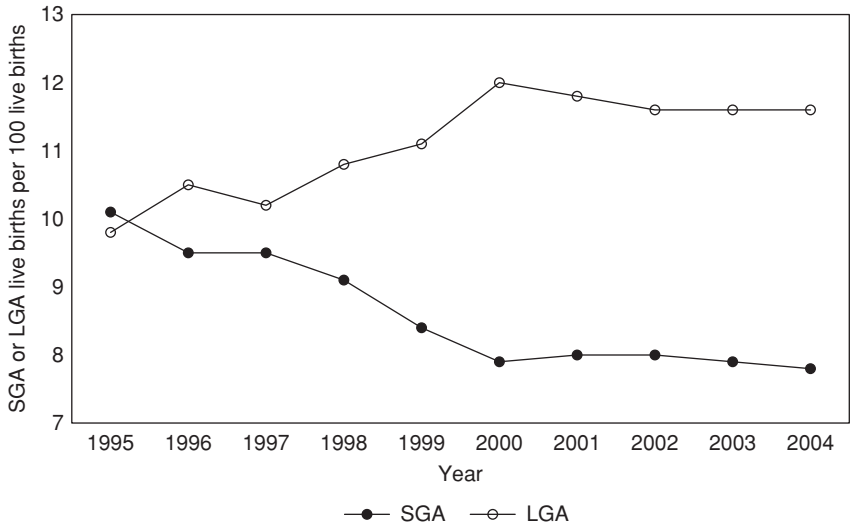


Figure 9.1 Temporal trends in rates of small-for-gestational age (SGA) and large-for-gestational age (LGA) live births in Canada (excluding Ontario), 1995 to 2004 (Public Health Agency of Canada, 2008).

REFERENCES

Alexander, G.R., Himes, J.H., Kaufman, R.B., Mor, R., and Kogan, M. 1996. A United States national reference for fetal growth. *Obstet Gynecol* 87: 163–68.

Ananth, C.V., Vintzileos, A.M., Shen-Schwarz, S., Smulian, J.C., and Lai, Y.L. 1998. Standards of birth weight in twin gestations stratified by placental chorionicity. *Obstet Gynecol* 91: 917–24.

Anderson, N.G., Jolly, I.J., and Wells, J.E. 2007. Sonographic estimation of fetal weight: Comparison of bias, precision and consistency using 12 different formulae. *Ultrasound Obstet Gynecol* 30: 173–79.

Beltrand, J., Alison, M., Nicolescu, R., Verkauskiene, R., Deghmoun, S., Sibony, O., et al. 2008. Bone mineral content at birth is determined both by birth weight and fetal growth pattern. *Pediatr Res* 64: 86–90.

Birkbeck, J.A., Billewicz, W.Z., and Thomson, A.M. 1975. Foetal growth from 50 to 150 days of gestation. *Ann Hum Biol* 2: 319–26.

Buckler, J.M.H. and Green, M. 1994. Birth-weight and head circumference standards for English twins. *Arch Dis Child* 71: 516–21.

Bukowski, R., Gahn, D., Denning, J., and Saade, G. 2001. Impairment of growth in fetuses destined to deliver preterm. *Am J Obstet Gynecol* 185: 463–67.

Bukowski, R., Smith, G.C., Malone, F.D., Ball, R.H., Nyberg, D.A., Comstock, C.H., et al. 2007. Fetal growth in early pregnancy and risk of delivering low birth weight infant: Prospective cohort study. *Br Med J* 334(7598): 836.

Catalano, P.M., Thomas, A.J., Avallone, D.A., and Amini, S.B. 1995. Anthropometric estimation of neonatal body composition. *Am J Obstet Gynecol* 173: 1176–81.

Chitty, L.S. and Altman, D.G. 2002. Charts of fetal size: Limb bones. *Br J Obstet Gynecol* 109: 919–29.

Cho, J.Y., Min, J.Y., Lee, Y.H., McCrindle, B., Hornberger, L.K., and Yoo, S.J. 2007. Diameter of the normal fetal thymus on ultrasound. *Ultrasound Obstet Gynecol* 29: 634–38.

Clausson, B., Gardosi, J., Francis, A., and Cnattingius, S. 2001. Perinatal outcome in SGA births defined by customised versus population-based birthweight standards. *BJOG* 108: 830–34.

Cunningham, F.G., Leveno, K.J., Bloom, S.L., Hauth, J.C., Gilstrap, L., III, Wenstrom, K.D., eds. 2005. *Williams obstetrics*, 22nd ed. Toronto, Ontario: McGraw-Hill.

Das, U.G. and Sysyn, G.D. 2004. Abnormal fetal growth: Intrauterine growth retardation, small for gestational age, large for gestational age. *Pediatr Clin North Am* 51: 639–54.

Davis, R.O., Cutter, G.R., Goldenberg, R.L., Hoffman, H.J., Cliver, S.P., and Brumfield, C.G. 1993. Fetal biparietal diameter, head circumference, abdominal circumference and femur length. A comparison by race and sex. *J Reprod Med* 38: 201–06.

Daya, S. 1993. Accuracy of gestational age estimation by means of fetal crown-rump length measurement. *Am J Obstet Gynecol* 168(3 Pt 1): 903–08.

Dietz, P.M., England, L.J., Callaghan, W.M., Pearl, M., Wier, M.L., and Kharrazi, M. 2007. A comparison of LMP-based and ultrasound-based estimates of gestational age using linked California livebirth and prenatal screening records. *Paediatr Perinat Epidemiol* 21(Suppl 2): 62–71.

Dilmen, G., Turhan, N.O., Toppare, M.F., Seçkin, N., Oztürk, M., and Göksin, E. 1995. Scapula length measurement for assessment of fetal growth and development. *Ultrasound Med Biol* 21: 139–42.

Dodd, J.M., Robinson, J.S., Crowther, C.A., and Chan, A. 2003. Stillbirth and neonatal outcomes in South Australia, 1991–2000. *Am J Obstet Gynecol* 189: 1731–36.

Drey, E.A., Kang, M.S., McFarland, W., and Darney, P.D. 2005. Improving the accuracy of fetal foot length to confirm gestational duration. *Obstet Gynecol* 105: 773–78.

Duck, F.A. 2008. Hazards, risks and safety of diagnostic ultrasound. *Med Eng Phys* 30: 1338–48.

Ego, A., Subtil, D., Grange, G., Thiebaugeorges, O., Senat, M.V., Vayssiere, C., and Zeitlin, J. 2006. Customized versus population-based birth weight standards for identifying growth restricted infants: A French multicenter study. *Am J Obstet Gynecol* 194: 1042–49.

El Kateb, A. and Ville, Y. 2008. Update on twin-to-twin transfusion syndrome. *Best Pract Res Clin Obstet Gynaecol* 22: 63–75.

Ellis, K.J. 2007. Evaluation of body composition in neonates and infants. *Semin Fetal Neonatal Med* 12: 87–91.

Ellis, K.J., Yao, M., Shypailo, R.J., Urlando, A., Wong, W.W., and Heird, W.C. 2007. Body-composition assessment in infancy: Air-displacement plethysmography compared with a reference 4-compartment model. *Am J Clin Nutr* 85: 90–95.

Embleton, N.D., Deshpande, S.A., Scott, D., Wright, C., and Milligan, D.W.A. 2001. Foot length, an accurate predictor of nasotracheal tube length in neonates. *Arch Dis Child Fetal Neonatal Ed* 85: F60–F64.

Gardosi, J., Chang, A., Kalyan, B., Sahota, D., and Symonds, E. 1992. Customised antenatal growth charts. *Lancet* 339: 283–87.

Gardosi, J., Mongelli, M., Wilcox, M., and Chang, A. 1995. An adjustable fetal weight standard. *Ultrasound Obstet Gynecol* 6: 168–74.

Gartner, A., Sarda, P., Dupuy, R.P., Maire, B., Delpeuch, F., and Rieu, D. 1994. Bioelectrical impedance analysis in small- and appropriate-for-gestational-age newborn infants. *Eur J Clin Nutr* 48: 425–32.

Gielen, M., Lindsey, P.J., Derom, C., Loos, R.J.F., Souren, N.Y., Paulussen, A.D.C., et al. 2008. Twin-specific intrauterine "growth" charts based on cross-sectional birthweight data. *Twin Res Hum Genet* 11: 224–35.

Glinianaia, S.V., Skjærven, R., and Magnus, P. 2000. Birthweight percentiles by gestational age in multiple births: A population-based study of Norwegian twins and triplets. *Acta Obstet Gynecol Scand* 79: 450–58.

Goldenberg, R.L., Nelson, K.G., Koski, J.F., and Cutter, G.R. 1985. Low birth weight, intrauterine growth retardation, and preterm delivery. *Am J Obstet Gynecol* 152: 980–84.

Goldstein, I., Reece, E.A., Pilu, G., Bovicelli, L., and Hobbins, J.C. 1987. Cerebellar measurements with ultrasonography in the evaluation of fetal growth and development. *Am J Obstet Gynecol* 156: 1065–69.

Goldstein, I., Tamir, A., Zimmer, E.Z., and Itskovitz-Eldor, J. 1998. Growth of the fetal orbit and lens in normal pregnancies. *Ultrasound Obstet Gynecol* 12: 175–79.

Gottlieb, A.G. and Galan, H.L. 2008. Nontraditional sonographic pearls in estimating gestational age. *Semin Perinatol* 32: 154–60.

Hadlock, F.P., Harrist, R.B., and Martinez-Poyer, J. 1991. In utero analysis of fetal growth: A sonographic weight standard. *Radiology* 181: 129–33.

Hadlock, F.P., Shah, Y.P., Kanon, D.J., and Lindsey, J.V. 1992. Fetal crown-rump length: Reevaluation of relation to menstrual age (5–18 weeks) with high-resolution real-time US. *Radiology* 182: 501–05.

Hediger, M.L., Luke, B., Gonzalez-Quintero, V.H., Martin, D., Nugent, C., Witter, F.R., et al. 2005. Fetal growth rates and the very preterm delivery of twins. *Am J Obstet Gynecol* 193: 1498–507.

Hediger, M.L., Scholl, T.O., Schall, J.I., Miller, L.W., and Fischer, R.L. 1995. Fetal growth and the etiology of preterm delivery. *Obstet Gynecol* 85: 175–82.

Henriksen, T.B., Wilcox, A.J., Hedegaard, M., and Secher, N.J. 1995. Bias in studies of preterm and postterm delivery due to ultrasound assessment of gestational age. *Epidemiology* 6: 533–37.

Hutcheon, J.A. and Platt, R.W. 2008. The missing data problem in birth weight percentiles and thresholds for "small-for-gestational-age." *Am J Epidemiol* 167: 786–92.

Hutcheon, J.A., Zhang, X., Cnattingius, S., Kramer, M.S., and Platt, R.W. 2008. Customised birthweight percentiles: Does adjusting for maternal characteristics matter? *BJOG* 115: 1397–404.

Joseph, K.S., Fahey, J., Platt, R.W., Liston, R.M., Lee, S.K., Sauve, R., et al. 2009. An outcome-based approach for the creation of fetal growth standards: Do singletons and twins need separate standards? *Am J Epidemiol* 169: 616–24.

Joseph, K.S., Liston, R.M., Dodds, L., Dahlgren, L., and Allen, A.C. 2007. Socioeconomic status and perinatal outcomes in a setting with universal access to essential health care services. *CMAJ* 177: 583–90.

Kaul, S.S., Babu, A., and Chopra, S.R.K. 1986. Fetal growth from 12 to 26 weeks of gestation. *Ann Hum Biol* 13: 563–70.

Konje, J.C., Abrams, K.R., Bell, S.C., and Taylor, D.J. 2002. Determination of gestational age after the 24th week of gestation from fetal kidney length measurements. *Ultrasound Obstet Gynecol* 19: 592–97.

Kramer, M.S. 1987. Determinants of low birth weight: Methodological assessment and meta-analysis. *Bull World Health Organ* 65: 663–737.

Kramer, M.S., Morin, I., Yang, H., Platt, R.W., Usher, R., McNamara, H., et al. 2002. Why are babies getting bigger? Temporal trends in fetal growth and its determinants. *J Pediatr* 141: 538–42.

Kramer, M.S., Platt, R.W., Wen, S.W., Joseph, K.S., Allen, A., Abrahamowicz, M., et al. 2001. A new and improved population-based Canadian reference for birth weight for gestational age. *Pediatrics 108*: *e35*. Online at: http://www.pediatrics.org/cgi/content/full/108/2/e35

Kuczmarski, R.J., Ogden, C.L., Grummer-Strawn, L.M., Flegal, K.M., Guo, S.S., Wei, R., et al. 2000. *CDC growth charts: United States. Advance data from vital and health statistics, No. 314. Hyattsville, MD*: *National Center for Health Statistics*. Online at:www.cdc.gov/growthcharts/

Lampl, M., Kusanovic, J.P., Erez, O., Espinoza, J., Gotsch, F., Goncalves, L., et al. 2009. Early rapid growth, early birth: Accelerated fetal growth and spontaneous late preterm birth. *Am J Hum Biol* 21: 141–50.

Landmann, E., Reiss, I., Misselwitz, B., and Gortner, L. 2006. Ponderal index for discrimination between symmetric and asymmetric growth restriction: percentiles for neonates from 30 weeks to 43 weeks of gestation. *J Matern Fetal Neonatal Med* 19: 157–60.

Larciprete, G., Valensise, H., Vasapollo, B., Novelli, G.P., Parretti, E., Altomare, F., et al. 2003. Fetal subcutaneous tissue thickness (SCTT) in healthy and gestational diabetic pregnancies. *Ultrasound Obstet Gynecol* 22: 591–97.

Lee, W., Balasubramaniam, M., Deter, R.L., Hassan, S.S., Gotsch, F., Kusanovic, J.P., et al. 2009. Fetal growth parameters and birth weight: their relationship to neonatal body composition. *Ultrasound Obstet Gynecol* 33: 441–46.

Luke, B., Minogue, J., Witter, F.R., Keith, L.G., and Johnson, T.R. 1993. The ideal twin pregnancy: patterns of weight gain, discordancy, and length of gestation. *Am J Obstet Gynecol* 169: 588–97.

Ma, G., Yao, M., Liu, Y., Lin, A., Zou, H., Urlando, A., et al. 2004. Validation of a new pediatric air-displacement plethysmograph for assessing body composition in infants. *Am J Clin Nutr* 79: 653–60.

Martin, R.H. and Higginbottom, J. 1971. A clinical and radiological assessment of fetal age. *J Obstet Gynaecol Br Commonw* 78: 155–62.

Mastrobattista, J.M., Pschirrer, E.R., Hamrick, M.A., Glaser, A.M., Schumacher, V., Shirkey, B.A., et al. 2004. Humerus length evaluation in different ethnic groups. *J Ultrasound Med* 23: 227–31.

Maulik, D. 2006. Fetal growth restriction: The etiology. *Clin Obstet Gynecol* 49: 228–35.

Meany, F.J. and Farrer, L.A. 1986. Clinical anthropometry and medical genetics: A compilations of body measurements in genetic and clinical disorders. *Am J Med Genet* 25: 343–59.

McCowan, L., Harding, J., and Stewart, A.W. 2005. Customised birthweight centiles predict SGA pregnancies with perinatal morbidity. *BJOG* 112: 1026–33.

Milani, S., Bossi, A., Bertino, E., Di Battista, E., Coscia, A., Aicardi, G., et al. 2005. Differences in size at birth are determined by differences in growth velocity during early prenatal life. *Pediatr Res* 57: 205–10.

Min, S.J., Luke, B., Gillespie, B., Min, L., Newman, R.B., Mauldin, J.G., et al. 2000. Birth weight references for twins. *Am J Obstet Gynecol* 182: 1250–57.

Morken, N.H., Källen, K., and Jacobsson, B. 2006. Fetal growth and onset of delivery: A nationwide population-based study of preterm infants. *Am J Obstet Gynecol* 195: 154–61.

Murphy, K.E., Hannah, M.E., Willan, A.R., Hewson, S.A., Ohlsson, A., Kelly, E.N., et al. 2009. Multiple courses of antenatal corticosteroids for preterm birth (MACS): A randomised controlled trial. *Lancet* 372: 2143–51.

Murphy, V.E., Smith, R., Giles, W.B., and Clifton, V.L. 2006. Endocrine regulation of human fetal growth: The role of the mother, placenta and fetus. *Endocr Rev* 27: 141–69.

Naeye, R.L., Benirschke, K., Hagstrom, J.W.C., and Marcus, C.C. 1966. Intrauterine growth of twins as estimated from liveborn birth-weight data. *Pediatrics* 37: 409–16.

Ott, W.J. 1993. Intrauterine growth retardation and preterm delivery. *Am J Obstet Gynecol* 168(6 Pt 1): 1710–15; discussion 1715–17.

Owen, P., Donnet, M.L., Ogston, S.A., Christie, A.D., Howie, P.W., and Patel, N.B. 1996. Standards for ultrasound fetal growth velocity. *BJOG* 103: 60–69.

Parker, A.J., Davies, P., Mayho, A.M., and Newton, J.R. 1984. The ultrasound estimation of sex-related variations of intrauterine growth. *Am J Obstet Gynecol* 149: 665–69.

Parretti, E., Carignani, L., Cioni, R., Bartoli, E., Borri, P., La Torre, P., et al. 2003. Sonographic evaluation of fetal growth and body composition in women with different degrees of normal glucose metabolism. *Diabetes Care* 26: 2741–48.

Platz, E. and Newman, R. 2008. Diagnosis of IUGR: Traditional biometry. *Semin Perinatol* 32: 140–47.

Public Health Agency of Canada. 2008. *Canadian perinatal health report*, 2008 edition. Ottawa: Minister of Public Works and Government Services Canada.

Sahota, D.S., Leung, T.Y., Leung, T.N., Chan, O.K., and Lau, T.K. 2009. Fetal crown–rump length and estimation of gestational age in an ethnic Chinese population. *Ultrasound Obstet Gynecol* 33: 157–60.

Salafia, C.M., Chales, A.K., and Maas, E.M. 2006. Placenta and fetal growth restriction. *Clin Obstet Gynecol* 49: 236–56.

Schluter, P.J. and Pritchard, G. 2005. Mid trimester sonographic findings for the prediction of Down syndrome in a sonographically screened population. *Am J Obstet Gynecol* 192: 10–16.

Sherer, D.M., Sokolovski, M., Dalloul, M., Khoury-Collado, F., Osho, J.A., Lamarque, M.D., and Abulafia, O. 2006. Fetal clavicle length throughout gestation: A nomogram. *Ultrasound Obstet Gynecol* 27: 306–10.

Shinwell, E.S. and Shlomo, M. 2003. Measured length of normal term infants changes over the first two days of life. *J Pediatr Endocrinol Metab* 16: 537–40.

Shipp, T.D., Bromley, B., Mascola, M., and Benacerraf, B. 2001. Variation in fetal femur length with respect to maternal race. *J Ultrasound Med* 20: 141–44.

Smith, A.P.M., Ong, S., Smith, N.C.S., and Campbell, D. 2001. A prospective longitudinal study of growth velocity in twin pregnancy. *Ultrasound Obstet Gynecol* 18: 485–87.

Smith, G.C.S. 2004. First trimester origins of fetal growth impairment. *Semin Perinatol* 28: 41–50.

Smith, G.C.S., Smith, M.F.S., McNay, M.B., and Fleming, J.E.E. 1998. First trimester growth and the risk of low birth weight. *N Engl J Med* 339: 1817–22.

Smith-Bindman, R., Chu, P.W., Ecker, J.L., Feldstein, V.A., Filly, R.A., and Bacchetti, P. 2002. US evaluation of fetal growth: prediction of neonatal outcomes. *Radiology* 223: 153–61.

Stetzer, B.P., Thomas, A., Amini, S.B., and Catalano, P.M. 2002. Neonatal anthropometric measurements to predict birth weight by ultrasound. *J Perinatol* 22: 397–402.

Taeusch, H.W., Ballard, R.A., and Gleason, C.A., eds. 2004. *Avery's diseases of the newborn*, 8th ed. Philadelphia, PA: WB Saunders.

UNICEF. 2007. *The state of the world's children 2008: Child survival*. New York: The United Nations Children's Fund (UNICEF).

WHO Multicentre Growth Reference Study Group. 2006. WHO Child Growth Standards based on length/height, weight and age. *Acta Paediatr Suppl* 450: 76–85.

Zhang, J. and Bowes, W.A., Jr. 1995. Birth-weight-for-gestational-age patterns by race, sex, and parity in the United State population. *Obstet Gynecol* 86: 200–08.

Zhang, J., Merialdi, M., Platt, L.D., and Kramer, M.S. 2010. Defining normal and abnormal fetal growth: promises and challenges. *Am J Obstet Gynecol* 202: 522–28.

Zhang, X., Joseph, K.S. and Kramer, M.S. 2010. Decreased term and postterm birthweight in the United States: impact of labor induction. *Am J Obstet Gynecol* 203: 124.e1-7.

Zhang, X., Platt, R.W., Cnattingius, S., Joseph, K.S., and Kramer, M.S. 2007. The use of customised versus population-based birthweight standards in predicting perinatal mortality. *BJOG* 114: 474–77.

10

Birth Defects

MARTHA M. WERLER

Occurrences of malformed organs, tissues, and structures have captivated humans since ancient times, spurring hypotheses about how they came about and shaping the modern fields of developmental biology, teratology, clinical and molecular genetics, and epidemiology. Indeed, as in many areas of science, it is through the search for causes of the abnormal that we begin to understand normal development, processes, and functions. In the field of reproductive and perinatal epidemiology, studies of birth defects have occupied the attention of investigators since Gregg first described congenital rubella syndrome in the 1940s (Dunn 2007); they gained more prominence in the 1960s when use of thalidomide, a sedative, was found to cause missing and malformed upper and lower limbs (Lenz 1962). As a group, major birth defects are not uncommon, occurring in approximately 3% of births. Understanding normal and abnormal development of fetal structures notwithstanding, compelling reasons to identify the patterns of occurrence of and risk factors for birth defects include their impact on mortality and morbidity. Birth defects are the most common cause of infant mortality, accounting for approximately one in five deaths (MMWR 1998). In terms of morbidity, they are associated with substantial costs for both medical care and special services. In 2009 dollars (adjusted from 1988 figures), estimated costs per child ranged from $224,000 for small intestinal atresia to $1,858,000 for Down syndrome (Waitzman 1994).

In addition to rubella infection and thalidomide exposure in pregnancy, other causes of birth defects have been identified. For example, cytomegalovirus is the most common congenital infection in the United States, and it can cause birth defects, but infections in pregnancy are rare (Kenneson and Cannon 2007), accounting for a small fraction of the overall occurrences of birth defects. Several medications are considered teratogenic, including thalidomide, warfarin, diethylstilbestrol, valproate, phenytoin, methotrexate, aminopterin, misoprostol, and tetracycline (Koren 1998). These medications are used to treat conditions that are rare in women of childbearing ages and, therefore, also account for only a small fraction of birth defects. Genetic studies have identified thousands of Mendelian-inherited causes, but most genetically associated birth defects appear to be new mutations, the causes of which remain unknown. The impact of more common causes of birth defects, such as folate deficiency and heavy alcohol intake, have

been reduced due to fortification of the cereal grain supply and public health campaigns, respectively, but even before these successes, the proportion of birth defects overall that were attributable to them was relatively small. In sum, the causes of the majority of birth defects remain unknown. This gap in knowledge will hopefully close as the complexities of developmental genetics and gene–environment interactions become better understood.

Separate from the above-named teratogens are *risk factors* for birth defects—those characteristics or exposures that have been associated with increased birth defect risks, but whose roles are unclear or the supporting evidence is not convincing enough to be considered causal. For example, some specific birth defects have been observed to occur more frequently in certain ethnic or race subgroups, such as the high prevalence of neural tube defects in Mexican and Irish populations (International Clearinghouse for Birth Defects 1991). The basis, however, for differences in the distributions of birth defects according to race or ethnicity is not entirely clear and could be due to cultural practices, genetics, or a combination of the two factors. Maternal age is another factor that is positively associated with some birth defects (e.g., Down syndrome, hypospadias) and inversely associated with other defects (e.g., gastroschisis) (Reefhuis 2004). In the case of Down syndrome, the association is presumably related to changes in hormones or the age of ova, but it is not clear whether biologic or behavioral reasons might explain other age-related risks of specific birth defects. High body mass index has been observed to increase the risk of several birth defects (Waller 1994, 2007), whereas low body mass index has been associated with gastroschisis risk (Lam 1999; Waller 2007). Dietary glycemic index has been positively associated with neural tube defects, suggesting hyperglycemia may play a role in neural tube closure (Yazdy 2010; Shaw 2003). Several studies have identified increased risks for birth defects overall in association with treatment for subfertility, the so-called assisted reproductive technologies (Hansen 2002; Reefhuis 2009). Because there are many different treatments and procedures for infertility, and these are constantly evolving, it is not known, at present, whether increased risks might result from any specific treatments, underlying conditions, or a combination of both. Cigarette smoking in pregnancy is contraindicated due to associations with increased risks of preterm delivery and low birth weight, but it is also linked to specific birth defects. Risks of oral clefts and gastroschisis, for example, have been reported in association with maternal smoking in pregnancy in many studies (Honein 2007; Werler 2009). Alcohol use is also contraindicated in pregnancy, but evidence to suggest that low to moderate alcohol consumption is associated birth defect risks is less clear, with many inconsistencies in findings and methodological weaknesses (Henderson 2007). Use of selective serotonin reuptake inhibitors in pregnancy for the treatment of depression or anxiety has been suspected as a risk factor with inconclusive evidence, but four studies have identified increased risks of congenital heart defects in association with use of one such antidepressant—paroxetine (Alwan 2007; Berard 2007; Louik 2007; Diav-Citrin 2008). Of course, this list of risk factors will change as subsequent studies are conducted, with some factors being added to or dropped from the list and perhaps others moving from "risk factor" to "known cause" status.

Other than through Mendelian-inheritance, paternal risk factors for birth defects have not been consistently established, although the results of many studies suggest that paternal occupational and environmental exposures might be important (Cordier 2008). Mechanisms by which paternal exposures might affect fetal development, beyond genetic mutations in sperm, include oocyte or fetal exposure through seminal fluid and epigenetic effects (Davis 1992; Barton 2005).

Although the inheritance of single-gene disorders makes up less than 10% of all birth defects (Kalter and Warkany 1983), genetic factors inevitably play a key, yet complicated, role in their development. The vast majority of birth defects are considered to be *complex*, meaning that multiple genes and interactions between gene and environmental factors are causally involved. Basic science has identified the pathways of genes involved in fetal development of limbs, craniofacial structures, the heart, and other structures that help inform the fields of human embryology and teratology. Genes involved in metabolism of xenobiotic exposures, such glutathione synthetase and cytochrome P450 enzymes, are just beginning to be studied in combination with relevant exposure histories as they relate to risks of birth defects (Lammer 2005; Torfs 2006). The role of genes in the folate pathway, such as methylene tetrahydrofolate reductase, has also been studied, but with inconclusive findings for neural tube defects, oral clefts, and cardiac defects (Amorim 2007; van Beynum 2007; Johnson 2008). Epigenetic regulation of gene activity undoubtedly plays a key role in fetal development. As these processes that turn on and off gene activity are better understood, genotyping techniques improve, molecular epidemiology methods gain sensitivity, and case-control series with biologic and exposure data accumulate, the largely unsolved mystery of what causes birth defects will begin to unfold.

DEFINITIONS

In epidemiologic studies, the term *birth defect*, which is used interchangeably with congenital anomaly or abnormality, should also be defined. In the broadest sense, *birth defect* can be considered synonymous with *developmental defect* and can be applied to nearly all adverse reproductive outcomes. Figure 10.1 shows that a defect in development can exhibit effects in the germ cell, embryo, or postnatally, meaning that outcomes such as infertility, spontaneous abortion, or autism might be considered birth defects.

Developmental defects of the fetus can include functional or metabolic problems, such as jaundice in the newborn, phenylketonuria, or abnormal growth. Typically, however, *birth defect* refers to malformations. Nevertheless, there are many different types of malformations, and more detailed definitions are in order. Some malformations result from constraint, and these are termed *deformations* or *positional anomalies* (e.g., plagiocephaly or flattened skull); they are in contrast to structural malformations that result from primary maldevelopment or disruption in utero of an otherwise normally forming structure. Of the structural malformations, the most commonly occurring are considered *minor*, such as webbing between the toes or low-set ears. *Major* structural malformations include those

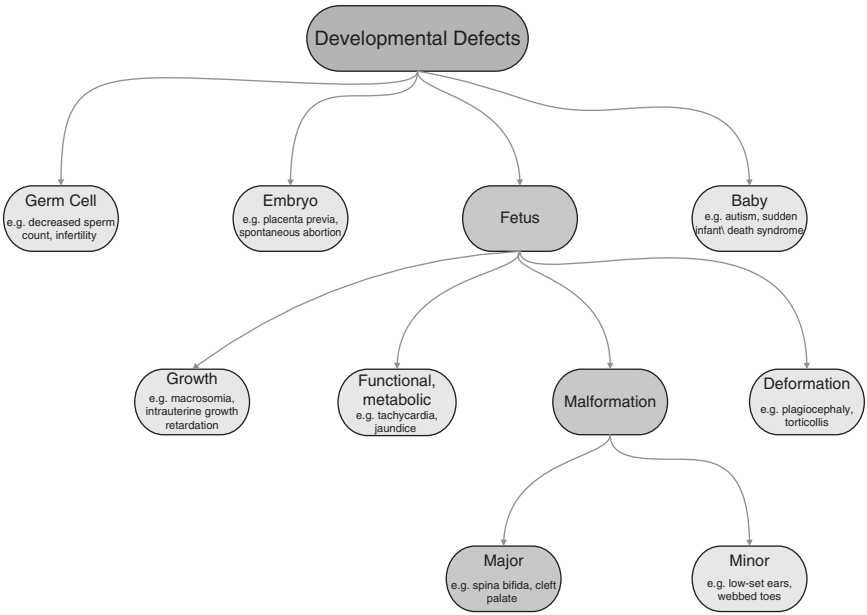


Figure 10.1 Different types of outcomes under the broad title of *Developmental Defects*.

that are medically, surgically, or cosmetically important, such as spina bifida, cleft palate, or port wine stain (Rasmussen and Moore 2001). A minor malformation may not itself be medically important, but clinical geneticists and teratologists pay close attention to patterns of minor malformations because they can be predictive of a major malformation or a syndrome (Holmes 1987; Leppig 1987). For example, the classic phenotype of Down syndrome includes several minor malformations, including a simian crease across the palm of the hand, webbed neck, and clinodactyly of the fifth finger.

Reports on the prevalence of birth defects can vary widely due to differences in definitions and inclusion criteria. Even when uniform definitions are applied, diagnostic issues can influence ascertainment, resulting in differences in prevalence rates. For example, some malformations are not symptomatic and come to diagnosis by prenatal sonogram or incidental finding. Renal agenesis is clearly a malformation, but unilateral cases may not be symptomatic, especially if the existing kidney is normal. Therefore, only bilateral renal agenesis meets the criteria of medically important. Obstructive renal anomalies, such as hydronephrosis, are vulnerable to the same criticism. With the common practice of prenatal sonogram screenings, many malformations that would otherwise go undiagnosed are reported to surveillance systems, resulting in reported increases in prevalence that may not be true. Some structural development occurs late in gestation or during parturition, such as descent of the testis and closure of the ductus arteriosus. Therefore, many surveillance systems and epidemiologic studies exclude cases

with undescended testis, patent ductus arteriosus, or patent foramen ovale among preterm births. Some surveillance systems include birth defects diagnosed up through early childhood, such as the Metropolitan Atlanta Congenital Defect Program, whereas others limit diagnoses to within the newborn period, such as the state of Maryland (NBDPN 2008). This criterion difference for inclusion is not likely to impact rates of visibly apparent birth defects, such as oral clefts, but has a significant impact of rates of internal defects, such as cardiac anomalies. Another complication is that the range of International Classification of Disease - 9 (ICD) codes for structural malformations (740–759) includes codes for many minor malformations and even some anomalies that are not considered structural. For all of these reasons, inclusion of major structural anomalies that are medically, surgically, or cosmetically important might require review of each case's diagnoses. Subsequent sections of this chapter pertain to *major structural birth defects*.

The definition of a *genetic* birth defect is becoming more confusing, as the ability to identify genetic aberrations has become more detailed. Decades ago, diagnostic testing was limited to the identification of extra or missing chromosomes or pieces of chromosomes; now, birth defects can be linked to specific single nucleotide polymorphisms within a gene. Epidemiologic studies looking for risk factors for birth defects typically exclude cases that are attributed to a single gene (i.e., Mendelian-inherited); however, mutations in single genes that cause birth defects provide clues for studies aimed at identifying genetic risk factors of so-called *complex* birth defects, or those that result from an interplay between genetic factors and environmental exposures. For example, mutations in the fibroblast growth factor receptor 2 gene were identified in children with Crouzon syndrome, a condition that involves craniosynostosis. Since then, somatic mutations of this same gene have been identified in some (but a minority of) nonsyndromic cases with craniosynostosis (Passos-Bueno 2008).

Known teratogens, such as thalidomide, isotretinoin, and valproate, increase the risk of specific defects or groups of defects, but not malformations overall. Based on this general pattern of teratogenicity, the effects of exogenous exposures on birth defect risks are expected to vary across the range of specific defects. In other words, a teratogenic exposure might only affect a small fraction of all birth defects, and that effect might be obscured if all birth defects are grouped together as one outcome. Thus, the most rigorous investigations of risk factors for birth defects focus on specific anomalies. This tenet is further supported by the fact that the descriptive epidemiology also varies across the range of specific defects, even within organ systems. For example, two of the more common gastrointestinal defects, tracheoesophageal fistula and small intestinal atresia, differ according to infant sex and race (Cragan 1993; Torfs 1995). When small intestinal atresia cases are further categorized, duodenal and jejunal subtypes also differ by infant sex and race (Cragan 1993). Therefore, for most birth defects, studies of more specific anomalies have a greater likelihood of identifying an effect, although this can be difficult to achieve without a very large population base because specific defects are rare. Table 10.1 shows prevalence estimates of selected specific defects.

Table 10.1 Estimated Prevalence of Specific Birth Defects (per 10,000 live births)

Birth Defect	*Prevalence*	
	*United States**	*Europe***
Anencephalus	0.98	0.40
Spina bifida	3.10	1.99
Encephalocele	1.05	0.33
Hydrocephalus	4.64	2.49
Transposition of the great arteries	2.46	2.79
Tetralogy of Fallot	3.87	2.43
Coarctation of the aorta	3.87	2.96
Atrial septal defect	13.55	19.48
Ventricular septal defect	45.28	26.35
Hypoplastic left heart syndrome	2.42	1.34
Aortic stenosis	1.11	1.00
Cleft lip with or without cleft palate	6.97	7.65
Cleft palate alone	4.06	4.90
Esophageal atresia/stenosis	1.90	1.88
Anal or rectal atresia/stenosis	2.13	2.22
Hirschsprung disease	2.29	
Hypospadias	26.51	13.29
Bilateral renal agenesis/dysgenesis	1.11	0.32
Any cystic kidney disease	5.43	4.08
Exstrophy of the bladder	0.15	
Limb deficiency	4.30	3.84
Clubfoot	9.87	7.52
Craniosynostosis	3.66	1.12
Omphalocele	1.50	1.16
Gastroschisis	2.48	1.84
Diaphragmatic hernia	3.14	1.96
Down syndrome	13.99	9.49

*From CDC. 2007. Metropolitan Atlanta Congenital Defects Surveillance Report. Birth Defects Research Part A 79(2):102,104–105.

**From Final Activity Report to European Commission March 2004 to August 2007. Online at: www.eurocat.ulster.ac.uk/pdf/EUROCAT-Final-EC-Report.pdf

ORGANOGENESIS

Most organs and structures are formed between the fourth and tenth week of gestation, counting from the time of conception. Thus, the first trimester is considered the period of organogenesis and the time frame of greatest focus for epidemiologic studies of environmental risk factors. Figure 10.2 illustrates the periods in gestation when different organ systems are developing and thus when those systems are most susceptible to environmental insults. For some organs or structures, such as the neural tube, the period of development is very short, meaning investigations into causal factors can be limited to a narrow window of time during pregnancy. In contrast, some parts of the cardiovascular system, such as the atrial septum, continue to form and mature throughout gestation and could be vulnerable to insults after the first trimester.

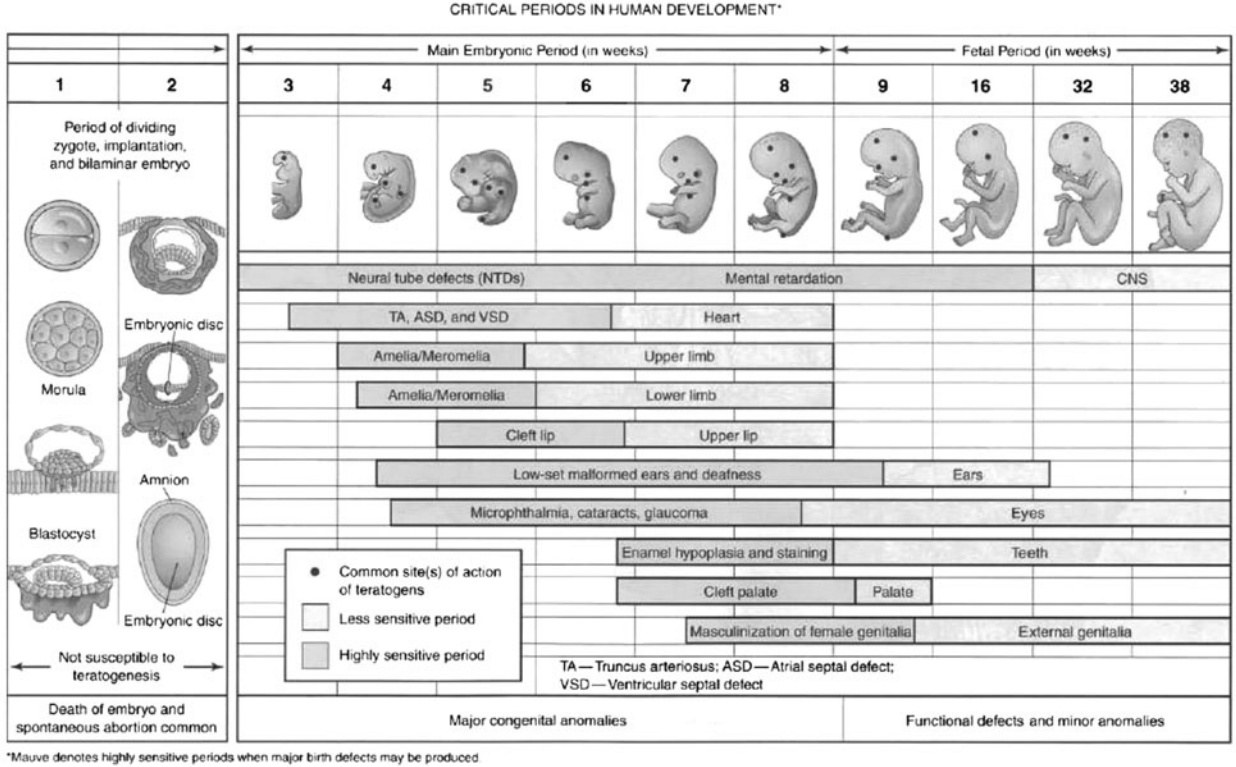


Figure 10.2 Developmental timing of selected defects during gestation. Reproduced from Moore, K.L. and Persaud, T.V.N., *The Developing Human, Clinically Oriented Embryology*, 7th edition, 2003, with permission of the publisher, Saunders/Elsevier.

METHODOLOGIC CONSIDERATIONS

Epidemiologic studies of birth defects are limited to prevalent cases because it is not currently possible to determine the birth-defect status of all individuals at risk—that is, all conceptions that are successfully implanted and survive to the embryonic stage of development. Since survivability might be related to the presence of a birth defect, it is inappropriate to consider birth defect proportions as incident proportions (rates). In epidemiologic circles, studying prevalent, rather than incident, cases can carry negative connotations, but in the context of reproductive and perinatal epidemiology and public health, prevalent cases of birth defects are likely to be the outcome of greater interest: Separate from wanting to know their risk of having a pregnancy loss, women generally want to know their risk of having a child with a birth defect, which is indeed what we measure in prevalence studies of birth defects. When ascertainment includes birth defects occurring among late fetal losses and therapeutic abortions, cases could be considered incident after a specified gestational age (e.g., 20 weeks).

An unresolved issue in both follow-up and case-control studies of birth defects concerns the study "unit": Are study subjects the mothers or their offspring? Cases are typically thought of as the offspring, but exposure measurement inevitably involves the mother for all but purely genetic risk factors. Mother–baby pairs have been identified as study subjects by some investigators, but multiple gestations further complicate matters. How are birth defects in twins or higher-order pregnancies counted when estimating risk? When pregnancies make up the denominator in a study, should twins who are concordant for a malformation each contribute to the numerator? Or, should discordant twins contribute 0.5 to the numerator? One approach is to consider the pregnancy as the unit of study, and affected pregnancies are those with any birth defect in offspring, regardless of the number of gestations. Malformations occur more often in multiple gestations than in singletons (Glinianaia 2008). This is true even in fraternal twins compared with siblings from separate pregnancies, who share the same amount of genetic information. On this basis, multiple gestation may be a confounder and is likely an effect modifier, prompting some investigators to exclude multiple gestations altogether. However, statistical models that can account for dependent outcomes, such as twins or higher-order gestations, allow inclusion of these potentially informative births. General estimating equations are particularly useful for this purpose (Ananth 2005).

Study designs

Because experimental study designs are limited within populations of pregnant women, studies of the effects of maternal exposure to toxicants rely, for the most part, on observational approaches. Prospective studies have contributed some findings to the field, but even the largest cohort studies suffer from small numbers of specific defects. For example, the Collaborative Perinatal Project (CPP), which collected information from over 50,000 pregnancies during gestation, resulted in fewer than 100 cases of even the more common major malformations, such as

neural tube defects, oral clefts, and atrial septal defects (Heinonen 1977). If one wished to study the association between a common maternal exposure, such as cigarette smoking with an approximate prevalence of 15%, and risk of cleft lip and palate, a cohort the size of the CPP would include approximately 7,500 exposed pregnancies. Given that cleft lip with or without cleft palate is prevalent in approximately 0.7 per 1,000 births, five exposed cases would be expected under an assumption of no association. If there was an observed doubling in risk, ten exposed cases would be expected, but the 95% confidence interval around that twofold relative risk would be 0.98 to 4.17. Statistical power would, of course, be even lower for the many specific anomalies that occur less frequently than cleft lip with or without cleft palate (see Table 10.1). The National Children's Study expects to enroll twice as many women as the CPP cohort, which will allow robust assessments of common exposures and the most common birth defects, but will lack statistical power to examine moderate effects of exposures that are less prevalent than 5% in relation to the majority of specific birth defects—those that occur less frequently than 1 in 1,000 births.

Prospective studies of birth defects have been conducted on women who contact telephone information services for advice on exposures in pregnancy (Felix 2004). The study design involves enrollment of women, regardless of the issues that prompted their calls; a standardized interview about exposures in early pregnancy; and follow-up of their births. Women who seek advice through these services may not be representative of all women, but comparisons of birth-defect risks between those exposed to a particular medication and those who were not exposed to any suspicious agents contribute important information, as long as the cohort excludes women who contacted the service because their pregnancies had already been identified as being affected with a birth defect. Unfortunately, these cohort studies suffer from low statistical power for most specific birth defects, as do cohort studies overall.

Follow-up of cohorts of exposed pregnant women, without their unexposed counterparts, is an approach that deserves special consideration. For example, environmental disasters (e.g., Chernobyl), occupational cohorts (e.g., pesticide workers), and medication registries (e.g., anticonvulsants) may offer important opportunities to assess risks of birth defects. In these situations, a group of unexposed women may not be available, and a historical comparison group may be utilized. For example, published rates from the Metropolitan Atlanta Congenital Defect Program (MADCP) readily serve as a historical comparison group (Wyszynski 2009), providing the same definition of outcomes is applied. However, the strong possibility of confounding bias challenges the validity of these comparisons. When tabulating outcomes of women who were known to be exposed to a suspected toxin, there is a temptation to allow a broad definition of "birth defect," which would bias effects in the positive direction if the historical comparison group used a stricter definition.

Case-control studies

Given the rarity of specific congenital anomalies, the design of choice is most often the case-control study, along with its inherent limitations and challenges.

Sources of cases include birth defect registries, vital records, hospital discharge diagnoses, insurance claims data, and prenatal diagnosis records. Completeness of any of these sources depends on the specific malformation or group of malformations comprising the case group. For example, neural tube defects, such as anencephaly and spina bifida, are easily identified at birth and, therefore, documented in medical records, thus suggesting that birth certificates and hospital discharge diagnoses would be good sources of neural tube defect cases. However, the vast majority of cases are prenatally diagnosed and a large fraction of pregnancies affected with these defects are terminated during the second trimester (Peller 2004). To capture all neural tube defects that survive past the first trimester, it is essential that ascertainment include terminated and early fetal loss cases, in addition to those occurring among live births and late pregnancy losses. Other congenital defects may not come to diagnosis until later in infancy or early childhood. Hemifacial microsomia, characterized by asymmetric underdevelopment of craniofacial structures, can be diagnosed at any time from the third trimester by ultrasound to early childhood at a first dental visit; some cases may go undiagnosed altogether. But the vast majority of cases come to diagnosis between 3 months and 3 years of age, indicating that birth certificates or hospital discharge diagnoses would not be a good source for this malformation. A preferable ascertainment source would then be where cases come to diagnosis, which, for hemifacial microsomia, would be craniofacial specialists.

Birth-defect registries exist in nearly every state in the United States and can be excellent sources of cases. The methods employed to ascertain cases vary from registry to registry, with some relying on passive reporting via birth certificates or physicians, some employing rigorous active surveillance, and others using a combination of various methods (NBDPN 2008). The latter type of registry offers more timely and complete ascertainment of cases by conducting reviews of birth records and adhering to detailed inclusion criteria. When considering a registry as a source of birth defect cases, it would be important to determine whether prenatally diagnosed cases are included and at what age newly diagnosed cases are included.

It is important to identify the study base for any case-control study, which typically depends on the source of cases. An advantage of identifying cases from a registry is its geographic base. Because controls should represent the population that gave rise to the cases, identifying the study base will facilitate the selection of an appropriate control group. In a study that uses cases from a birth-defect registry, controls could be births from that same geographic region. Cases ascertained from hospitals or clinics or a mix of sources can complicate identification of the study base. When cases are ascertained from birth hospitals, controls might be births at those same institutions. However, prenatal diagnosis or high-risk pregnancies might result in referrals away from some hospitals and into other, typically large, tertiary care hospitals. Birth-defect cases that are referred to such birth hospitals may not arise from the same background population as the remainder of births at that same institution. The study base for cases ascertained at specialty clinics in pediatric hospitals may be defined as the population of pregnant women who would have gone to the same pediatric hospital if their pregnancy had been

similarly affected. How would such a population then be identified for control ascertainment? Persons who were admitted or had clinic visits for reasons other than the case diagnosis likely were referred through a process similar to that of the cases; in other words, they would likely have been also been referred there if they'd had the same diagnosis as that of the case group. However, this control series would have other conditions that brought them to the hospital or clinic, which might be related to the exposure or exposures to be studied. One would need to be convinced that no such associations exist before selecting other conditions as the control group. However, evidence may not be available, one way or the other, to inform the decision about controls. In the absence of such evidence, a control group comprising a wide variety of different diagnoses reduces concerns of bias under the assumption that no single exposure would be associated with each of the many conditions. Another approach to identifying control subjects for hospital or clinic-based case series is to use the referring health care provider as the source. Cases seen at a pediatric hospital were likely referred by their primary care provider, who can offer other patients as potential controls. Similarly, for cases born to women who were referred to a tertiary birth hospital, controls could come from the referring provider.

Case-control studies of birth defects have utilized various sources of exposure information. Preferential sources would have detailed, specific, accurate, and complete documentation of exposure on all study subjects, as is desirable in studies of other reproductive outcomes. In studies of birth defects, exposure information should be especially detailed and specific with regard to timing, because the developmentally relevant period is typically a few weeks in early pregnancy, when many changes in behavior and exposure occur. Accurate and complete measurement of exposure would eliminate misclassification, although it is rarely achieved.

If the "exposure" is purely genetic—say, the presence of a particular mutation—then measurement, even years later, can approach a high level of detail and accuracy (as long as rigorous genotyping methods are employed), and specificity of timing is not an issue. Measurement of environmental exposures presents many more challenges, and trade-offs are inevitable. Because of concerns about reporting or diagnostic biases, documentation that precedes the diagnosis of a birth defect is optimal, such as obstetric records, pharmacy records, and environmental databases, but these sources often suffer from lack of detail on timing, dose, and potential confounders. For example, an obstetric record might state that a patient was prescribed an antiemetic to relieve severe nausea and vomiting, but whether she filled it or actually took the medication or how often is not known. Biologic samples collected after birth for measurement of exposures are not subject to reporting or diagnostic biases, but may be poor indicators of exposure during the developmentally relevant time frame. Often, the most accessible source of information on timing, frequency, and dose of exposure is the mother herself, but interperson variability of detailed, specific, accurate, and complete reporting is undoubtedly high.

Although maternal report of many exposures is typically far from perfect, it can be the only source on certain details. Therefore, mothers are the primary source of exposure information in several large-scale case-control studies of birth defects,

including the Boston University Slone Epidemiology Pregnancy Health Interview Study (Louik 2007), the National Birth Defects Prevention Study (Alwan 2007), the Baltimore-Washington Infant Study (Ferencz 1997), and the Spanish Collaborative Study of Congenital Malformations (Martinez-Frias 2007). Exposure measurement that relies solely on maternal recall has always been suspect, although accuracy depends on several factors, such as the type of exposure. For example, recall of parity, age at menarche, cigarette smoking, or history of gallbladder disease are likely to be relatively high (Paganini-Hill 1982; Sanderson 1998; Yawn 1998; Tomeo 1999; Must 2002), whereas recall of occurrences of headache, use of antacids, or number of servings of broccoli during the first trimester are likely to be lower. The social or public health stigma attached to an exposure might also be important. A study of alcohol, cocaine, and marijuana use in pregnancy found that women under-reported exposures at the antenatal interview compared with the postpartum interview. The amount of under-reporting was 44% for alcohol, 57% for marijuana, and 70% for cocaine, suggesting that the amount of denial varies according to its negative perception in society (Jacobson 1991). Social stigma can change over time, depending on what is in the news immediately preceding the time of the data collection. Another potentially important influence on recall, which seems to be underappreciated in epidemiologic circles, but has been given attention in the field of psychology, is personality. Also, a longer period of time between data collection and the event that is being recalled has been shown to negatively impact accuracy of reporting (Lewis 2006), as intuition would suggest. If these factors are balanced between cases and controls, their impact is considered *random misclassification* of exposure, which tends (but not always) to result in biasing risk estimates toward no effect (Jurak 2005). On the other hand, accuracy of reporting that is dependent on case-control status, can introduce *recall bias* in either direction, and deserves further consideration.

Recall bias

Retrospective studies of risk factors for birth defects seem to inevitably invoke accusations of recall bias. Concerns of recall bias were founded by clinicians caring for children with birth defects because mothers so often ask how the birth defect could have come about. Such questions are often framed as, "Could it be the X or Y that I was exposed to in pregnancy?" Intuitively, it makes sense that mothers of birth-defect cases would review the events and exposures of their pregnancies, searching for an explanation, whereas mothers of "normal" controls would be more focused on the present and future. Recall bias in this scenario would result in an overestimation of the relative risk. Hence, reports of increased risks for birth defects in association with exposures that were reported retrospectively are often greeted with suspicion of recall bias.

Empirical evidence in support of this type of recall bias is scant, however. This is due, in part, to the difficulty of measuring recall bias in the setting of birth-defect case-control studies. Assessment of reporting accuracy requires validation of retrospective reports, using a gold standard. Data in vital records, medical records, biologic markers, and prospective studies have been used as gold standards, but

each source carries its own limitations. Exposures documented in vital records, primarily birth certificates, may be less accurate than a mother's report and, like maternal report, may be vulnerable to bias because the outcome of pregnancy is already known. Many exposures are not documented in medical records and even those that tend to be, such as illnesses and treatments, are not recorded with specific details on timing, severity, or dose. Also, the lack of a noted illness or treatment could represent no occurrence or unknown information. Biologic markers of exposure that are collected close to the first trimester can be an excellent gold standard if genetic variation in metabolism isn't a factor. For example, serum folate levels are a function of both folate intake and genotype of many different enzymes. A one-time biologic sample may also represent a narrow, and possibly etiologically irrelevant, time frame of exposure, especially if it is not fat-soluble and varies over time.

Prospective data collection from pregnant women is a far superior gold standard, but cohorts with both prospective and retrospective data collection have not been large enough to allow robust comparisons between mothers of children with birth defects and mothers of healthy children. Hence, validation studies of maternal retrospective reports have compared mothers of healthy children with those having a variety of adverse reproductive outcomes, such as prematurity, intrauterine growth retardation, neonatal intensive care admittance, sudden infant death syndrome, miscarriages, stillbirths, and neonatal deaths (Klemetti and Saxon 1967; Mackenzie and Lippman 1989; Drews 1990). In terms of searching for a causal exposure, the mindset of mothers whose children have a birth defect may well be different from mothers who experienced these other reproductive outcomes. Nevertheless, an upward bias of relative risk estimates was not observed for postpartum reports of most exposures (Mackenzie and Lippman 1989; Drews 1990). Of note, however, is that repetition of an interview could aid recall, resulting in underestimation of true bias.

Traditionally, recall bias is considered a possible explanation for a spurious increased risk estimate. However, differential recall could operate in the opposite direction, in which mothers of cases deny exposure and, therefore, under-report relative to control mothers. In this scenario, there would be a downward bias of the risk estimate.

Indirect evidence of recall bias comes from a comparison of two studies of cardiovascular birth defects in relation to use of Bendectin (pyridoxine/doxylamine), an antiemetic medication (Rothman 1979; Zierler and Rothman 1985). The first study identified an increased risk based on data derived from retrospective interviews in which mothers were asked a general question about drug use in pregnancy. The second study also employed a case-control design, but asked mothers standardized questions about Bendectin medication use, and no association was observed. The authors concluded that the results of the first study were likely due to recall bias. Indirect evidence of recall bias was also apparent in a study of folic acid supplementation and risk of neural tube defects, when case mothers were stratified according to their knowledge of the study hypothesis (Werler 1993). In retrospective interviews, women were asked an open-ended question, "Have you heard there are any vitamins, minerals, or anything else that may cause

birth defects?" and "Any birth defect in particular?" Women who reported vitamins or folic acid in response to the first question and any type of neural tube defect in response to the second question were considered to have knowledge of the study hypothesis. The now well-documented reduction in neural tube defect risk associated with periconceptional folic acid supplementation was observed for women without knowledge of the hypothesis, but the odds ratio for women with such knowledge was closer to no effect.

Information bias—both *random misclassification* and *recall bias*—can be reduced by use of a standardized questionnaire and decreasing the interval between pregnancy and data collection. Another approach to minimize the possibility of recall bias is the use of a control group comprising mothers of children with malformations other than those of the case group (Werler 1999, 2002). The idea here is that reporting accuracy would be similar for case and control mothers, with the major caveat that there is no association between exposure and the birth defects included in the control group. Inclusion of a wide variety of different specific malformations should reduce the possibility of introducing a selection bias by diluting the impact of any unidentified associations with some specific defects on the control group as a whole. In support of this approach is that most teratogens are not linked to all types of malformations. On the contrary, obesity, diabetes, and heavy alcohol consumption are examples of maternal exposures that appear to affect many different developing organs and tissues in the fetus. Studies that have utilized both malformed and non-malformed control groups have shown remarkably similar prevalences of multivitamin supplementation, obesity, and decongestant and analgesic use (Werler 1996, 1999, 2002), thus providing further indirect evidence against recall bias.

CONCLUSION

In summary, much work is yet to be done to identify the causes of major structural anomalies. Advances in this field are most likely to come from studies that incorporate accurate information on exogenous exposures and genetic influences, with careful consideration of birth-defect diagnoses. Given the rarity of specific anomalies, case-control studies represent the standard. Newer designs such as case only (for gene–environment interactions) might help when no suitable control group can be identified. Although the challenges of conducting retrospective data collection are great, they are not insurmountable. Many risk factors of specific anomalies have been confirmed in case-control studies, including the protective effect of maternal periconceptional folic acid supplementation on neural tube defects in offspring (Mulinare 1988; MRC 1991;Werler 1993; Shaw 1995).

REFERENCES

Alwan, S., Reefhuis, J., Rasmussen, S.A., Olney, R.S., and Friedman, J.M. 2007. Use of selective serotonin-reuptake inhibitors in pregnancy and the risk of birth defects. *N Engl J Med* 356(26): 2684–92.

Amorim, M.R., Lima, M.A., Castilla, E.E., and Orioli, I.M. 2007. Non-Latin European descent could be a requirement for association of NTDs and MTHFR variant 677C > T: A meta-analysis. *Am J Med Genet A* 143A(15): 1726–32.

Ananth, C.V., Platt, R.W., and Savitz, D.A. 2005. Regression models for clustered binary responses: Implications of ignoring the intracluster correlation in an analysis of perinatal mortality in twin gestations. *Ann Epidemiol* 15(4): 293–301.

Barton, T.S., Robaire, B., and Hales, B.F. 2005. Epigenetic programming in the preimplantation rat embryo is disrupted by chronic paternal cyclophosphamide exposure. *Proc Natl Acad Sci USA* 102(22): 7865–70.

Berard, A., Ramos, E., Rey, E., Blais, L., St-Andre, M., and Oraichi, D. 2007. First trimester exposure to paroxetine and risk of cardiac malformations in infants: The importance of dosage. *Birth Defects Res B Dev Reprod Toxicol* 80(1): 18–27.

Buka, S.L., Goldstein, J.M., Spartos, E., and Tsuang, M.T. 2004. The retrospective measurement of prenatal and perinatal events: Accuracy of maternal recall. *Schizophr Res* 71(2–3): 417–26.

Cordier, S. 2008. Evidence for a role of paternal exposures in developmental toxicity. *Basic Clin Pharmacol Toxicol* 102(2): 176–81.

Cragan, J.D., Martin, M.L., Moore, C.A., and Khoury, M.J. 1993. Descriptive epidemiology of small intestinal atresia, Atlanta, Georgia. *Teratology* 48(5): 441–50.

Davis, D.L., Friedler, G., Mattison, D., and Morris, R. 1992. Male-mediated teratogenesis and other reproductive effects: Biologic and epidemiologic findings and a plea for clinical research. *Reprod Toxicol* 6(4): 289–92.

Diav-Citrin, O., Shechtman, S., Weinbaum, D., Wajnberg, R., Avgil, M., Di Gianantonio, E., et al. 2008. Paroxetine and fluoxetine in pregnancy: A prospective, multicentre, controlled, observational study. *Br J Clin Pharmacol* 66(5): 695–705.

Drews, C.D., Kraus, J.F., and Greenland, S. 1990. Recall bias in a case-control study of sudden infant death syndrome. *Int J Epidemiol* 19(2): 405–11.

Dunn, P.M. 2007. Perinatal lessons from the past: Sir Norman Gregg, ChM, MC, of Sydney (1892–1966) and rubella embryopathy. *Arch Dis Child Fetal Neonatal Ed* 92(6): F513–14.

Felix, R.J., Jones, K.L., Johnson, K.A., McCloskey, C.A., and Chambers, C.D. 2004. Postmarketing surveillance for drug safety in pregnancy: The Organization of Teratology Information Services project. *Birth Defects Res A Clin Mol Teratol* 70(12): 944–47.

Ferenz, C., Loffredo, C., Correa-Villasenor, A., and Wilson, P. 1997. Genetic and environmental risk factors of major cardiovascular malformations: The Baltimore-Washington Infant Study 1981–1989. In *Perspectives in pediatric cardiology*, Vol. 5, R. Anderson, ed. Armonk, NY: Future Publishing Co, Inc.

Glinianaia, S.V., Rankin, J., and Wright, C. 2008. Congenital anomalies in twins: A register-based study. *Hum Reprod* 23(6): 1306–11.

Hansen, M., Kurinczuk, J.J., Bower, C., and Webb, S. 2002. The risk of major birth defects after intracytoplasmic sperm injection and in vitro fertilization. *N Engl J Med* 346(10): 725–30.

Heinonen, O., Slone, D., and Sharpiro, S. 1977. *Birth defects and drugs in pregnancy*. Littleton, MA: Publishing Sciences Group, Inc.

Henderson, J., Gray, R., and Brocklehurst, P. 2007. Systematic review of effects of low-moderate prenatal alcohol exposure on pregnancy outcome. *BJOG* 114(3): 243–52.

Holmes, L.B., Kleiner, B.C., Leppig, K.A., Cann, C.I., Munoz, A., and Polk, B.F. 1987. Predictive value of minor anomalies: II. Use in cohort studies to identify teratogens. *Teratology* 36(3): 291–97.

Honein, M.A., Rasmussen, S.A., Reefhuis, J., Romitti, P.A., Lammer, E.J., Sun, L., and Correa, A. 2007. Maternal smoking and environmental tobacco smoke exposure and the risk of orofacial clefts. *Epidemiology* 18(2): 226–33.

International Clearinghouse for Birth Defects Monitoring Systems. 1991. *Congenital malformations worldwide*. Amsterdam: Elsevier.

Jacobson, S.W., Jacobson, J.L., Sokol, R.J., Martier, S.S., Ager, J.W., and Kaplan, M.G. 1991. Maternal recall of alcohol, cocaine, and marijuana use during pregnancy. *Neurotoxicol Teratol* 13(5): 535–40.

Johnson, C.Y. and Little, J. 2008. Folate intake, markers of folate status and oral clefts: Is the evidence converging? *Int J Epidemiol* 37(5): 1041–58.

Jurek, A.M., Greenland, S., Maldonado, G., and Church, T.R. 2005. Proper interpretation of non-differential misclassification effects: Expectations vs observations. *Int J Epidemiol* 34(3): 680–87.

Kalter, H. and Warkany, J. 1983. Medical progress. Congenital malformations: Etiologic factors and their role in prevention (first of two parts). *N Engl J Med* 308(8): 424–31.

Kenneson, A. and Cannon, M.J. 2007. Review and meta-analysis of the epidemiology of congenital cytomegalovirus (CMV) infection. *Rev Med Virol* 17(4): 253–76.

Klemetti, A. and Saxen, L. 1967. Prospective versus retrospective approach in the search for environmental causes of malformations. *Am J Public Health Nations Health* 57(12): 2071–75.

Koren, G., Pastuszak, A., and Ito, S. 1998. Drugs in pregnancy. *N Engl J Med* 338(16): 1128–37.

Korones, S.B. 1986. Congenital rubella: An encapsulated review. In *Teratogen update: Environmentally induced birth defect risks*, pp. 77–80. New York: Alan R. Liss, Inc.

Krall, E.A., Valadian, I., Dwyer, J.T., and Gardner, J. 1988. Recall of childhood illnesses. *J Clin Epidemiol* 41(11): 1059–64.

Lam, P.K., Torfs, C.P., and Brand, R.J. 1999. A low pregnancy body mass index is a risk factor for an offspring with gastroschisis. *Epidemiology* 10(6): 717–21.

Lammer, E.J., Shaw, G.M., Iovannisci, D.M., and Finnell, R.H. 2005. Maternal smoking, genetic variation of glutathione s-transferases, and risk for orofacial clefts. *Epidemiology* 16(5): 698–701.

Lenz, W. 1962. Thalidomide and congenital abnormalities. *Lancet* 1: 271–72.

Leppig, K.A., Werler, M.M., Cann, C.I., Cook, C.A., and Holmes, L.B. 1987. Predictive value of minor anomalies. I. Association with major malformations. *J Pediatr* 110(4): 531–37.

Lewis, J.D., Strom, B.L., Kimmel, S.E., Farrar, J., Metz, D.C., Brensinger, C., et al. 2006. Predictors of recall of over-the-counter and prescription non-steroidal anti-inflammatory drug exposure. *Pharmacoepidemiol Drug Saf* 15(1): 39–45.

Louik, C., Lin, A.E., Werler, M.M., Hernandez-Diaz, S., and Mitchell, A.A. 2007. First-trimester use of selective serotonin-reuptake inhibitors and the risk of birth defects. *N Engl J Med* 356(26): 2675–83.

Mackenzie, S.G. and Lippman, A. 1989. An investigation of report bias in a case-control study of pregnancy outcome. *Am J Epidemiol* 129(1): 65–75.

Martinez-Frias, M.L. 2007. Postmarketing analysis of medicines: Methodology and value of the spanish case-control study and surveillance system in preventing birth defects. *Drug Saf* 30(4): 307–16.

MMWR. 1998. Trends in infant mortality attributable to birth defects: United States, 1980–1995. *MMWR Morb Mortal Wkly Rep* 47(37): 773–78.

MRC. 1991. Prevention of neural tube defects: Results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet* 338(8760): 131–37.

Mulinare, J., Cordero, J.F., Erickson, J.D., and Berry, R.J. 1988. Periconceptional use of multivitamins and the occurrence of neural tube defects. *JAMA* 260(21): 3141–45.

Must, A., Phillips, S.M., Naumova, E.N., Blum, M., Harris, S., Dawson-Hughes, B., and Rand, W.M. 2002. Recall of early menstrual history and menarcheal body size: After 30 years, how well do women remember? *Am J Epidemiol* 155(7): 672–79.

NBDPN. 2007. State birth defects surveillance program directory. *Birth Defects Res A Clin Mol Teratol* 79(12): 815–73.

Paganini-Hill, A. and Ross, R.K. 1982. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol* 116(1): 114–22.

Passos-Bueno, M.R., Serti Eacute, A.E., Jehee, F.S., Fanganiello, R., and Yeh, E. 2008. Genetics of craniosynostosis: genes, syndromes, mutations and genotype-phenotype correlations. *Front Oral Biol* 12: 107–43.

Peller, A.J., Westgate, M.N., and Holmes, L.B. 2004. Trends in congenital malformations, 1974–1999: Effect of prenatal diagnosis and elective termination. *Obstet Gynecol* 104(5 Pt 1): 957–64.

Petrini, J., Damus, K., and Johnston, R.B. 1998. Trends in infant mortality attributable to birth defects: United States, 1980–1995. *MMWR Morb Mortal Wkly Rep* 47(37): 773–78.

Rasmussen, S.A. and Moore, C.A. 2001. Effective coding in birth defects surveillance. *Teratology* 64(Suppl 1): S3–S7.

Reefhuis, J. and Honein, M.A. 2004. Maternal age and non-chromosomal birth defects, Atlanta, 1968–2000: Teenager or thirty-something, who is at risk? *Birth Defects Res A Clin Mol Teratol* 70(9): 572–79.

Reefhuis, J., Honein, M.A., Schieve, L.A., Correa, A., Hobbs, C.A., and Rasmussen, S.A. 2009. Assisted reproductive technology and major structural birth defects in the United States. *Hum Reprod* 24(2): 360–66.

Rothman, K.J., Fyler, D.C., Goldblatt, A., and Kreidberg, M.B. 1979. Exogenous hormones and other drug exposures of children with congenital heart disease. *Am J Epidemiol* 109(4): 433–39.

Sanderson, M., Williams, M.A., White, E., Daling, J.R., Holt, V.L., Malone, K.E., et al. 1998. Validity and reliability of subject and mother reporting of perinatal factors. *Am J Epidemiol* 147(2): 136–40.

Shaw, G.M., Schaffer, D., Velie, E.M., Morland, K., and Harris, J.A. 1995. Periconceptional vitamin use, dietary folate, and the occurrence of neural tube defects. *Epidemiology* 6(3): 219–26.

Shaw, G.M., Quach, T., Nelson, V., Carmichael, S.L., Schaffer, D.M., Selvin, S., and Yang, W. 2003. Neural tube defects associated with maternal periconceptional dietary intake of simple sugars and glycemic index. *Am J Clin Nutr* 78(5): 972–78.

Tomeo, C.A., Rich-Edwards, J.W., Michels, K.B., Berkey, C.S., Hunter, D.J., Frazier, A.L., et al. 1999. Reproducibility and validity of maternal recall of pregnancy-related events. *Epidemiology* 10(6): 774–77.

Torfs, C.P., Christianson, R.E., Iovannisci, D.M., Shaw, G.M., and Lammer, E.J. 2006. Selected gene polymorphisms and their interaction with maternal smoking, as risk factors for gastroschisis. *Birth Defects Res A Clin Mol Teratol* 76(10): 723–30.

Torfs, C.P., Curry, C.J., and Bateson, T.F. 1995. Population-based study of tracheoesophageal fistula and esophageal atresia. *Teratology* 52(4): 220–32.

Van Beynum, I.M., den Heijer, M., Blom, H.J., and Kapusta, L. 2007. The MTHFR 677C->T polymorphism and the risk of congenital heart defects: A literature review and meta-analysis. *QJM* 100(12): 743–53.

Waitzman, N.J., Romano, P.S., and Scheffler, R.M. 1994. Estimates of the economic costs of birth defects. *Inquiry* 31(2): 188–205.

Waller, D.K., Mills, J.L., Simpson, J.L., Cunningham, G.C., Conley, M.R., Lassman, M.R., and Rhoads, G.G. 1994. Are obese women at higher risk for producing malformed offspring? *Am J Obstet Gynecol* 170(2): 541–48.

Waller, D.K., Shaw, G.M., Rasmussen, S.A., Hobbs, C.A., Canfield, M.A., Siega-Riz, A.M., et al. 2007. Prepregnancy obesity as a risk factor for structural birth defects. *Arch Pediatr Adolesc Med* 161(8): 745–50.

Werler, M.M., Hayes, C., Louik, C., Shapiro, S., and Mitchell, A.A. 1999. Multivitamin supplementation and risk of birth defects. *Am J Epidemiol* 150(7): 675–82.

Werler, M.M., Louik, C., Shapiro, S., and Mitchell, A.A. 1996. Prepregnant weight in relation to risk of neural tube defects. *JAMA* 275(14): 1089–92.

Werler, M.M., Mitchell, A.A., Moore, C.A., Honein, M.A., and the National Birth Defects Prevention Network. 2009. Is there epidemiologic evidence to support vascular disruption as a pathogenesis of gastroschisis? *Am J Med Genet A* 149A(7): 1399–406.

Werler, M.M., Shapiro, S., and Mitchell, A.A. 1993. Periconceptional folic acid exposure and risk of occurrent neural tube defects. *JAMA* 269(10): 1257–61.

Werler, M.M., Sheehan, J.E., and Mitchell, A.A. 2002. Maternal medication use and risks of gastroschisis and small intestinal atresia. *Am J Epidemiol* 155(1): 26–31.

Wyszynski, D.F. 2009. Pregnancy exposure registries: Academic opportunities and industry responsibility. *Birth Defects Res A Clin Mol Teratol* 85(1): 93–101.

Yawn, B.P., Suman, V.J., and Jacobsen, S.J. 1998. Maternal recall of distant pregnancy events. *J Clin Epidemiol* 51(5): 399–405.

Yazdy, M.M., Liu, S., Mitchell, A.A., and Werler, M.M. 2010. Maternal dietary glycemic intake and the risk of neural tube defects. *Am J Epidemiol* 171(4):407–14.

Zierler, S. and Rothman, K.J. 1985. Congenital heart disease in relation to maternal use of Bendectin and other drugs in early pregnancy. *N Engl J Med* 313(6): 347–52.

11

Maternal Mortality and Severe Maternal Morbidity

K.S. JOSEPH

MATERNAL MORTALITY

Although maternal mortality rates have declined dramatically over the last century, especially in industrialized countries, maternal deaths remain an important perinatal concern worldwide.

Definition

According to the *International Statistical Classification of Diseases and Related Health Problems*, version 10 (ICD-10), a maternal death is defined as "… . the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and the site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management, but not from accidental or incidental causes" (World Health Organization (WHO) 1993). This implies a causal link between the pregnancy and the death. Maternal deaths are divided into direct and indirect obstetric deaths. *Direct obstetric deaths* refer to "… maternal deaths resulting from obstetric complications of the pregnant state (pregnancy, labor, and puerperium), from interventions, omissions, incorrect treatment, or from a chain of events resulting from any of the above" (WHO 1993). Examples of direct obstetric deaths include those due to ectopic pregnancy, complications of anesthesia during labor, and postpartum hemorrhage. *Indirect obstetric deaths*, on the other hand, refer to maternal deaths "…resulting from previous existing disease or disease that developed during pregnancy and which was not due to direct obstetric causes but which was aggravated by the physiologic effects of pregnancy" (WHO 1993). Maternal death due to rheumatic heart disease aggravated by pregnancy and sudden death in pregnancy due to epilepsy are examples of indirect obstetric deaths.

The labelling of direct and indirect obstetric deaths is sometimes a source of controversy, however. Most suicides occurring during pregnancy or the puerperium are classified as indirect maternal deaths by the Confidential Enquiry into Maternal Deaths (CEMACH) in the United Kingdom because they may be due to puerperal

mental illness (CEMACH 2004, 2007). Similarly, CEMACH classifies some deaths from cancer as indirect maternal deaths, if pregnancy (with its altered hormonal status) is believed to have altered disease progression. Neither suicides nor cancer deaths are classified as indirect obstetric deaths under the ICD coding rules.

The ICD-10 also has defined two other types of death that have a bearing on maternal mortality, namely pregnancy-related death and late maternal death. A *pregnancy-related death* is "… . the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the cause of death" (WHO 1993). The difference between this entity and direct and indirect obstetric death is that the ambit of pregnancy-related deaths is broader and includes death from accidental and incidental causes, such as those due to domestic violence, road traffic collision, and suicide. The ICD-10 defines a *late maternal death* as "… . the death of a woman from direct or indirect obstetric causes more than 42 days but less than 1 year after termination of pregnancy" (WHO 1993). Intensive care units and other modern technologies can postpone death for weeks, and late maternal deaths fulfill the need to capture such deaths even though they occur more than 42 days after termination of the pregnancy.

The ICD-10 recommends that the denominator for calculating maternal mortality be specified as either the number of live births or the number of total births. The term "rate" is acknowledged to be a misnomer in this context, and the *maternal mortality rate* (MMR) is also referred to as the *maternal mortality ratio*. Other options for the MMR denominator include: (a) the number of maternities, i.e., number of mothers delivered of a live birth or stillbirth (differs numerically from births in that women with multifetal pregnancy are counted once only); (b) the number of women aged 15–44 years (maternal mortality rates calculated using this denominator will partly reflect fertility rates); and (c) the estimated number of pregnancies. This third option constitutes the most appropriate denominator and is especially relevant when calculating rates of maternal death in early pregnancy (e.g., death following an ectopic pregnancy or early pregnancy termination). However, it is also the most challenging denominator to obtain accurately, with counts of spontaneous abortions being particularly unreliable.

Sources of data and under-reporting issues

The systems and sources of data for estimating MMR vary widely by country. Civil registration systems, with routine registration of birth and deaths, exist is several countries, although only about 14% of all deliveries in the world occur in such jurisdictions (WHO 2007). Other methods for estimating MMR worldwide include the sisterhood methods (survey methods which determines age, age at death, and year of death of the respondent's adult sisters), reproductive age mortality studies (identification and investigation of all deaths of women of reproductive age in a defined population using multiple sources of data including interviews, vital registration, and hospital records), sample registration systems, census studies, and special studies. The World Health Organization (WHO) and related agencies estimate that 24.5% of births occur in countries where none of these systems are extant (WHO 2007).

The challenge of estimating MMR accurately is highlighted by the fact that even in countries with good civil registration systems, carefully conducted ad hoc studies show substantial underestimation in routine maternal mortality estimates (Turner et al. 2002; Deneux-Tharaux et al. 2005; Lang and King 2008). Death certificates do not provide accurate counts of maternal deaths for several related reasons: death certificates are not specifically designed to identify maternal death; unrecognized early pregnancy or unknown recent delivery at the time of death; death miscoded as being due to a medical cause (e.g., cardiac disease); and death misclassified as being due to "multiple organ failure" after prolonged hospitalization in an intensive care unit.

In the United States, two systems provide information on the number of maternal deaths (MacKay et al. 2005; Lang and King 2008). First, the National Center for Health Statistics (NCHS) reports maternal mortality using death certificate information from the National Vital Statistics System (NVSS). Second, the CDC's National Center for Chronic Disease Prevention and the Pregnancy-Related Mortality Surveillance System (PMSS) conducts epidemiologic surveillance of pregnancy-related deaths using death certificate information augmented by data from other sources, including through a linkage to birth certificate data. The implementation of the ICD–10 in 1999 resulted in about a 13% increase in the number of deaths identified as maternal deaths. The introduction of a separate pregnancy question on the U.S. Standard Certificate of Death increased the rate again between 2002 and 2003 (Hoyert 2007). In Canada, the Canadian Perinatal Surveillance System, through the Maternal Health Study Group, endorses enhanced surveillance of maternal mortality and severe morbidity as priority areas. Surveillance data sources include information from death certificates, coroners' reports, and hospitalization data (Health Canada 2004).

The Confidential Enquiry into Maternal Deaths in the United Kingdom is carried out every 3 years as per a government requirement. All health care providers have a duty to provide relevant information to the Enquiry. This surveillance model is termed "sentinel event surveillance," and a highly organized system of information reporting and collection is in place in the United Kingdom. The Enquiry and its reports (CEMACH 2004, 2007) are considered the "gold standard" for maternal mortality surveillance (Australia Institute of Health and Welfare 2001). Information from CEMACH reports provide a sense of the magnitude by which routine vital registration systems underestimate maternal mortality rates (Fig. 11.1). The report also shows how improvements in data collection methods sometime results in a rising (presumably spurious) secular trend in maternal mortality and other adverse events.

International comparisons

The most recent estimates of the global MMR were 402 (95% confidence interval 216 to 654) per 100,000 live births for 2005 (WHO 2007), 251 (95% confidence interval 221 to 289) per 100,000 live births for 2008 (Hogan et al. 2010) and 260 (95% confidence interval 200 to 370) per 100,000 live births for 2008 (WHO 2010). Although these may appear to be widely dissimilar estimates, the wide confidence

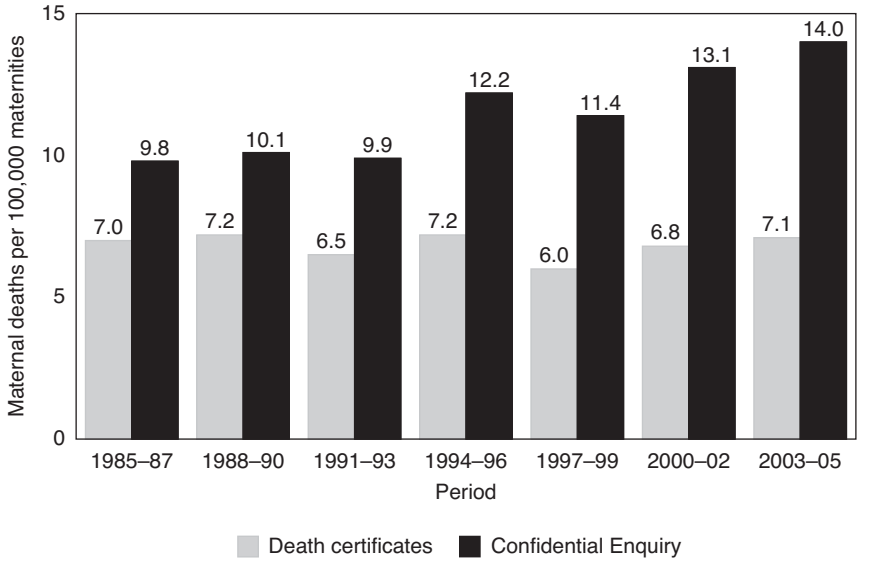


Figure 11.1 Rates of maternal death (direct and indirect) in the United Kingdom, ascertained from death certificates and by the Confidential Enquiry into Maternal Deaths, 1985–1987 to 2003–2005 (CEMACH 2007). *Note*: The underestimation of the maternal death rate in death certificates and the rising rate of maternal deaths as determined by the Enquiry, likely reflect improved data capture and reporting.

intervals mean that they are in fact not significantly different (P value >0.05). More than 99% of all maternal deaths occur in less-industrialized countries. The MMR was highest in Africa, followed by Oceania, Asia, and Latin America and the Caribbean (Table 11.1). However, these figures hide large intercountry variations within continents. With the exception of Afghanistan, all 12 countries with MMRs in excess of 750 were in Africa (WHO 2010). The MMR in some remote, rural regions of Afghanistan was estimated to be 6,507 (95% confidence interval [CI] 5,026–7,988) per 100,000 live births in 1999–2002 (Bartlett et al. 2005). Nevertheless, differences in population size and the number of births mean that the largest absolute number of maternal deaths occurred in India (63,000 annually) (WHO 2010).

Maternal mortality rates per 100,000 live births in selected countries in 2008 include: Sweden 5; Australia 8; Japan 6; Canada 12; France 8; United Kingdom 12; United States 24; Russia 39; China 38; Brazil 58; Mexico 85; India 230; Haiti 300; South Africa 410; Sierra Leone 970; Afghanistan 1400 (WHO 2010).

Millennium development goals

The international community adopted the Millennium Declaration in 2000, and set the year 2015 as the year for reducing maternal mortality by 75% (i.e., from 430 per 100,000 live births in 1990 to 108 per 100,000 live births by 2015). Unfortunately, sufficient progress has not been made toward achieving this goal.

Table 11.1 Worldwide maternal death rates (per 100,000 live births) by region, 1990 and 2008

Region/country	*Deaths in 1990*		*Deaths in 2008*	
	Number	*Rate*	*Number*	*Rate*
World	546,000	400	358,000	260
Industrialized	2,000	16	1,700	14
Less-industrialized	540,000	450	355,000	290
Africa	208,000	780	207,000	590
Asia	315,000	390	139,000	190
Latin America, Caribbean	17,000	140	9,200	85
Oceania	540	290	550	230

From World Health Organization. 2010. *Trends in maternal mortality: 1990 to 2008. Estimates developed by WHO, UNICEF, UNFPA, and the World Bank*. Geneva: World Health Organization.

Between 1990 and 2008, the MMR decreased by 2.3% per year to 260 maternal deaths per 100,000 live births, whereas a 5.5% decline annually is required to achieve this Millennium Development Goal (United Nations 2008). Progress toward this goal has been particularly lacking in sub-Saharan Africa. The much lower estimate of MMR made for 2008 (Hogan et al. 2010) does not alter this conclusion regarding dismal progress. Millenium Devlopment Goal 5 was based on a 75% reduction in maternal mortality from 1990 levels and the study which lowered the 2008 estimate also revised the 1990 estimate downwards to 320 per 100,000 live births (Hogan et al. 2010).

Historical trends

National statistics on maternal mortality show that this rate was as high as 900 per 100,000 live births in Sweden in 1751, when national statistics were first complied (Högberg 2004). At the beginning of the 20th century, the MMR in England and Wales was 440 per 100,000 live births, whereas that in the United States was between 520 and 850 maternal deaths per 100,000 live births. This compared with a rate of 230 per 100,000 live births in Sweden. The substantially lower MMR in Sweden in the early 1900s has been attributed to an extensive collaboration between physicians and highly competent trained midwives, who attended almost 80% of deliveries (Loudon 2000; Högberg 2004). It is estimated that in the years between 1861 and 1900, use of antiseptic techniques reduced maternal mortality due to puerperal sepsis by about 50%, and midwifery reduced maternal deaths due to nonseptic causes by another 40%–50% (Högberg 2004).

Maternal mortality rates declined dramatically in industrialized countries from the mid-1930s onward. The main interventions responsible for this change in maternal survival were antibacterial agents, blood transfusions, ergometrine (an ergot alkaloid used to reduce postpartum hemorrhage), and improvements in the organization of obstetric services (Loudon 2000; Högberg 2004). Prontosil, the trade name of the first sulfonamide patented, was discovered in the early 1930s and proved effective against an important cause of puerperal sepsis (*Streptococcus*

pyogenes). The first hospital blood bank in the United States was established in 1937, with blood transfusions and antibacterial agents/antibiotics (such as sulfonamides, penicillin, and streptomycin) becoming widely used after World War II.

Causes of maternal deaths

The major causes of maternal death globally (WHO 2003) include hemorrhage; hypertensive disorders of pregnancy, such as preeclampsia and eclampsia; sepsis; obstructed labor; deaths following unsafe abortion; anemia; and HIV/AIDS (Table 11.2). However, the leading causes of maternal death in industrialized countries are considerably different (Khan et al. 2006). In the United Kingdom, the most frequent cause of direct maternal death in 2003–2005 was thrombosis and thromboembolism (1.94 per 100,000 maternities), whereas preeclampsia/eclampsia (0.85), sepsis (0.85), amniotic fluid embolism (0.80), and hemorrhage (0.66) were the other leading causes (CEMACH 2007). The leading causes of indirect maternal death included deaths from cardiac causes (2.27 per 100,000 maternities), psychiatric causes (0.85), and malignancy-related deaths (0.47) (CEMACH 2007). The leading causes of maternal death in Canada between 1999 and 2004 (Public Health Agency of Canada 2008) were hypertensive disorders (6.0 per 1,000,000 live births), obstetrical pulmonary embolism (5.0), and cerebrovascular disorders (6.5). In the United States (1991–1999), embolism, hemorrhage, hypertension, and infection were the leading causes of pregnancy-related death (MacKay et al. 2005). Some of these between-country differences in the leading causes of death in industrialized countries appear to be due to differences in the classification and categorization schemes used with respect to causes of death.

Table 11.2 Estimated proportion of maternal deaths by cause and region, 2000

Cause of death	*Proportion of all maternal deaths by cause*			
	Industrialized countries	*Africa*	*Asia*	*Latin America and the Caribbean*
Hemorrhage	13.4	33.9	30.8	20.8
Hypertensive disorders	16.1	9.1	9.1	25.7
Sepsis/infection	2.1	9.7	11.6	7.7
Abortion	8.2	3.9	5.7	12.0
Obstructed labor	0.0	4.1	9.4	13.4
Anemia	0.0	3.7	12.8	0.1
HIV/AIDS	0.0	6.2	0.0	0.0
Ectopic pregnancy	4.9	0.5	0.1	0.5
Embolism	14.9	2.0	0.4	0.6
Other causes	35.7	21.6	14.1	7.7
Unclassified	4.8	5.4	6.1	11.7

From Khan, K.S., Wojdyla, S., Say, L., Gülmezoglu, M., and Van Look, P.F. 2006. WHO analysis of causes of maternal death: a systematic review. *Lancet* 367:1066–74.

METHODOLOGIC ISSUES IN THE STUDY OF MATERNAL MORTALITY

Study size and design issues

Approximately 20 maternal deaths occur in Canada every year (Public Health Agency of Canada 2008), while in the United Kingdom, there are less than 100 maternal deaths (CEMACH 2007). This relatively small number of maternal deaths has led many countries to carry out their reporting of maternal death on a periodic basis. Maternal death reporting tends to be a detailed numerator analysis, supplemented with information on the overall patterns of pregnancy and childbirth. Can causal conclusions be drawn on the basis of such death reviews? In fact, in recent years, several consequential policy decisions have been made based on information from the Confidential Enquiry into Maternal Deaths in the United Kingdom. The temporal trends in deaths from thrombosis and thromboembolism illustrate how CEMACH used Enquiry information to change policy and reduce maternal mortality in the United Kingdom.

By the early 1990s, several reports had noted the increasing number of deaths after cesarean delivery, and the need for thromboprophylaxis following gynecologic surgery (CEMACH 2004). In 1995, the Royal College of Obstetricians and Gynaecologists (RCOG) published recommendations for risk assessment and thromboprophylaxis following cesarean delivery. This recommendation was endorsed in the Confidential Enquiry report published in 1996 (based on data from the 1991–1993 triennium). In 1994–1996, there were 46 deaths from pulmonary embolism, of which 15 occurred after cesarean delivery and 10 occurred after vaginal delivery. With RCOG recommendations widely publicized and implemented, deaths due to pulmonary embolism following cesarean delivery fell sharply to four in the subsequent triennium (1997–1999), while the number of deaths due to pulmonary embolism following vaginal delivery did not decrease (Fig. 11.2). The subsequent 2000–2002 report included guidelines requiring physicians to determine each women's risk for thromboembolism (e.g., based on factors such as previous venous thromboembolism, thrombophilia, obesity, parity) and, depending on risk, advocated interventions such as early mobilization, leg stockings, and/or heparin prophylaxis for women delivering by the cesarean or the vaginal route (CEMACH 2004). However, maternal deaths from thrombosis and thromboembolism in the United Kingdom have increased in recent years, at least partly due to changes in population characteristics (e.g., age, body mass index [BMI]) and obstetric practices. Thrombosis and thromboembolism remain the leading causes of maternal death. This example of combining numerator information with that from other sources illustrates a synthetic and less formal method of formulating inference and policy, given fragmented information.

Although maternal mortality reviews such as the Confidential Enquiry may appear to be a numerator analysis, the Confidential Enquiry may be viewed as a cohort study that allows the estimation of relative risks by maternal age, plurality, ethnicity, BMI, and other factors, with the denominator information obtained from an independent data source (such as the Health Survey of England). For instance,

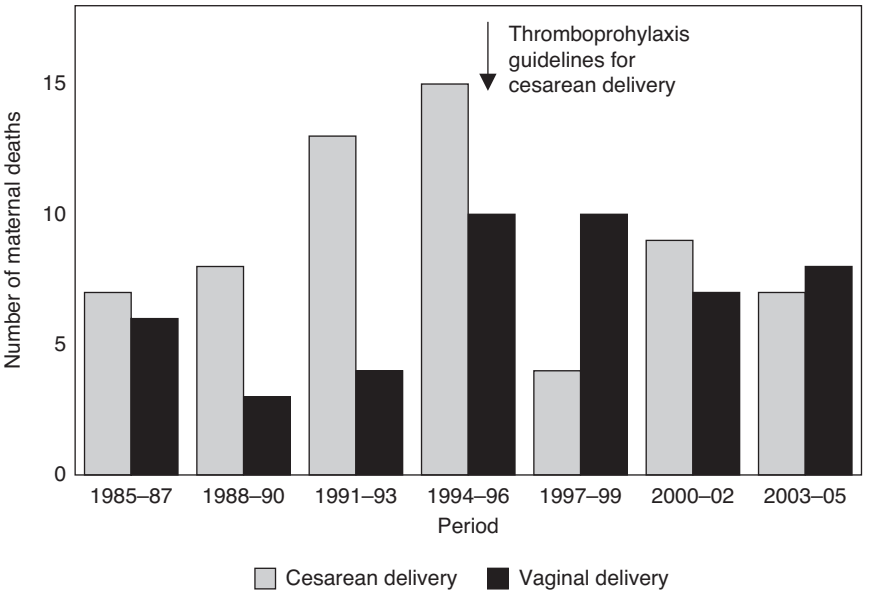


Figure 11.2 Numbers of maternal deaths due to pulmonary embolism in the United Kingdom, 1985–2005. *Note*: The effect of thromboprophylaxis guidelines (introduced in 1995–1996 for the prevention of pulmonary embolism associated with cesarean delivery) was initially evident on cesarean but not vaginal deliveries.

postal code information was used to determine that women from the most socioeconomically deprived areas of England were 5.1 (95% CI 3.2–8.1) times as likely to suffer a maternal death compared to women from the least-deprived areas (CEMACH 2007). The main limitation of such estimates is that adjustment for other factors is not feasible in individual-level analyses. The Confidential Enquiry could also serve as a "population-based" case-control study (primary study base with secondary scheme for case ascertainment, in which the challenge is complete case ascertainment, not sampling of study base [Miettinen 1985]). With the Confidential Enquiry achieving complete case ascertainment, it should be a relatively simple matter to sample the population of the United Kingdom for a representative base series.

Confounding by indication

In 2000–2002, the MMR among women in the United Kingdom who delivered by the vaginal route was 4.8 (95% CI 3.8–6.0) per 100,000 maternities, compared with a maternal death rate of 17.2 (95% CI 13.4–21.6) per 100,000 maternities among women who delivered by the cesarean route (CEMACH 2004). The maternal death rate among emergency and urgent cesarean deliveries was even higher (20.8; 95% CI 15.1–27.9, per 100,000 maternities), whereas that among scheduled and elective cesarean deliveries was relatively lower (13.6; 95% CI 9.1–19.5, per 100,000

maternities). Although some of the excess deaths among women who had a cesarean delivery may have occurred due to complications of anesthesia or surgery, inference regarding the safety of the mode of delivery based on the above rates would be seriously compromised because the association is confounded by the indication for cesarean delivery. As the CEMACH report stated "… it is almost impossible to disentangle the consequences of caesarean section from the indication for the operation. True, there are occasional deaths from anaesthesia or hemorrhage that result directly from the procedure…… For the large majority of deaths that followed caesarean section, however, there were serious prenatal complications or illness that, in many cases, precipitated the caesarean section. Perimortem caesarean section is the starkest example of this" (CEMACH 2004).

Confounding by indication is, in fact, a well-recognized problem, especially in pharmacoepidemiology (Miettinen 1983; McMahon 2003). It is the underlying phenomenon that explains why death rates among patients in an intensive care unit are higher than among patients in the non–intensive care setting, why rates of stroke are higher among people on antihypertensive medication compared with people who are not on such medication, why death rates are higher among those receiving a blood transfusion in the emergency room or other setting compared with those not receiving blood transfusion, why perinatal mortality rates are higher among hospital births compared with home births, and why perinatal mortality rates among women cared for by obstetricians are typically higher than those cared for by nonobstetricians.

The indication for a particular therapy implies a high risk for the particular outcome, which the therapy is expected to prevent. Often the indication can be graded (e.g., severity of hypertensive disease, degree of blood loss following trauma). However, it is not possible to precisely quantify disease severity (if this were feasible, we could adjust away relevant differences in nonexperimental studies). This is perhaps the most important reason why efficacy of therapies is best assessed through a randomized trial (with randomization creating intervention and placebo groups that are comparable with regard to known and unknown risk factors for the outcome).

Mortality among pregnant and nonpregnant women

Pregnancy, and especially childbirth, are sometimes referred to as dangerous events since they may pose unique risks to the mother (e.g., due to eclampsia or hemorrhage). Nevertheless, the term "dangerous" has to be qualified with reference to a group of known risk. The medical literature shows that suicide rates are much lower in pregnant women and in the first year after birth, compared with women not recently pregnant (CEMACH 2004, 2007). In fact, overall rates of death are much lower among pregnant women and in the year following childbirth, compared with women not recently pregnant (CEMACH 2004, 2007). This appears to be a "healthy-woman" effect, analogous to a healthy-worker effect. Nevertheless, in particular circumstances that are notable in terms of socioeconomic deprivation, high levels of violence, or a lower social status for women, the pregnant state can in fact be associated with higher death rates (Dannenberg et al. 1995; Bartlett et al. 2005).

SEVERE MATERNAL MORBIDITY

In recent years, an increasing emphasis has been placed on issues surrounding severe maternal morbidity (and the subset of life threatening situations referred to as "near misses"). There are several reasons for this emphasis, including the higher frequency of this outcome as compared with maternal mortality. Moreover, life-threatening illnesses are a legitimate focus for study and are often associated with long-term sequelae and disability.

Definition

Severe maternal morbidity is defined using disease-specific, intervention-specific, and organ system–specific definitions. *Disease-specific severe maternal morbidity* includes entities such as severe postpartum hemorrhage, eclampsia, severe preeclampsia, thromboembolism, and amniotic fluid embolism (among others). The specific conditions included in a comprehensive list varies from study to study, and at present, there is no international consensus on the components of severe maternal morbidity.

Blood transfusion, admission to an intensive care unit, and emergency hysterectomy are some of the interventions used to identify severe maternal morbidity. Such definitions largely avoid the problem of mild cases of specific diseases being included in disease-specific definitions of severe maternal morbidity, although indications/criteria for specific interventions may vary by place and time. One limitation of intervention-specific criteria, such as admission to an intensive care unit, arises because medical care of high intensity is increasingly delivered outside the intensive care unit (thereby, leading to an underestimation of severe maternal morbidity). Also, various regional and practice considerations can determine how specific interventions are used. For instance, major obstetric hemorrhage rates in Scotland increased from 3.4 (95% CI 3.0–4.0) per 1,000 maternities in 2003 to 4.4 (95% CI 3.8–5.0) per 1,000 maternities in 2005, whereas rates of hysterectomy for postpartum hemorrhage decreased (Brace et al. 2007; CEMACH 2007; Knight et al. 2009). On the other hand, in Canada, where rates of atonic postpartum hemorrhage increased from 29.4 to 39.5 per 1,000 deliveries from 1991 to 2004, rates of postpartum hemorrhage with hysterectomy increased by 73% (Joseph et al. 2007). The increase in hysterectomy rates in Canada was probably because of a relative preference for hysterectomy over blood transfusion in the management of severe postpartum hemorrhage. The reluctance of Canadian physicians to transfuse patients may be traced to a crisis in the blood banking system following blood transfusion–related transmission of HIV and hepatitis C infection in the early 1990s (Gray 1998).

Severe maternal morbidity may also be classified according to the organ system that fails and causes a life-threatening event (e.g., cardiac, renal) (WHO 2004). Recently, the WHO proposed a definition and an organ dysfunction–based scheme for classifying and identifying severe maternal morbidity (Pattison et al. 2009; Say et al. 2009). The proposal includes a list of potentially life-threatening conditions, namely, hemorrhagic disorders (e.g., abruptio placentae), hypertensive disorders

(e.g., severe preeclampsia), other systemic disorders (e.g., endometritis), and "severe management indicators" (e.g., blood transfusion) to optimize prospective surveillance for organ dysfunction. Organ dysfunction is to be ascertained based on various clinical (e.g., acute cyanosis), laboratory-based (e.g., oxygen saturation <90% for ≥60 minutes), and management criteria (e.g., continuous use of vasoactive drugs).

Most schemes for identifying and classifying severe maternal morbidity use both diseases and interventions in their lists of severe maternal morbidity. An index called the *maternal morbidity outcome indicator* (MMOI) has been designed (and validated) to measure major maternal morbidity in routinely collected population health data (Roberts et al. 2008). The MMOI is based on diagnoses (such as severe preeclampsia, uterine rupture), procedures (such as blood transfusion, hysterectomy, mechanical ventilation, general anesthesia), and duration of hospital stay over 10 days.

Frequency of severe maternal morbidity

A Scottish study (Brace et al. 2004) determined the prevalence of severe maternal morbidity in 2001–2002 to be 3.8 (95% CI 3.3–4.4) per 1,000 deliveries, a Canadian study (Wen et al. 2005) using a different definition found 4.4 (95% CI 4.3–4.5) cases of severe maternal morbidity per 1,000 deliveries in 1991–2001, and a recent study from the United States (Callaghan et al. 2008) reported a rate of 5.9 per 1,000 deliveries for the period 1999–2003. These rates contrast with a rate of severe maternal morbidity of 12.5 per 1,000 deliveries documented in New South Wales between 1999 and 2004 (Roberts et al. 2009) and a rate of 13.8 (95% CI 13.6–14.0) per 1,000 deliveries in Canada for the period 2003 to 2007, based on a different definition of severe maternal morbidity (Joseph et al. 2010). A nine-country European study (Zhang et al. 2005) conducted between 1995 and 1998, which assessed the rate of severe preeclampsia, hemorrhage, and sepsis showed that overall rates varied from 6.0 (95% CI 4.1–8.3) per 1,000 deliveries in Austria, to 14.7 (95% CI 12.9–16.6) per 1,000 deliveries in Belgium. Some of this variation in rates of severe maternal morbidity is not unexpected, given the widely varying definitions used. Nevertheless, two outstanding features include the relatively large contribution of hemorrhage to some of the rates (50% in Scotland, 67% in New South Wales, and 60% in the United States) and the magnitude of the overall rates. The latter appear sufficiently large to accord severe maternal morbidity the same status as other perinatal events, such as stillbirth and neonatal death. However, the emphasis placed on surveillance and research related to severe maternal morbidity, at least historically, does not appear to be consistent with its relative high frequency and substantial impact on the mother (and the baby).

Non–life threatening morbidity

Maternal morbidity that is serious but not life threatening is also receiving increasing emphasis in recent years. For example, the lack of evidence supporting the need for routine episiotomy and the potential effects of the mode of delivery on pelvic floor

function have received considerable public attention in recent years. Similarly, the physical and psychosocial morbidity associated with severe perinatal laceration and incontinence (urinary, fecal, and flatal) has become the subject of active research.

METHODOLOGIC AND SUBSTANTIVE ISSUES IN THE STUDY OF SEVERE MATERNAL MORBIDITY

General considerations

Whereas the need for monitoring stillbirth, neonatal death, and infant mortality rates has been historically addressed using vital statistics data complemented by ad hoc studies, obtaining detailed information on severe maternal morbidity within populations requires large hospital discharge databases. These data sources have much in common with routine vital statistics data, especially because their primary function often tends to be of an administrative or financial nature. Understanding data quality issues and methods to optimize inferences from such data is a critical issue.

Other important methodologic caveats in the study of severe maternal morbidity include the recognition that patterns and trends in maternal mortality may not necessarily be reflected in patterns and trends in severe maternal morbidity. This is illustrated by the example of postpartum hemorrhage, which is stable or declining as a cause of maternal death in Canada (Public Health Agency of Canada 2008). On the other hand, postpartum hemorrhage and severe postpartum hemorrhage (e.g., hysterectomy for postpartum hemorrhage) rates have been increasing over the last several years (Joseph et al. 2007). This is true for many industrialized countries, where obstetric hemorrhage is a relatively small contributor to maternal death, despite constituting a substantially larger fraction of severe maternal morbidity. This disconnection illustrates how medical technology and management can disrupt the expected correlation between patterns of maternal mortality and morbidity. A second methodologic issue relates to the categorization of heterogenous conditions of varying etiology into a single severe maternal morbidity index. The rationale for studying the entire gamut of severe maternal morbidity versus focusing on specific entities has to be informed by clinical understanding. The sections below discuss three specific types of severe maternal morbidity and highlight some epidemiologic features of these diseases.

Postpartum hemorrhage

The province of Victoria, Australia, documented an increase in postpartum hemorrhage in the late 1990s (Haynes et al. 2004). This was attributed to a local change in the definition of postpartum hemorrhage, from a blood loss of 600 mL to a blood loss of 500 mL. However, an increase in hysterectomy for postpartum hemorrhage in the early 2000s remained unexplained (Haynes et al. 2004). New South Wale, Australia, reported an increase in postpartum hemorrhage from 4.7% in 1994 to 6.0% in 2002 (Cameron et al. 2006; Ford et al. 2007). This increase was

observed among women delivering by the vaginal route and by the cesarean route. The Confidential Enquiry into Maternal and Child Health (2000–2002 triennium [CEMACH 2004]) reported that maternal deaths from placental abruption, placenta previa, and postpartum hemorrhage in the United Kingdom had increased in 2000–2002 for the first time since 1988 (from 3.3 per million maternities in 1997–1999 to 8.5 per million maternities in 2000–2002). This increase was due to a rise in maternal deaths due to postpartum hemorrhage (1 death in 1997–1999 and 10 deaths in 2000–2002). The more recent CEMACH report (CEMACH 2007) documented a lower rate of maternal death from placental abruption, placenta previa, and postpartum hemorrhage (6.6 per million maternities in 2003–2005), although deaths due to postpartum hemorrhage have remained high (9 in 2003–2005).

The Canadian Perinatal Surveillance System reported an increase in postpartum hemorrhage associated with hysterectomy in its 2003 report (Public Health Agency of Canada 2008). A subsequent investigation (Joseph et al. 2007) showed that postpartum hemorrhage rates in Canada had increased from 4.1% in 1991 to 5.1% in 2004, whereas atonic postpartum hemorrhage rates had increased from 2.9% in 1991 to 3.9% in 2004. No explanation for the temporal increase in atonic postpartum hemorrhage was forthcoming, despite accounting for maternal age, previous cesarean delivery, labor induction, epidural anaesthesia, prolonged second stage, and other factors. Maternal mortality due to postpartum hemorrhage did not increase. Detailed analyses of hospitalization data from the United States have also shown a substantial increase in the frequency of cases with atonic postpartum hemorrhage, very similar to the patterns observed in Canada (Knight et al. 2009).

There appears to be an epidemic of postpartum hemorrhage in several industrialized countries in recent years (Knight et al. 2009). The increase in postpartum hemorrhage coincided with large changes in maternal characteristics (e.g., increases in older maternal age and obesity) and obstetric practice (e.g., increasing use of epidural anesthesia, labor induction, and cesarean delivery). Although it is tempting to attribute the increase in atonic postpartum hemorrhage to one or more of these changes, a pressing issue that needs to be addressed relates to data quality. The diagnosis of postpartum hemorrhage is difficult to make, since estimation of blood loss following delivery is challenging. Although it seems unlikely that hospital discharge databases in several countries would make similar errors, such a possibility cannot be excluded, given the similarity in their health information systems. For instance, their use of ICD codes, which cannot clearly identify atonic postpartum hemorrhage from other causes of immediate postpartum hemorrhage, is one obvious point of concern (Knight et al. 2009). Nevertheless, clinicians have to be alert to this potential increase in postpartum hemorrhage frequency, and further research should help to ascertain the cause.

Amniotic fluid embolism

Amniotic fluid embolism is a rare disease characterized by the abrupt onset of hypotension, hypoxia, and disseminated intravascular coagulation among women in labor or in the immediate postpartum period. The frequency is estimated to be about 1 in 20,000 deliveries (Gilbert and Danielsen 1999). Despite the low frequency,

the high case fatality rate makes it one of the leading causes of maternal death in industrialized countries. Although amniotic fluid entering the circulation has been proposed as the causative mechanism for this condition, it is well recognized that such an event is common. The alternative explanation proposed is that the characteristic collapse observed is a consequence of an anaphylactic reaction to amniotic fluid entering the maternal circulation (Clark et al. 1995).

A recent epidemiologic study based on a hospital discharge database (Kramer et al. 2006) provides some interesting risk-factor information about amniotic fluid embolism. One noteworthy aspect of this study is that the risk-factor analysis was carried out on all cases of amniotic fluid embolism and repeated on all fatal cases of amniotic fluid embolism. The similarity of the results from the two analyses provides some assurance that the database information was probably reliable (since the case fatality rate among women with amniotic fluid embolism was unexpectedly low). The findings of the study were also of interest from a substantive point of view. Many of the significant risk factors (e.g., multiple pregnancy, labor induction, cesarean delivery, placenta previa/abruption, polyhydramnios, cervical laceration, uterine rupture) and the protective (e.g., dystocia) factors identified were consistent with the presumed causal roles of strong uterine contractions, excess amniotic fluid, and disruption of the uterine vasculature. Other studies have confirmed some but not all of these findings (Abenhaim et al. 2008; Spiliopoulos et al. 2009).

Childbirth and incontinence

The relation between vaginal delivery and incontinence began to receive increased attention in the 1990s, with endosonographic and related studies showing occult sphincter damage during delivery (Sultan et al. 1993). Such pelvic floor concerns soon became a focal point in the debate around the benefits/risks of elective cesarean delivery. More recently, numerous studies have examined the relation between urinary and other incontinence and childbirth. For instance, a survey of a random sample of over 4,000 women aged 25–84 years contrasted women who had delivered vaginally versus those who had delivered by cesarean (Lukacz et al. 2006). The adjusted odds ratio (OR) for stress urinary incontinence given vaginal delivery was 1.81 (95% CI 1.25–2.61), that for anal incontinence was 1.72 (95% CI 1.27–2.35), and that for more than one pelvic floor disorder was 1.85 (95% CI 1.42–2.41). The authors estimated that seven cesarean deliveries would prevent one pelvic floor disorder.

Similar risks associated with vaginal delivery have been identified in other studies. A longitudinal follow-up of women at 3 months and 6 years after birth showed lower rates of persistent urinary incontinence (OR 0.46, 95% CI 0.32–0.68) and long-term urinary incontinence (0.50, 95% CI 0.40–0.63) among women who had delivered exclusively by cesarean, compared with women who had delivered vaginally (MacArthur et al. 2006). Women who had had both cesarean and vaginal deliveries were not similarly protected (OR for persistent incontinence 0.93, 95% CI 0.67–1.23).

A follow-up of the Term Breech Trial (Hannah et al. 2002) also provided an interesting insight into this issue. This nonblinded trial randomized women with breech presentation at term to planned vaginal or planned cesarean delivery.

Although the trial was designed primarily to ascertain the effects of the mode of delivery on the infant, the effects on urinary and other types of incontinence were also assessed. After 3 months of follow-up, 4.5% of women in the planned cesarean group and 7.3% of women in the planned vaginal group reported experiencing urinary incontinence (relative risk [RR] 0.62, 95% CI 0.41–0.93), 0.8% versus 1.5% experienced fecal incontinence (RR 0.54, 95% CI 0.18–1.62), and 10.7% versus 9.1% experienced flatal incontinence (RR 1.10, 95% CI 0.79–1.54).

Although these studies represent substantial evidence favoring an adverse effect of vaginal birth on urinary incontinence, other approaches to the issue cast the problem in a different light. A study on postmenopausal (nulliparous) nuns with an average age of 68 years showed that 50% (95% CI 41%–58%) reported having a problem with bladder control, with 30% of affected women indicating stress incontinence (problem when coughing, sneezing, or exercise), 24% indicating urge incontinence (problem during sleep), 35% responding that they had a mixed problem, and 11% unclear on type of incontinence (Buchsbaum 2002). Similarly, another study contrasting the experience of nulliparous women and their parous sisters (average age 61 years) showed negligible differences (48% vs. 50%) in rates of urinary incontinence (Buchsbaum 2005). This study had two other equally informative findings. First, there was a high concordance in continence status within biological sisters (63% of sisters were concordant, 17% of sisters were discordant, with the nulliparous sister being incontinent and 20% of sisters were discordant, with the parous sister being incontinent). The second interesting finding of this study was the substantial (20%) disagreement between symptom-based diagnosis of urinary incontinence versus diagnosis based on clinical evaluation by observation, pad test, and multichannel urodynamic testing. In contrast to this study, other reports (Abramov et al. 2005; Goldberg et al. 2005) on twin sisters (mean age 47 years) showed an association between vaginal delivery and urinary stress incontinence (OR 2.3, 95% CI 1.1–4.6, compared with cesarean delivery).

Numerous other research studies have examined the effects of childbirth and the mode of delivery on incontinence and other pelvic floor dysfunction. A preliminary observation one can make based on the above-cited studies is that even if cesarean delivery does protect against incontinence in the short term, age may be an important modifier of the relation (i.e., vaginal nulliparity appears to "buy time"). More careful studies, which resolve issues related to the definition of incontinence (symptom-based vs. urodynamic evaluation) and address issues of short- versus long-term effects, are needed.

CONCLUSION

In summary, maternal mortality remains an important global concern in many parts of the world. Severe maternal morbidity is also receiving increasing attention in epidemiologic circles. For the perinatal epidemiologist, the study of maternal mortality and severe maternal morbidity can provide a stimulating area for research, given its challenges, the historical inattention accorded to the field, and the potential for improving the health status of mothers.

REFERENCES

Abenhaim, H.A., Azoulay, L., Kramer, M.S., Leduc, L. 2008. Incidence and risk factors of amniotic fluid embolisms: A population-based study on 3 million births in the United States. *Am J Obstet Gynecol* 199: 49.e1–8.

Abramov, Y., Sand, P.K., Botros, S.M., Gandhi, S., Miller, J.J., Nickolov, A., and Goldberg, R.P. 2005. Risk factors for female anal incontinence: New insight through the Evanston-Northwestern twin sisters study. *Obstet Gynecol* 106: 726–32.

Australia Institute of Health and Welfare. *Report on maternal deaths in Australia*, 1994–1996.

Australia Institute of Health and Welfare. 2001. *Cat No. PER 13*. Canberra: Australia Institute of Health and Welfare.

Bartlett, L.A., Mawji, S., Whitehead, S., Crouse, C., Dalil, S., Ionete, D., et al. 2005. Where giving birth is a forecast of death: Maternal mortality in four districts of Afghanistan, 1999–2002. *Lancet* 365: 864–70.

Brace, V., Kernaghan, D., and Penney, G. 2007. Learning from adverse clinical outcomes: Major obstetric hemorrhage in Scotland, 2003–2005. *BJOG* 114: 1388–96.

Brace, V., Penney, G., and Hall, M. 2004. Quantifying severe maternal morbidity: A Scottish population study. *BJOG* 111: 481–84.

Buchsbaum, G.M., Chin, M., Glantz, C., and Guzick, D. 2002. Prevalence of urinary incontinence and associated risk factors in a cohort of nuns. *Obstet Gynecol* 100(2): 226–29.

Buchsbaum, G.M., Duecy, E.E., Kerr, L.A., Huang, L.S., and Guzick, D.S. 2005. Urinary incontinence in nulliparous women and their parous sisters. *Obstet Gynecol* 106: 1253–58.

Callaghan, W.M., Mackay, A.P., and Berg, C.J. 2008. Identification of severe maternal morbidity during delivery hospitalizations, United States, 1991–2003. *Am J Obstet Gynecol* 199: 133.e1–8.

Cameron, C.A., Roberts, C.L., Olive, E.C., Ford, J.B., and Fischer, W.E. 2006. Trends in postpartum haemorrhage. *Aust N Z J Public Health* 30: 151–56.

Clark, S.L., Hankins, G.D., Dudley, D.A., Dildy, G.A., and Porter, T.F. 1995. Amniotic fluid embolism: Analysis of the national registry. *Am J Obstet Gynecol* 172: 1158–67.

Confidential Enquiry into Maternal and Child Health. 2004. *Why mothers die 2000–2002: The sixth report of the confidential enquiries into maternal deaths in the United Kingdom*. London: RCOG Press.

Confidential Enquiry into Maternal and Child Health. 2007. *Saving mothers' lives: Reviewing maternal deaths to make motherhood safer, 2003–2005*. The seventh report of the confidential enquiries into maternal deaths in the United Kingdom: Reviewing maternal deaths to make motherhood safer, 2003–2005. London: ROG Press.

Dannenberg, A.L., Carter, D.M., Lawson, H.W., Ashton, D.M., Dorfman, S.F., and Graham, E.H. 1995. Homicide and other injuries as causes of maternal death in New York City, 1987 through 1991. *Am J Obstet Gynecol* 172: 1557–64.

Deneux-Tharaux, C., Berg, C., Bouvier-Colle, M.H., Gissler, M., Harper, M., Nannini, A., et al. 2005. Underreporting of pregnancy-related mortality in the United States and Europe. *Obstet Gynecol* 106: 684–92.

Ford, J.B., Roberts, C.L., Simpson, J.M., Vaughan, J., and Cameron, C.A. 2007. Increased postpartum hemorrhage rates in Australia. *Int J Gynaecol Obstet* 98: 237–43.

Gilbert, W.M. and Daneilsen, B. 1999. Amniotic fluid embolism: Decreased mortality in a population-based study. *Obstet Gynecol* 93: 973–77.

Goldberg, R.P., Abramov, Y., Botros, S., Miller, J.J., Gandhi, S., Nickolov, A., et al. 2005. Delivery mode is a major environmental determinant of stress urinary incontinence:

Results of the Evanston-Northwestern Twin Sisters Study. *Am J Obstet Gynecol* 193: 2149–53.

Gray, C. 1998. Long before Krever's report, blood scare had changed face of medicine. *CMAJ* 158: 89–91.

Hannah, M.E., Hannah, W.J., Hodnett, E.D., Chalmers, B., Kung, R., Willan, A., et al. 2002. Outcomes at 3 months after planned cesarean vs planned vaginal delivery for breech presentation at term: The international randomized Term Breech Trial. *JAMA* 287: 1822–31.

Haynes, K., Stone, C., and King, J. 2004. *Major morbidities associated with childbirth in Victoria: Obstetric haemorrhage and associated hysterectomy*. Melbourne: Public Health Group, Department of Human Services.

Health Canada. 2004. *Special report on maternal mortality and severe morbidity in Canada - enhanced surveillance: The path to prevention*. Ottawa: Minister of Public Works and Government Services Canada.

Hogan M.C., Foreman K.J., Naghavi M., Ahn S.Y., Wang M., Makela S.M, et al. Maternal mortality for 181 countries, 1980-2008: a systematic analysis of progress towards Millennium Development Goal 5. 2010. *Lancet* 375: 1609–23.

Högberg, U. 2004. The decline in maternal mortality in Sweden: The role of community midwifery. *Am J Public Health* 94: 1312–20.

Hoyert, D.L. 2007. Maternal mortality and related concepts. National Center for Health Statistics. *Vital Health Stat* 3(33): 1–13.

Joseph, K.S., Rouleau, J., Kramer, M.S., Young, D.C., Liston, R.M., Baskett, T.F., and the Maternal Health Study Group of the Canadian Perinatal Surveillance System. 2007. Investigation of an increase in postpartum haemorrhage in Canada. *BJOG* 114: 751–59.

Joseph, K.S., Liu, S., Wu, J., Rouleau, J., Kirby, R., Kramer, M.S., et al. Severe maternal morbidity in Canada, 2003 to 2007: Surveillance using routine hospitalization data and ICD10 codes. *J Obstet Gynaecol Can* 32: 837–46.

Khan, K.S., Wojdyla, S., Say, L., Gülmezoglu, M., and Van Look, P.F. 2006. WHO analysis of causes of maternal death: A systematic review. *Lancet* 367: 1066–74.

Knight, M., Callaghan, W., Berg, C., Alexander, S., Bouvier-Colle, M., Ford, J., et al. 2009. Trends in post-partum haemorrhage in high resource countries: A review and recommendations from the International Postpartum Haemorrhage Collaborative Group. *BMC Pregnancy Childbirth* 9: 55.

Kramer, M.S., Rouleau, J., Baskett, T.F., Joseph, K.S., and the Maternal Health Study Group of the Canadian Perinatal Surveillance System. 2006. Amniotic-fluid embolism and medical induction of labour: A retrospective, population-based cohort study. *Lancet* 368: 1444–48.

Lang, C.T. and King, J.C. 2008. Maternal mortality in the United States. *Best Pract Res Clin Obstet Gynaecol* 22: 517–31.

Loudon, I. 2000. Maternal mortality in the past and its relevance to developing countries today. *Am J Clin Nutr* 72: 241S–46S.

Lukacz, E.S., Lawrence, J.M., Contreras, R., Nager, C.W., and Luber, K.M. 2006. Parity, mode of delivery, and pelvic floor disorders. *Obstet Gynecol* 107: 1253–60.

MacArthur, C., Glazener, C.M., Wilson, P.D., Lancashire, R.J., Herbison, G.P., and Grant, A.M. 2006. Persistent urinary incontinence and delivery mode history: A six-year longitudinal study. *BJOG* 113: 218–24.

MacKay, A.P., Berg, C.J., Duran, C., Chang, J., and Rosenberg, H. 2005. An assessment of pregnancy-related mortality in the United States. *Paediatr Perinat Epidemiol* 19(3): 206–14.

McMahon, A.D. 2003. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf* 12: 551–58.

Miettinen, O.S. 1983. The need for randomization in the study of intended effects. *Stat Med* 2: 267–71.

Miettinen, O.S. 1985. The "case-control" study: Valid selection of subjects. *J Chronic Dis* 38: 543–48.

Pattinson, R., Say, L., Souza, J.P., Broek, N., Rooney, C., and the WHO Working Group on Maternal Mortality and Morbidity Classifications. 2009. WHO maternal death and near-miss classifications. *Bull World Health Organ* 87: 734.

Public Health Agency of Canada. 2008. *Canadian perinatal health report, 2008 edition*. Ottawa: Minister of Public Works and Government Services Canada.

Roberts, C.L., Cameron, C.A., Bell, J.C., Algert, C.S., and Morris, J.M. 2008. Measuring maternal morbidity in routinely collected health data: Development and validation of a maternal morbidity outcome indicator. *Med Care* 46: 786–94.

Roberts, C.L., Ford, J.B., Algert, C.S., Bell, J.C., Simpson, J.M., and Morris, J.M. 2009. Trends in adverse maternal outcomes during childbirth: A population-based study of severe maternal morbidity. *BMC Pregnancy Childbirth* 9: 7.

Say, L., Souza, J.P., Pattinson, R.C., and the WHO working group on Maternal Mortality and Morbidity Classifications. 2009. Maternal near miss: Towards a standard tool for monitoring quality of maternal health care. *Best Pract Res Clin Obstet Gynaecol* 23: 287–96.

Spiliopoulos, M., Puri, I., Jain, N.J., Kruse, L., Mastrogiannis, D., and Dandolu, V. 2009. Amniotic fluid embolism-risk factors, maternal and neonatal outcomes. *J Matern Fetal Neonatal Med* 22: 439–44.

Sultan, A.H., Kamm, M.A., Hudson, C.N., Thomas, J.M., and Bartram, CI. 1993. Anal-sphincter disruption during vaginal delivery. *N Engl J Med* 329: 1905–11.

Turner, L.A., Cyr, M., Kinch, R.A., Liston, R., Kramer, M.S., Fair, M., Heaman, M., for the Maternal Mortality and Morbidity Study Group of the Canadian Perinatal Surveillance System. 2002. Under-reporting of maternal mortality in Canada: a question of definition. *Chronic Dis Can* 23: 22–30.

United Nations. 2008. *The Millennium Development Goals Report 2008*. United Nations. New York.

Wen, S.W., Huang, L., Liston, R., Heaman, M., Baskett, T., Rusen, I.D., et al. for the Maternal Health Study Group, Canadian Perinatal Surveillance System. 2005. Severe maternal morbidity in Canada, 1991-2001. *CMAJ* 173: 759–64.

World Health Organization. 1993. *International statistical classification of diseases, and related health problems*, 10th revision (ICD-10). Geneva: World Health Organization.

World Health Organization. 2003. *Maternal mortality in 2000: Estimates developed by WHO, UNICEF and UNFPA*. Geneva: World Health Organization.

World Health Organization. 2004. *Beyond the numbers: Reviewing maternal deaths and disabilities to make pregnancy safer*. Geneva: World Health Organization.

World Health Organization. 2007. *Maternal mortality in 2005: Estimates developed by WHO, UNICEF, UNFPA, and the World Bank*. Geneva: World Health Organization.

World Health Organization. 2010. *Trends in maternal mortality: 1990 to 2008. Estimates developed by WHO, UNICEF, UNFPA and the World Bank*. Geneva: World Health Organization.

Zhang, W.H., Alexander, S., Bouvier-Colle, M.H., Macfarlane, A., and the MOMS-B Group. 2005. Incidence of severe pre-eclampsia, postpartum haemorrhage and sepsis as a surrogate marker for severe maternal morbidity in a European population-based study: The MOMS-B survey. *BJOG* 112: 89–96.

12

Fetal and Infant Mortality and Morbidity

K.S. JOSEPH

FETAL MORTALITY

The death of a fetus in utero and its subsequent delivery (stillbirth) constitute an unfortunate endpoint of pregnancy and represent an important outcome in perinatal research.

Definition

The *International Statistical Classification of Diseases, and Related Health Problems*, version 10 (ICD-10), defines a fetal death as "… death prior to the complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of pregnancy; the death is indicated by the fact that after such separation the fetus does not breathe or show any other evidence of life, such as beating of the heart, pulsation of the umbilical cord, or definitive movement of voluntary muscles" (World Health Organization [WHO] 1993). An important feature of this definition is the absence of birth weight and gestational age (GA) criteria in defining fetal death. However, from a clinical and public health perspective, there is particular interest in preventable fetal deaths—that is, deaths among viable fetuses. Toward this end, the ICD-10 recommends that perinatal mortality statistics focus on deaths of fetuses weighing at least 500 g or when birth weight is unavailable, after 22 completed weeks of gestation (in addition to early neonatal deaths). The fetal death or stillbirth rate is typically expressed per 1,000 total births (i.e., live births plus stillbirths).

In Canada, the definition of stillbirth includes all fetal deaths with a birth weight of 500 g or more or a GA of 20 weeks or more (Public Health Agency of Canada 2008). This GA criterion contrasts with that stipulated by ICD-10, which uses a 22-week criterion for defining fetal death when birth weight is not recorded. The interchangeable use of GA and birth weight criteria for defining births of perinatal interest is problematic because a birth weight of 500 g is the approximate median birth weight observed among live births at 22 weeks (Alexander 1996; Kramer 2001). Thus, a substantial fraction of in utero deaths at 22 weeks' gestation will

not be labeled as stillbirths if the 500 g birth weight criterion is used. In the United States, the National Center for Health Statistics defines a fetal death as an in utero death that occurs at 20 or more weeks of gestation, whereas a fetal death after 28 weeks is considered a late fetal death. In some European countries and elsewhere, registration of fetal deaths is restricted to those at a GA of 24 weeks or greater (e.g., Italy and the United Kingdom), whereas in other countries only fetal deaths at 28 or more weeks of gestation are registered (e.g., Sweden and Denmark) (Lack et al. 2003; Macfarlane et al. 2003). These differences in fetal death registration make international comparisons of crude fetal death rates less meaningful (Public Health Agency of Canada 2008).

On a related note, recent increases in the uptake of prenatal diagnosis and selective termination of pregnancies with major congenital anomalies have led to some confusion regarding fetal death registration following pregnancy termination at 22 or 23 weeks' gestation. Such deaths are required to be registered as fetal deaths under the ICD-10 criterion for GA, despite their iatrogenic origins. However, registration procedures vary among countries, and some are more definition-based while others are more pragmatic (Public Health Agency of Canada 2008). This further compromises international comparisons of crude stillbirth rates.

Subtypes

Fetal deaths may be categorized into early (20–27 weeks' gestation or <1,000 g birth weight) and late fetal deaths (≥28 weeks' gestation or ≥1,000 g birth weight). Early fetal deaths are considered relatively less preventable, whereas late fetal deaths are considered more preventable. Such distinctions are technology- and time-dependent, however.

Another important categorization of stillbirths is based on the timing of death in relation to labor, with those that occur prior to the onset of labor labeled *antepartum stillbirths* and those that occur during labor labeled *intrapartum stillbirths*. The majority of stillbirths (85%) occur in the antepartum period. The motivation for the categorization is the unique set of essentially preventable causes of death that operate to cause intrapartum stillbirth. These include birth trauma (e.g., due to obstructed labor, shoulder dystocia), cord accidents, and intrapartum asphyxia.

INFANT MORTALITY

Definition

The period of infancy extends from (live) birth to the end of the first year after birth (364 days), and death during this period is termed an *infant death*. The definition of a live birth is critical to defining an infant death. The ICD-10 defines a live birth as "… the complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of the pregnancy, which, after such separation, breathes or shows any other evidence of life, such as beating of the heart, pulsation of the umbilical cord, or definite movement of the voluntary

muscles, whether or not the umbilical cord has been cut or the placenta is attached; each product of such a birth is considered liveborn" (WHO 1993). Note the absence of either birth weight or GA criteria for defining a live birth.

Subtypes

Infancy is divided into periods, with the interval from birth to the end of the first week after birth (0–6 days) termed the early neonatal period, the period from the end of first week to the end of first month after birth (7–27 days) labeled the late neonatal period, and that from the end of first month to end of first year after birth (28–364 days) referred to as the postneonatal period (WHO 1993). Infant deaths can be categorized, depending on the period when death occurs, into neonatal deaths (early and late) and postneonatal deaths. Note that the count of days following birth, according to the ICD/WHO system, begins with day 0 (i.e., the day of birth). Thus, a period of 24 hours has to elapse after birth before day 1 arrives. Similarly, the first 7 days after birth are completed at the end of day 6, and the first year after birth is completed at the end of day 364. This is analogous to the ICD/WHO scheme for counting GA, with the first day of the last menstrual period deemed to be day 0 (WHO 1993). Some alternative schemes recommend that the day of birth be numbered day 1 (with infant death defined as death between 1 and 365 days) and a mixing up of the two schemes may lead to confusion.

Causes of neonatal death differ from causes of postneonatal death. Temporal proximity ensures that pregnancy factors (e.g., complications, such as preeclampsia) have a greater impact on neonatal mortality than on postneonatal mortality. Congenital anomalies and pregnancy-related causes are the leading causes of neonatal death in most industrialized countries, whereas congenital anomalies and sudden infant death syndrome (SIDS) are the leading cause of postneonatal death (Kung et al. 2008; Public Health Agency of Canada 2008). In less industrialized countries, infectious diseases (e.g., diarrhea and lower respiratory tract infection) are among the leading causes of postneonatal death. Postneonatal deaths predominate among infant deaths in less-industrialized countries, whereas most infant deaths in industrialized countries occur in the neonatal period. For instance, in 2000, when the infant mortality rate in the least-industrialized countries was 102 per 1,000 live births, the postneonatal mortality rate was approximately 59 per 1,000 live births. This was in contrast to industrialized countries, whose infant and postneonatal mortality rates in 2000 were 6 and 2 per 1,000 live births, respectively (UNICEF 2002). In the calculation of such age at death–specific mortality rates, live births constitute the denominator for all types of neonatal deaths. For the postneonatal death rate, the denominator options include all live births or infants who survive the neonatal period. Although the latter restriction to infants at risk for postneonatal death may be preferable from an epidemiologic standpoint, it is not uncommon to encounter the former traditional calculation in the literature.

International infant mortality rates

The infant mortality rate is often used as an indicator of general health development and health status of populations (e.g., for comparing countries or periods).

It is typically defined as the number of infant deaths in a calendar year divided by the number of live births in the same calendar year and expressed per 1,000 live births. UNICEF estimates that the infant mortality rate for the world in 2007 was 47 per 1,000 live births; in industrialized countries, the rate was 5 per 1,000 live births, whereas in developing countries and in the least-developed countries, it was 51 and 84 per 1,000 live births, respectively (UNICEF 2009). Infant mortality rate estimates for 2007 for selected countries include Afghanistan 165, Sierra Leone 155, India 54, China 19, United States 7, Canada 5, Sweden 3, Japan 3, and Singapore 2 per 1,000 live births (UNICEF 2009).

The Millennium Development Goals, adopted by the United Nations in 2000, included a commitment to reduce child mortality under 5 years by two-thirds by the year 2015 (from 93 per 1,000 live births in 1990). This implies a reduction in under-5 mortality from 10 in 1990 to approximately 4 per 1,000 live births in 2015 in industrialized countries, and from 103 in 1990 to 34 per 1,000 live births in 2015 in less-industrialized countries. In 2007, the rate of under-5 mortality was 6 per 1,000 live births in industrialized countries, 74 per 1,000 live births in less-industrialized countries, and 68 per 1,000 live births worldwide (United Nations 2009). Approximately 37% of global under-5 mortality occurs in the neonatal period.

Period and cohort infant mortality rates

The infant mortality rate as defined above (period infant mortality) is a misnomer, and this index is in fact a ratio. Infant deaths in a given calendar year could have occurred among live births during that or the previous calendar year (e.g., live birth occurs in December of the previous year and the 2-month-old infant dies in February of the year of interest). Similarly, not all live births included in the denominator may result in an infant death in that calendar year depending upon month of birth. Alternatively, the infant mortality rate can be formulated as the experience of a birth cohort. For this calculation, a birth cohort (defined by the event of birth in a particular calendar year) is followed forward for 1 year, and the number of infant deaths is documented.

These two formulations of the infant mortality rate serve different functions. The ratio measure (i.e., the period infant mortality rate) is a simple index that can be obtained in a timely manner by collating birth and (infant) death registrations at the end of each calendar year. The infant mortality rate based upon the experience of a birth cohort (for a given calendar year) takes more effort to estimate. Also, since all live births have to be followed for 1 year to document their survival status, this index only becomes available a year later, at the earliest. The advantage of the cohort-type infant mortality rate is that it allows calculation of indices such as birth weight– and GA-specific mortality (not possible with the ratio measure, given the disconnection between the numerator and the denominator). In countries such as Canada and the United States, the cohort-type infant mortality rate is routinely estimated through a linkage of all live birth and infant death registrations. In addition to this linked live birth–infant death cohort file, the United States also creates a linked *period infant mortality file*. This contains all

live births and infant deaths in a given calendar year, with infants deaths linked to live birth registrations in the year of interest and in the previous year. This somewhat unconventional construct enables an examination of birth weight–specific, GA-specific and other analyses of infant mortality using the period file, which is available well before the birth cohort follow-up is completed.

PERINATAL MORTALITY RATE

The ICD-10 defines the perinatal mortality rate as "… the number of deaths of fetuses weighing at least 500 g (or, when birth weight is unavailable, after 22 completed weeks of gestation or a crown–heel length of 25 cm or more), plus the number of early neonatal deaths, per 1,000 total births" (WHO 1993).

Rationale for combining fetal deaths and early neonatal deaths

There are significant physiologic differences between fetuses and infants. Since the fetus receives oxygen via the placenta, the transition after birth includes an expansion of the lungs. The fetal circulatory system also undergoes substantial remodeling after birth, including a closure of the two shunts that transfer blood in utero from the right to the left sides of the circulation (Taeusch et al. 2005). Despite these and other differences between fetuses and infants, there is substantial overlap in causes of death. If the typical perinatal death is viewed as being the terminal event of a serious pregnancy complication, then the distinction between fetal and early neonatal death depends on whether spontaneous labor or obstetric intervention (e.g., cesarean delivery) occurred relatively early or late in the terminal process. However, some experts insist that causes of fetal and neonatal mortality differ and warrant separate classification schemes for causes of death (Hey et al. 1986), and the debate related to the rationale for combining fetal and neonatal deaths persists (Kramer et al. 2002).

Obstetric definition of perinatal mortality

Obstetricians define perinatal deaths as stillbirths plus neonatal deaths (Cunningham et al. 2005), based on the premise that early and late neonatal deaths typically share a common etiology with stillbirths. Perinatal death, thus defined, is routinely used as part of the primary outcome in many randomized clinical trials of therapeutic efficacy in obstetric research (e.g., Hannah et al. 2000).

Relation between fetal and infant deaths

The availability and utilization of prenatal diagnostic services and selective termination of pregnancies with serious congenital malformations demonstrates the closely related and sometimes reciprocal nature of the relation between population fetal and infant death rates (Liu et al. 2002; Public Health Agency of Canada 2008). Prenatal diagnosis and selective termination of affected

pregnancies has resulted in an increase in congenital anomaly–related stillbirths and neonatal deaths in the less than 500 g birth weight category (pregnancy termination of serious congenital anomalies may occasionally result in a live birth at very early gestation). On the other hand, it has led to substantial reductions in neonatal deaths due to congenital anomalies among live births of 500 g or more birth weight (Fig. 12.1). Widespread availability and uptake of prenatal diagnosis in some countries, such as Canada, has resulted in congenital anomaly–related infant death falling from the preeminent position as a cause of infant death to second place behind immaturity and related causes (Public Health Agency of Canada 2008).

TEMPORAL AND SPATIAL COMPARISONS OF FETAL AND INFANT MORTALITY

Figure 12.2 shows the rates of fetal and infant mortality in Canada from 1921 to 2007 (Statistics Canada 1993, 2008). The decline in the rates of late fetal death has been impressive and has continued in recent years. On the other hand, stillbirths of 20 weeks gestation or longer, only recorded since 1985 in Canada, did not show a similar decline. This is because the frequency of stillbirths with a birth weight of less than 500 g (and a gestation of about 20–23 weeks) have increased in recent years (see below). Infant mortality rates and neonatal mortality rates also declined exponentially, but the decline appears to have reached a plateau in the most recent decade for reasons that are similar to those responsible for the plateau in rates of stillbirth of 20 weeks or more (see below). On the other hand, postneonatal deaths, which are generally unaffected by births at the borderline of viability, have continued their exponential decline. It is encouraging that extrapolation of the postneonatal mortality rate suggests that it may fall below 1 per 1,000 live births in Canada in the near future.

In recent years, both Canada (Joseph and Kramer 1996) and the United States (MacDorman et al. 2005) have experienced temporal increases in infant mortality (i.e., the infant mortality rate in a given year was higher than that in the previous year). When the first such change occurred in Canada in 1993 (the first such increase in over three decades), there was considerable speculation in the lay media as to potential causes (with an emphasis on environmental factors) (Mitchell 1995). It has since become evident that these changes in infant mortality rates are a product of opposing forces, including improvements in infant health and health care services and increases in the registration of live births at the borderline of viability (e.g., those of <500 g birth weight) (Joseph and Kramer 1996; Public Health Agency of Canada 2008). The reasons for the increase in the birth registration of live births of less than 500 g include more complete registration, given better survival of such babies, changes in social attitudes, and an increasing availability and uptake of prenatal diagnosis and pregnancy termination for serious congenital anomalies. The increases in live births with a birth weight of less than 500 g do not appear to be the product of deteriorating maternal health, since rates of low birth weight (LBW) have not increased simultaneously. The LBW (< 2,500 g)

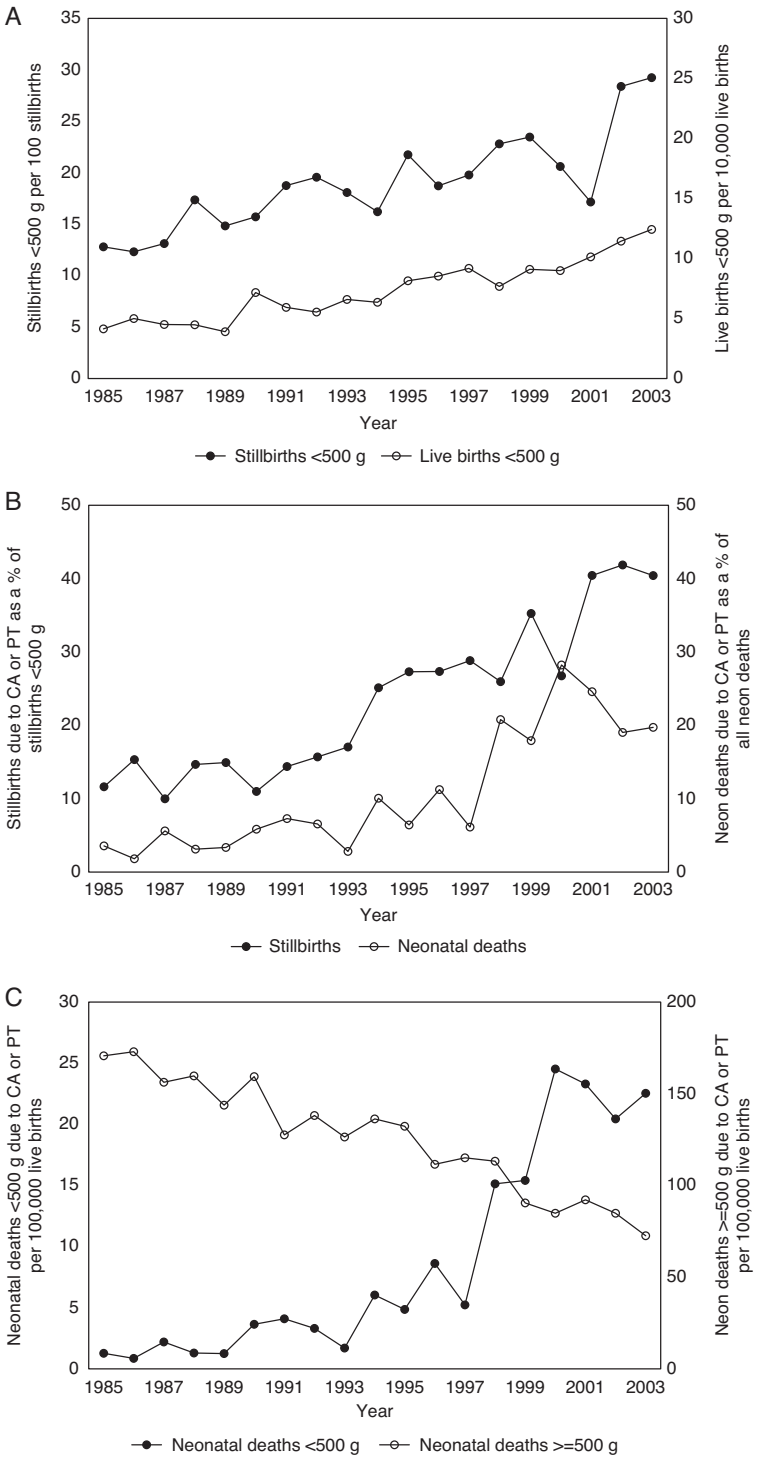


Figure 12.1 Changes in live births and stillbirths of <500 g and in congenital anomaly–related fetal and neonatal deaths, Canada (excluding Ontario, Newfoundland, and Labrador) 1985 to 2003. **A**: Stillbirths of <500 g per 100 stillbirths (primary Y-axis) and live births

Figure 12.1 (*continued*)
<500 g per 10,000 live births (secondary Y-axis). **B**: Stillbirths of <500 g due to congenital anomalies (CA) or pregnancy termination (PT) as a percent of stillbirths <500 g (primary Y-axis) and neonatal deaths <500 g due to CA or PT as a percent of neonatal deaths <500 g (secondary Y-axis). **C**: Neonatal deaths (ND) <500 g due to CA or PT per 100,000 live births (primary Y-axis) and neonatal deaths ≥500 g due to CA or PT per 100,000 live births (secondary Y-axis). Adapted from the Canadian Perinatal Health Report 2008, with permission.

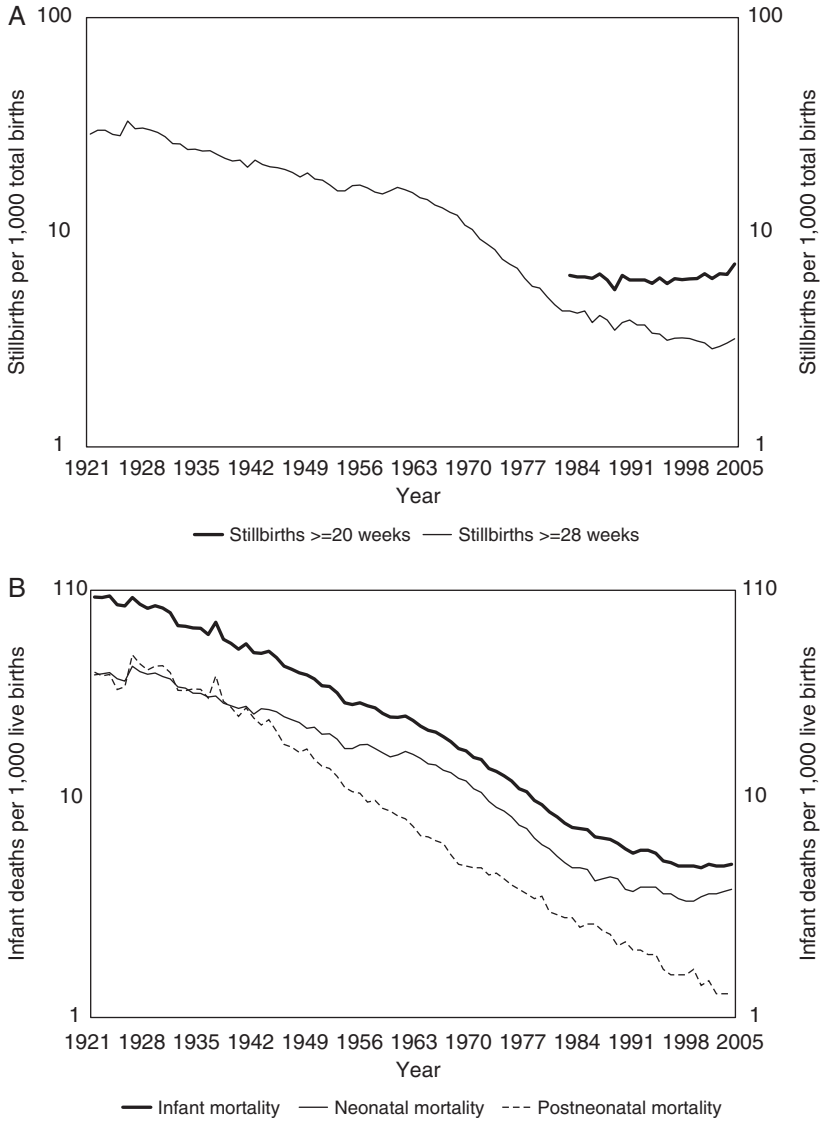


Figure 12.2 Fetal and infant death rates in Canada, 1921–2007. Fetal death rates among births of ≥20 weeks' gestation and among births of ≥28 weeks' gestation (**A**) and infant, neonatal, and postneonatal death rates (**B**), Canada, 1921–2007.

rate in Canada was essentially unchanged (5.7% in 1995 and 5.9% in 2004), whereas the rate of live births of less than 500 g increased by over 50% over the same period (from 8.2 to 12.4 per 10,000 live births) (Public Health Agency of Canada 2008).

The problems caused by temporal changes in birth registration (better counting) related to births at the borderline of viability has a counterpart in spatial contrasts of infant mortality. Many organizations, such as UNICEF and the Organization for Economic Cooperation and Development, rank countries on the basis of infant mortality rates (or related indices like the under-5 years mortality rate). This becomes a source of nationalistic pride, and not infrequently, serves as the basis of political rhetoric. In fact, large international variations in the birth registration of stillbirths and live births at the borderline of viability mean that such international ranking are generally meaningless, at least for industrialized countries. As mentioned previously, in connection with fetal death, birth registration in many countries is pragmatic rather than definition-based, and definitions also vary widely (Lack et al. 2003; Macfarlane et al. 2003; Public Health Agency of Canada 2008;). Whereas the ICD-9 and ICD-10 definitions of live birth do not include any birth weight or GA criteria, some European countries exclude live births of less than 500 g from their birth registers. For instance, birth registration in Finland is limited to live births with a birth weight of 500 g or more and a GA of 22 weeks or greater, whereas Sweden requires a minimum GA of 28 weeks or more (Lack et al. 2003; Macfarlane et al. 2003).

Numerous studies (Howell and Blondel 1994; Sepkowitz 1995; Sachs et al. 1995; Kramer et al. 2002) have shown large differences in the proportion of live births in different countries according to birth weight category. In one study, the birth proportions for live births with a birth weight of less than 500 g were: Sweden 0.6, Canada 4.5, and United States whites 9.1 per 10,000 live births; for live births with a birth weight of 500–749 g: Sweden 7.5, Canada 13.4, and United States whites 16.2 per 10,000 live births; and for live births with a birth weight 750–999 g: Sweden 14.6, Canada 17.4, and United States whites 18.9 per 10,000 live births (Kramer et al. 2002). Such differences were not observed in other LBW categories (e.g., the birth proportion for live births with a birth weight between 1,000–1,499 g were Sweden 49.6, Canada 48.9, and United States whites 49.3 per 10,000 live births [Kramer et al. 2002]), suggesting that differences in birth registration (and not differences in maternal health status) were responsible for the observed differences in the frequency of live births at the borderline of viability.

The World Health Organization (WHO) recommends that international infant mortality comparisons be restricted to live births with a birth weight of 1,000 g or more, given the limitations of crude comparisons of infant mortality (WHO 1993). This is reasonable from an epidemiologic standpoint (i.e., restriction as a method of dealing with spatial or temporal confounding by variable birth registration at the borderline of viability), although the approach is not without problems. For instance, such birth weight–specific infant mortality statistics are not widely available even for many industrialized countries. Moreover, in many industrialized countries, like Canada, over 40% of infant deaths occur among infants with a birth weight of less than 1,000 g. International comparisons of infant mortality

among live births of 1,000 g or more will thus exclude assessments of care services for this particularly high-risk category of infants.

BIRTH WEIGHT AND GESTATIONAL AGE OF STILLBIRTHS

A recognized inaccuracy in the measurement of birth weight and GA with regard to stillbirths arises because of the difference between the time of fetal death and the time of delivery (stillbirth). Postmortem changes in the fetus can diminish or increase weight (e.g., due to fluid loss or retention). Gestational age at delivery will always be greater than or equal to GA at death, and records in perinatal databases may routinely reflect time of delivery rather than time of death. This latter overestimation may be less of an issue in recent years, given improved access to medical care. Elective labor induction is used to affect delivery given fetal demise, if labor does not begin spontaneously; studies suggest that approximately 40%–60% of women choose induction within days of fetal death (Stringham et al. 1982; Kellner et al. 1984).

CAUSES OF DEATH

There have been numerous attempts at developing classification schemes for categorizing the causes of fetal and infant death. For instance, a simple clinical classification scheme for fetal death (Cunningham et al. 2005) categorizes causes into those that are fetal (e.g., chromosomal anomalies, other birth defects, infections), placental (e.g., premature placental separation, placental insufficiency, cord accident, placenta previa, chorioamnionitis), maternal (e.g., diabetes, hypertensive disorders, trauma, abnormal labor, uterine rupture, sepsis), and unexplained. Given that multiple causes are typically associated with death, hierarchical rules have been proposed to identify one underlying cause of death (e.g., a lethal congenital anomaly takes precedence over other causes). The cardinal rule of such schemes is to identify, if possible, an underlying preventable cause (Cole et al. 1986). Various classification schemes have been proposed for this purpose (Wigglesworth 1980; Fretts and Usher 1997). Accurate classification of the cause of perinatal death depends on where the death occurred and whether an autopsy was performed (Goldenberg et al. 2004). Death certificates often provide inaccurate assessments of the underlying cause of death. For instance, studies have shown that only 60% of congenital malformations identified by a routine cause of fetal death investigation program were identified correctly in death certificates (Greb et al. 1987). However, causes of death from death certificates may be useful for identifying temporal trends in causes of fetal or infant death. A recently proposed classification scheme categorizes causes of death into ten main levels (i.e., fetal, neonatal, maternal, congenital anomalies, intrapartum, cord, placenta, infection, termination of pregnancy, and unknown) and promises to better organize information for health care, policy, and research purposes (Frøen et al. 2009).

SERIOUS FETAL AND INFANT MORBIDITY

Fetal and infant outcomes of interest include severe morbidity that requires extended hospitalization and may result in disability. The focus on serious morbidity is particularly relevant because, in recent decades, rates of perinatal death have declined substantially, both in industrialized and less-industrialized countries.

Diseases of preterm gestation

Preterm birth (birth prior to 37 completed weeks of gestation), especially very preterm birth (birth at <32 weeks), is strongly associated with serious morbidity of several different types. Such morbidity can affect the respiratory system (e.g., respiratory distress syndrome [RDS], which can lead to chronic lung disease), the gastrointestinal system (e.g., necrotizing enterocolitis [NEC], involving damage and perforation of the colon), the central nervous system (e.g., intraventricular hemorrhage and periventricular leukomalacia [PVL], two types of neurologic injury that can lead to cerebral palsy, mental retardation, hearing loss, and other neurodevelopmental delays), and the eyes (e.g., retinopathy of prematurity [ROP] leading to blindness) (Taeusch et al. 2005).

Respiratory distress syndrome occurs due to an insufficiency of pulmonary surfactant in the lungs of preterm infants. Endogenous surfactant, a mixture of several compounds, including lipids and proteins, first appears in the lung surfaces at about 23–24 weeks' gestation, with optimal concentrations reached at term gestation. Respiratory distress syndrome is characterized by the onset of tachypnea, retractions, and grunting soon after birth, and severe disease requires assisted ventilation. Antenatal corticosteroid therapy for accelerating fetal lung maturation among women at risk of preterm birth and exogenous surfactant administered to the infant are prophylactic and therapeutic interventions, respectively, that prevent/treat RDS. These treatment modalities have been in routine clinical use since the early 1990s.

Necrotizing enterocolitis, which tends to occur in clusters against a background endemic rate, has an unclear etiology, with infection, intestinal ischemia, immune factors, and enteral feeding included among the hypothesized causes. Management of such cases includes discontinuation of enteral feeds, antibiotics, and surgery, if necessary. Necrotizing enterocolitis is associated with a significant mortality risk, and long-term complications include intestinal strictures and complications that sometimes follow surgical resection of the bowel.

Hemodynamic changes in the immature brain of the preterm infant that occur after birth, or in association with medical procedures or seizures, can lead to intraventricular hemorrhage, whose severity is graded on a scale from I to IV. Such bleeding can be asymptomatic, or present as gradual clinical deterioration, or catastrophically as stupor, coma, and seizures. Milder forms of intraventricular hemorrhage (grades I and II, in which there is no distention of cerebral ventricles with blood) tend to resolve without chronic disability, whereas the more serious forms (grades III and IV, in which hemorrhage leads to ventricular dilatation and

parenchymal infarction) are associated with high rates neurologic sequelae such as seizure disorders, cerebral palsy, blindness, and deafness (Taeusch et al. 2005). Periventricular leukomalacia, a form of injury to the white matter of the brain, is characterized by necrosis, cyst formation, and diffuse injury. A high proportion of affected children develop motor deficits and developmental disabilities, such as spastic diplegia and visual and auditory impairment (Taeusch et al. 2005).

Retinopathy of prematurity results from a neovascularization of the immature retinas of the preterm infant, which lack the ability to deal with hyperoxia following birth. The condition, previously called *retrolental fibroplasia*, was first recognized in the 1940s as being associated with the use of high concentrations of oxygen to treat preterm infants. Laser photocoagulation is the treatment used to arrest disease progression and involves ablation of the peripheral avascular retina, with central vision being preserved at the expense of peripheral vision.

Neonatal encephalopathy

This clinical entity, seen in term or near-term infants, is defined on the basis of abnormal consciousness, tone, reflexes, feeding, or respiration, and seizures (ACOG and AAP 2003). Among infants with a cluster of perinatal events including a low 5-minute Apgar score, abnormal signs, and seizures, follow-up studies show that 55% of infants develop chronic motor disability and 70% suffer death or disability in childhood (Ellenberg and Nelson 1988). However, most infants with mild or moderate neonatal encephalopathy develop normally. A majority of cases of neonatal encephalopathy arise secondarily to causes that operate before the onset of labor, including growth restriction, genetic disorders, infection, placental bleeding or other abnormalities, preeclampsia, and other factors (ACOG and AAP 2003). Only a small fraction of such cases occur secondarily to intrapartum hypoxia. This understanding, which has emerged in recent years, has been important in clarifying the role of obstetric care in cases of neonatal encephalopathy. In previous years, the term *hypoxic-ischemic encephalopathy* was used to describe essentially the same clinical syndrome (variously referred to as postasphyxial encephalopathy, *birth asphyxia*, or *perinatal asphyxia*), with the cause ascribed entirely to asphyxia during labor or delivery (ACOG and AAP 2003). The term hypoxic-ischemic encephalopathy is currently restricted to the small subset of neonatal encephalopathy that is suspected to have an intrapartum cause (ACOG and AAP 2003).

Cerebral palsy

Cerebral palsy (CP) refers to a heterogeneous group of central nervous system disorders that manifest aberrant control of movement or posture, typically begin in utero, and are not the result of recognized progressive disease. Diagnosis of CP is typically made at 18 to 24 months of age, when motor functioning can be assessed, although recent advances in brain imaging increasingly permit the visualization of the underlying neuropathologic lesions at a much earlier age. Cerebral palsy is commonly classified by the type of neurologic dysfunction (spastic, in

which the muscles exhibit excessive tone; dyskinetic, characterized by involuntary movements; ataxic, characterized by uncoordinated movements; and mixed), and the number and distribution of limbs involved (quadriplegia, diplegia, hemiplegia, or monoplegia). The major types and frequencies of cerebral palsy include spastic quadriplegia (20%, has a strong association with mental retardation and seizures), diplegia (30%, common in preterm and LBW infants), hemiplegia (30%), choreoathetoid types (15%), and mixed varieties. Studies have shown that the important antepartum antecedents of cerebral palsy include congenital malformations, birth weight of less than 2000 g, GA at birth of less than 32 weeks, and infection (Nelson and Ellenberg 1986; Nelson and Grether 1999; Nelson and Willoughby 2000).

The prevalence of CP is approximately 1.0–2.3 per 1,000 live births. Studies from some regions that have carefully documented population rates of CP suggest that its frequency has changed in complex ways (Himmelmann et al. 2005). Although CP shows little change in rates between the mid-1950s and more recently, its stable temporal trends mask important changes that have occurred within subgroups of affected term and preterm infants and within specific subtypes (e.g., declines in diplegia, increase in dyskinetic cerebral palsy). These patterns suggest that changes in the population prevalence of risk factors (such as maternal diabetes, multifetal pregnancy) and, possibly, obstetric and neonatal interventions, may be partly responsible for the complex temporal patterns (Himmelmann et al. 2005).

Fetal morbidity

Morbidity in the fetus has long been recognized. For example, in the era before Rh immune globulin was available, Rh sensitization in the mother commonly led to hemolytic anemia in the fetus, and evidence of severe anemia was noted radiographically as signs of fetal congestive heart failure (e.g., halo due to dependent edema). More recently, improved imaging techniques have allowed the visualization of various types of fetal morbidity. Fetal stroke, for example, has been increasingly documented in utero using ultrasound and magnetic resonance imaging (Ozduman et al. 2004).

RECENT INTERVENTIONS ASSOCIATED WITH IMPROVED FETAL AND INFANT OUTCOMES

Perinatal outcomes have improved substantially over the last several decades. The interventions that have enabled this include advances in neonatal care that have led to better care of the preterm infant in the nursery. The improved ability to care for preterm infants has permitted obstetricians to deliver fetuses in compromised intrauterine environments at earlier GAs. Obstetric advances include antenatal corticosteroid therapy to accelerate fetal lung maturity, improved fetal monitoring, and increased use of cesarean delivery for preterm delivery at less than 28 weeks' gestation, whereas advances in neonatal care include the use of exogenous

surfactant for improving lung function, indomethacin to prevent or treat patent ductus arteriosus, assisted ventilation, and parenteral nutrition, among others. Use of some therapies has been reconsidered and sharply reduced, including use of postnatal steroid therapy (used to prevent or treat chronic lung disease, but found to be associated with an increase in CP) and endotracheal assisted ventilation (reduced due to concerns regarding barotrauma leading to chronic lung disease) (Hack et al. 2008).

The effects of these interventions can be appreciated based on studies that have documented temporal changes in intervention rates and outcomes among infants with a birth weight of between 500 and 999 g (Wilson-Costello et al. 2005, 2007). Although these studies document the experience of a tertiary perinatal care center in Cleveland, Ohio, they likely represent a fair approximation of the population experience, given its status as the referral center for high-risk pregnancies in the region. As Figure 12.3 shows, rates of antenatal corticosteroid use, cesarean delivery, and surfactant use increased dramatically from the 1980s to the 1990s, and increased further in the early years of the 21st century (except for assisted ventilation rates, which declined between the 1990s and 2000–2002) (Wilson-Costello et al. 2007). At the same time, rates of death declined from 54% to 43% to 31%, while rates of survival with impairment increased from 13% to 23% and then declined to 16% (Wilson-Costello et al. 2005). This figure shows that, between the 1980s and the 1990s, deaths among infants with birth weights between 500 and 999 g declined, while survival with or without impairment(s) both increased. On the other hand, between 1990–1999 and 2000–2002, both death and survival with impairment decreased, while impairment-free survival increased.

CHOICE OF OUTCOME IN CLINICAL AND EPIDEMIOLOGIC STUDIES

The Term Breech Trial (Hannah et al. 2000), which attempted to determine the optimal delivery mode for breech presentation at term, randomized 2,088 women to planned cesarean delivery or to planned vaginal delivery. The first paper from the study focused on a composite outcome, namely, perinatal mortality or serious neonatal morbidity (e.g., birth trauma, seizures, Apgar score <4 at 5 minutes, etc.) and showed that planned cesarean was the preferred method of delivery (relative risk [RR] 0.33, 95% confidence interval [CI] 0.19–0.56). A later paper (Whyte et al. 2004) focused on the outcome of death or neurodevelopmental delay at 2 years of age and showed that planned cesarean delivery was not superior to planned vaginal delivery (RR 1.09, 95% CI 0.52–2.30). Which of these clinical entities represents the definitive outcome in clinical and epidemiologic studies? Should such studies use perinatal mortality/serious neonatal morbidity as the primary outcome, or should they use death or neurodevelopmental delay at 2 years of age?

These seemingly straightforward questions appears to have been responsible for some confusion in the contemporary literature. On a related note, investigators reporting on a recent randomized trial on aggressive versus conservative

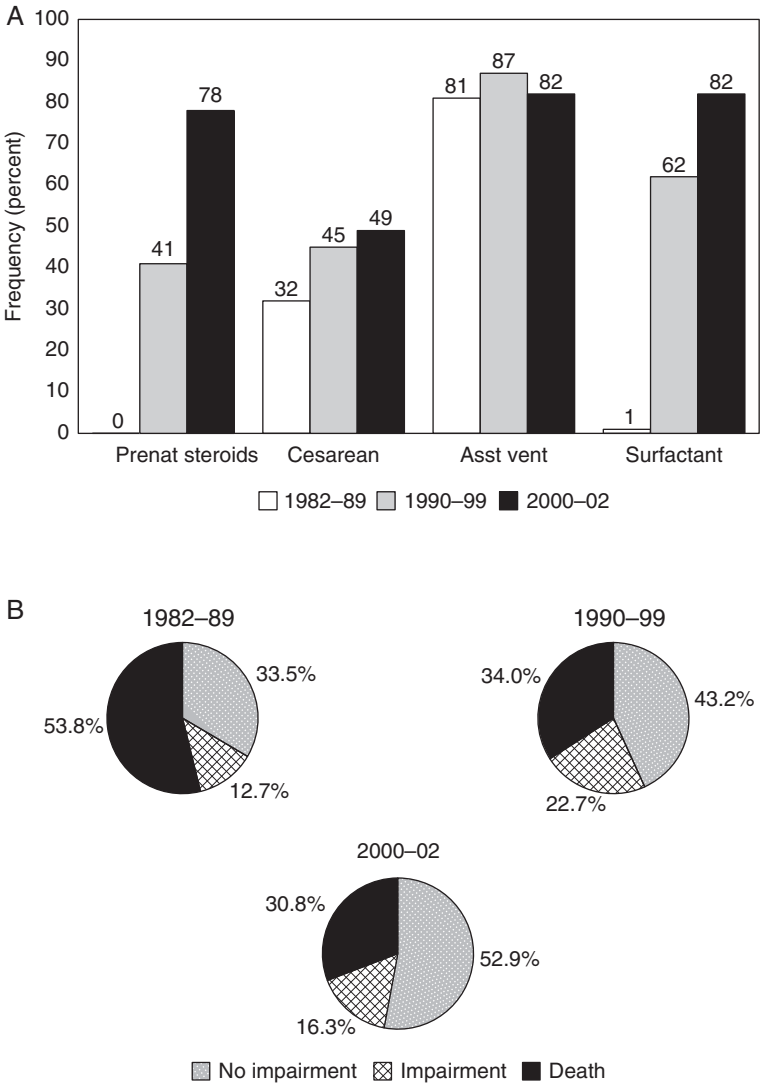


Figure 12.3 Temporal trends in obstetric and neonatal care interventions for infants 500–999 g birth weight in three epochs (1980–89, 1990–99, and 2000–2002, **A**) and changes in death rates and rates of survival with impairment and impairment-free survival (**B**). From Wilson-Costello, D., Friedman, H., Minich, N., Siner, B., Taylor, G., Schluchter, M., et al. 2007. Improved neurodevelopmental outcomes for extremely low birth weight infants in 2000–2002. *Pediatrics* 119: 37–45.

phototherapy for extremely LBW infants (Morris et al. 2008) provided post hoc analyses for the outcome "profound disability," in addition to their primary outcome, which was a composite of death or neurodevelopmental impairment at 18–22 months of corrected age. The two outcomes provided different results; aggressive phototherapy was not significantly different from conservative phototherapy

(RR 0.94, 95% CI 0.87–1.02) when death or neurodevelopmental impairment at 18–22 months was the outcome, whereas aggressive phototherapy reduced profound impairment (RR 0.68, 95% CI 0.52–0.89).

A similar situation was observed in a randomized trial of magnesium sulfate for the prevention of CP among women at imminent risk of preterm birth (Rouse et al. 2008). The composite primary outcome (stillbirth, infant death, or moderate or severe CP at or beyond 2 years of corrected age) was not significantly different in the magnesium sulfate and placebo groups (RR 0.97, 95% CI 0.77–1.23), but a prespecified secondary outcome (moderate or severe CP) occurred less frequently in the magnesium sulfate group (RR 0.55, 95% CI 0.32–0.95). The authors concluded that "… magnesium sulfate may reduce the chance that cerebral palsy will subsequently be diagnosed in a child who was at high risk for preterm birth" and that "….the rate of cerebral palsy was reduced among survivors" (Rouse et al. 2008).

Although the above-mentioned question regarding the definitive fetal/infant outcome is framed with the context of randomized trials, the issue also applies to nonexperimental epidemiologic studies. A related question arises as to the status and relative merit of other, much-studied outcomes, such as preterm birth at less than 37 weeks, preterm birth at less than 32 weeks, and growth-restriction (small for GA [SGA] <10th percentile or <3rd percentile). Are preterm birth and SGA definitive outcomes in perinatology, or do they constitute intermediate or proxy outcomes that may be studied for explanatory or efficiency purposes? These questions are best addressed by considering the characteristics of an ideal perinatal outcome in terms of issues related to relevance, validity, and feasibility.

Relevance

A prime consideration when choosing the definitive adverse outcome is the relevance of the outcome to the question under consideration. For questions that address therapeutic efficacy, it is generally recognized that death fulfils this criterion, with disability a close second. Although there may be a temptation to consider specific disease entities, these may or may not be relevant, given mild, moderate, and severe variants of the disease and a less than perfect association with what matters most from a social standpoint (i.e., death and disability). On the other hand, if the issue under consideration pertains to the elucidation of disease mechanisms, an intermediate outcome, such as preterm birth or fetal growth, may become the primary outcome of interest.

Validity

A second equally critical criterion requires that the ideal outcome should allow for a valid assessment of therapeutic or other effects. In the context of a randomized trial, and in other contexts as well, it is not difficult to see how outcomes such as "profound disability" can distort effects (because they require survival up to a stage at which assessment of disability is possible). For instance, in studies involving very preterm infants, in which a substantial fraction can die before disability can be assessed, an intention-to-treat analysis (that respects the groups created by

the randomization) requires that the definitive outcome be a composite that includes both death and disability.

Feasibility versus relevance and validity

It is not uncommon for feasibility issues to lead to a compromise that balances competing priorities. For instance, in the multicenter Term Breech Trial (Hannah et al. 2000), which recruited 2,088 subjects from 121 centers in 26 countries, follow-up for 2 years was only planned (where feasible) for 1,159 children in 85 centers in 18 countries. Of the 1,159 children being followed-up for the 2-year outcome, information could only be obtained on 920 children (Whyte et al. 2004). Thus, feasibility issues dictated that the primary outcome in that study was perinatal death or serious neonatal morbidity (99.5% of randomized subjects were included in the intention-to-treat analysis). For this reason, it is generally accepted that the Term Breech Trial showed planned cesarean delivery to be the preferred management option for breech presentation at term (since rates of perinatal mortality/serious neonatal morbidity were significantly lower with planned cesarean delivery in comparison with vaginal delivery). However, feasibility is place- and time-dependent; several recent studies have achieved 2-year follow-up rates that are consistent with the best state-of-the-art randomized trials (Morris et al. 2008; Rouse et al. 2008).

Feasibility, or related issues such as size or cost efficiency, may require that the ideal outcome be replaced with an intermediate outcome that is highly correlated with the ideal definitive outcome. For example, preterm birth at less than 37 weeks, preterm birth at less than 32 weeks, or SGA (<10th percentile or <3rd percentile) may constitute such proxy outcomes that are closely correlated with death or neurodevelopmental delay. This general principle can be illustrated by the effect of older maternal age, which increases the risk of adverse perinatal outcomes such as preterm birth, SGA live birth, perinatal mortality, and serious neonatal morbidity by approximately the same magnitude (Joseph et al. 2005). Initial studies with proxy outcomes can be followed by other more definitive studies with the ideal outcomes, if researchers deem the preliminary results sufficiently encouraging. An example of the use of a proxy intermediate outcome preceding the definitive evaluation was seen with studies evaluating progesterone for preventing preterm birth. The first few recent studies used preterm birth as the primary outcome to assess if progesterone could prevent preterm birth among women with a past history of preterm birth (da Fonseca et al. 2003; Meiss et al. 2003). The demonstrated efficacy of progesterone in these studies led to other studies that assessed whether progesterone had an effect on reducing perinatal mortality and serious neonatal morbidity among women at risk of recurrent preterm birth (Lim et al. 2007).

A NOTE ON SEMANTIC ISSUES

As with any science, there are miscellaneous terms peculiar to perinatology. For instance, the jargon of perinatal epidemiology makes a clear distinction between

the *fetus*, whose existence is entirely in utero, and the *infant* who arrives at birth. *Total births* (i.e., live births plus stillbirths) constitute the denominator for perinatal mortality, whereas *live births* make up the denominator for infant mortality. *All births* is a term sometimes used by demographers to refer to singleton and multiple births. The word "*premature*" is not defined under the ICD system, and its use is discouraged. Preferred terms include *preterm, small-for-gestational age*, and *low birth weight*. Traditionally, the rates associated with these latter terms are calculated among live births only. Another issue that sometimes leads to confusion relates to the terminological difference between the event experienced by the mother versus that experienced by the fetus/infant during the process of childbirth. Some perinatologists make the distinction by referring to the mother's experience of childbirth as a "delivery," whereas the fetus/infant's experience is termed a "birth." This provides clarity in some circumstances, including in multifetal pregnancies that end with one delivery and more than one birth.

One aspect of lay writing and even the scientific literature that sometimes seems incongruent to the perinatologist is the use of the term "the first year of life," which is sometimes used synonymously with "the first year after birth." The former term implies that life begins at birth, whereas the latter term merely specifies birth as the anchor for the chronologic age time scale. Temporal ambiguity between birth and the point at which life begins arises in the most diverse situations. For instance, the sentence, "Medicine is a secular and scientific profession that… must still contend with the sacred matters of *birth, life, and death*" (Anon 2005) seems to suggest a sequence that may appear disordered to the perinatal epidemiologist (whose routine focus extends from fetal *life, to birth, to infancy,* and beyond). This issue has serious legal and moral dimensions and is raised here merely to highlight semantic and terminological issues and to suggest a literal approach to defining chronologic age (e.g., first year after birth).

REFERENCES

Anon. 2005. The sacred and the secular: The life and death of Terri Schiavo. *CMAJ* 172: 1149.

ACOG and AAP. 2003. *Neonatal encephalopathy and cerebral palsy: Defining the pathogenesis and pathophysiology*. Washington, DC: The American College of Obstetricians and Gynecologists and the American Academy of Pediatrics.

Alexander, G.R., Himes, J.H., Kaufman, R.B., Mor, J., and Kogan, M. 1996. A United States national reference for fetal growth. *Obstet Gynecol* 87: 163–68.

Cole, S.K., Hey, E.N., and Thomson, A.M. 1986. Classifying perinatal death: An obstetric approach. *Br J Obstet Gynaecol* 93: 1204–12.

Cunningham, F.G., Leveno, K.J., Bloom, S.L., Hauth, J.C., Gilstrap, L., III, and Wenstrom, K.D., 2005. eds. *Williams obstetrics,* 22nd ed. Toronto: McGraw-Hill.

da Fonseca, E.B., Bittar, R.E., Carvalho, M.H., and Zugaib, M. 2003. Prophylactic administration of progesterone by vaginal suppository to reduce the incidence of spontaneous preterm birth in women at increased risk: A randomized placebo-controlled double-blind study. *Am J Obstet Gynecol* 188: 419–24.

Ellenberg, J.H. and Nelson, K.B. 1988. Cluster of perinatal events identifying infants at high risk for death or disability. *J Pediatr* 113: 546–52.

Fretts, R.C. and Usher, R.H. 1997. Causes of fetal death in women of advanced maternal age. *Obstet Gynecol* 89: 40–45.

Frøen, J.F., Pinar, H., Flenady, V., Bahrin, S., Charles, A., Chauke, L., et al. 2009. Causes of Death and Associated Conditions (CODAC) - a utilitarian approach to the classification of perinatal deaths. *BM Pregnancy Childbirth* 9: 22.

Goldenberg, R.L., Kirby, R., and Culhane, J.F. 2004. Stillbirth: A review. *J Matern Fetal Neonatal Med*. 16: 79–94.

Greb, A.E., Pauli, R.M., and Kirby, R.S. 1987. Accuracy of fetal death reports: Comparison with data from an independent stillbirth assessment program. *Am J Public Health* 77: 1202–06.

Hack, M. and Costello, D.W. 2008. Trends in the rates of cerebral palsy associated with neonatal intensive care of preterm children. *Clin Obstet Gynecol* 51: 763–74.

Hannah, M.E., Hannah, W.J., Hewson, S.A., Hodnett, E.D., Saigal, S., and Willan, A.R. 2000. Planned caesarean section versus planned vaginal birth for breech presentation at term: A randomized multicentre trial. *Term Breech Trial Collaborative Group. Lancet* 356: 1375–83.

Hey, E.N., Lloyd, D.J., and Wigglesworth, J.S. 1986. Classifying perinatal death: Fetal and neonatal factors. *Br J Obstet Gynaecol* 93: 1213–23.

Himmelmann, K., Hagberg, G., Becklung, E., Hagberg, B., and Uvebrant, P. 2005. The changing panorama of cerebral palsy in Sweden. IX. Prevalence and origin in the birth–year period 1995–1998. *Acta Paediatrica* 94: 287–94.

Howell, E.M. and Blondel. B. 1994. International infant mortality rates: Bias from reporting differences. *Am J Pub Health* 84: 850–52.

Joseph, K.S. and Kramer, M.S. 1996. Recent trends in Canadian infant mortality rates: Effect of changes in the registration of liveborns weighing less than 500 g. *CMAJ* 155: 1047–52.

Joseph, K.S., Allen, A.C., Dodds, L., Turner, L.A., Scott, H., and Liston, R. 2005. The perinatal effects of delayed childbearing. *Obstet Gynecol* 105: 1410–18.

Kellner, K.R., Donnelly, W.H., and Gould, S.D. 1984. Parental behavior after perinatal death: Lack of predictive demographic and obstetric variables. *Obstet Gynecol* 63: 809–14.

Kramer, M.S., Platt, R.W., Wen, S.W., Joseph, K.S., Allen, A., Abrahamowicz, M., et al. 2001. A new and improved population-based reference for birth weight for gestational age. *Pediatrics* 108: e35.

Kramer, M.S., Platt, R.W., Yang, H., Haglund, B., Cnattingius, S., and Bergsjo, P. 2002. Registration artifacts in international comparisons of infant mortality. *Paediatr Perinat Epidemiol* 16: 16–22.

Kramer, M.S., Liu, S., Luo, Z., Yuan, H., Platt, R.W., and Joseph, K.S. 2002. Analysis of perinatal mortality and its components: Time for a change? *Am J Epidemiol* 156: 493–97.

Kung, H.C., Hoyert, D.L., Xu, J.Q., and Murphy, S.L. 2008. *Deaths: Final data for 2005*. National vital statistics reports, vol. 56 no. 10. Hyattsville, MD: National Center for Health Statistics.

Lack, N., Zeitlin, J., Krebs, L., Kunzel, W., and Alexander, S. 2003. Methodological difficulties in the comparison of indicators of perinatal health across Europe. *Eur J Obstet Gynecol Reprod Biol* 111 Suppl 1: S33–44.

Lim, A.C., Bloemenkamp, K.W., Boer, K., Duvekot, J.J., Erwich, J.J., Hasaart, T.H., et al. 2007. Progesterone for the prevention of preterm birth in women with multiple pregnancies: The AMPHIA trial. *BMC Pregnancy Childbirth* 7: 7.

Liu, S., Joseph, K.S., Kramer, M.S., Allen, A.C., Sauve, R., Rusen, I.D., and Wen, S.W. 2002. Relationship of prenatal diagnosis and pregnancy termination to overall infant mortality in Canada. *JAMA* 287: 1561–67.

MacDorman, M.F., Martin, J.A., Mathews, T.J., Hoyert, D.L., and Ventura, S.J. 2005. *Explaining the 2001-02 infant mortality increase: Data from the linked birth/infant death data set*. National vital statistics reports, vol. 53 no. 12. Hyattsville, Maryland: National Center for Health Statistics.

Macfarlane, A., Gissler, M., Bolumar, F., and Rasmussen, S. 2003. The availability of perinatal health indicators in Europe. *Eur J Obstet Gynecol Reprod Biol* 111(Suppl 1): S15–32.

Meis, P.J., Klebanoff, M., Thom, E., Dombrowski, M.P., Sibai, B., Moawad, A.H., et al. 2003. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. *N Engl J Med* 348: 2379–2385.

Mitchell, A. 1995. Rising deaths among infants stun scientists. *Globe and Mail (Toronto)* June 2: A4.

Morris, B.H., Oh, W., Tyson, J.E., Stevenson, D.K., Phelps, D.L., O'Shea, T.M., et al. 2008. Aggressive vs. *conservative phototherapy for infants with extremely low birth weight. N Engl J Med* 359: 1885–96.

Nelson, K.B. and Ellenberg, J.H. 1986. Antecedents of cerebral palsy: Multivariate analysis of risk. *N Engl J Med* 315: 81–86.

Nelson, K.B. and Grether, J.K. 1999. Causes of cerebral palsy. *Curr Opin Pediatr* 11: 487–91.

Nelson, K.B. and Willoughby, R.E. 2000. Infection, inflammation and the risk of cerebral palsy. Curr Opin Neurol 13: 133–39.

Ozduman, K., Pober, B.R., Barnes, P., Copel, J.A., Ogle, E.A., Duncan, C.C., and Ment, L.R. 2004. Fetal stroke. *Pediatr Neurol* 30: 151–62.

Public Health Agency of Canada. 2008. Canadian Perinatal Health Report, 2008 ed. Ottawa: Minister of Public Works and Government Services Canada.

Rouse, D.J., Hirtz, D.G., Thom, E., Varner, M.W., Spong, C.Y., Mercer, B.M., et al. 2008. A randomized, controlled trial of magnesium sulfate for the prevention of cerebral palsy. *N Engl J Med* 359: 895–905.

Sachs, B.P., Fretts, R.C., Gardner, R., Hellerstein, S., Wampler, N.S., and Wise, P.H. 1995. The impact of extreme prematurity and congenital anomalies on the interpretation of international comparisons of infant mortality. *Obstet Gynecol* 85: 941–46.

Sepkowitz, S. 1995. International rankings of infant mortality and the United States' vital statistics natality data collection system—failure and success. *Int J Epidemiol* 24: 583–88.

Statistics Canada. 1993. *Selected infant mortality and related statistics, Canada 1921-1990*. Catalogue 82–549. Ottawa: Statistics Canada.

Statistics Canada. 2008. *Summary tables: Infant mortality*. *Ottawa*: *Statistics Canada*. Online at: http: //www40.statcan.gc.ca/l01/cst01/health21a-eng.htm

Stringham, J.G., Riley, J.H., and Ross, A. 1982. Silent birth: Mourning a stillborn baby. *Soc Work* 27: 322–27.

Taeusch, H.W., Ballard, R.A., and Gleason, C.A., eds. 2005. *Avery's diseases of the newborn*, 8th ed. Philadelphia: Elsevier Saunders.

UNICEF. 2002. *The state of the world's children 2002: Leadership*. New York: The United Nations Children's Fund (UNICEF).

UNICEF. 2009. *The state of the world's children 2008: Maternal and newborn health*. New York: The United Nations Children's Fund (UNICEF).

United Nations. 2008. *The millennium development goals report 2008*. *New York*: *United Nations*. Online at: www.un.org/millenniumgoals/pdf/The%20Millennium%20Development%20Goals%20Report%202008.pdf">http://www.un.org/millenniumgoals/pdf/The%20Millennium%20Development%20Goals%20Report%202008.pdf

Whyte, H., Hannah, M.E., Saigal, S., Hannah, W.J., Hewson, S., Amankwah, K., et al. 2004. Outcomes of children at 2 years after planned cesarean birth versus planned vaginal birth for breech presentation at term: The International Randomized Term Breech Trial. *Am J Obstet Gynecol* 191: 864–71.

Wigglesworth, J.S. 1980. Monitoring perinatal mortality: A pathophysiological approach. *Lancet* 2(8196): 684–86.

Wilson-Costello, D., Friedman, H., Minich, N., Fanaroff, A.A., and Hack, M. 2005. Improved survival rates with increased neurodevelopmental disability for extremely low birth weight infants in the 1990s. *Pediatrics* 115: 997–1003.

Wilson-Costello, D., Friedman, H., Minich, N., Siner, B., Taylor, G., Schluchter, M., et al. 2007. Improved neurodevelopmental outcomes for extremely low birth weight infants in 2000–2002. *Pediatrics* 119: 37–45.

World Health Organization. 1993. *International statistical classification of diseases, and related health problems,* 10th revision (ICD-10). Geneva: World Health Organization.

13

The Fetuses-at-Risk Approach: Causal and Noncausal Models

K.S. JOSEPH

TRADITIONAL CONCEPTS IN PERINATOLOGY

Many traditional concepts in perinatology are intuitive. For instance, the concept of birth weight–specific declines in perinatal mortality (i.e., the strong inverse relationship between birth weight and fetal and neonatal death, Fig. 13.1) has been well understood for centuries. Similarly, the exponential decline in perinatal mortality that is observed with increasing gestational age (GA) has long been self-evident. Birth weight– and GA-specific perinatal mortality calculations and patterns are, thus, central concepts in traditional perinatal epidemiology and also in obstetrics and neonatology.

Birth weight–specific perinatal mortality rates

The calculation of birth weight–specific perinatal mortality rates is illustrated here using data from the United States. There were 1,551 stillbirths, 141,599 live births, and 962 neonatal deaths among singletons in the United States with a birth weight between 2,000 and 2,499 g in 2004. The birth weight–specific perinatal mortality rate is calculated as follows:

$$\text{Birth weight–specific perinatal death rate} = \frac{\text{Perinatal deaths in birth weight category}}{\text{Total births in birth weight category}} \times 1000$$

$$\text{Perinatal mortality rate at 2,000–2,499 g} = \frac{(1{,}551 + 962)}{(1{,}551 + 141{,}599)} \times 1{,}000 = 17.6 \text{ per } 1{,}000 \text{ total births}$$

Birth weight–specific stillbirth rates are calculated similarly, with total births in the denominator, whereas neonatal mortality rates use live births as the denominator.

In the above instance, the stillbirth rate among those 2,000–2,499 g was 10.8 per 1,000 total births and the neonatal death rate was 6.8 per 1,000 live births.

Gestational age–specific perinatal mortality rates

In the United States, 567 stillbirths, 13,122 live births, and 282 neonatal deaths occurred among singletons born at 32 weeks' gestation in 2004. The GA-specific stillbirth rate, neonatal mortality rate, and perinatal mortality rates have the same form as the corresponding birth weight–specific mortality rates. The GA-specific perinatal mortality rate is calculated as follows:

$$\text{GA-specific perinatal death rate} = \frac{\text{Perinatal deaths in GA category}}{\text{Total births in GA category}} \times 1{,}000$$

$$\text{Perinatal mortality rate at 32 weeks} = \frac{(567 + 282)}{(567 + 13{,}122)} \times 1{,}000 = 62.0 \text{ per } 1{,}000 \text{ total births}$$

The stillbirth rate among births at 32 weeks was 41.4 per 1,000 total births, and the neonatal death rate was 21.5 per 1,000 live births.

Time scales and anchors

All time scales have anchors (e.g., the commonly used Gregorian calendar is anchored to the estimated year of Christ's birth, and 5-year survival rates given cancer treatment are anchored to the time of treatment initiation). The two time scales commonly used in perinatology, which measure the duration of life in utero (scale referred to as GA) and the duration of life after birth (scale referred to as chronologic age) use different anchors, namely, the first day of the last menstrual period and the date of birth. Gestational age typically exceeds postconceptional age by approximately 2 weeks (assuming menstrual cycles with regular 28-day cycles and midcycle ovulation). The two scales "overlap" because the chronologic age scale can begin at almost any point on the GA scale.

Quantitative (duration) issues related to time scales

In specific situations, the use of these two overlapping time scales creates problems, especially when infants born at different GAs are compared by chronologic age without consideration of GA. Clearly, the status of a 10-day-old infant depends very much on whether birth occurred at 26 weeks' gestation or at 36 weeks' gestation. Clinicians solve the above-mentioned problem of dual overlapping time scales by speaking of "postmenstrual age" or "corrected GA." For instance, this understanding is reflected in the evolution of the neonatal disease bronchopulmonary dysplasia (BPD), a serious respiratory condition seen in extremely preterm newborns. Initially, BPD was defined as a requirement for oxygen at 28 days after

birth. The current definition of BPD refers to a requirement for oxygen or ventilatory support at 36 weeks of postmenstrual age (Taeusch et al. 2005).

Qualitative (labeling) issues related to time scales

An important aspect of the use of dual time scales that has nothing to do with duration is the qualitative label that gets assigned to death, depending on whether death occurs before or after the second time scale becomes operational. Thus, a fetus who dies in utero at 38 weeks is a stillbirth, but another who dies at 2 weeks of chronologic age after birth at 36 weeks is a neonatal death. This scheme, in which birth has a preeminent position in qualifying life events, has important legal and social implications. Also, birth is an important event from the biologic point of view, since dramatic cardiorespiratory remodeling occurs with the transition from the intrauterine to the extrauterine environment, among other transformations. Nevertheless, the importance accorded to birth is probably sociologic in origin.

Conundrums in perinatal epidemiology

Several aspects of traditional perinatal epidemiology appear somewhat incoherent or at least inconsistent with other perinatal concepts. Some of these are described below.

The paradox of modern obstetrics

Medically indicated early delivery (i.e., labor induction and/or cesarean delivery given fetal compromise or maternal indication) is the cornerstone of modern obstetrics (Joseph 2007). With advances in the diagnosis of fetal compromise (e.g., biophysical profile, umbilical artery Doppler velocimetry), treatments for accelerating fetal lung maturity (e.g., antenatal corticosteroid therapy), and neonatal care (e.g., surfactant, assisted ventilation), rates of medically indicated labor induction and cesarean delivery have increased substantially in recent decades. This has resulted in increases in iatrogenic preterm birth (PTB), earlier birth in the term range, and declines in post-term birth (i.e., a "left shift" in the population GA distribution; Fig. 13.1A). Such obstetric intervention appears misguided under the traditional perinatal model, which suggests that a left shift in the GA distribution, especially at preterm and term gestation, will lead to higher rates of perinatal mortality (since perinatal mortality declines exponentially with increasing pregnancy duration, Fig. 13.1B). Paradoxically, population rates of perinatal death have been declining steadily as rates of medically indicated early delivery at preterm, term, and post-term gestation have increased (Joseph et al. 1998; Ananth et al. 2005).

Disconnect between gestational age–specific growth restriction and perinatal mortality

The use of the 10th percentile (or the 3rd percentile) to define small-for-gestational age (SGA) live births at each gestational week suggests that SGA rates are constant across gestational duration. If growth restriction causes perinatal death, or if growth restriction is a pathway to perinatal death, one would expect concordance in patterns of growth restriction and perinatal death (not a stable growth restriction

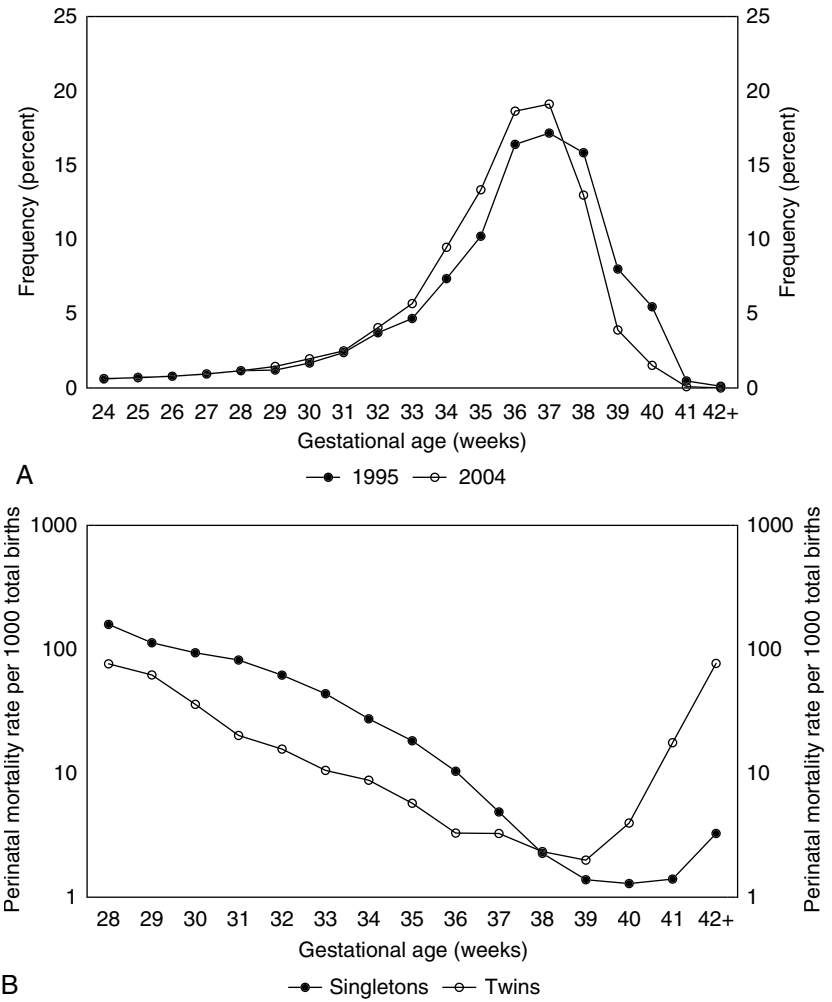


Figure 13.1 Gestational age (GA) distribution of twin live births in the United States, 1995 versus 2004 (**A**) and the traditional calculation of GA-specific perinatal mortality rates per 1,000 total births among singletons and twins, United States 2004 (**B**).

rate and an exponentially declining perinatal mortality rate across gestation [Joseph 2005]). In fact, the strong, clinically observed association between growth restriction in utero and perinatal mortality is decidedly at odds with the disconnection in the traditional perinatal model between constant growth restriction rates and declining perinatal mortality across gestation.

The paradox of intersecting perinatal mortality curves

Over 30 years ago, Yerushalmy identified a paradoxical relationship between maternal smoking and birth weight–specific neonatal mortality (Yerushalmy 1971). He showed that neonatal death rates among infants of smokers were lower

than among infants of nonsmokers at birth weights of 3000 g or less; the reverse was true at higher birth weights. In the last four decades, this observation has been confirmed as a general phenomenon that emerges when birth weight– or GA-specific perinatal mortality rates are compared across plurality, race, parity, infant sex, and country, among other variables (Fig. 13.1). These findings appear to suggest that low-birth-weight (LBW) infants born to mothers who smoke are healthier than LBW infants born to nonsmokers. A number of explanations have been proposed to resolve this paradox and include biological hypotheses and explanations related to "relative" (subpopulation specific) birth weight (Wilcox 1993; Wilcox et al. 1995; Cheung et al. 2000). Nevertheless, the issue continues to be a source of debate and controversy (Joseph 2009).

Candidates for cerebral palsy

In recent decades, understanding regarding the critical period for the development of cerebral palsy (CP) has shifted from the intrapartum period to the antepartum period of pregnancy. Whereas complications of labor and delivery were previously viewed as causal factors, more recent evidence has implicated prenatal factors, as the in utero cause of neurologic injury. The latter includes congenital malformations, vascular insults, and maternal infection (Nelson and Ellenberg 1986; Nelson and Grether 1999; Nelson and Willoughby 2000). If CP occurs among fetuses, then fetuses (rather than live-born infants) would appear to be the appropriate candidates (denominator) for calculating the GA-specific rate of CP. However, the traditional GA-specific rate of CP is calculated exactly as the GA-specific neonatal mortality rate is calculated (with live-born infants as the candidates for CP), and shows that CP rates decline exponentially with increasing GA.

RELATIVE PAUCITY OF EPIDEMIOLOGIC CONCEPTS WITHIN PERINATAL EPIDEMIOLOGY

Although the general epidemiologic literature places a premium on time-related issues, incidence measures, and the distinction between causal and noncausal models, these concepts appear to be less than central in traditional perinatology.

Incidence measures in perinatology

As discussed, there is a tendency to view birth as the starting point for life. Nevertheless, epidemiologists have long recognized that some seemingly new outcomes that reveal themselves at birth may in fact constitute "birth prevalence." Thus, the frequency of congenital malformations at birth is considered birth prevalence, since numerous cases of congenital malformation are lost at a stage when the pregnancy outcome would have been classified as a miscarriage (Rothman 1986). Antenatal testing resulting in the termination of pregnancy also impacts the birth prevalence.

The above recognition notwithstanding, there appears to be a relative lack of emphasis on incidence measures in perinatal epidemiology. This assertion can be justified by examining the general understanding of the concept of PTB in perinatal epidemiology. Although most students of perinatal epidemiology are familiar with the definition of PTB, few recognize it as the epidemiologic measure of disease frequency that it represents. Is PTB a point prevalence, period prevalence, cumulative incidence, incidence density, or ratio? The surprising failure to recognize the epidemiologic flavor of this key perinatal index is not surprising, since formal courses and texts on PTB universally refrain from discussing such issues.

Status of gestational age: determinant versus survival time

Gestational age is mostly treated as a predictor in perinatal epidemiology, just as chronologic age is often treated as a predictor in general epidemiologic studies. For instance, many studies model perinatal death and other outcomes using logistic regression models, with GA treated as an independent variable in the model. Gestational age can also be treated as follow-up time or survival time, as per epidemiologic principles. There are good reasons for treating GA as survival time, as noted in the following two scenarios. First, GA represents the time scale on which the duration of life in utero is measured, and this makes it survival time for a fetus/pregnancy. Thus, the "incidence" of preeclampsia can be viewed as something other than the proportion of pregnancies that are afflicted with this complication (as is commonplace in contemporary obstetric texts). In a bona fide epidemiologic sense, the incidence of preeclampsia can refer to the incidence density with which this complication is encountered with advancing gestational duration. Second, GA and birth weight are variables in the causal pathway between many determinants and many outcomes. For example, pregnancy complications (such as preeclampsia and multifetal gestation) often cause fetal compromise and/or lead to early onset labor at preterm gestation. Adjusting for GA in such contexts is contraindicated and will eliminate the effect operating through this causal pathway (Rothman 1986; see also Chapter 15).

In fact, GA qualifies both as a highly useful predictor and as survival time. The issue of whether GA is survival time or just another determinant can only be answered depending on the question that is being asked. Specifically, the answer revolves around the purpose of the model being proposed and whether it seeks to serve a causal or predictive (noncausal) function.

Risk of cerebral palsy

Although it is widely accepted that PTB is an important risk factor for CP, it has been known for decades that less than half of CP cases occur at preterm gestation, with the remaining occurring among term and post-term births (Jarvis S. et al. 2003; Clark et al. 2008). The belief that preterm gestation is a high-risk period for CP does not incorporate the temporal aspects of the situation: term gestation encompasses a mere 6 weeks or so (37 to 42 weeks), whereas preterm gestation extends for over twice as many weeks (22 to 36 weeks).

INDICES OF INCIDENCE

The fetuses-at-risk approach

In 1987, Yudkin et al. proposed a revolutionary new formulation for calculating GA-specific stillbirth rates. Their proposition was that all fetuses that reach a particular GA are at risk of stillbirth at that GA (not just those born at that gestation). Yudkin and colleagues' *fetuses-at-risk* (FAR) proposition is illustrated in Table 13.1 data from the United States on singletons born in 2004 (see also Figure 13.2).

The calculation of the GA-specific stillbirth rate at 32 weeks under the traditional calculation mentioned previously results in a rate of = (567) × 1,000/(567+13,122) or 41.4 per 1,000 total births. Under the FAR approach, however, the candidates for stillbirth at 32 weeks are not restricted to those fetuses who delivered at 32 weeks, but include all fetuses who survived until 32 weeks' gestation (see last column in Table 13.1, which cumulates all fetuses delivered at 31 weeks and beyond, 32 weeks and beyond, etc.). The stillbirth rate under this FAR formulation equals 567/3,387,623 or 0.167 per 1,000 fetuses at risk.

Details regarding the FAR calculation include (a) the number of fetuses at risk at 20 weeks' gestation will equal the number of total births at 20 weeks of gestation or greater; (b) births with an unknown GA cannot be included in such FAR calculations; (c) the number of fetuses at risk in the category at 42+ weeks will be the same as the number of total births in the 42+ weeks category. Therefore, the stillbirth rate under the traditional and FAR approaches at 42+ weeks will be identical.

The two models of stillbirth provide very distinct patterns of stillbirth occurrence across gestational duration, with stillbirth rates declining exponentially under the traditional model and increasing gradually with increasing gestation under the FAR formulation. Both models are widely used in the literature (Hilder et al. 1998; Hartley et al. 2001; Kramer et al. 2002; Cunningham et al. 2005).

Cumulative incidence

The stillbirth rate at 32 weeks calculated above using the FAR approach (0.167 per 1,000 fetuses at risk) is a cumulative incidence over a 1-week period. This is because the number of fetuses at the beginning of the risk period (3,387,623) was used as the denominator (Boulvain et al. 2000; Hilder et al. 2000; Yudkin and Redman 2000; Joseph 2004).

Table 13.1 Numbers of singleton live births, stillbirths and neonatal deaths at 31 to 33 weeks gestation, United States, 2004

Gestational Age (weeks)	*Stillbirths*	*Live Births*	*Neonatal Deaths*	*Fetuses at Risk**
31	489	8,299	232	3,396,411
32	567	13,122	282	3,387,623
33	562	18,372	271	3,373,934

*Cumulative number of fetuses delivered at that gestational week or later

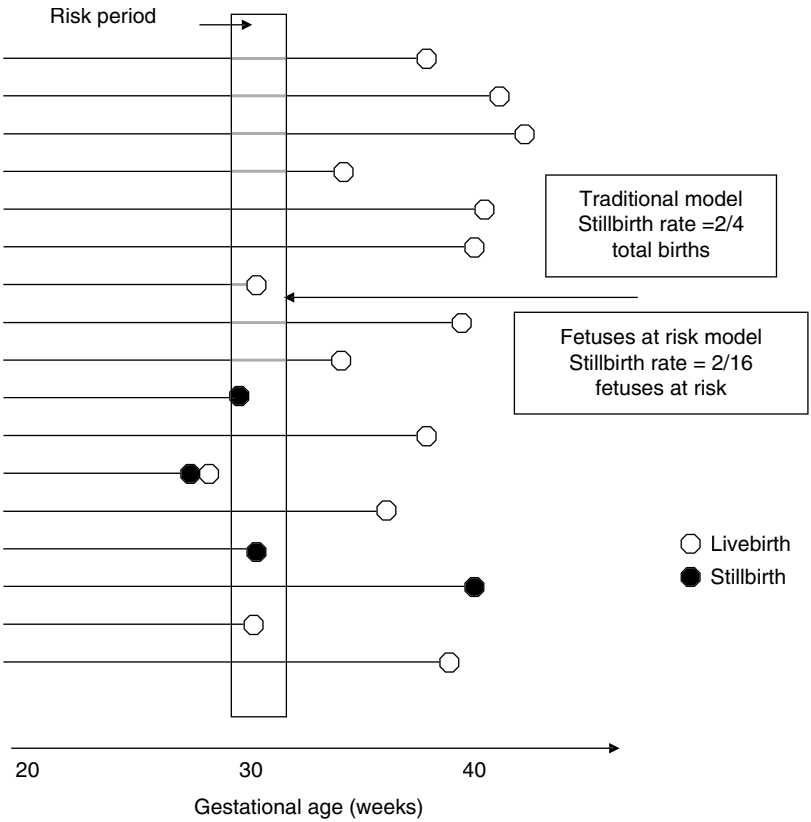


Figure 13.2 Schematic depiction of pregnancy course and options for calculating the gestational age (GA)-specific stillbirth rate (Joseph 2007). *Traditional calculation*: Number of stillbirths at any gestational week/Number of total births at that gestational week = 2/4 = 500 per 1,000 total births. *Fetuses-at-risk calculation*: Number of stillbirths at any gestational week/Number of fetuses at risk of stillbirth at that gestational week = 2/16 = 125 per 1,000 fetuses at risk.

Incidence density

The cumulative incidence of stillbirth calculated above approximates an incidence density. It is an approximation because not all fetuses contribute 1 week of follow-up. Those who are born at 32^{+0} week (exactly 32 completed weeks), 32^{+1} weeks (32 completed weeks plus 1 day), 32^{+2} weeks, etc., contribute less than 1 week of follow-up. Since the period of follow-up is short, the cumulative incidence estimated above provides a reasonable approximation of the incidence density at most GAs. However, a better way to obtain the incidence density rate would require using the mid-week number of fetuses as the denominator, as an approximation of the "person time" (Boulvain et al. 2000; Hilder et al. 2000; Yudkin and Redman 2000; Joseph 2004). This can be done by averaging the number of fetuses at risk at 32 and 33 weeks (assuming that losses because of birth are evenly

distributed across the week), and would yield a stillbirth incidence density rate at 32 weeks = 567/((3,387,623 + 3,373,934)/2) = 0.168 per 1,000 fetus-weeks.

Although the difference between cumulative incidence and incidence density is negligible at early gestation, it becomes more substantial at later GAs, when birth is more frequent. For instance, at 40 weeks' gestation, the cumulative incidence of stillbirth among singletons (United States 2004) was 0.549 per 1,000 fetuses at risk (625/1,137,595), whereas the incidence density of stillbirth was 0.89 per 1,000 fetus weeks (625/(1,137595 + 268,714)/2). Note that this approximation of incidence density is also biased toward providing underestimates or overestimates of the true incidence density because relatively more fetuses will be born in the later part of the week at 37 weeks (i.e., more births at 37^{+6} weeks compared to 37^{+0} weeks), whereas the opposite is true at 41 weeks and beyond (i.e., more births at 41^{+0} weeks compared to 41^{+6} weeks). Ideally, measuring incidence density would require counting the exact number of fetal days of follow-up (Boulvain et al. 2000; Hilder et al. 2000; Yudkin and Redman 2000; Joseph 2004).

INCIDENCE OF PREGNANCY COMPLICATIONS

Few studies describe incidence patterns of preeclampsia and other pregnancy complications, such as placental abruption and chorioamnionitis, but, in general, risk of complications increases with gestation, especially from term to post-term gestation (Caughey et al. 2003; Caughey and Musci 2004). Such studies use ongoing pregnancies in the denominator, as illustrated in Figure 13.3, which is an appropriate strategy for studying maternal complications. For fetal outcomes, however, the FAR denominator is preferable as it accounts for the plurality of fetuses in a multifetal pregnancy (Joseph 2008).

INCIDENCE OF LABOR INDUCTION AND CESAREAN DELIVERY

Pregnancies and fetuses at risk for delivery have rarely been used to estimate the incidence of labor induction and cesarean delivery. Such calculations show that the incidence of labor induction and cesarean delivery increase with increasing pregnancy duration, as would be expected given increases in the incidence of pregnancy complications (Joseph 2007).

INCIDENCE OF BIRTH

The incidence of birth is calculated by dividing the number of births at any gestation by the number of fetuses at risk of birth at that specific GA (Joseph 2004, 2007). As previously discussed, this index is a cumulative incidence (per gestational week), which approximates the incidence density of birth. The other cumulative incidence that is of interest to the perinatal health community is the familiar index of preterm birth, with the latter defined as encompassing live birth

(or stillbirth and live birth) in the period up to 28, 32, or most typically 37 weeks of GA (Joseph 2004).

INCIDENCE OF FETAL GROWTH RESTRICTION

Concept

The incidence of fetal growth restriction is defined as the occurrence of new cases of growth restriction (SGA) at a particular GA among a cohort of fetuses followed prospectively (Joseph 2004, 2007; Hutcheon and Platt 2008; Paneth 2008). Determining the incidence of intrauterine growth restriction is a contemporary challenge, because this requires the identification of all growth-restricted fetuses at each GA (see Figure 13.3). Serial ultrasound assessment of fetal weight and growth restriction among continuing pregnancies is neither fully valid (Nahum and Stanislaw 2003; Lerner 2004; Dudley 2005) nor routinely available at the population level.

Operationalization

The above-mentioned limitation means that contemporary estimation of the incidence of growth restriction must rely on alternative proxy measures, such as the rate of "revealed growth restriction" (i.e., by using SGA live births at any GA as the proxy for the frequency of growth restriction at that GA). This requirement (of birth to occur for SGA to be diagnosed) is not as serious a limitation as it might first appear, since seriously growth-restricted fetuses tend to die or to be born either spontaneously or following obstetric intervention. Revealed growth restriction rates are useful for understanding the general pattern of growth restriction in pregnancy: Does the rate increase, decrease, or remain constant with advancing gestation? Their main limitation arises from the fact that they are influenced by patterns of birth. Further developments in ultrasound technology and availability of information on fetal growth throughout pregnancy will help refine this index of perinatal health.

The incidence of revealed growth restriction can be calculated by dividing the number of SGA births at any GA by the number of fetuses at risk of being born SGA at that GA (Fig. 13.3). This calculation shows that growth restriction rates increase with advancing GA. A consequence of contemporary methods of constructing fetal growth standards (e.g., using the 10th percentile of birth weight among live births at a particular gestation as the SGA cutoff) is that for the predominant population used in the construction of the fetal growth curve (e.g., singletons; twins if a twin-specific standard is used), the SGA rate at any gestation will be approximately one-tenth the birth rate at that gestation.

INCIDENCE OF DEATH

The form of the model for describing the incidence of death depends on the function for which it is to be used; that is, if the focus is to be on the fetus exclusively,

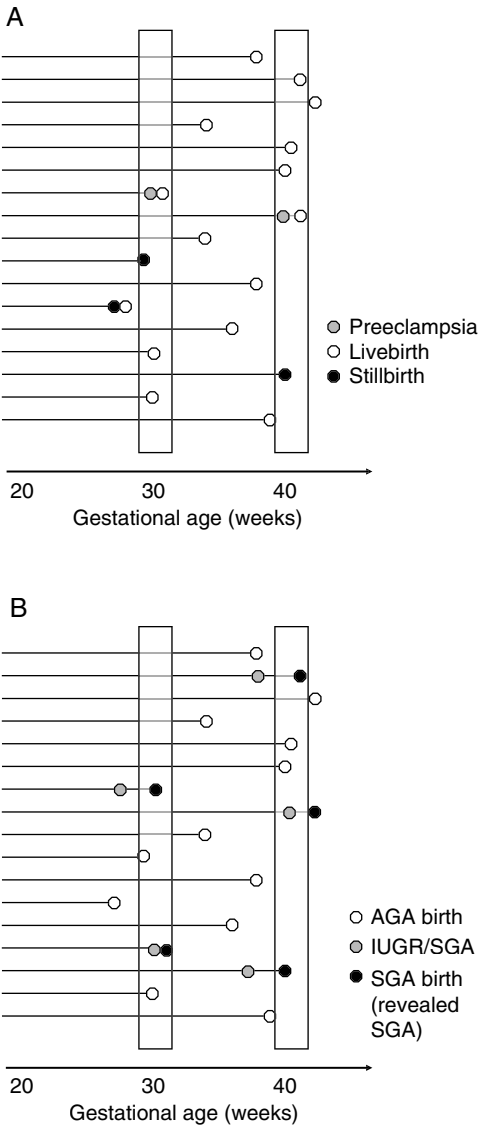


Figure 13.3 Schematic depiction of pregnancy course and options for calculating the incidence of pregnancy complications (**A**), small-for-gestational age (SGA), and revealed SGA (**B**) according to the fetuses-at-risk model (Joseph 2007). **A**: *Traditional calculation*: Number of deliveries with preeclampsia at any gestational week/Number of deliveries at that gestational week = 1/4 = 250 per 1,000 deliveries for the first period and 1/5 = 200 per 1,000 deliveries for the second period. *Fetuses-at-risk calculation:* Number of new cases of preeclampsia at any gestational week/Number of pregnancies at risk of preeclampsia at that gestational week = 1/16 = 63 per 1,000 pregnancies at risk in the first period and 1/6 = 167 per 1,000 fetuses at risk in the second period. **B**: *Traditional calculation for the SGA rate*: Assumed to be uniform 10% or 3% at each gestation depending on cutoff used (10th percentile or

(continued)

Figure 13.3 (*continued*) 3rd percentile). *Fetuses-at-risk calculation* for the SGA rate: Number of new SGA cases at any gestational week/Number of fetuses at risk of SGA at that gestational week = 1/15 = 67 per 1,000 fetuses at risk for the first risk period and 1/4 = 250 per 1,000 fetuses at risk for the second risk period. *Fetuses-at-risk calculation* for the revealed SGA rate: Number of revealed SGA cases at any gestational week/Number of fetuses at risk of SGA birth at that gestational week = 2/16 = 125 per 1,000 fetuses at risk in the first risk period and 2/6 = 333 per 1,000 fetuses at risk in the second period.

as is the case in obstetrics, or on both the fetus and the infant, as may be the case more generally.

Obstetric model of death

Fetuses at any gestation are considered the candidates at risk of death at that gestation from an obstetric standpoint (Joseph 2004). Estimating the incidence of stillbirth, neonatal death, and perinatal mortality rates represents a survival analysis model with censoring of subjects (fetuses) at death or birth (which ever occurs earlier) and the assignment of neonatal death to the point of birth (Fig. 13.4). The time between birth and neonatal death is ignored, because the pathology responsible for death is typically present at birth (i.e., neonatal death *typically* has its origins in pregnancy/labor complications). It is for this reason that most state-of-the-art obstetric intervention trials, such as those assessing the effects of multiple course antenatal steroids (Roberts and Dalziel 2006), progesterone (Dodd et al. 2008), or magnesium sulfate (Rouse et al. 2008) use a composite outcome that includes stillbirth and neonatal death.

The incidence of perinatal death (and serious neonatal morbidity, which is typically pregnancy- or labor-related as well) under the FAR model is a critical concept for theoretical obstetrics. The *increase* in stillbirth and neonatal death that is seen with increasing GA under this formulation provides the foundation for developing obstetric theory to guide intervention in obstetric practice (i.e., it justifies selective, carefully timed early delivery given fetal compromise or a maternal indication). The formulation also provides insight into the recent iatrogenic increases in preterm birth, the overall left shift in the GA distribution, and the associated declines in fetal and neonatal mortality (Joseph 2007).

Comprehensive model of death

The comprehensive model of death requires abandoning the dual time scales of GA and chronologic age altogether and avoiding the distinction between death immediately before birth versus death immediately after birth (Joseph 2004). This represents a survival analysis model with censoring of subjects (whether fetuses or infants) at death only. Birth is ignored as an event for anchoring age and classifying type of death (stillbirth, neonatal, and postneonatal), and time is measured on the original scale whether postconceptional or postmenstrual. The incidence of fetal or infant death is measured over postconceptional/postmenstrual age using

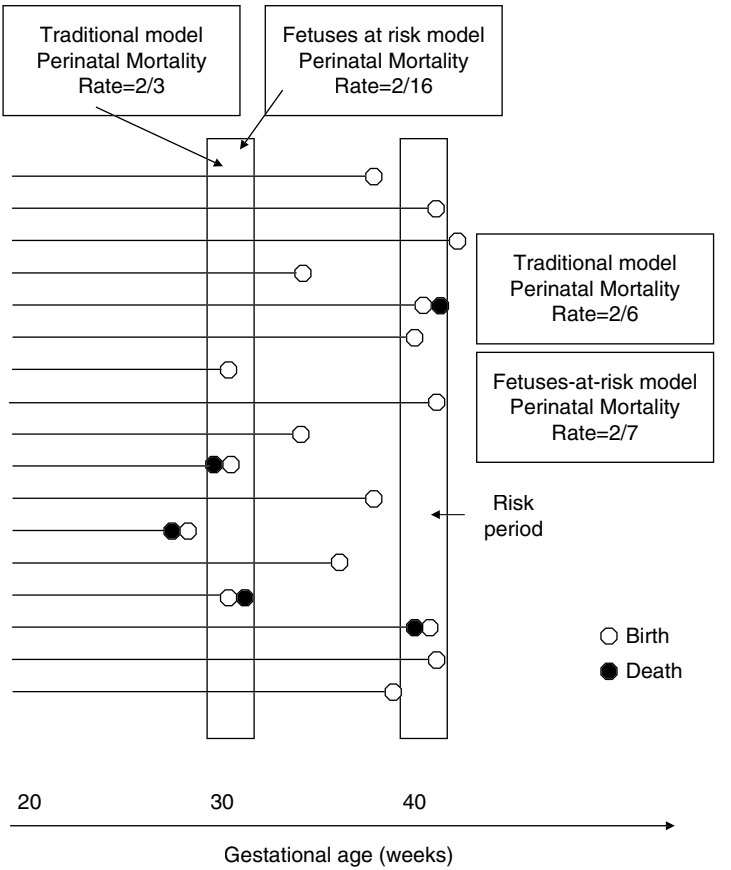


Figure 13.4 Schematic depiction of the survival analysis (obstetric) for perinatal death under the fetuses-at-risk model (Joseph 2007), with censoring at death or birth (whichever occurs earlier). Perinatal death is assigned to the point of birth. In the first risk period, there are 16 fetuses at risk of perinatal death, three births, one stillbirth, and one neonatal death. In the second risk period, there are seven fetuses at risk, six births, one stillbirth, and one neonatal death. Under the conventional calculation, with perinatal mortality defined as the number of perinatal deaths within any period divided by the number of total births in that period, the perinatal mortality rate is 2/3 in the first risk period and 2/6 in the second. Note increase in denominator and decrease in rate (from 67% to 33%). Under the fetuses-at-risk formulation, with perinatal mortality defined as the number of perinatal deaths in any period divided by the number of fetuses at risk of perinatal death in that period, the perinatal mortality is 2/16 in the first risk period and 2/7 in the second risk period. Note decrease in denominator and increase in rate (from 13% to 29%).

the number of fetuses and infants at risk of death (at any given age) as the denominator. At any point in time, up to 40-odd weeks postconception, the risk set would include fetuses and infants, and deaths could include stillbirths to fetuses or neonatal or postneonatal deaths to infants born days to weeks previously. Birth (whether spontaneous or induced) and growth faltering (in utero and also after

birth) become objects of study or factors whose effect on death may be modeled (Joseph 2004).

Such a model has utility in a comprehensive examination of death, but is complex because it must simultaneously incorporate intrauterine influences and also factors that operate in infancy (e.g., factors associated with death due to injury). The comprehensive model of death may be integrated into a survival analysis framework (Platt et al. 2004).

Choice of obstetric versus comprehensive model

The FAR model, with a focus on perinatal death (but not the comprehensive model using the fetuses/infants-at-risk approach) is appropriate for obstetric purposes because obstetric decision-making focuses on the unborn fetus/pregnant woman at a particular GA. Integrating serious morbidity diagnosed in the perinatal period (e.g., bronchopulmonary dysplasia, brain injury, severe retinopathy of prematurity) into the obstetric model permits the inclusion of outcomes that may lead to disability rather than death, but which are nevertheless highly relevant for obstetric management. On the other hand, excluding deaths that occur after the first month of chronologic age and perinatal deaths due to congenital anomalies ensures that extraneous deaths, unrelated to obstetric management (e.g., due to injury, homicide) are not part of the outcome.

One important issue to recognize is that the obstetric model incorporates biological concepts, such as a latent period, by assigning a prenatal or intrapartum etiology to pregnancy-related outcomes even if they are diagnosed remote from birth. Thus neonatal death that is thought to be of prenatal origin is attributed to pregnancy-related determinants. Survival analysis, based on the comprehensive model of death, needs to explicitly incorporate such biological concepts. Ignoring latent period issues and associating contemporaneous determinants and outcomes is a common failing of epidemiology that is increasingly recognized as an issue of methodologic concern. Nevertheless, this particular construct of the extended FAR approach (i.e., its extension beyond stillbirth to neonatal death and serious neonatal morbidity) represents the most controversial aspect of the model. Perhaps the best example of the need for incorporating latent-period concepts into such models is seen with CP (Joseph et al. 2003); this condition typically manifests 18–24 months after birth, although it is generally accepted that the neurologic injury responsible for the motor deficit typically occurs in utero (Nelson and Ellenberg 1986; Nelson and Grether 1999; Nelson and Willoughby 2000).

Growth restriction and mortality under the traditional versus fetuses-at-risk models

Stillbirth, neonatal death, and perinatal mortality rates decline exponentially under the traditional model and paradoxical crossovers are observed when mortality is plotted by GA or birth weight (Yerushalmy 1971). Under the FAR model, these same rates increase gradually with increasing GA duration, and no crossovers in mortality curves are observed over survival time (GA). The patterns of birth, revealed

SGA, and perinatal death are consistent as well and increase in frequency as gestational duration increases (see Figure 13.5).

CONCLUSION

Causal and noncausal models in perinatal epidemiology

Models have to be evaluated against the research questions they attempt to answer. The conventional and FAR models of perinatal death are a case in point. The fact that mortality declines exponentially with duration of gestation under the conventional model and increases under the FAR model shows that the two models provide different perspectives and do not address the same question.

In this context, it is instructive to recall Miettinen's classification of models into causal and noncausal categories (Miettinen 1985). Noncausal models serve a predictive function, and determinant selection for such models is without any view to causal interpretation. Such models, for example the scoring system for the diagnosis of group A streptococcal sore throat (using fever >38°C and other determinants [McIsaac et al. 1998]) and the 5-minute Apgar score for predicting neonatal death (using color, reflex irritability, and other determinants [Casey et al. 2001]), are particularly useful for diagnosis and descriptive prognosis. The noncausal nature of the relationships within such models does not detract from the model's utility for diagnosis or prognosis. Required, however, is a strong relation between the determinant and the outcome (Miettinen 1985). Causal models, on the other hand, provide etiologic insight and serve as the basis for clinical intervention even when the effects of intervention are modest (Miettinen 1985). Drug and other therapies within a clinical management model constitute examples.

The conventional model of perinatal death falls into the noncausal category. The exponentially declining relation between gestational duration and perinatal death implies that one can reasonably predict outcome given pregnancy duration. Similarly, the strong relationship between birth weight and neonatal death provides a useful prediction model for setting prognosis at birth. However, the conventional model of perinatal death ignores the biologic continuum (by using separate, overlapping time scales for life in utero and after birth) and treats gestational duration as a determinant and not as follow-up (survival) time. In fact, short gestational duration and LBW are better viewed as markers for serious pregnancy complications and other risk factors, which are the causal determinants in this situation.

The FAR model, on the other hand, is a causal model. It respects the biologic continuum and, in keeping with general epidemiologic principles, treats gestational duration as follow-up (survival) time. This permits the estimation of incidence rates, including the incidence of birth, growth restriction, and death. The model provides a theoretical justification for selective, carefully timed early delivery, given fetal compromise, which is the cornerstone of modern obstetrics (Joseph 2007). Such early delivery involves a balancing of benefits and harms, including formal and bedside considerations of therapeutic indices like the number needed

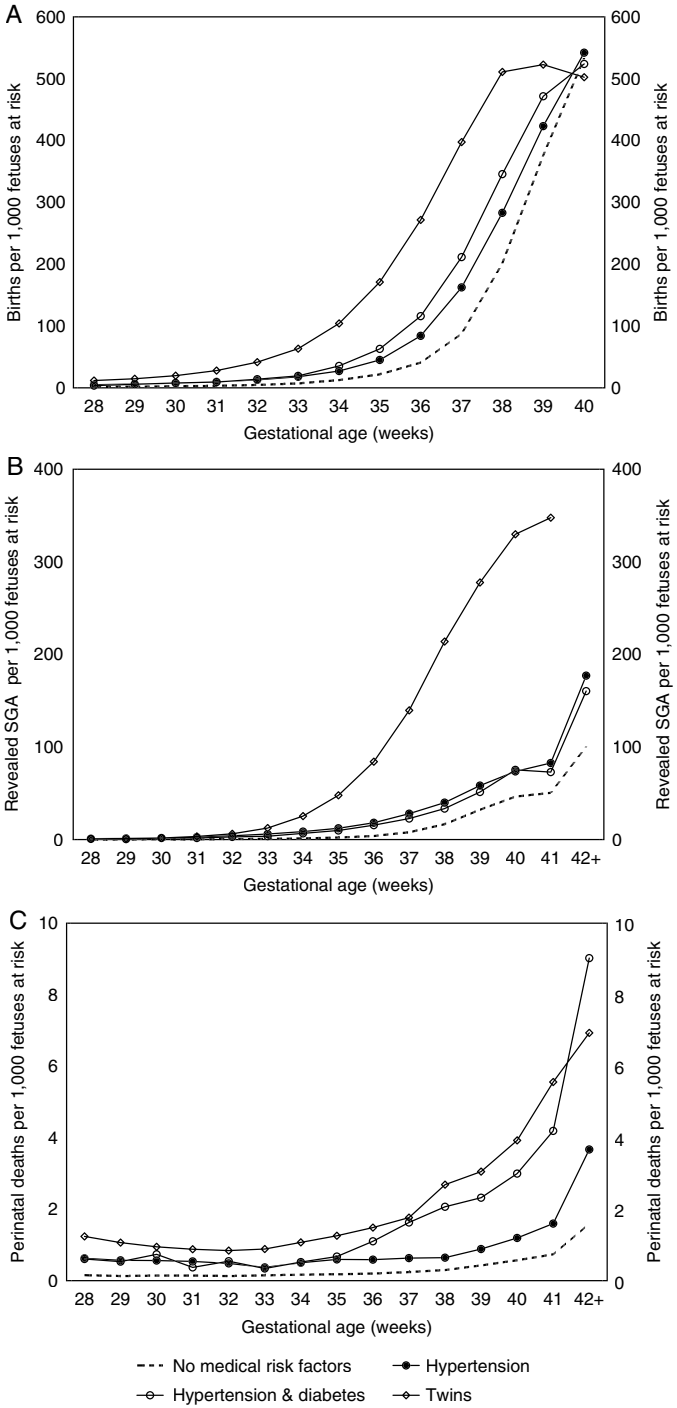


Figure 13.5 Incidence of birth (**A**), incidence of revealed growth restriction (**B**) and incidence of perinatal death (**C**) at 28 weeks' gestation and over, among pregnancies with no medical risk factors, hypertension, hypertension and diabetes, and twins, United States 1999–2000 (Joseph 2007).

to treat. Notably, medically indicated early delivery is predicated on short-term risks that are reassessed at short, periodic intervals, because risks in pregnancy can change dramatically over relatively short periods (Joseph 2008).

The FAR model brings coherence to the full range of perinatal phenomena by yielding congruent incidence patterns of birth, growth restriction, pregnancy complications, labor induction/cesarean delivery, perinatal death, and serious neonatal morbidity. Diverse perinatal issues converge to the fetuses/pregnancies-at-risk model when candidates for the perinatal outcome of interest are identified in a manner consistent with fundamental epidemiologic principles. As mentioned, however, these considerations apply when the focus is on models that seek to address causal questions related to biology or obstetric intervention, rather than models oriented toward prediction and which serve a diagnostic or prognostic function (Additional information on causal analysis may be found in Chapter 15).

REFERENCES

Ananth, C.V., Joseph, K.S., Oyelese, Y., Demissie, K., and Vintzileos, A.M. 2005. Trends in preterm birth and perinatal mortality among singletons: United States, 1989 through 2000. *Obstet Gynecol* 105: 1084–89.

Boulvain, M., Faltin, D., Ibecheole, V., and Irion, O. 2000. Randomised trials of earlier induction of labour are needed. *BMJ* 320: 445.

Casey, B.M., McIntire, D.D., and Leveno, KJ. 2001. The continuing value of the Apgar score for the assessment of newborn infants. *N Engl J Med* 344: 467–71.

Caughey, A.B., Stotland, N.E., and Escobar, G.J. 2003. What is the best measure of maternal complications of term pregnancy: Ongoing pregnancies or pregnancies delivered? *Am J Obstet Gynecol* 189: 1047–52.

Caughey, A.B. and Musci, T.J. 2004. Complications of term pregnancies beyond 37 weeks of gestation. *Obstet Gynecol* 103: 57–62.

Cheung, Y.B., Yip, P., and Karlberg, J. 2000. Mortality of twins and singletons by gestational age: A varying-coefficient approach. *Am J Epidemiol* 152: 1107–16.

Clark, S.M., Ghulmiyyah, L.M., and Hankins, G.D. 2008. Antenatal antecedents and the impact of obstetric care in the etiology of cerebral palsy. *Clin Obstet Gynecol* 51: 775–86.

Cunningham, F.G., Leveno, K.J., Bloom, S.L., Hauth, J.C., Gilstrap, L.C., III, and Wenstrom, K.D., eds. 2005. *Williams' obstetrics*, 22nd ed., p 678. New York: McGraw-Hill.

Dodd, J.M., Flenady, V.J., Cincotta, R., and Crowther, C.A. 2008. Progesterone for the prevention of preterm birth: A systematic review. *Obstet Gynecol* 12: 127–34.

Dudley, N.J. 2005. A systematic review of the ultrasound estimation of fetal weight. *Ultrasound Obstet Gynecol* 25: 80–89.

Hartley, R.S., Emanuel, I., and Hitti, J. 2001. Perinatal mortality and neonatal morbidity rates among twin pairs at different gestational ages: Optimal delivery timing at 37 to 38 weeks' gestation. *Am J Obstet Gynecol* 184: 451–58.

Hilder, L., Costeloe, K., and Thilaganathan, B. 1998. Prolonged pregnancy: Evaluating gestation-specific risks of fetal and infant mortality. *Br J Obstet Gynaecol* 105: 169–73.

Hilder, L., Costeloe, K., and Thilaganathan, B. 2000. Prospective risk of stillbirth. *BMJ* 320: 444.

Hutcheon, J.A. and Platt, R.W. 2008. The missing data problem in birth weight percentiles and thresholds for "small-for-gestational-age." *Am J Epidemiol* 167: 786–92.

Jarvis, S., Glinianaia, S.V., Torrioli, M.G., Platt, M.J., Miceli, M., Jouk, P.S., et al. 2003. Cerebral palsy and intrauterine growth in single births: European collaborative study. *Lancet* 362: 1106–11.

Joseph, K.S. 2004. Incidence-based measures of birth, growth restriction and death can free perinatal epidemiology from erroneous concepts of risk. *J Clin Epidemiol* 57: 889–97.

Joseph, K.S. 2007. Theory of obstetrics: An epidemiologic framework for justifying medically indicated early delivery. *BMC Pregnancy Childbirth* 7: 4.

Joseph, K.S. 2008. The fetuses-at-risk approach: Semantic and conceptual misapprehension. *BMC Pregnancy Childbirth* 8: 11.

Joseph, K.S. 2009. Exegesis of effect modification–Biologic or spurious? *Paediatr Perinat Epidemiol* 23: 417–20.

Joseph, K.S., Allen, A.C., Lutfi, S., Murphy-Kaulbeck, L., Vincer, M.J., and Wood, E. 2003. Does the risk of cerebral palsy increase or decrease with increasing gestational age? *BMC Pregnancy Childbirth* 3: 8.

Joseph, K.S., Kramer, M.S., Marcoux, S., Ohlsson, A., Wen, S.W., Allen, A.C., and Platt, R. 1998. Determinants of preterm birth rates in Canada from 1981 through 1983 and from 1992 through 1994. *N Engl J Med* 339: 1434–39.

Kramer, M.S., Liu, S., Luo, Z., Yuan, H., Platt, R.W., and Joseph, K.S. 2002. Analysis of perinatal mortality and its components: Time for a change? *Am J Epidemiol* 156: 493–97.

Lerner, J.P. 2004. Fetal growth and well-being. *Obstet Gynecol Clin North Am* 31: 159–76.

McIsaac, W.J., White, D., Tannenbaum, D., and Low, D.E. 1998. A clinical score to reduce unnecessary antibiotic use in patients with sore throat. *CMAJ* 158: 75–83.

Miettinen, O.S. 1985. *Theoretical epidemiology: Principles of occurrence research in medicine*, pp. 11–12. Toronto: John Wiley & Sons.

Nahum, G.G. and Stanislaw, H. 2003. Ultrasonographic prediction of term birth weight: How accurate is it? *Am J Obstet Gynecol* 188: 566–74.

Nelson, K.B. and Ellenberg, J.H. 1986. Antecedents of cerebral palsy: Multivariate analysis of risk. *N Engl J Med* 315: 81–86.

Nelson, K.B. and Grether, J.K. 1999. Causes of cerebral palsy. *Curr Opin Pediatr* 11: 487–91.

Nelson, K.B. and Willoughby, R.E. 2000. Infection, inflammation and the risk of cerebral palsy. *Curr Opin Neurol* 13: 133–39.

Paneth, N. 2008. Invited commentary: The hidden population in perinatal epidemiology. *Am J Epidemiol* 167: 793–96.

Platt, R.W., Joseph, K.S., Ananth, C.V., Grondines, J., Abrahamowicz, M., and Kramer, M.S. 2004. A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death. *Am J Epidemiol* 160: 199–206.

Roberts, D. and Dalziel, S. 2006. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database Syst Rev* 3: CD004454. DOI: 10.1002/14651858.CD004454.pub2.

Rouse, D.J., Hirtz, D.G., Thom, E., Varner, M.W., Spong, C.Y., Mercer, B.M., et al. 2008. A randomized, controlled trial of magnesium sulfate for the prevention of cerebral palsy. *N Engl J Med* 359: 895–905.

Rothman, K.J. 1986. *Modern epidemiology,* p. 34. *Toronto*: Little Brown and Company.

Taeusch, H.W., Ballard, R.A., and Gleason, C.A., eds. 2005. *Avery's disease of the newborn*, 8th ed. Philadelphia: Elsevier Saunders.

Wilcox, A.J. 1993. Birth weight and perinatal mortality: The effect of maternal smoking. *Am J Epidemiol* 137: 1098–1104.

Wilcox, A.J., Skjœrven, R., Buekens, P., and Kiely, J. 1995. Birth weight and perinatal mortality: A comparison of the United States and Norway. *JAMA* 272: 709–11.

Yerushalmy, J. 1971. The relationship of parents' cigarette smoking to outcome of pregnancy: Implications as to the problem of inferring causation from observed associations. *Am J Epidemiol* 93: 443–56.

Yudkin, P.L., Wood, L., and Redman, C.W.G. 1987. Risk of unexplained stillbirth at different gestational ages. *Lancet* 1: 1192–94.

Yudkin, P. and Redman, C. 2000. Impending fetal death must be identified and pre-empted. *BMJ* 320: 445.

14

The Fetuses-at-Risk Approach: An Evolving Paradigm

ROBERT W. PLATT

This chapter expands and generalizes the concept of fetuses-at-risk (FAR), as fully discussed in Chapter 13, especially with regard to the methodologic aspects of the FAR paradigm for perinatal epidemiology. Implicit in the FAR approach is the acceptance of the "comprehensive model" of mortality, in which fetuses are conceptualized as being at risk of death from conception through birth and on into early life, with all deaths falling onto this time axis, starting at conception. The comprehensive model can be thought of as addressing the following etiologic question: Given that a woman is pregnant and is exposed to an exposure (E) at a specific gestational age (k) as measured in weeks, will the exposure increase or decrease the offspring's chance of surviving more than a year from k weeks? This question is necessarily formulated in a general way; we are now considering all deaths from week k forward, and essentially ignoring birth in the process. This chapter reviews methods appropriate for addressing this research question while comparing and contrasting them with other approaches.

Perinatal mortality comprises two etiologically heterogeneous events—stillbirth and early neonatal death—and failure to address this heterogeneity when using perinatal mortality as an outcome may be shortsighted (Kramer et al. 2002). In addition, an arbitrary restriction of early neonatal mortality to the first 7 days after birth seems questionable. This definition is problematic, since it arbitrarily determines that an infant born at 250 days' gestation and dying at 7 days of life (257 days post last menstrual period [LMP]) represents a perinatal death, but an infant born at 249 days' gestation and dying at 8 days of life (also 257 days post LMP) is a perinatal survivor. This infant would be part of the denominator but not the numerator in estimating the rate of perinatal mortality. This definition, based in historical definitions of the perinatal period, is legitimate in consideration of short-term risk, but may not make sense from an epidemiologic perspective.

This chapter argues that the temporal nature of gestational age (GA) is an important reason to consider it as a time axis in a time-to-event analysis rather than as a conventional covariate. Accordingly, we develop a measure of mortality based on time since conception that extends methods based on the FAR paradigm described in Chapter 13. We propose survival analytic techniques and examine

the risk of mortality post- LMP as a function of race and maternal smoking, while investigating the possible time-dependence of these associations. We begin by discussing the cohorts we typically study, and noting some limitations.

COHORTS OF BIRTHS AS DENOMINATORS

Typically, studies of perinatal outcomes collect cohorts of births, or pregnant women in varying stages of gestation. This aspect of study design reflects convenience, given the challenges in recruiting women prior to conception or during the earliest stages of pregnancy, when the women may not recognize their pregnancies. Studies that recruit early in pregnancy can suffer from selection bias as well (Kramer et al. 2009), because subjects who present early and are willing to consent to study may differ on important risk factors for outcome from others who participate. Many pregnancies are lost before being recognized (Wilcox et al. 1988), and women enter prenatal care and studies at varying points in gestation. Finally, the events under study (e.g., perinatal mortality) are (reassuringly) rare, so large sample sizes are typically required for study, thus prompting widespread reliance on vital statistics data. An immediate question associated with this practice is the acceptability of such an approach.

To address the issue, we typically start counting time at birth and adjust for GA. Why is this a problem? Birth is an important event, and with advancing GA there is a huge shift in risk. Figure 14.1 describes the association between GA and perinatal mortality using the U.S. vital statistics data from 1999–2001 (MacDorman and Atkinson 1999); at early GAs almost all births result in fetal or neonatal death, whereas at term death is very infrequent. A cohort of births is, evidently, selected based on birth (GA). Given this, consideration of what to do with GA in a statistical analysis becomes paramount, and it becomes essential to understand the associations between GA, birth, and birth outcomes.

USING GESTATIONAL AGE TO DEFINE THE AT-RISK COHORT

Let us apply basic epidemiologic principles to these cohorts of births. First, we need a few simplifying assumptions. We assume that exposure takes place either before or during early pregnancy but does not affect conception; in our assumed notation, given above, this means that k (the week at which exposure occurs) is close to zero. Many exposures have effects on conception and on pregnancy outcomes (e.g., smoking, alcohol consumption [Shiono et al. 1986; Wilcox 1993]); see Chapter 3 for a discussion of the role of risk factors in fecundity and fertility. Attempting to estimate the effect of exposure on mortality when exposure affects both conception and mortality complicates things considerably and is beyond the scope of this book. This problem is the subject of current research.

Now, given that we have an exposure that is initiated near the time of conception, it is clear that, at least theoretically, the cohort of interest for the exposure we are considering would be those subjects who initiated (or not) the exposure, followed

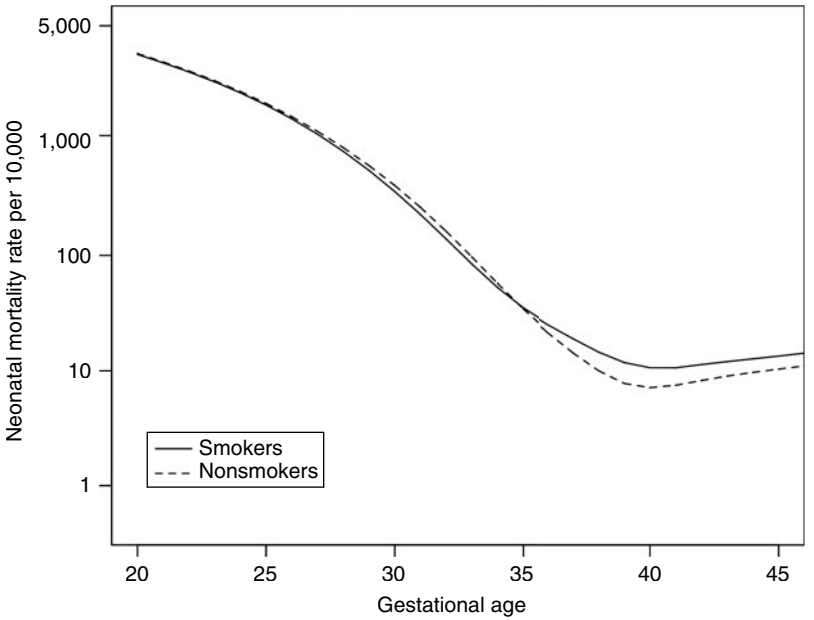


Figure 14.1 Neonatal mortality rate as a function of gestational age, stratified by smoking status, U.S. live births 1999–2001.

from the start of exposure and risk (conception) to ascertainment of the outcome (mortality) (Rothman et al. 2008). Any cohort initiated after the exposure onset is at least theoretically subject to selection bias (Flanders and Klein 2007). To guarantee that selection is not a problem for exposure onset, it is clear that, for maternal exposures early in pregnancy, we would want to recruit at, or prior to, conception. The utility and feasibility of preconception enrollment of women/couples has been previously reviewed (Buck et al. 2004) as further described in the next section.

Cohorts of conceptions

As discussed, ideally, one could recruit very early in pregnancy. Figure 14.2 is a schematic of this cohort. In Figure 14.2, open circles represent live births, and black dots represent deaths. Subject 1 is a stillbirth at 37 completed weeks, subject 2 a live birth and neonatal death, and subjects 3 to 5 are live births who survived the early neonatal period. Other subjects reflect these patterns with varying lengths of follow-up.

Cohorts of births as a sample from a cohort of conceptions

Consider again Figure 14.2. A cohort of births can be viewed as a sample from the cohort of conceptions described in the figure. This sampled cohort, however, is now a structured sample of the underlying time at risk based on GA, with the structure of the sample determined by the times of birth.

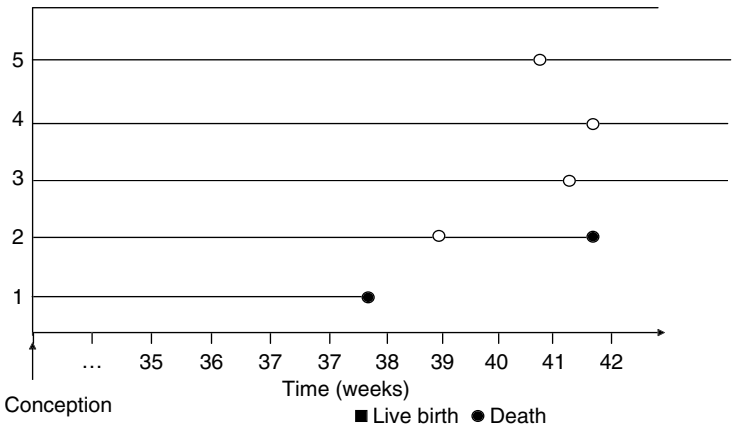


Figure 14.2 Schematic of a conception cohort.

Two problems arise with this sample. Those infants born earlier will contribute more high-risk time to the cohort of births (because of the steep gradient of risk of mortality with GA). Very preterm infants in particular have a very high risk of negative outcomes and contribute events to this sample disproportionate to their numbers in the sample. Second, birth is not a random event; numerous risk factors for preterm birth and shorter GAs have been established (Kramer et al. 2001). Thus, infants with these risk factors are born earlier and contribute disproportionately to the part of the cohort at highest risk for mortality.

This gives rise to an important question: What is the causal effect of interest in perinatal epidemiology when GA is of interest? Exposures that shorten GA may increase mortality *indirectly*; that is, they lower GA, which leads to higher risk. Exposures may also have a *direct* effect on mortality; that is, they directly cause death irrespective of the value of the intermediate variables. Finally, one could consider the *total* effect of an exposure, which is defined as the effect through all possible direct and indirect pathways. What is the effect of typical interest in perinatal epidemiology? The strong association of GA with outcome, and the known associations between GA and exposures, lead many authors to use statistical adjustment to control for GA, either because they believe it is a confounder of the association, or because they want to estimate the direct effect of exposure, controlling for the indirect effect through GA. Chapter 15 discusses direct and indirect effects, and the challenges that arise with attempts to estimate these effects, in particular when GA and birth weight are the intermediates in question. For the remainder of this chapter, we focus on the total effect; this is, essentially, the question described in the introduction of this chapter.

Problems with definitions of neonatal and perinatal mortality

Our use of total mortality as an outcome requires us to revisit some of the currently used outcome measures and definitions. Consider three pregnancies, each

conceived on the same day. The first is born on day 266 post-LMP and dies on day 280 post-LMP; thus, a late neonatal death (14 days after birth). The second is born on day 273, and dies on day 280; an early neonatal death (7 days after birth). The third is stillborn on day 280; a stillbirth or fetal death. By standard definitions of outcomes, these are three different events, all among term fetuses. Is this sensible? From the perspective of a cohort of conceptions and the outcome total mortality, these are all the same event (Platt et al. 2004, 2005). All entered the cohort on the same day and left the cohort via death on the same day; the fact that they were born on different days does not matter to this calculation. On the other hand, in the study of neonatal death or perinatal death, all three are classified differently, as described above (Kramer et al. 2002).

Measurement issues

We assume in these discussions that GA and the time of conception is known with certainty. Gestational age represents a complicated measurement problem; it is typically not directly observable and is estimated using several different approaches (Chapter 8). Since time of conception is unobservable under most circumstances (other than some assistive reproductive technique settings), GA is defined as the time from start of the LMP to birth. If the date of the start of the LMP is known, this is not a problem, but significant error in LMP dates is typical in vital statistics data (Platt et al. 2001). Other methods (ultrasound dating, clinical estimates of GA [Joseph et al. 2007]) use characteristics of the fetus or infant to back-calculate the date of LMP. Improved measurement of GA and better methods for understanding errors in GA are the subject of current research (Parker and Schenker 2007). For now, we assume that this is not a problem.

MULTIVARIABLE MODELS

Given these challenges and limitations of approaches that ignore survival time, what approaches are possible?

Fetuses-at-risk (time from conception to birth)

Joseph, in a series of papers, proposed the FAR paradigm as a way to model fetal and infant mortality. As stated in Chapter 13, Yudkin (Yudkin et al. 1987) suggested that the cohort of fetuses in utero prior to a stillbirth should be the denominator in calculation of stillbirth rates, rather than the cohort of births at the same GA as the stillbirth. This was termed *fetuses-at-risk*. Yudkin further noted that the risk of stillbirth calculated using births at the same GA would be biased relative to the underlying risk. Joseph (Joseph et al. 2003, 2004) extended this model to other events, including neonatal and infant mortality. Briefly, the denominator for all deaths was considered to be all fetuses at risk. A baby born at day 273 post-LMP (39 weeks) who died anytime within the first year of life (i.e., an infant death) was compared to the denominator of FAR at day 273, which is all births at day

273 or later. This is equivalent to all fetuses in utero just prior to the *birth* of the index case. Babies born prior to day 273 are removed from this denominator.

Fetuses-at-risk, while identifying the cohort in utero as a relevant denominator, does not completely address the issue pointed out earlier in this chapter, in the section on measurement issues. Recall that in that example, there were three births, at days 266, 273, and 280 days post-LMP, and that all three resulted in death at 280 days post-LMP. All three births were fetuses at-risk at day 266, so were in the denominator for the late neonatal death of that fetus. At day 273, the latter two were in the FAR denominator, whereas at day 280, only the final birth was part of the FAR denominator. So, although they all three suffered events on the same day post-LMP, they are compared to each other only in the first denominator. This problem has been recognized (Joseph 2004) and ascribed to the "obstetrical perspective" of FAR calculations. An obstetrician only has opportunity to monitor and intervene on fetuses at-risk, and once a baby is born it ceases to be in the risk set for obstetric intervention. So, for example, at day 280, the infant born at day 263 is not at risk for obstetric intervention or obstetric outcome; outcomes for this infant that are related to his or her birth are ascribed (in the FAR approach) to day 263.

In the next section, we extend the FAR model to a comprehensive model of conception and mortality; this model addresses the research question proposed at the beginning of this chapter and also addresses many of the challenges that arise in perinatal epidemiology.

Time from conception to death

Time from conception to death is a somewhat unnatural endpoint from the perspective of traditional perinatal epidemiology (Klebanoff and Schoendorf 2004; Wilcox and Weinberg 2004). It requires that we ignore birth as a mediating event and that we consider all deaths (fetal and infant) as equivalent if they occur at the same time. However, it lends itself to directly being able to answer the question posed at the beginning of the chapter. In the example above, all three conceptions have the same time zero and have events on the same day (day 280) and, therefore, the same survival time.

Cox and other survival models for multivariable analyses

The *Cox proportional hazards model for survival data* (Cox 1972) is a well-known standard approach to the analysis of survival data. We do not provide a detailed discussion here, but refer the reader to Chapter 16 and the original Cox paper or textbooks (e.g., Klein and Moeschberger 1997) for details. Briefly, the hazard, or instantaneous rate, of death, is modeled as a function of an unspecified baseline hazard and covariates as

$$h(t,\vec{x}) = h_0(t)e^{\sum \beta_i x_i}$$

where $h(t,\vec{x})$ represents the instantaneous hazard at time t for a subject with characteristics $\vec{x}$, $h_0(t)$ represents the unspecified baseline hazard, and β_i relates the hazard to covariates x_i.

Extended Cox models

Therneau and Grambsch (2000) described extensions of the Cox model, in particular with regard to the β_i. The key assumption of the Cox model is that hazards are proportional; that is, that the ratio of the hazard for two levels of x_i is constant over time. This assumption may be violated, and these authors show how the Cox model can be extended to $\beta_i(t)$, so that the function varies over time. These models can be applied to our problem in the following way: Time from conception is the outcome, and exposure and other covariates are entered in the model, as with any Cox model. In the next subsection, we discuss examples from the literature of this approach.

Examples

Several authors have presented models similar to those proposed in this chapter, for a variety of outcomes.

Smith (2001) used life-table methods to estimate the incidence of perinatal outcomes based on varying denominators, including considering stillbirths as a function of all ongoing pregnancies. He and co-authors (Smith et al. 2003) considered previous cesarean section delivery as a risk factor for unexplained stillbirths with Cox models. Hernandez-Diaz et al. (2002) studied the association between folic acid supplementation and the time to gestational hypertension using the Cox model and computed relative risks. Platt et al. (2004) used a Cox model to assess associations between maternal smoking and maternal race and mortality, showing that time-dependent hazard ratios for these two exposures could be estimated, and that results were consistent with biological and clinical considerations. The Cox model considered included smoking and maternal race as time-fixed covariates, and birth as a time-varying covariate indicating whether the fetus/infant had been born yet.

Time to birth as a covariate

Wilcox and Weinberg noted a number of problems with the above examples using birth as a covariate (Wilcox and Weinberg 2004). They noted that entering birth as a dichotomous variable in a model, as proposed by Platt et al. (2004), ignored the fact that a 38-week infant born yesterday has a higher risk than the 36-week infant born 2 weeks ago because the 2-week-old 36-week infant has survived through the test of labor and delivery. The proposed model, which assumes that they both have the same risk, ignores the peak of mortality near birth and its steep decline thereafter. Platt et al. (2005) proposed a second model, ignoring time of birth and simply modeling the hazard as a function of exposure and covariates. Results of this model, without the covariate for birth, are presented in Figures 14.3 and 14.4, compared to the results including the covariate for birth in Figures 14.5 and 14.6. It is interesting to note some additional conclusions from this figure: The association between race and mortality differs depending on whether birth is controlled for in the model, whereas the association between smoking and mortality is relatively similar across the two models. This suggests different results with respect

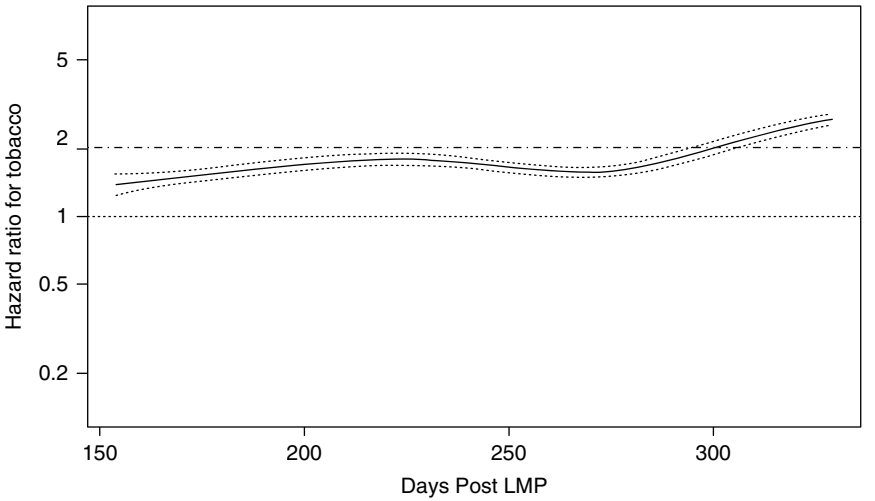


Figure 14.3 Estimated hazard ratio (curved solid line) (with 95% confidence interval [*two dotted lines*]) for black race (relative to white race) as a function of time post–last menstrual period; 1998 U.S. National Center for Health Statistics data, unadjusted for time of birth. The horizontal dashed line represents the estimated fixed hazard ratio.

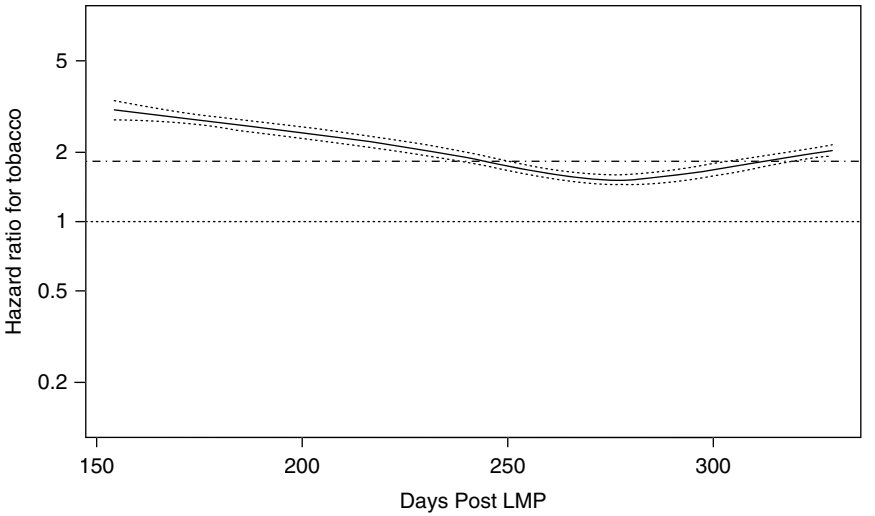


Figure 14.4 Estimated hazard ratio (curved solid line) (with 95% confidence interval [*two dotted lines*]) for maternal smoking at the time of the last menstrual period (LMP) (relative to mothers who did not smoke) as a function of time post-LMP; 1998 U.S. National Center for Health Statistics data, unadjusted for time of birth. The horizontal dashed line represents the estimated fixed hazard ratio.

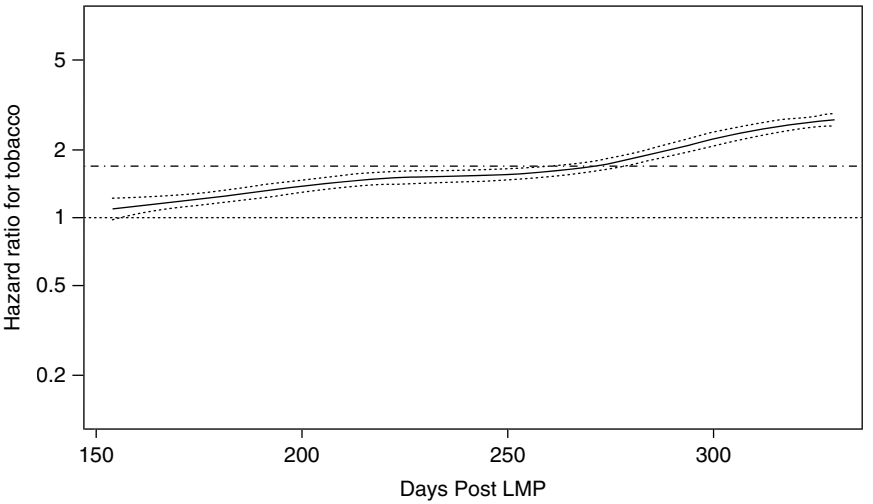


Figure 14.5 Estimated hazard ratio (*curved solid line*) (with 95% confidence interval [*two dotted lines*]) for black race (relative to white race) as a function of time post–last menstrual period; 1998 U.S. National Center for Health Statistics data. The horizontal dashed line represents the estimated fixed hazard ratio.

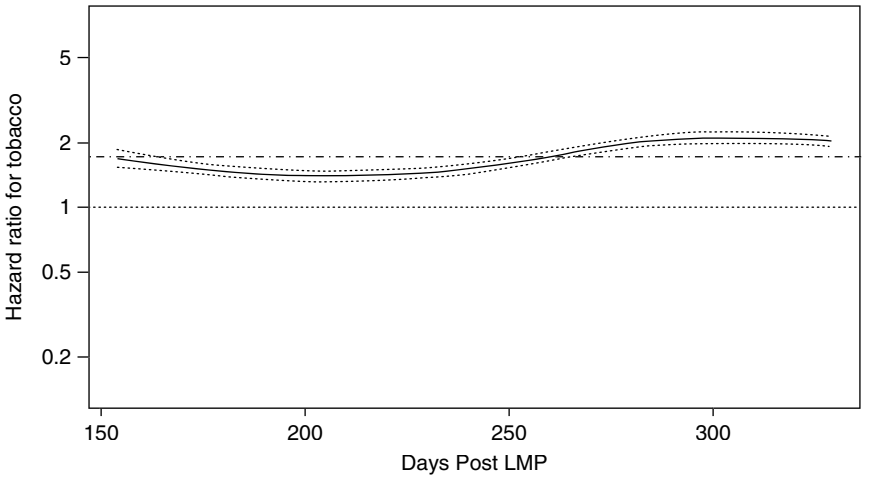


Figure 14.6 Estimated hazard ratio (*curved solid line*) (with 95% confidence interval [*two dotted lines*]) for maternal smoking at the time of the last menstrual period (LMP) (relative to mothers who did not smoke) as a function of time post-LMP; 1998 U.S. National Center for Health Statistics data. The horizontal dashed line represents the estimated fixed hazard ratio.

to mediation by GA of the effect of exposure on outcome; the effect of maternal race may be mediated through a shift in GA, whereas the effect of smoking may be primarily a direct effect.

Other multivariable modeling approaches

One might consider other approaches to analysis of the time from conception to death. Cox models and other hazard-based approaches suffer from important limitations (Hernán et al. 2004; Flanders and Klein 2007) because, in effect, all predictors of outcome must be accounted for in the model. Accelerated failure time models, such as structural nested failure time models (Hernán et al. 2005), provide an alternative approach. Briefly, these models, rather than modeling the hazard function, model the survival time directly. Exposures that increase risk *accelerate failure*—they reduce the time to death. These models can be parametric or semiparametric; examples can be found elsewhere (Hernán et al. 2005). However, these approaches have not been directly applied to perinatal outcomes.

OTHER ISSUES

The problems arising from conditioning on birth that arise in the treatment of mortality as an outcome arise in other areas as well. In this section, I outline some other, closely related problems.

Intrauterine growth–based small-for-gestational age rates

Since their first publication by Lubchenco and colleagues (Lubchenco et al. 1963) over 40 years ago, weight-for-gestational age charts (Kallen 1995; Alexander et al. 1996; Kramer et al. 2001) have been a cornerstone of screening for infants with poor intrauterine development. In these charts, the weight distributions of live births at each week of gestation are converted to percentiles, and any infant whose weight falls below a certain statistical threshold of the population, typically the 10th percentile, is labelled as *small-for-gestational age* (SGA) and considered to be at increased risk of perinatal morbidity and mortality due to intrauterine, or fetal, growth restriction (Arnold et al. 1991). Since fetal growth restriction is typically not measurable in population-based data, the majority of research to identify risk factors for fetal growth restriction comprises comparisons of the risks of SGA, as established through weight-for-GA charts, among exposed and unexposed groups (or alternatively, the case-control approach comparing the characteristics of SGA infants to their appropriate-for-gestational age [AGA] peers). Although case definitions for "small-" or "appropriate-for-gestational age" established using conventional weight-for-GA charts are well accepted in perinatal epidemiology, their validity according to general epidemiologic principles has rarely been considered. In this section, we summarize arguments that size-for-GA charts created only from the weight distributions of live births have serious shortcomings because they are based on the cohort of births; from another perspective, they suffer from

missing data from the weights of fetuses who remain in utero at each GA (Hutcheon and Platt 2008).

Figure 14.7 shows a comparison of the distribution of ultrasound-estimated fetal weights for those fetuses in utero at 32 weeks, to the distribution of birth weights for live births at 32 weeks. The ultrasound-estimated fetal weights are derived from an unselected obstetrical population at the Royal Victoria Hospital, a McGill University teaching hospital in Montreal, Canada (Hutcheon and Platt 2008), and the live birth distribution is derived from a Canadian birth weight reference (Kramer et al. 2001). The median estimated weight of the fetuses still in utero is over 120 g heavier than that of live births, whereas the 10th percentile cutoffs differ by more than 300 g. This discrepancy between the 10th percentile cutoffs means that when the national birth weight reference is applied to the intrauterine population (which, at this age, constitutes >99.7 of the total conception cohort), it will not identify 10% of the remainder of the conception cohort as "SGA." Instead, since the 10th percentile of the national birth weight reference is

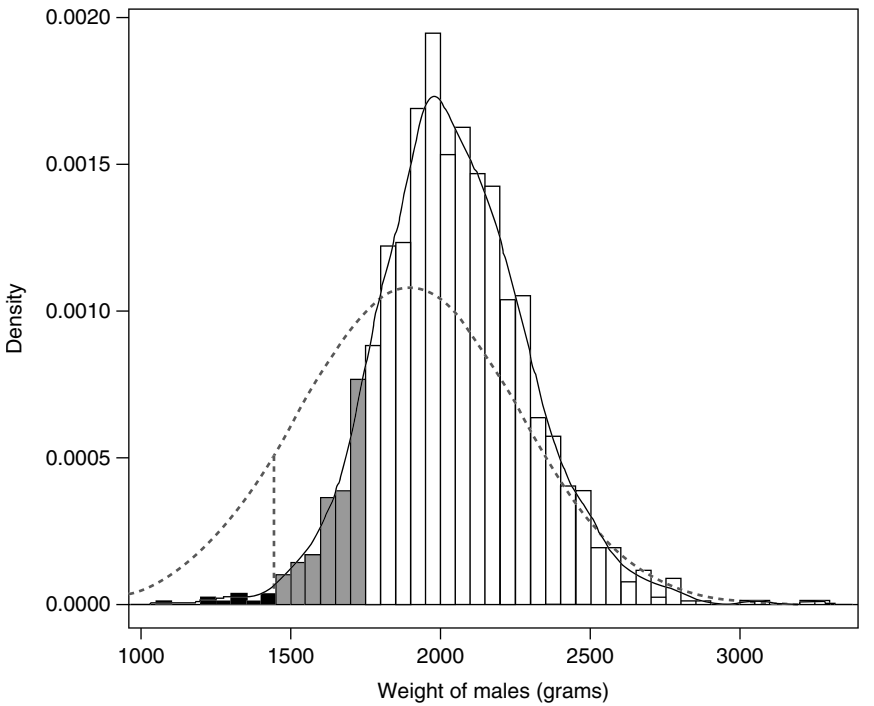


Figure 14.7 Comparison of the estimated fetal weight distributions at 32 weeks of a general obstetrical population at the Royal Victoria Hospital in Montreal, Canada for males (n = 1,540) with the distribution of Canadian live birth weight references for 32 weeks (*dashed lines*). Gray shading indicates the smallest 10% of the in utero population; black shading indicates those in the smallest 10th percent of the in utero population identified as SGA by the live birth reference (SGA threshold indicated by vertical dashed line).

much lower than the 10th percentile of the total cohort, the live birth weight–based reference will identify less than 0.1% of the total cohort as SGA.

This would not pose a problem if those who were born at 32 weeks were a random sample of those in utero just prior to the event. Hutcheon and Platt (2008) noted that this is, essentially, a statistical missing data problem. However, it is not a simple one. At preterm ages, the missing data in neonatal weight references that arise due to the birth process and GA are clearly not *missing-completely-at-random* (MCAR), making the current approach of using available data inappropriate. This is analogous to a "complete-case" approach in typical missing data problems, which is known to be inappropriate unless the missingness is MCAR. If the distribution of missing, or hidden (Paneth 2008) intrauterine weights were similar to that of the available birth weight data within strata of measured predictors the data would be *missing-at-random* (MAR). With MAR data, approaches such as multiple imputation (Schafer 1999) are appropriate and could be used to account for the hidden data. However, since our ability to explain the hidden data, which is equivalent to predicting birth among a cohort in utero, is generally agreed to be poor, even considering all known social and medical risk factors (Buekens and Klebanoff 2001), there likely remain important unknown predictors of GA, and the hidden data are likely not MAR. The hidden data in neonatal weight references are therefore likely *missing-not-at-random* (MNAR), meaning that the missingness process depends on unobserved variables, and any weight-for-GA reference must take this missing data mechanism into account. On the other hand, although using the live birth data alone is unlikely to be fruitful, an approach using data on fetal size measures may be of use.

REFERENCES

Alexander, G.R., Himes, J.H., Kaufman, R.B., Mor, J., and Kogan, M. 1996. A United States national reference for fetal growth. *Obstet Gynecol* 87(2): 163–68.

Arnold, C.C., Kramer, M.S., Hobbs, C.A., McLean, F.H., and Usher, R.H. 1991. Very low birth weight: A problematic cohort for epidemiologic studies of very small or immature neonates. *Am J Epidemiol* 134(6): 604–13.

Buck, G.M., Lynch, C.D., Stanford, J.B., Sweeney, A.M., Schieve, L.A. Rockett, J.C., Selevan, S.G., and Schrader, S.M. 2004. Prospective pregnancy study designs for assessing reproductive and developmental toxicants. *Environ Health Perspect* 112(1): 79–86.

Buekens, P. and Klebanoff, M. 2001. Preterm birth research: From disillusion to the search for new mechanisms. *Paediatr Perinat Epidemiol* 15(Suppl 2): 159–61.

Cox, D.R. 1972. Regression models and life-tables. *J R Stat Soc B-Statistical Methodology* 34(2): 187.

Flanders, W.D., and Klein, M. 2007. Properties of 2 counterfactual effect definitions of a point exposure. *Epidemiology* 18(4): 453–60.

Hernán, M.A., Cole, S.R., Margolick, J., Cohen, M. and Robins, J.M. 2005. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol Drug Saf* 14(7): 477–91.

Hernán, M.A., Hernandez-Diaz, S., and Robins, J.M. 2004. A structural approach to selection bias. *Epidemiology* 15(5): 615–25.

Hernandez-Diaz, S., Werler, M.M., Louik, C., and Mitchell, A.A. 2002. Risk of gestational hypertension in relation to folic acid supplementation during pregnancy. *Am J Epidemiol* 156(9): 806–12.

Hutcheon, J.A. and Platt, R.W. 2008. The missing data problem in birth weight percentiles and thresholds for small-for-gestational age. *Am J Epidemiol* 167(7): 786–92.

Joseph, K.S. 2004. Incidence-based measures of birth, growth restriction, and death can free perinatal epidemiology from erroneous concepts of risk. *J Clin Epidemiol* 57(9): 889–97.

Joseph, K.S., Huang, L., Liu, S., Ananth, C.V., Allen, A.C., Sauve, R., and Kramer, M.S. 2007. Reconciling the high rates of preterm and postterm birth in the United States. *Obstet Gynecol* 109(4): 813–22.

Joseph, K.S., Liu, S., Demissie, K., Wen, S.W., Platt, R.W., Ananth, C.V., et al. 2003. A parsimonious explanation for intersecting perinatal mortality curves: Understanding the effect of plurality and of parity. *BMC Pregnancy Childbirth* 3 (1): 3.

Kallen, B. 1995. A birth weight for gestational age standard based on data in the Swedish Medical Birth Registry, 1985–1989. *Eur J Epidemiol* 11: 601–606.

Klebanoff, M.A. and Schoendorf, K.C. 2004. Invited commentary: What's so bad about curves crossing anyway? *Am J Epidemiol* 160(3): 211–12.

Klein, J.P. and Moeschberger, M.L. 1997. *Survival analysis: Techniques for censored and truncated data*. New York: Springer-Verlag.

Kramer, M.S., Goulet, L., Lydon, J., Seguin, L., McNamara, H., Dassa, C., et al. 2001. Socio-economic disparities in preterm birth: Causal pathways and mechanisms. *Paediatr Perinat Epidemiol* 15(Suppl 2): 104–23.

Kramer, M.S., Liu, S., Luo, Z., Yuan, H., Platt, R.W., and Joseph, K.S. 2002. Analysis of perinatal mortality and its components: Time for a change? *Am J Epidemiol* 156(6): 493–97.

Kramer, M.S., Platt, R.W., Wen, S.W., Joseph, K.S., Allen, A., Abrahamowicz, M., et al., for the Fetal/Infant Health Study Group of the Canadian Perinatal Surveillance. 2001. A new and improved population-based Canadian reference for birth weight for gestational age. *Pediatrics* 108(2): E35.

Kramer, M.S., Wilkins, R., Goulet, L., Seguin, L., Lydon, J., et al. 2009. Investigating socio-economic disparities in preterm birth: Evidence for selective study participation and selection bias. *Paediatr Perinat Epidemiol* 23(4): 301–09.

Lubchenco, L.O., Hansman, C., Dressler, M., and Boyd, E. 1963. Intrauterine growth as estimated from liveborn birth-weight data at 24 to 42 weeks of gestation. *Pediatrics* 32: 793–800.

MacDorman, M.F. and Atkinson, J.O. 1999. Infant mortality statistics from the 1997 period linked birth/infant death data set. *Natl Vital Stat Rep.* 47(23): 1–23.

Paneth, N. 2008. Invited commentary: The hidden population in perinatal epidemiology. *Am J Epidemiol* 167(7): 793–96; author reply 797–98.

Parker, J.D. and Schenker, N. 2007. Multiple imputation for national public-use datasets and its possible application for gestational age in United States Natality files. *Paediatr Perinat Epidemiol* 21(Suppl 2): 97–105.

Platt, R.W., Abrahamowicz, M., Kramer, M.S., Joseph, K.S., Mery, L., Blondel, B., et al. 2001. Detecting and eliminating erroneous gestational ages: A normal mixture model. *Statistics Med* 20: 3491–03.

Platt, R.W., Joseph, K., Ananth, C.V., and Kramer, M.S. 2005. *Mechanism underlying effects of smoking and race on fetal and infant mortality: Application of the extended Cox model*. Society for Paediatric and Perinatal Epidemiology annual meeting, Toronto, Ontario.

Platt, R.W., Joseph, K.S., Ananth, C.V., Grondines, J., Abrahamowicz, M., and Kramer, M.S. 2004. A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death. *Am J Epidemiol* 160(3): 199–206.

Rothman, K.J., Greenland, S., and Lash, T. 2008. *Modern epidemiology*. Philadelphia: Lippincott Williams & Wilkins.

Schafer, J.L. 1999. Multiple imputation: A primer. *Stat Methods Med Res* 8(1): 3–15.

Shiono, P.H., Klebanoff, M.A., and Rhoads, G.G. 1986. Smoking and drinking during pregnancy. Their effects on preterm birth. *JAMA* 255(1): 82–84.

Smith, G.C. 2001. Life-table analysis of the risk of perinatal death at term and post term in singleton pregnancies. *Am J Obstet Gynecol* 184(3): 489–96.

Smith, G.C., Pell, J.P., and Dobbie, R. 2003. Caesarean section and risk of unexplained stillbirth in subsequent pregnancy. *Lancet* 362(9398): 1779–84.

Therneau, T.M. and Grambsch, P.M. 2000. *Modeling survival data: Extending the Cox model*. New York: Springer.

Wilcox, A.J. 1993. Birth weight and perinatal mortality: The effect of maternal smoking. *Am J Epidemiol* 137(10): 1098–1104.

Wilcox, A.J. and Weinberg, C.R. 2004. Invited commentary: Analysis of gestational-age-specific mortality—on what biologic foundations? *Am J Epidemiol* 160(3): 213–14.

Wilcox, A.J., Weinberg, C.R., O'Connor, J.F., Baird, D.D., Schlatterer, J.P., Canfield, R.E., et al. 1988. Incidence of early loss of pregnancy. *N Engl J Med* 319(4): 189–94.

Yudkin, P.L., Wood, L., and Redman, C.W.G. 1987. Risk of unexplained stillbirth at different gestational ages. *Lancet* 1(8543): 1192–94.

15

Causal Inference in Perinatal Epidemiology: Revisiting the Birth Weight Paradox

ENRIQUE F. SCHISTERMAN AND ROBERT W. PLATT

Does smoking cause or prevent neonatal mortality? What does it mean when we say that a mother's smoking during pregnancy causes her baby increased mortality risk? These are examples of questions of interest, and to answer these one has to consider the mortality risk of a baby whose mother smoked. What if we could go back in time and change only the mother's smoking status, leaving all other variables unchanged? This is the simplified definition of a *counterfactual*, and under this counterfactual (i.e., no smoking during pregnancy, but all else equal), what would the baby's mortality risk be? If the mortality risk when the mother did not smoke is higher than the mortality risk when the mother smoked, we say that smoking caused reduced mortality risk. Note that causation involves changes in time. A change in an exposure causes a change in an outcome that occurs after the exposure.

Understanding causation—how a change in one variable leads to a subsequent change in another variable—is essential in reproductive and perinatal epidemiology, because an understanding of causation informs the creation of effective interventions aimed at improving human health. A classic question in epidemiology is whether maternal smoking actually causes a reduced mortality risk in a baby born with a low birth weight. Should smoking be recommended for mothers who might have low-birth-weight (LBW) babies? Causal analysis can address this issue directly. Thus, pitfalls in causal analysis can have real-world implications.

Great developments in the field of causal thinking have occurred during the last 20 years. In particular, causal diagrams in the form of *directed acyclic graphs* (DAGs) have changed the way we approach causal inference by helping investigators to more clearly define causal questions. This chapter provides an overview of causal thinking and, more specifically, the use of DAGs in helping to design etiologically oriented reproductive and perinatal epidemiologic research. By design, the chapter does not provide complete details on performing causal analysis. Such analyses are considered elsewhere (Hernán and Robins 2010).

To illustrate the relevance of causal analysis to reproductive and perinatal epidemiology, we use the classic *birth weight paradox* (Yerushalmy 1972; Wilcox 2001),

or the so-called crossing of perinatal or infant mortality curves (Chapter 12). We use DAGs to describe a causal paradigm for this problem, inclusive of a framework for illustrating results.

Birth weight has long been identified as a robust predictor of perinatal and infant mortality (Wilcox and Russell 1983), as discussed in Chapter 12. When investigators are interested in assessing causal effects of other risk factors for mortality, the issue of how to correctly model birth weight is immediately raised, given its known strong inverse relation with infant mortality and morbidity. This has prompted many investigators to automatically control for birth weight when assessing other risk factors for infant mortality (Wilcox and Russell 1990; Buekens and Wilcox 1993; Wilcox 1993; Ananth and Wilcox 2001) without considering the inherent causal assumptions that such adjustment introduces into the estimation process.

A sizable body of evidence has reported a crossover in the birth weight–specific mortality curves, regardless of the risk factor of interest, when adjusting or stratifying by birth weight, as depicted in Figure 15.1A based upon data from the 1995 U.S. linked birth and infant death file (Macdorman and Atkinson 1998). The crossing of curves represents the so-called birth weight paradox, the phenomenon in which LBW infants in high-mortality-risk populations have lower mortality than do LBW infants in low-risk populations, or the phenomenon in which two populations have similar mortality risks but different mortality rates among LBW infants. This paradox has been a source of controversy for decades (Yerushalmy 1972; Wilcox 2001), in part, given the lack of a plausible biological explanation for the apparent paradox. For example, the Colorado neonatal mortality rate is very similar to the United States (U.S.) neonatal mortality rate overall, so one

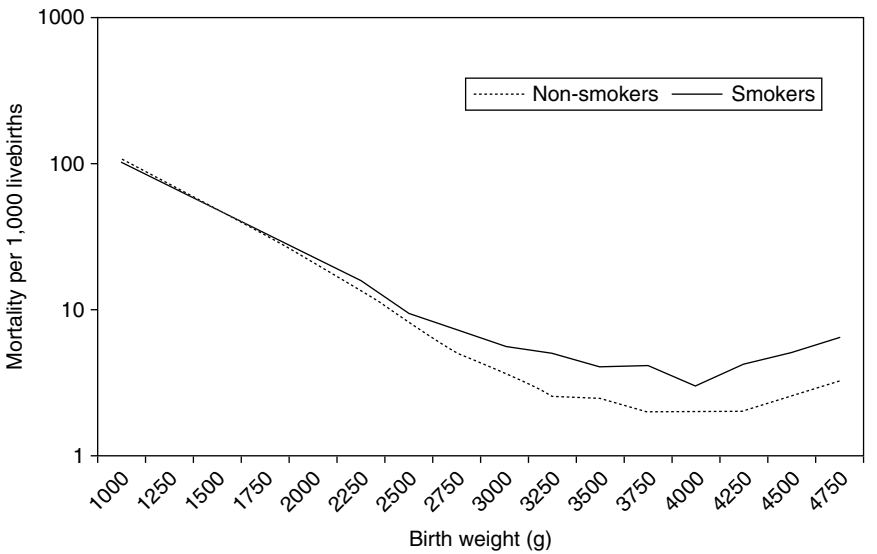


Figure 15.1A Birth weight–specific infant mortality curves for smokers and nonsmokers.

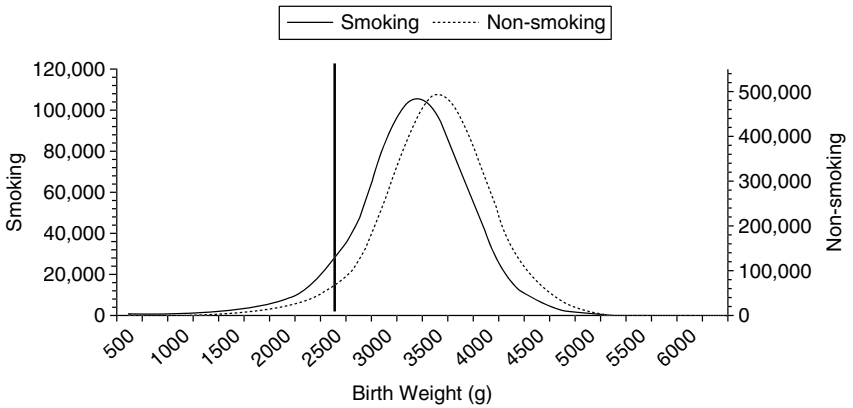


Figure 15.1B Birth weight distribution for smokers and nonsmokers. The line at 2,500 g indicates the cutoff point used for low birth weight.

might expect the Colorado LBW infant mortality rate to be similar to the U.S. LBW infant mortality rate. However, Colorado's LBW infant mortality rate is lower than that of the U.S. (Wilcox 1993). Similarly, twins have higher perinatal mortality rates than do singletons, and yet LBW twins are reported to have lower mortality than LBW singletons (Buekens and Wilcox 1993). In addition, LBW infants born to mothers who smoked cigarettes have been reported to have lower mortality than LBW infants born to nonsmokers (Wilcox 1993). Some authors have suggested that the effect of maternal smoking is modified by birth weight in such a way that smoking is beneficial for LBW babies (Yerushalmy 1972).

Various approaches have been suggested for addressing the birth weight paradox, including replacing absolute birth weight with a relative measure of birth weight. For example, some authors advocate using birth weight percentiles rather than birth weights, allowing each population to have its own birth weight distribution, and each member of a population to have his or her population-specific percentile birth weight calculated. When using percentiles, mortality curves for African-Americans and European-Americans no longer show the crossover pattern (Hertz-Picciotto and Din-Dzietham 1998). Joyce and Peacock (2003) evaluated several proposed strategies for adjusting a population's mortality rate for its birth weight distribution to determine which, if any, methods yielded adjusted mortality rates that were uncorrelated with characteristics of birth weight distributions. Methods included indirect standardization by birth weight, by percentiles, or by z-scores, and regression methods. The authors reported indirect standardization by birth weight to be optimal relative to the other approaches considered. Hertz-Picciotto (2003) questioned the effectiveness of this comparison, given the rather large percentile groupings. The lingering question regarding the ideal birth weight adjustment method may be less of an issue with the use of DAGs, whose application has demonstrated that bias may entirely account for the crossing of the curves.

In the following sections of this chapter, we introduce the causal theory of DAGs and *collider stratification bias*. We then discuss the impact of adjusting for variables when estimating total effects and when estimating direct and indirect effects. Later, we return to the birth weight paradox, applying DAGs to help estimate total, direct, and indirect effects and to explain why *z*-scores remove the paradoxical crossing of curves. By explaining the gestational age paradox using DAGs, we demonstrate that issues relevant to the birth weight paradox can arise again with other outcome or exposure variables. We further use DAGs to explain a causal definition of overadjustment. Finally, we discuss the limitations of DAGs by demonstrating how DAGs and other graphical methods complement each other.

USING CAUSAL DIAGRAMS TO FRAME RESEARCH QUESTIONS

Directed acyclic graphs (DAGs) are diagrams that represent causal relations between variables (Pearl 1995; Greenland et al. 1999; Hernán et al. 2002). Investigators use their expert knowledge to create DAGs, proposing hypothetical causal networks linking risk factors and outcomes, such as maternal cigarette smoking, birth weight, and infant mortality. Directed acyclic graphs have their own terminology, with precise meaning. Specifically, DAGs link variables (nodes) by arrows (directed edges) that represent the hypothesized direct causal effects (protective or causative) of one variable on another. Directed acyclic graphs are directed, meaning that their edges have direction. Sometimes a graph may have an undirected edge, as noted by a dashed line linking variables, which represents a noncausal association between the variables. Strictly speaking, such graphs are not directed, but in this text we will call them DAGs, provided that they include directed edges. A *path* is a sequence of adjacent connected nodes, regardless of the direction of edges connecting the nodes. A *directed path* has only edges that point in the same direction, whereas a *backdoor path* has edges that include an arrow to the initiating variable. Finally, DAGs are acyclic, meaning that there is no directed path from any given variable back to itself, so that cause precedes effect.

The absence of an arrow between two variables indicates that the investigator believes there is no direct effect (i.e., a causal effect not mediated through other variables in the DAG) of one variable on the other (Greenland et al. 1999; Hernandez-Diaz et al. 2006). For a DAG to represent a causal diagram, all common causes of variables in the DAG must be included in the DAG; the absence of such a common cause indicates the absence of confounding (Hernán et al. 2002). Last, a square around a node in a DAG indicates control for the variable represented by the node through adjustment, restriction, or stratification.

The causal relations in a DAG indicate which pairs of variables are expected to be statistically associated. Paths between two variables that contribute to the association are deemed *open paths*, whereas those that do not contribute to the association are called *closed paths*. In the absence of control variables, only

directed paths between two variables are open, contributing to observed associations between those variables. Controlling for a node in a directed path removes its contribution to the association between any variable preceding the node and any variable after the node. Removing its contribution to an association is called *blocking the path*, whereas inducing a path to contribute to association (as described in the next section) is called *opening the path*.

One key feature of DAGs is that they depend on the investigator's assumptions; therefore, there is no "correct" DAG for any research question. Rather, various DAGs are plausible, depending upon what is known about the question of interest. The key element is that DAGs encourage the investigator to formalize his or her set of assumptions. Each DAG reflects a set of assumptions that the investigator is willing to accept. Investigators may legitimately disagree on assumptions and, therefore, on the structure of the DAG. Such assumptions impact the design and implementation of the analytic plan, as well as the interpretation of the findings.

COLLIDER STRATIFICATION BIAS

When variables A and B each cause a third variable C, C becomes a collider on any path that includes A-C-B. In such a DAG, a collider must have at least two arrows pointing to it, as demonstrated by variable C in Figure 15.2. In Figure 15.3A, for example, birth weight is a collider on the path labeled "altitude-birth weight-smoking" and on the path labeled "altitude-birth weight-smoking-neonatal mortality."

A collider blocks the path it is on, whereas controlling for a collider opens the path by inducing an association between the collider's direct causes, the variables that have arrows pointing to the collider (Greenland et al. 1999). The induced association is equivalent to mistakenly adding a path between the collider's causes when no causal relation exists between the causes. When controlling for a collider leads to such a spurious path between exposure and outcome, or from another perspective opens a backdoor path between the collider and outcome, the estimator of the exposure's effect on outcome becomes biased. This bias has been referred to as *selection bias* (Hernán et al. 2004) or collider stratification bias (Greenland 2003). In Figure 15.3B, controlling for birth weight induces an association between altitude and smoking, creating a spurious altitude-smoking-mortality path; that is, it opens the birth weight-smoking-mortality backdoor path that biases the estimator of smoking's effect on mortality.

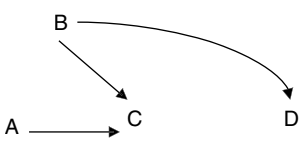


Figure 15.2 Collider C on Paths A-C-B and A-C-B-D.

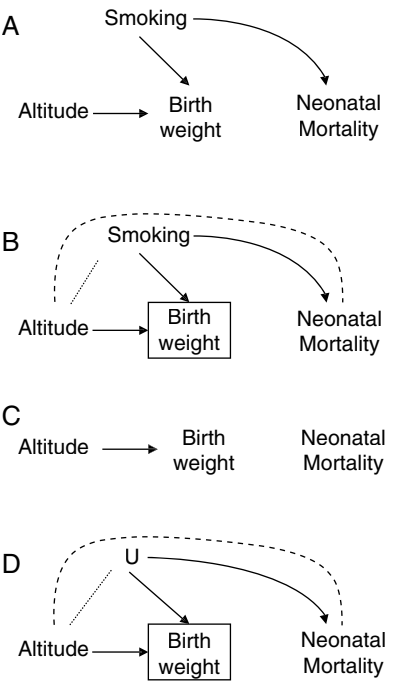


Figure 15.3 Estimating the total effect of altitude on mortality.

ESTIMATING EFFECTS

The total effect of a variable A on a variable C is the change in C caused by a change in A, through all paths that lead from A to C. An indirect effect of A on C arises due to a directed path from A to C that contains at least one other variable, whereas the direct effect of A on C (relative to the indirect path through B) is due to a directed path from A to C that contains only A and C. The paths leading to the total effect of A on C, the direct effect of A on C, or the indirect effect of A on C are illustrated in Figure 15.4.

Estimating Total Effects

Inhabitants of the state of Colorado reside in areas of high altitude. Given the inverse relation between altitude and birth weight, and assuming that lower birth weight leads to a higher infant mortality risk, one may erroneously predict that mortality would be higher in the Colorado birth cohort in comparison to the U.S. birth cohort. Some investigators have postulated that altitude is not causally related to mortality, but has a direct effect on birth weight, which is not causally related to neonatal mortality as depicted in Figure 15.3C.

Under this scenario, if one is interested in estimating the total effect (or crude effect) of altitude on neonatal mortality, one should be able to estimate the effect using common regression techniques like log-linear models. Under this causal

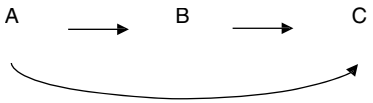


Figure 15.4 Total effect of A on C, direct effect of A on C, and indirect effect of A on C through B.

DAG, no association is expected between altitude and neonatal mortality. Like the crude estimates, the inclusion of the adjustment for birth weight, although unnecessary, should yield unbiased estimates of the total effect of altitude on mortality, which in this case would be null.

Empirical evidence suggests that smoking has been linked to neonatal mortality by increasing the rate from 805 per 100,000 live births to 1,235 per 100,000 live births (i.e., by 430 per 100,000 live births; Figure 15.1A) (Hernandez-Diaz et al. 2006). In addition, smoking reduces birth weight by about 200 g on average (Figure 15.1B) (Hernandez-Diaz et al. 2006) This scenario can be depicted by the DAG in Figure 15.3A. Under this scenario, if one is interested in estimating the total effect (or crude effect) of altitude on neonatal mortality, one should be able to estimate this effect using common regression techniques like logistic log-linear models. In contrast to the previous scenario, adjusting only for birth weight will introduce collider stratification bias for the altitude–mortality association. Adjusting for birth weight induces an association between altitude and smoking in at least one stratum of birth weight (e.g., LBW infants born to mothers who did not smoke are more likely to reside at high altitude). Adjusting for the collider birth weight opens the birth weight-smoking-mortality backdoor path, thereby, creating a spurious altitude-smoking-mortality directed path from altitude to mortality. Controlling for smoking blocks this spurious path. If only birth weight is accounted for, however, the results would be biased away from the null, as depicted in the DAG in Figure 15.3B.

The model adjusting for birth weight would yield biased estimates for the total effect of altitude on mortality, whereas the crude model (unadjusted) would have yielded unbiased estimates. Alternatively, adjusting for smoking status and birth weight would also yield unbiased estimates, because adjusting for smoking status would remove the collider stratification bias introduced by adjusting for birth weight. In modeling terms, where we assume linear relationships among variables and μ represents the rate of neonatal mortality, either,

$$\log(\mu) = \alpha + \beta_1 \, Altitude \qquad \text{(Eq. 1)}$$

or

$$\log(\mu) = \alpha + \beta_1 \, Altitude + \beta_2 \, Birth\ weight + \beta_3 \, Smoking, \qquad \text{(Eq. 2)}$$

would yield unbiased estimates of the total effect of altitude on neonatal mortality. However, the model

$$\log(\mu) = \alpha + \beta_1 \, Altitude + \beta_2 \, Birth\ weight, \qquad \text{(Eq. 3)}$$

would yield biased estimates and would make it appear as if altitude had a direct effect on mortality not mediated by birth weight.

It is worth noting that most investigators recognize that many relevant covariates are unmeasured; hence, the DAG may not be fully specified. In this scenario, assuming that smoking went unmeasured, adjusting for it (applying Equation 2) would have been impossible and, therefore, it would have been infeasible to remove the bias when also adjusting for birth weight. This would be true for any variable that causes both birth weight and neonatal mortality, as exemplified by *U* in the DAG in Figure 15.3D. Thus, the only realistic and parsimonious option for estimating the total effect of altitude on neonatal mortality, or any other variable of interest, would be to model the total effect using the crude or unadjusted approach (Equation 1).

Estimating Direct and Indirect Effects

Alternatively, other investigators might believe that a biological causal link exists between birth weight and neonatal mortality. In this case, any variable that affects both birth weight and neonatal mortality could have a direct and indirect effect (via birth weight) on neonatal mortality. In DAG terms, this would be represented in Figure 15.5A.

Given this hypothesized causal structure, if the question of interest is the total effect (the sum of the indirect and direct effects) of smoking on neonatal mortality, then the estimators of β_1 via the model:

$$\log(\mu) = \alpha + \beta_1 \, Smoking, \qquad \text{(Eq. 4)}$$

would offer the most appropriate answer. On the other hand, if the question of interest is the effect of smoking not mediated through birth weight, then in the model:

$$\log(\mu) = \alpha + \beta_1 \, smoking + \beta_2 \, birth\ weight, \qquad \text{(Eq. 5)}$$

the parameter β_1 represents the direct effect of smoking on neonatal mortality, and β_2 helps characterize the indirect effect of smoking on neonatal mortality. However, in the very possible yet untestable scenario in which an unmeasured variable is present, as represented in Figure 15.5B, the estimation of direct and indirect effects becomes biased (Cole and Hernán 2002).

To estimate direct and indirect effects of smoking on mortality, one would typically adjust for birth weight. However, birth weight is a collider in this case (between smoking and *U*), so that, when adjusted for it creates a spurious path between the unmeasured variable *U* and smoking and, hence, a spurious path between smoking and mortality, as depicted in Figure 15.5C. Collider stratification biases the estimator of smoking's direct effect on mortality. Use of DAGs demonstrates that the direct and indirect effects cannot be estimated with the available data, irrespective of adjusting for birth weight, without strong and untestable assumptions about the absence of common causes (Robins and Greenland 1992; Cole and Hernán 2002; Kaufman et al., 2004, 2005).

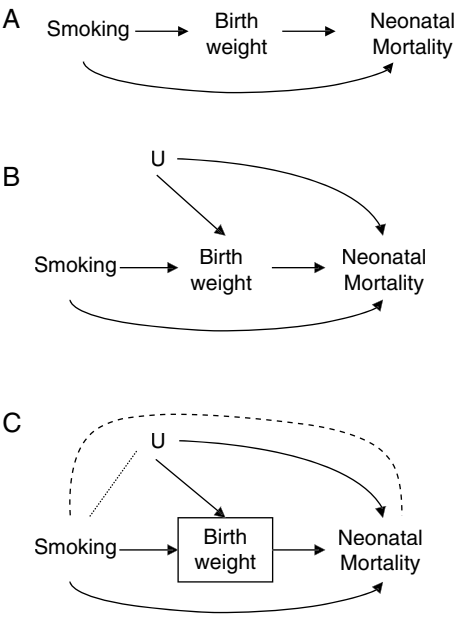


Figure 15.5 Estimating direct and indirect effects of smoking on mortality through birth weight.

APPLYING DAGS TO UNDERSTAND BIRTH WEIGHT Z-SCORES' ROLE IN THE BIRTH WEIGHT PARADOX

Recognizing that adjustment for birth weight was problematic, and some authors adjusted for *z*-scores within strata of another risk factor of interest, such as smoking (Wilcox and Russell 1990; Wilcox 2001). *Z*-scores are calculated within strata in the following way:

$$z\text{-}score\ bw_i = \frac{bw_{ij} - \overline{bw_j}}{SD\left(bw_j\right)}, \qquad \text{(Eq. 6)}$$

where *j* represents strata 1 through *k* of another risk factor (like smoking); bw_{ij} is the birth weight of the *i*th individual in the *j*th stratum; $\overline{bw_j}$ is the estimated mean birth weight for the *j*th stratum; and SD(bw_j) is the standard deviation of birth weight in the *j*th stratum. Use of this approach presumably removes the crossing of the curves and produces unbiased results, as displayed in Figure 15.6. A modified approach of this method also has been proposed, in which the within-strata percentiles (rather than *z*-scores) are utilized to uncross the curves (Hertz-Picciotto and Din-Dzietham 1998). However, no analytical or empirical evaluation of this approach has been undertaken. Do *z*-scores that uncross the curves really yield unbiased estimators of the direct, indirect, and/or the total effects? Let us revisit the maternal smoking, birth weight, and neonatal mortality scenarios using DAGs to understand how birth weight *z*-scores uncross mortality curves.

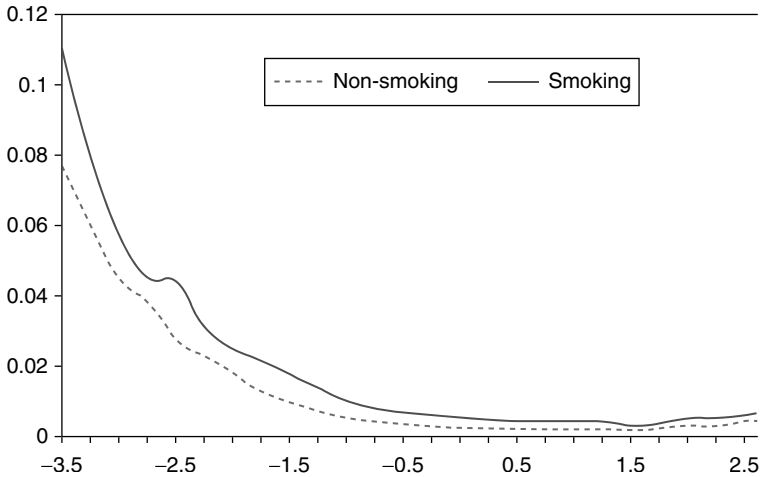


Figure 15.6 Standardized birth weight and neonatal mortality by smoking status.

Let us consider the scenario in Figure 15.7A, in which smoking has a direct effect on birth weight and mortality, but birth weight itself does not directly affect mortality, and *z*-scores are a function of smoking status and birth weight. If one is interested in estimating the total effect (or crude effect) of smoking on neonatal mortality, one should be able to estimate this effect using common regression techniques like log-linear models. Adjusting for birth weight *z*-scores will not introduce bias, because adjustment for *z*-scores does not open or close any paths between smoking and mortality.

In modeling terms, the total effect can be estimated without bias either by the unadjusted model or by the model adjusted for the birth weight *z*-scores, as shown in Equations 7 and 8, respectively.

$$\log\left(\mu\right) = \alpha + \beta_1 \; smoking \tag{Eq. 7}$$

$$\log\left(\mu\right) = \alpha + \beta_1 \; smoking + \beta_2 \; z\text{-}score\; birth\; weight \tag{Eq. 8}$$

Therefore, when one is interested in estimation of the total effect of smoking—the sum of the indirect effect through birth weight and the direct effect—on neonatal mortality, neither birth weight nor *z*-scores need to be in the model. Thus, we now limit our focus to whether inclusion of a regression term for birth weight *z*-scores in models of neonatal mortality allows for decomposition of total effects into direct and indirect effects.

Using *z*-scores is comparable to variable transformation. Although it is clear that both birth weight and exposure (e.g., smoking) have a deterministic relationship with individual *z*-scores, as shown in Figure 15.7B, the effect of the transformation is to create a birth weight variable that is independent of exposure status.

It may be demonstrated in datasets with information on smoking, birth weight, and neonatal mortality, and in which a relation between birth weight and neonatal

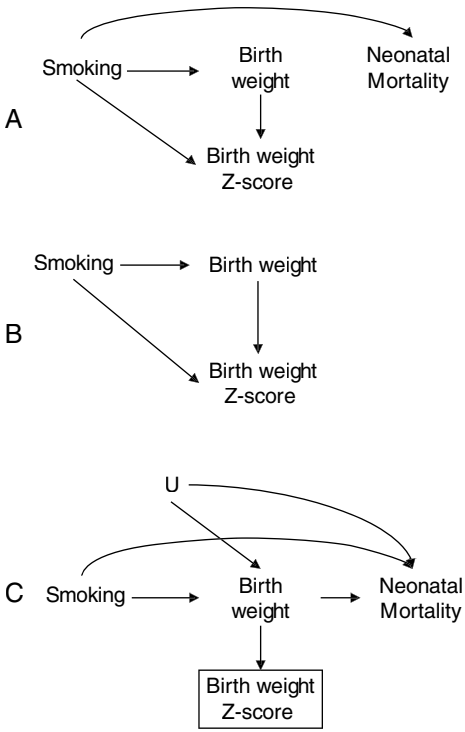


Figure 15.7 Birth weight z-scores are actually independent of smoking.

mortality is observed, that z-scores retain an association with birth weight and neonatal mortality but are independent of smoking status. To understand the independence of z-scores and smoking, one must assume that birth weight distributions for mothers who do and do not smoke are normally distributed. The transformation applied to birth weight to obtain birth weight z-scores makes the birth weight z-score a normally distributed variable with a mean of 0 and a standard deviation of 1. The birth weight z-score for smokers would then have the same distribution as the birth weight z-score for nonsmokers; the z-scores would both be normally distributed with a mean of 0 and a standard deviation of 1. A given infant would have a z-score that is the same distance from the mean smoking birth weight z-score as it is from the mean nonsmoking birth weight z-score, so that knowing an infant's birth weight z-score would not reveal any information about whether the infant's mother smoked during the pregnancy. Since the z-score is a measure of distance from the mean, it remains independent of the infant's smoking stratum. Figure 15.7C shows this independence, as well as the possibility that birth weight is a collider.

Figure 15.7C reflects that the z-score retains its relation with birth weight and will be related to the outcome whenever birth weight is related to the outcome. Effect estimates for z-scores will differ from those of birth weight as a matter of

the different scale due to transformation. One can calculate the exact relation between the *z*-score coefficient in the regression of outcome on *z*-score and the birth weight coefficient in the regression of outcome on birth weight (Schisterman et al. 2009).

Note that Figure 15.7C represents a situation that has been described as *unfaithfulness*, which is said to occur when the DAG does not represent the true causal structure (Pearl 2000). In this case, the smoking variable was used to perform the transformation, implying the presence of an arrow between smoking and the *z*-score, as shown in Figure 15.7B, which represents the calculation of the *z*-score. However, the transformed birth weight variable *z*-score is independent of smoking, as if there were no arrow between the two nodes, as shown in Figure 15.7C.

The fact that smoking is uncorrelated with birth weight *z*-score has important implications. First, the independence of smoking and birth weight *z*-scores suggests that estimates from smoking-only models and those adjusted for *z*-scores should be equal, as noted in the following formula:

$$\text{smoking} \perp\!\!\!\perp \text{z-score} \Rightarrow \mathrm{E}\left(\mathrm{Y}\middle|\text{smoking}\right) = \mathrm{E}\left(\mathrm{Y}\middle|\text{smoking, z-score}\right)$$

where $\perp\!\!\!\perp$ denotes independence. Accordingly, the DAG suggests that the use of *z*-scores should be effective in avoiding collider stratification bias. Specifically, the backdoor pathway opened by stratifying on birth weight no longer exists when stratifying by *z*-score. Hence, stratifying by *z*-score does not bias smoking's total effect estimate. Second, because the *z*-score and smoking are independent, controlling for the *z*-score does not allow for estimation of smoking's direct effect.

Adjusting for *z*-scores in regression models yields unbiased estimates for the total effect of the primary factor of interest, whether or not birth weight represents a collider. These results match results from unadjusted models, and *z*-score adjustment offers no advantage over unadjusted models. We have shown that neither birth weight nor *z*-scores may be used for effect decomposition, and that the true utility of *z*-scores is in estimating the unbiased total effects of exposures, even when collider stratification would adversely impact estimates from birth weight–adjusted models. The ability of *z*-scores to successfully remove the paradoxical crossing of the curves is due to an alteration in the causal parameter being estimated. Instead of adjusting for confounding or estimating a direct effect of smoking on mortality, using *z*-scores estimates the total effect of exposure on outcome.

APPLYING DAGS TO UNDERSTAND THE GESTATIONAL AGE PARADOX

Our lack of understanding regarding the complex relations between birth weight, gestational age, and perinatal mortality stems from a theoretical blending of two distinct etiologic pathways—biologic maturity and fetal growth (Wilcox 2001; Hertz-Picciotto 2003; Ananth and Platt 2004), as discussed in Chapter 8 and 9. Thus, some investigators have attempted to adjust for gestational age at birth (GA) rather than birth weight or *z*-scores when assessing the effect of maternal smoking

on neonatal mortality (Hertz-Picciotto and Din-Dzietham 1998; Wilcox 2001). The reasoning behind using GA is that it is a more accurate measure of fetal maturity, rather than fetal growth which birth weight is thought to represent. Although GA is a major determinant of birth weight, it is only one of several contributors, including those that are difficult to measure or are currently unknown, all of which produce variation in birth weight within each GA stratum.

Gestational age and birth weight are among the strongest predictors of fetal and infant mortality, and both are affected by many common exposures such as maternal smoking. Analyses of the effect of smoking on mortality are often adjusted for GA, even though GA is affected by smoking and shares a common cause with the outcome. Adjustment for GA is often done to avoid comparing the neonatal mortality between infants born at varying ages (e.g., 25 and 40 weeks), since such comparison may not be interesting given the substantial differences in the absolute risk of mortality and different etiologic pathways. As with birth weight, adjusting for GA may result in a GA "paradox" (Joseph 2007; Platt et al. 2010). This GA paradox also may be explained through the use of causal diagrams, as illustrated in Figure 15.8.

To identify the causes of neonatal mortality, one needs to measure the effect of the exposure of interest, such as maternal smoking, on death within the first week of life. Typically, investigators may simply adjust for GA to compare infants born at similar GAs. This approach may be perceived as estimating the direct effect of smoking on neonatal mortality, in part, by blocking the indirect effect through GA. However, this situation is a close analogue to that described above for birth weight when there are common causes of GA and neonatal mortality. Specifically, GA is a common effect of smoking and U (an unmeasured confounder), with selection bias introduced if one conditions on GA even when using proxies, such as time of delivery. This can be easily generalized to other types of mortality such as stillbirth, and perinatal or infant mortality.

Similar to adjustment for birth weight and z-scores, GA-specific mortality curves cross. Fetuses and infants born to mothers who smoked during pregnancy are reported to have lower mortality risks at the earliest GAs than are infants born to mothers who did not smoke. Such findings have no biological explanation. Through the use of causal DAGs, we can argue that GA cannot possibly confound the effect of prenatal exposures, for it cannot cause a prenatal exposure and it is an unlikely proxy for preexposure variables that confound the effect. Thus, adjustment for GA

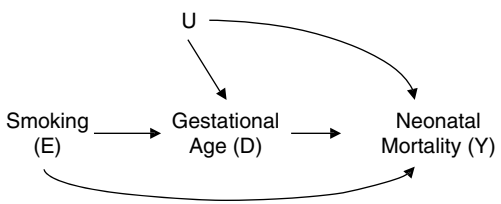


Figure 15.8 Estimating the effect of smoking on mortality. Controlling for gestational age leads to the gestational age paradox because gestational age is a collider between smoking and U.

cannot generally be justified as an attempt to adjust for confounding of the effect of prenatal exposures on neonatal mortality. Rather than correcting for confounding, one might adjust for GA in an attempt to find the direct effect of prenatal exposure on mortality, but such an analysis is biased. The GA paradox is most likely a statistical artifact arising from selection bias introduced by controlling for a collider—GA. Analyses conditioned on GA may suffer from selection bias, but analyses not conditioned on GA may address a very different causal question, for instance estimating the total rather than the direct effect of the exposure.

OVERADJUSTMENT

Following Schisterman et al., we define overadjustment bias as arising from the incorrect control of an intermediate variable (or a descending proxy for an intermediate variable) on the causal path from exposure to outcome (Schisterman et al. 2009). Controlling for an intermediate variable or its descending proxy biases the total effect estimate; controlling for an imperfect descending proxy biases direct and indirect effect estimates; but controlling for an intermediate does not always bias the direct and indirect effect estimates. We will first consider controlling for an intermediate variable and then controlling for a descending proxy of an intermediate variable.

Figure 15.9A provides a causal diagram representing the simplest case of overadjustment bias. Bodnar and colleagues (2005) postulated the mediating role of triglycerides (M in our notation) in the association between prepregnancy body mass index (E in our notation) and preeclampsia (D in our notation). In this scenario, one can consistently estimate the total causal effect of exposure E on outcome D using common regression techniques by ignoring the intermediate variable M. However, if one controls (i.e., adjusts, stratifies, restricts) for the intermediate variable M, which is on a causal pathway between exposure and outcome, the total causal effect of the exposure on the outcome cannot be consistently estimated. Controlling for M will typically bias the estimate of the total causal effect of E on D toward the null. When the only causal path between exposure E and outcome D is the path mediated through M (i.e., no direct effect of E on D, which requires a perturbation of the DAG in Figure 15.9A), the observed association between exposure E and outcome D will typically be null in expectation, conditional on the intermediate M. Controlling for an intermediate variable can correctly estimate the direct causal effect, although with added assumptions, such as the absence of a common cause of intermediate and outcome, and the absence of exposure interactions with the intermediate variable (Cole and Hernán 2002; Kaufman et al. 2004, 2005; Robins and Greenland 1992).

Figure 15.9B provides a second causal diagram representing a common case of overadjustment bias (Weinberg 1993). This diagram illustrates the assumption that exposure E and unmeasured intermediate U both affect the outcomes D and M. The measured variable M is a "descending" imperfect proxy for the variable U. In this case, one can think of M as an inaccurately measured version of U under a classic measurement error model, or as an event caused by U. One could place a half-square around U to imply the partial adjustment for the unmeasured U that occurs upon adjustment for the measured M.

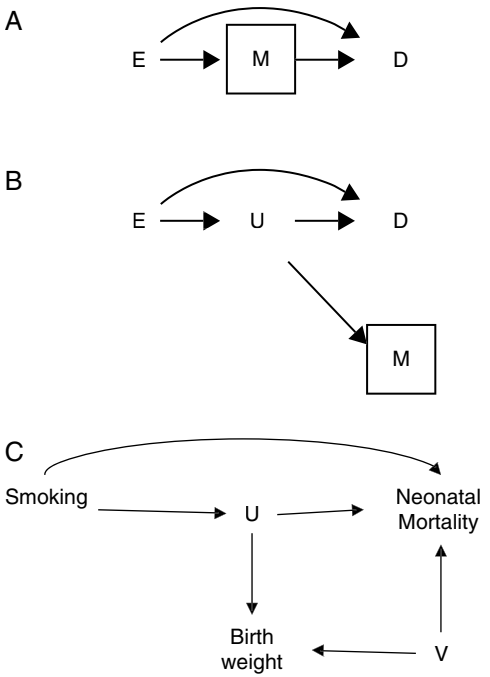


Figure 15.9 Overadjustment is control for an intermediate variable or its descending proxy that also leads to bias in an effect estimate.

One may consistently estimate the total causal effect of exposure E on outcome D using common regression techniques by ignoring M. However, if one controls (i.e., adjusts, stratifies, restricts) for the variable M, which is a proxy for variable U that is on a causal pathway between exposure E and outcome D, the total causal effect of the exposure on the outcome cannot be consistently estimated. Moreover, since M is an imperfect rather than a perfect proxy, controlling for M is imperfect control for U, which results in a biased estimate of the direct and indirect effects of exposure on outcome. If controlling for M leads to collider stratification bias, as it would if there were a common cause V of M and outcome D, the total, direct, and indirect effects would be further biased.

To further illustrate the effect of controlling for an intermediate variable or its descending proxy on estimation of total, direct, or indirect effects, consider the following example. Basso and colleagues (2006) postulated a causal diagram to illustrate the association between birth weight and neonatal mortality, assuming its relation was due to an unmeasured confounder such as a malformation, fetal or placental aneuploidy, infection, or an imprinting disorder, as illustrated in Figure 15.9C. In this situation, U represents some unmeasured aspect of fetal development whereas V represents an unmeasured confounder of the association between birth weight and neonatal mortality.

If the intermediate U were measured, controlling for it would bias the estimate of the total but not the direct effect of smoking on mortality. Controlling for an

intermediate *U* could bias a direct effect of smoking on mortality if *U* were a collider; for instance, if there were an arrow from *V* to *U*. Controlling for *U* would then open the smoking-*U*-*V*-mortality path and thereby introduce collider stratification bias. This collider stratification bias would constitute overadjustment bias, since *U* is an intermediate variable whose control leads to the bias. Since *U* is unmeasured, it might seem more appropriate to control for birth weight rather than *U* if one is interested in estimating the direct effect of smoking on mortality. We will see that controlling for birth weight, here a descending proxy of an intermediate variable between exposure and outcome, is invalid under some common conditions.

Let us first consider the implications of controlling for birth weight in the estimation of the total effect of smoking on neonatal mortality. Since birth weight is an imperfect descending proxy of the intermediate variable *U*, controlling for birth weight partially controls for *U*, which biases the total effect of smoking on mortality. Moreover, this overadjustment bias is exacerbated by collider stratification bias, since adjusting for the collider birth weight would open the smoking-*U*-birth weight-*V*-mortality path. This scenario exemplifies that it is incorrect to control for a variable (or its descending proxy) on the causal path between exposure and disease when estimating the total effect. Such control inherently controls the intermediate variable either totally or partially, leading to biased total effect estimates.

Let us consider the implications of controlling for birth weight in the estimation of direct and indirect effects of smoking on neonatal mortality. Controlling for the imperfect descending proxy birth weight biases the direct effect estimate and, consequently, the indirect effect estimate of smoking on mortality. This overadjustment bias arises because of collider stratification bias and because birth weight is an imperfect proxy for the intermediate *U*, so that controlling for birth weight only partially controls for *U*.

Overadjustment bias is not induced by estimating direct and indirect effects, per se, when the proper statistical methods are applied for this purpose. Previous authors (Joffe and Colditz 1998; Robins and Greenland 1992) have independently provided the conditions under which estimation of direct and indirect effects can be separated.

In conclusion, overadjustment bias is bias that arises from controlling for an intermediate variable (or a descending but not ascending proxy for an intermediate variable) on a causal path from exposure to outcome when estimating any effect (total, direct, or indirect) under realistic conditions. These realistic conditions include the presence of collider stratification bias and the imperfect association between a variable and its proxy. Using causal diagrams, we can identify overadjustment bias and thus utilize appropriate analytic strategies for estimating the causal questions of interest.

COMPARING GRAPHICAL TOOLS IN CAUSAL INFERENCE

The birth weight paradox provides a good example for comparing graphical tools, including DAGs. As a reminder, the common version of the paradox is that LBW

infants in high-mortality-risk populations have lower mortality than LBW infants in low-risk populations. Figure 15.10 illustrates the paradoxical crossing of neonatal mortality versus birth weight curves for smoking and nonsmoking mothers (Hernandez-Diaz et al. 2008). In this section we will compare DAGs and mortality curves and specifically address the advantages of each.

Directed acyclic graphs specify the existence of causal relations and their direction, whereas mortality curves do not. In Figure 15.11, the DAG specifies that maternal smoking causes birth weight, and neither smoking nor birth weight causes mortality. The DAG is consistent with the accompanying mortality-versus-birth weight curves. However, the curves themselves do not specify the causal relations that the DAG specifies. One cannot deduce from the curves whether causal relations exist, such as whether or not birth weight causes infant mortality. The lack of association between birth weight and mortality could be due to bias by an unconsidered variable that confounds the association between birth weight and mortality. Furthermore, one cannot deduce from the curves the direction of causal relations. For example, if one suspected that maternal smoking causes LBW, with only the curves and knowledge that maternal smoking precedes birth weight, one could suspect that maternal smoking causes LBW, but not vice versa. Here, it is the knowledge of the temporal relationship between variables that allows one to suspect the direction of causality. The curve itself does not motivate this suspicion. Without knowledge of the temporal relationship between variables—for example, whether diet affects disease status or vice versa—one cannot know the direction of possible causation between those variables by simply looking at the curve.

When thinking about DAGs and mortality curves, several important underlying differences emerge. Directed acyclic graphs convey the existence and direction of

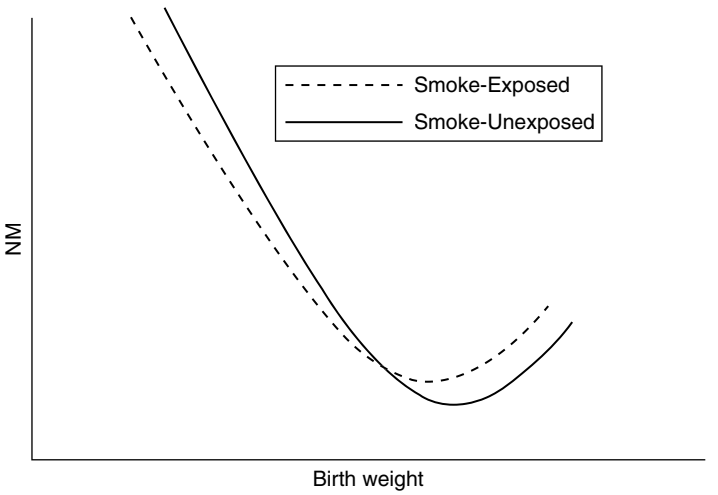


Figure 15.10 Neonatal mortality (NM) versus birth weight curves.

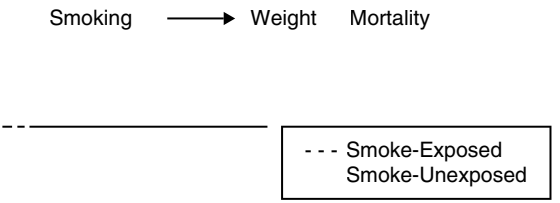


Figure 15.11 Directed acyclic graphs specify the direction of causal relations, whereas mortality versus birth weight curves do not.

causal relations, whereas mortality curves quantify the magnitude and/or direction of an *association*. This important difference explains why one DAG may be consistent with multiple sets of mortality curves, as shown in Figure 15.12. If smoking increased mortality at every birth weight, the DAG in Figure 15.12A and the mortality curves in Figure 15.12B would be consistent. If smoking decreased mortality at every birth weight, the DAG and the mortality curves in Figure 15.12C would be consistent. If smoking decreased mortality for LBW babies but increased mortality for high-birth-weight (HBW) babies, the DAG and the mortality curves in Figure 15.12D would be consistent, since DAGs cannot convey the magnitude and direction of smoking's association with mortality at each birth weight category.

Although mortality curves can convey the magnitude and direction of hypothesized causal effects, they cannot distinguish between interaction and collider stratification bias. Hence, one set of mortality curves can be consistent with multiple

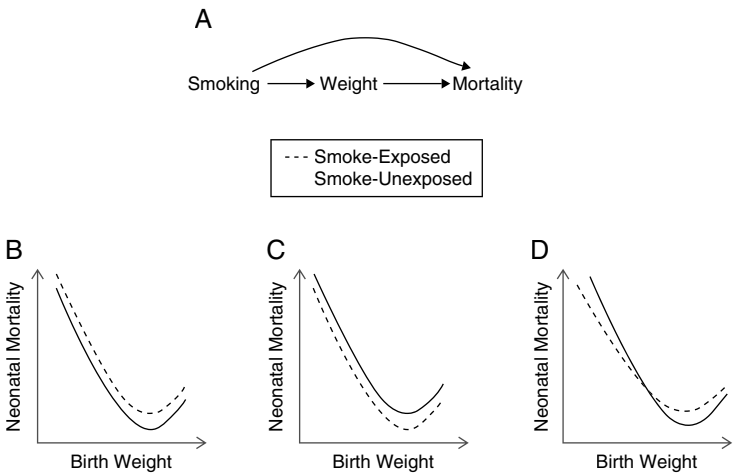


Figure 15.12 One directed acyclic graph (DAG) can be consistent with multiple mortality curves. A DAG cannot quantify the magnitude and direction of association.

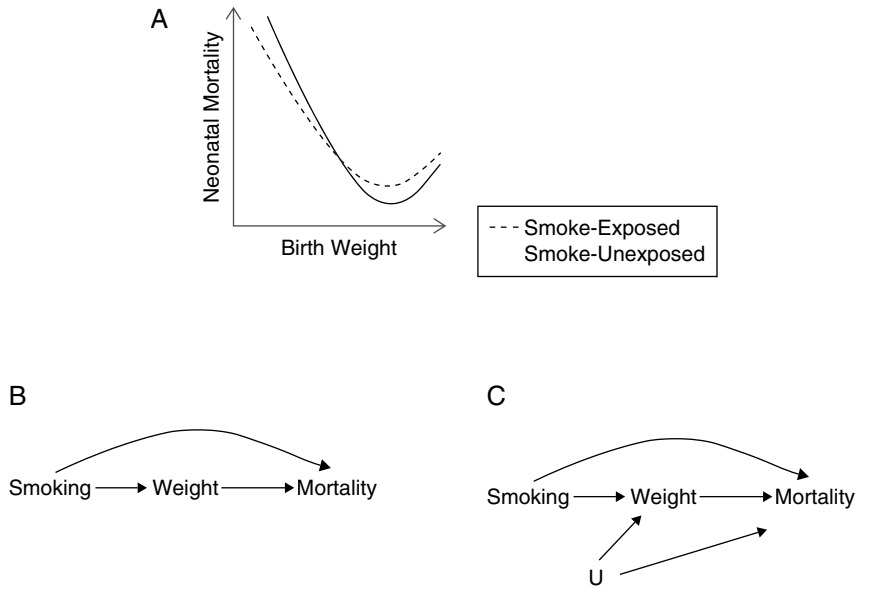


Figure 15.13 One mortality versus birth weight curve can be consistent with multiple directed acyclic graphs. Mortality curves cannot distinguish between collider stratification bias and interaction.

DAGs. For example, if smoking and birth weight were to interact, such that smoking reduced mortality for LBW babies but increased mortality for HBW babies, then the mortality curves in Figure 15.13A would be consistent with the DAG in Figure 15.13B. On the other hand, if in Figure 15.13C there were no interaction between smoking and birth weight, variable *U* could induce an association (collider stratification bias) between smoking and mortality conditional on birth weight, resulting in the crossing of the mortality curves. Similar to mortality curves, DAGs cannot identify interaction, as was described earlier in the discussion of Figure 15.12A. However, unlike mortality curves, DAGs can reveal sources of bias, as in Figure 15.13C.

The respective strengths and limitations of these two graphical tools underscore their complementary nature. Directed acyclic graphs specify the existence and direction of causal relations, thus revealing sources of bias, whereas mortality curves quantify the magnitude and direction of association. Neither method unambiguously identifies the presence of interaction. These characteristics of DAGs and mortality curves explain why there is not a one-to-one correspondence between DAGs and mortality curves. Because a DAG can be consistent with only certain exposure–outcome graphs and an exposure–outcome graph can be consistent with only certain DAGs, a DAG can narrow down the possible exposure–outcome graphs and an exposure–outcome graph can narrow down the possible DAGs. These two graphical methods are complementary.

CONCLUSION

Knowledge of plausible causal relations allows for the creation of DAGs. By specifying causal relations among variables, including revealing colliders or other variables whose adjustment leads to bias, DAGs encourage awareness of what effect of exposure on outcome is being estimated—the total, direct, or indirect effect—and indicate the choice of variables to adjust (or not adjust) for. When examining a DAG to decide which variables to adjust for, one must be aware of which paths lead from exposure to outcome and which paths would be opened or closed by adjustment. Adjusting for an intermediate variable closes a path, removing its contribution to the association between exposure and outcome, whereas adjusting for a collider induces association between the collider's immediate causes, potentially opening a spurious path from exposure to outcome. We illustrated the applicability of DAGs by showing how they can resolve past enigmas in epidemiology, including the birth weight paradox, where LBW babies in a high-mortality-risk population paradoxically have lower mortality than LBW babies in a low-mortality-risk population. While it has been argued that the crossing of the curves could be explained by unmeasured confounding, or could be resolved by adjusting for birth weight z-scores, we demonstrated that this paradox can be entirely explained by collider stratification bias of the direct effect of smoking on mortality, a bias introduced by controlling for the collider birth weight. Unmeasured confounding could result in bias either towards or away from the null, but could not cause the crossing of the curves. Rather, the crossing of the curves is a direct consequence of adjustment for a collider. Further, the proposition to use z-scores to remove the crossing of the curves, and the bias due to adjustment for birth weight, is incorrect. Adjustment using z-scores does not result in estimation of direct and indirect effects, but rather an estimate of the total effect. The ability of z-scores to successfully remove the paradoxical crossing of the curves is due to an alteration in the causal parameter being estimated, and does not remove the bias due to collider stratification.

In short, there is no replacing the importance of substantive input in designing DAGs. Appropriate adjustment requires careful consideration of the assumptions regarding the proposed underlying biological mechanisms and a clearly defined causal question. To prevent selection bias and misinterpretation of effect, care should be given to using the appropriate crude or adjusted model that will accurately answer the causal question of interest. Directed acyclic graphs provide epidemiologists with an additional graphic tool to aid in the formulation of a research question and its accompanying analytic plan. Although they do not quantify the magnitude and direction of an exposure–outcome association, they specify the existence and direction of causal relations, and hence help ensure that we answer the intended question and interpret the science within a formalized causal paradigm.

REFERENCES

Ananth, C.V. and Wilcox, A.J. 2001. Placental abruption and perinatal mortality in the United States. *Am J Epidemiol* 153(4): 332–37.

Buekens, P. and Wilcox, A. 1993. Why do small twins have a lower mortality rate than small singletons? *Am J Obstet Gynecol* 168(3 Pt 1): 937–41.

Cole, S.R. and Hernán, M.A. 2002. Fallibility in estimating direct effects. *Int J Epidemiol* 31(1): 163–65.

Greenland, S. 2003. Quantifying biases in causal models: Classical confounding vs collider-stratification bias. *Epidemiology* 14(3): 300–06.

Greenland, S., Pearl, J., and Robins, J.M. 1999, Causal diagrams for epidemiologic research. *Epidemiology* 10(1): 37–48.

Hernán, M.A., Hernandez-Diaz, S., and Robins, J.M. 2004. A structural approach to selection bias. *Epidemiology* 15(5): 615–25.

Hernán, M.A., Hernández-Díaz, S., Werler, M.M., and Mitchell, A.A. 2002. Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology. *Am J Epidemiol* 155(2): 176–84.

Hernandez-Diaz, S., Schisterman, E.F., and Hernán, M.A. 2006. The birth weight "paradox" uncovered? *Am J Epidemiol* 164(11): 1115–20.

Hernandez-Diaz, S., Wilcox, A.J., Schisterman, E.F., and Hernán, M.A. 2008. From causal diagrams to birth weight-specific curves of infant mortality. *Eur J Epidemiol* 23(3): 163–66.

Hertz-Picciotto, I. 2003. Is it time to abandon adjustment for birth weight in studies of infant mortality? *Paediatr Perinat Epidemiol* 17(2): 114–16.

Hertz-Picciotto, I. and Din-Dzietham, R. 1998. Comparisons of infant mortality using a percentile-based method of standardization for birthweight or gestational age. *Epidemiology* 9(1): 61–67.

Joseph, K.S. 2007, Theory of obstetrics: An epidemiologic framework for justifying medically indicated early delivery. *BMC Pregnancy Childbirth* 7: 4.

Kaufman, J.S., MacLehose, R.F., and Kaufman, S. 2004. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. *Epidemiol Perspect Innov* 1(1): 4.

Kaufman, S., Kaufman, J.S., MacLehose, R.F., Greenland, S., and Poole, C. 2005. Improved estimation of controlled direct effects in the presence of unmeasured confounding of intermediate variables. *Stat Med* 24(11): 1683–702.

Macdorman, M.F. and Atkinson, J.O. 1998. Infant mortality statistics from the linked birth/infant death data set – 1995 period data. *Monthly Vital Stat Rep* 46(6).

Miguel, A., Hernán, J., and Robins, M. 2010. *Causal inference*. Boca Raton, FL: CRC.

Pearl, J. 1995. Causal inference from indirect experiments. *Artif Intell Med* 7(6): 561–82.

Platt, R.W., Schisterman, E.F., Hernandez-Diaz, S., and Hernán, M.A. 2010, unpublished. *The perils of adjusting for gestational age when trying to estimate the effect of prenatal exposures on neonatal mortality*.

Robins, J.M. and Greenland, S. 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3(2): 143–55.

Schisterman, E.F., Cole, S.R., and Platt, R.W. 2009. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 20(4): 488–95.

Schisterman, E.F., Whitcomb, B.W., Mumford, S.L., and Platt, R.W. 2009. Z-scores and the birthweight paradox. *Paediatr Perinat Epidemiol* 23(5): 403–13.

Weinberg, C.R. 1993. Toward a clearer definition of confounding. *Am J Epidemiol* 137(1): 1–8.

Whitcomb, B.W., Schisterman, E.F., Perkins, N.J., and Platt, R.W. 2009. Quantification of collider-stratification bias and the birthweight paradox. *Paediatr Perinat Epidemiology* 23(5): 394–402.

Wilcox, A. and Russell, I. 1990. Why small black infants have a lower mortality rate than small white infants: The case for population-specific standards for birth weight. *J Pediatr* 116(1): 7–10.

Wilcox, A.J. 1993. Birth weight and perinatal mortality: The effect of maternal smoking. *Am J Epidemiol* 137(10): 1098–104.

Wilcox, A.J. 2001. On the importance—and the unimportance—of birthweight. *Int J Epidemiol* 30(6): 1233–41.

Wilcox, A.J. and Russell, I.T. 1983. Birthweight and perinatal mortality: II. On weight-specific mortality. *Int J Epidemiol* 12(3): 319–25.

Yerushalmy, J. 1972. Infants with low birth weight born before their mothers started to smoke cigarettes. *Am J Obstet Gynecol* 112(2): 277–84.

16

Unique Methodologic Challenges in Reproductive and Perinatal Epidemiology

ROBERT W. PLATT AND GERMAINE M. BUCK LOUIS

A number of general methodologic challenges are unique to reproductive and perinatal epidemiology. We outline some of these issues in this chapter, with reference to the concerns they raise about inference from existing studies and to issues for the design and analysis of future studies. We first review some conceptual challenges related to the study of reproductive and perinatal outcomes. Clustering of outcomes within groups of individuals is an issue highly relevant to epidemiology, but especially so for this subspecialty, in that couples typically have more than one pregnancy, the phenomenon of multiple births, and the need for multiple treatment cycles for couples undergoing infertility treatment. This chapter provides an overview of the unique methodologic challenges, modeling approaches that may be useful strategies for study design and analysis, and remaining challenges. This chapter is not intended to be a comprehensive review of the issues or available methodologic approaches, as provided elsewhere (Savitz et al. 2006).

CONCEPTUAL CHALLENGES

Critical and sensitive windows

Reproductive and perinatal epidemiology has long recognized the importance of conceptualizing its research within critical windows of human development, in recognition of the timed and highly interrelated processes underlying human reproduction and development. For example, epidemiologic investigations focusing on the etiologic determinants of birth defects painstakingly measure exposures during the relevant embryonic window for a particular structural birth defect, since exposures outside the window may confer no additional risk. Similarly, when studying pregnancy loss, epidemiologists ensure that purported etiologic exposures are measured prior to and not following the loss (which can be challenging to precisely identify). Since the exact timing of embryonic or fetal death

is often unknown and can vary considerably in relation to days since conception, collection of daily exposures is required for the accurate estimation of effect.

Most recently, considerable effort is under way to identify sensitive (and not just critical) windows of human development that may be relevant for changes in function rather than structural defects, per se. For example, toxicants that affect semen quality need to be measured prior to or during the sensitive window relevant for spermatogenesis (≈72 days) to ensure the accurate estimation of an effect. The rapidly evolving epigenetic literature supports continued research efforts aimed at identifying exposures across the spectrum of sensitive windows—from preconception through pregnancy and the early neonatal period—that may permanently reprogram the human organism for alterations in function across the lifespan (Barker 2002) or, possibly, generations (Anway and Skinner 2006). Examples of exposures during early critical or sensitive windows that are associated with a range of adverse structural and/or functional effects include bisphenol A (BPA) and diethylstilbestrol (DES), an environmental agent and pharmaceutical estrogen, respectively. The extent to which environmental agents account for the purported decline in human fecundity remains to be established (Hamilton and Ventura 2006; Jensen et al. 2008), but will require prospective cohort designs with longitudinal measurement of exposures during critical or sensitive windows. A more complete description of the conceptual basis of these windows is provided in Chapter 2; the methodologic implications are discussed below.

Parental exposures

Historically, much of the reproductive and perinatal epidemiologic research has focused on maternal characteristics, behaviors, or exposures in relation to a spectrum of reproductive and perinatal outcomes, most likely because pregnancy is limited to women. Traditionally, study designs were largely restricted to women, ignoring the couple-dependent nature of human reproduction. To this end, less is known about male- or parentally mediated exposures and human reproduction and development relative to female-mediated effects. Of late, several authors have called for the study of parentally mediated exposures and human reproduction and development, particularly those commencing during the pre- or periconception sensitive window (Chapin et al. 2004; Hassan and Killick 2004; Louis et al. 2008). Of the prospective pregnancy cohort studies conducted to date, including those with preconception enrollment of study participants, only a few have captured exposures from both partners of the couple (France et al. 1984; de Mouzon et al. 1988; Zinaman et al. 1996; Bonde et al. 1998), while the remainder have relied upon women, as previously summarized (Buck et al. 2004). Such a dearth of information makes it challenging to fully assess maternally mediated effects in the absence of male effects or, possibly, the interaction of couple effects.

Given the relatively short interval for most reproductive and perinatal outcomes—ranging from days (e.g., conception, menstruation) to weeks (e.g., embryonic or fetal periods of development) or months (e.g., infancy)—prospective cohort designs are best suited for etiologic or prediction research. Moreover, longitudinal data collection should be targeted to purported critical and/or

sensitive windows to ensure the capture of acute and chronic exposures of interest to the investigators. Again, capture of exposures for both partners of the couple is ideal for the assessment of couple-dependent reproductive or perinatal outcomes. Such a strategy may not be feasible for some outcomes involving assisted reproductive technologies (ART) that involve gamete donation, gestational carriers, or surrogates. When studying successive pregnancies (as described later in this chapter) or complete reproductive performance, it is important to plan accordingly for change in paternity, consistent with the demographic structure of study populations.

Although the types of exposures to be measured in a study will vary in relation to research aims, the timing of exposures remains a challenging task, given that many of our essential outcomes (such as ovulation, conception, or embryonic demise) rely upon proxy markers, as the actual events are difficult to empirically measure, particularly at the population level. Coupled with the missing denominator issue, as described in Chapter 14, a number of unique methodologic issues underlie reproductive and perinatal epidemiology and need to be carefully considered in designing research in response to critical data gaps.

CLUSTERING OUTCOMES (IMPORTANCE OF PRIOR REPRODUCTIVE HISTORY)

The clustering of observations (nonindependence between observations in groups or clusters, or between repeated observations for the same individual) is a significant challenge for epidemiologists and, especially so for those interested in reproductive and perinatal epidemiology (Breslow and Clayton 1993; Greenland 2000). Briefly, when observations within defined clusters or groups within the total sample are dependent, the information in the total sample is less than that which would be present in a similar-sized sample of independent subjects. This is the case because, in effect, knowing information about one observation gives information about a subsequent one. Perhaps the most common occurrence of clustering arises from the number of pregnancies per a woman or couple participating in the study, since pregnancies to the same woman or couple tend to be more alike than those occurring to other women or couples. For example, consider birth weights in clusters of siblings. Birth weights among siblings are strongly correlated. So, if two out of three siblings have birth weights of over 4,000 g, it is reasonable to expect that the third sibling also will have a relatively high birth weight. An infant sampled from another cluster (mother) for whom we have no information about siblings would have an expected birth weight nearer the mean of the distribution of the sample.

Several problems in reproductive and perinatal epidemiology give rise to clustered data including (a) infertility treatment, in that many couples require multiple cycles for pregnancy (in which cycles within the same couple are clearly nonindependent); (b) successive pregnancies (in which multiple characteristics, from infant size to socioeconomic status, cluster within families); and (c) twins and higher-order multiple births (in which characteristics are generally very

tightly clustered). In this section, we consider the specifics of each of these cases in turn, from a conceptual perspective. In a later section on modeling strategies, we discuss some analytic approaches that are relevant for clustered outcomes.

Multiple treatment cycles

Methodologic challenges have arisen in response to the marked biotechnologic advances in the treatment of male, female, and couple-based infertility. Infertility treatment cycles are heterogeneous and may include ovulation induction, intrauterine insemination, or any of the ARTs, which are defined as the in vitro handling of both human oocytes and sperm or embryos (Zegers-Hochschild et al. 2009), such as occurs with in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI). Two key methodologic considerations are needed in designing research focusing on male, female, or couple determinants of fecundity and fertility such as fertilization, cleavage, implantation, embryonic quality, and pregnancy loss or live birth. These considerations include the hierarchical and dependent data structure characteristic of research focusing on infertility treatment (Buck Louis et al. 2005). Figure 16.1 illustrates the hierarchical data structure ranging from couple- to individual- to cycle- to treatment-level factors that may be relevant for predicting treatment success, however defined, or in assessing the impact of ART on child health and development. Couple-level factors might include use of cigarettes, alcoholic or caffeinated beverages, or the body mass indices of both partners (Hassan and Killick 2004). Partner-level factors may include semen quality or hormonal milieu, whereas treatment-level factors reflect the varying types of

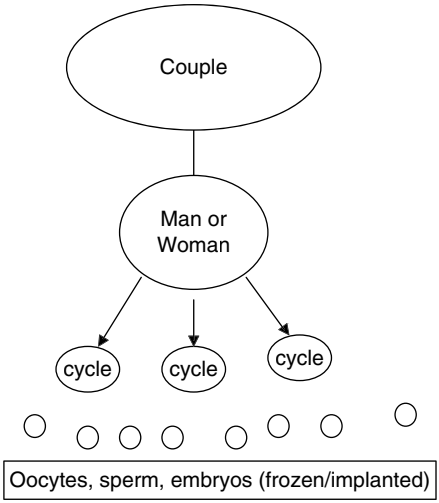


Figure 16.1 Illustration of the possible hierarchical structure of reproductive data.

treatment, typically moving from the least to most invasive procedures over time. Oocyte-level factors include the number retrieved or fertilized.

Infertility treatment cycles are highly correlated, given that they are administered to the same man, woman, or couple even if treatment is altered over time. In fact, couples only seek a second treatment cycle if they fail to become pregnant during the first cycle. Hence, selecting only the first cycle for analysis assumes that it is representative of all cycles. This assumption is rarely plausible. Much of the research focusing on the analysis of ART data has ignored this inherent clustering, thus making it difficult to fully interpret much of the published data. Recently, clinicians have recognized the need to model the inherent correlation within treatment cycles by ensuring that all treatment cycles are arrayed per woman or couple, and not assumed to be independent events. The latter assumption will overinflate statistical power, resulting in possible incorrect inferences. Analytic techniques must be responsive and appropriate for the data structure to avoid artificially inflating statistical power. If we are to empirically assess critical data gaps regarding the prediction of treatment success or its effect on children's growth and development, appropriate analytic techniques must be used.

Successive pregnancies

It is common knowledge that reproductive history is informative in the study of subsequent pregnancies (Weinberg 1995; Buck Louis et al. 2006; Hutcheon et al. 2006; Hutcheon and Platt 2007), leading to the correlation of pregnancy outcomes when studying more than one pregnancy from an individual woman (i.e., the likelihood that negative outcomes cluster within a woman). This problem has been addressed from the statistical and epidemiologic perspective.

Studies of pregnancy typically select their sample from a population of individual pregnancies. However, one can think of this sample of pregnancies as nested within, or as a subsample of, a hypothetical cohort of reproductive histories of women. Many of the sampled pregnancies will have come from reproductive histories with one, two, or three pregnancies; some from histories with more than three; while other women will have no pregnancies during their reproductive lives. These latter women do not factor into the selection of pregnancies, but their lack of reproductive history will be informative for some exposure–outcome associations (e.g., for exposures that have an effect on time-to-pregnancy). This hypothetical cohort follows each woman throughout her reproductive life, containing her entire collection of pregnancy experiences (and notably, the full cohort includes women who contribute no pregnancies). The resulting collection of pregnancies and pregnancy outcomes ultimately included in the study is a randomly or arbitrarily selected subsample of the cohort of pregnancies generated by the reproductive histories. We now consider several potential samples from within this cohort, represented schematically in Figure 16.2.

Most studies select pregnancies at random by recruiting from prenatal care providers or using birth registries, or other registries. The ensuing cohort of pregnancies can be conceived of as being generated from a sample from the full cohort of

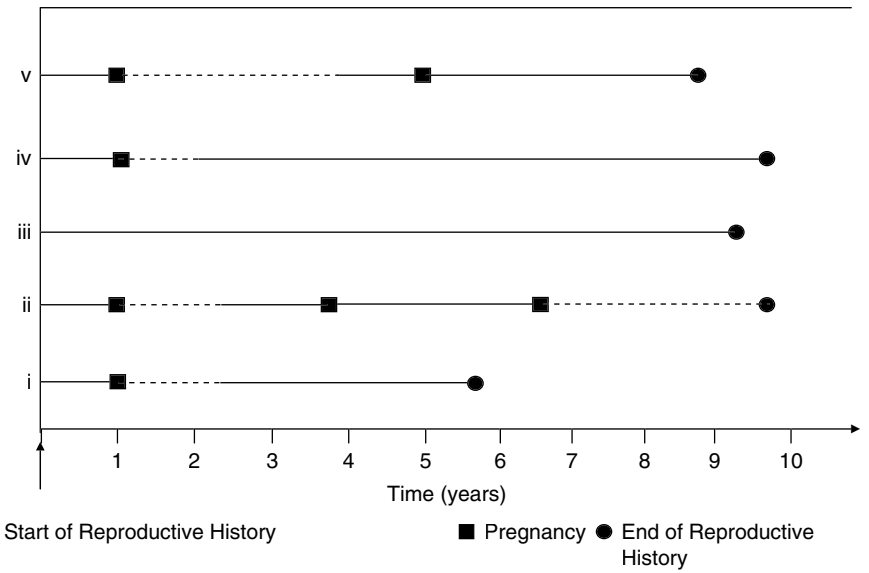


Figure 16.2 Schematic of a cohort of women and their pregnancy history. Solid lines represent time at risk for pregnancy; dashed lines time not at risk.

reproductive histories. There are several obvious problems with this design. First, a woman whose reproductive history includes several pregnancies is more likely to appear in the sample because she has more pregnancies to be sampled from; she may occur more than once if sampling takes place over several years. A woman with few pregnancies, spaced far apart in time, will be much less likely to appear in the sample. This phenomenon can be considered as analogous to the concept of length-biased sampling (Zelen 2004); in the study of chronic diseases, those with the disease for longer periods of time are more likely to be sampled in a prevalent sample. Similarly, women with longer pregnancy histories are more likely to have one of their pregnancies randomly selected. At the other extreme, women whose reproductive histories include no pregnancies will have no possibility of being selected for the cohort; hence, the sample is not really a random sample of the referent population of women. This may not seem a problem in the study of pregnancy outcomes, but if an exposure affects conception as well as pregnancy outcome, or if, for example, a socioeconomic factor has an effect on the number of children a couple has, the total effect on pregnancy outcome may be biased due to selection bias induced by ignoring those women with no pregnancies and oversampling those who have multiple pregnancies. For example, when studying exposures related to pregnancy in women aged 25 years (which may come about through stratification or adjustment for age), what is the appropriate cohort? Selecting only pregnant 25-year-old women, which is what happens if we start with a sample of pregnancies, ignores the 25-year-old women who did not get pregnant. If an exposure reduces the probability of pregnancy at age 25 years, but in the pregnant group has a positive effect on fetal outcomes, one could erroneously conclude that the exposure is beneficial.

An alternative strategy is to sample a specific single pregnancy from each woman, for example, her first pregnancy. This is equivalent to sampling a single point from each reproductive history (except in those with no pregnancies). It is evident that the cohort of first pregnancies represents a well-defined subset of the larger cohort. In such a study of primigravid women, there is no reason or need to adjust for pregnancy history, because there is none at that point. Further, there is (obviously) no clustering, as this includes only one pregnancy per woman. This restrictive group of women forms a well-defined population, results are generalizable to the population of first pregnancies, and there is no length-biased sampling, although the selection bias due to exposures affecting conception remains. Olsen (1994) recommends that in any study involving a sample of pregnancies, a subanalysis be done using only the first pregnancies. Although this analysis is not generalizable to the entire reproductive history, it does represent something well-defined, and the degree to which the randomly selected pregnancy analysis is biased can be illustrated by comparison to this analysis. Buck Louis et al. (Buck Louis et al. 2006) further point out that selecting a random pregnancy from each woman creates a nonrepresentative sample that overemphasizes first pregnancies and does not generalize to pregnancies or to reproductive histories.

Several authors have discussed the difficulties associated with the automatic adjustment for pregnancy history when the study cohort comprises multiparous women (Bakketeig and Hoffman 1983; Weinberg 1993; Hernán et al. 2002); that is, when the cohort is a random sample of pregnancies, some of which may be clusters from the same woman. In particular, it is very important to consider the causal question of interest, and the need to adjust for truly confounding variables while avoiding adjustment for consequences of the exposure. Weinberg (1995) considered adjustment for prior pregnancy outcome and demonstrated that this could introduce bias when the exposure of interest or a correlate was also a potential cause of the prior outcome. Olsen (1994) noted that, although authors typically adjust for parity or gravidity in regression models, there is limited or no grounds for doing so to adjust for confounding. It may seem reasonable to adjust for gravidity, so that comparisons are between women with similar numbers of past pregnancies, but (a) gravidity may not fit modern definitions of a confounder (Hernán et al. 2002), and (b) if gravidity is affected by the exposure (i.e., if the exposure increases or decreases the number of pregnancies), then selection bias may be induced by adjusting for gravidity (see Chapter 14 for details). This latter situation is not implausible; if an exposure causes spontaneous abortion, then it may cause increased gravidity because couples with the exposure may attempt to conceive again after the spontaneous abortion. This association could lead to potential selection bias. Both of these cases illustrate that adjustment for pregnancy history, if done without consideration of the causal structure of the research question, may introduce rather than eliminate bias.

Another issue to consider in successive pregnancies is that effects may be different between women versus within a single woman's reproductive history. For example, Neuhaus and Kalbfleisch (1998) studied the association between maternal age and low birth weight, and noted substantial differences between

coefficients measuring the effect across women and those measuring the effect within women. The effect of within-woman *average* age at the time of pregnancy was negative, where the average age is the average across all pregnancies for a given woman. However, the effect of deviation from this average for an individual pregnancy was very close to null, indicating that there was no effect of age within woman. The model ignoring the potential differences between these coefficients produced an effect that was close to the average of the negative effect between women and the null effect within woman. Hutcheon et al. (2006) studied the effect of maternal prepregnancy weight and maternal weight gain on birth weight in a cohort of siblings and noted substantial differences in the association between prepregnancy weight and birth weight when considering between- versus within-woman effects separately.

In the end, the most important factor in dealing with successive pregnancies is the careful specification of the etiologic research question under study. If, for example, the question is limited to first pregnancies, there is no issue related to correlated data. Hutcheon and Platt utilized causal diagrams to show that adjustment for successive pregnancies was unnecessary when the effects of prior pregnancies on current outcome were entirely mediated through measured confounders for the current pregnancy (Hutcheon and Platt 2007). Howards and colleagues (2007) considered a number of settings and demonstrated the need for methods, such as marginal structural models, that can adjust for variables that are simultaneously confounders and intermediate variables in certain circumstances when pregnancy outcome in a first pregnancy is an intermediate and a confounding variable for subsequent pregnancies (Robins et al. 2000).

Multiple pregnancies

Multiple pregnancies (twins, triplets, and higher-order multiples) present several obvious challenges and some less evident ones. The most obvious is the correlation that exists between multiples. This is a relatively straightforward problem to solve, as methods exist for the analysis of clustered data that are straightforward to implement. However, it is important to be clear regarding the question under study, as this may change the analytic method that is most appropriate. Ananth et al. (2005) reviewed this issue and noted that methods that do not account for clustering can grossly underestimate variance when studying twins. Carlin et al. (2005) clearly distinguished between effects that can occur between twins in the same pair and effects that occur at the level of the pair, analogous to those differences in successive pregnancies. There are two principal approaches for clustered-data scenarios, conditional and marginal methods. Conditional methods can be used either in the design phase (matching on characteristics of the twin pair) or analytic phase (random effects, or multilevel models) of research. Increasingly, marginal methods, such as generalized estimating equations, are used for averaging across twin pairs. Matching and fixed-effects methods (Hutcheon and Platt 2007) are conditional methods, based on creating groups (or matched sets) of subjects in which all subjects have the same levels of confounding variables. Confounding is then accounted for when the analysis is done

conditionally on the matching; subjects are compared only to others with the same characteristics.

MODELING STRATEGIES

Several reviews have been written on methods for clustered pregnancy data (successive pregnancies [Buck Louis et al. 2006] and twins [Ananth et al. 2005] in particular). In this section, we outline two key groups of methods: hierarchical models and their Bayesian counterparts, and semiparametric methods based on estimating equations.

Hierarchical models and Bayesian approaches

When clustering is present in data, statistical methods must account for it to ensure that variance estimates reflect the increased uncertainty caused by the reduced effective sample size (because observations within a cluster do not provide independent information). A common approach to this problem is referred to as *hierarchical modeling*, *multilevel modeling*, or *random effects modeling* (Breslow and Clayton 1993; Burton et al. 1998). A comprehensive discussion of random effects models can be found in Verbeke and Molenberghs (2001) or Singer and Willett (2003). We briefly summarize the methods here.

Consider a regression model of the form

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \varepsilon_{ij}$$

where *i* indexes clusters, and *j* indexes subjects within cluster. The standard regression model is based on the assumption that $\varepsilon_{ij} \sim N(0,\sigma^2)$ and that the ε_{ij} are independent of each other. When the data are clustered, this is not a reasonable assumption. If *Y* is birth weight, clusters are the collection of pregnancies for individual women, and *X* is a (possibly vector-valued) covariate, then it is expected that the ε_{ij} are correlated (those for some women will more likely be all above zero than expected, and those for others will more likely be all below zero). A more reasonable assumption is that the model for these data is of the form

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + b_i + \varepsilon_{ij} \qquad \text{(Eq. 1)}$$

where $b_i = N\left(0,\sigma_b^2\right)$. The b_i are referred to as random effects and represent deviations between clusters. It is necessary to assume a distribution for these random effects; otherwise, they would, in effect, be indicator or dummy variables for each cluster. This latter setting would cause the number of parameters to increase with sample size, which violates asymptotic theory in statistics, which states that the number of parameters should stay fixed as the sample size increases (Casella and Berger 2001). These models are called random effects models because the "effect" of each cluster is a random term; they are sometimes also referred to as

hierarchical because random effects exist at the cluster level and at the individual level. The hierarchy can be generalized to multiple levels; one could envision pregnancies clustered within a woman, and women within a hospital or community, and a community within a region. When this occurs, the approach is to simply extend the random effects. These models are sometimes referred to as multilevel (Goldstein et al. 2002), especially when the clustering is across several levels. Models such as that given in Equation 1 can be fit using a variety of algorithms and software packages. In linear models (when the response is continuous) the results are straightforward (Burton et al. 1998), but there are a number of important differences between approaches when the response is discrete; the optimal choice of estimation procedure remains a subject of research (Breslow 2004).

Bayesian hierarchical methods (Gelman et al. 2003) can be viewed as a generalization of the hierarchical models presented above. A Bayesian approach can be easily implemented for models of the type given by Equation 1; the difference is that prior distributions are placed on each of the parameters in the model. Bayesian approaches can be made much more flexible than frequentist methods due to the availability of *Markov chain Monte Carlo* (MCMC) methods, which allow straightforward use of, for example, non-normal distributions for the random effects. These models have been used effectively in perinatal and reproductive research, for example in the study of time-to-pregnancy and early reproductive effects (Dunson et al. 2005; Dunson and Stanford 2005), in which the structure of the model is difficult to construct in a frequentist framework.

Semiparametric models and estimating equations

A model is semiparametric when it is not completely specified by parameters. Some aspects of the model are specified nonparametrically. The most well-known semiparametric model is probably the Cox model, in which the hazard function is given by

$$h(t,\vec{x}) = h_0(t)e^{\sum \beta_i x_i}$$

so that part of the hazard function is specified by the β parameters. The component $h_0(t)$ is referred to as the baseline hazard because it is the hazard function when all of the x_i are zero. Semiparametric models are also used in causal inference (see Robins 1998; Robins et al. 2000 for a discussion of marginal structural models) and in the analysis of clustered data (e.g., generalized estimating equations [Zeger et al. 1988; Liang and Zeger 1993]).

The generalized estimating equation (GEE) methodology is a semiparametric alternative to hierarchical models for clustered data. Rather than fully modeling the correlation between observations on the same cluster, GEE only requires that the researcher specifies a functional form for the marginal mean

$$E\left[Y_{ij} \mid X_{ij}\right] = \beta_0 + \beta_1 X_{ij} \qquad \text{(Eq. 2)}$$

and a form for the correlation between observations within a cluster, $\rho = Corr(Y_{ij}, Y_{ik})$. The specification of these two components defines a set of estimating equations from which parameter estimates can be computed. Unlike random effects models, the generalization of the model using Equation 2 to other outcomes and link functions in the family of generalized linear models (e.g., logistic and Poisson regression) is straightforward. Another advantage of these methods is that the specification of ρ need not be correct. Efficiency is gained if ρ is correctly specified, but variance estimates for β are correct even if the specification is incorrect. However, these methods require more assumptions with regard to missing data.

DISTINCTIONS BETWEEN HIERARCHICAL AND MARGINAL MODELS

An important distinction between hierarchical models and GEE is in the interpretation of the coefficients (Neuhaus et al. 1991). In a random effects (i.e., hierarchical) model, the β coefficients are interpreted conditionally on the b_i, meaning that they have a *cluster-specific* interpretation. This means that both between- and within-subject comparisons are conditional on b_i , and so can be interpreted as comparing women with the same unmeasured characteristics. In a GEE, the coefficients are interpreted marginally, or at a *population-average* level. They can be interpreted as comparing the average in, for example, an exposed group versus an unexposed group. This distinction between coefficients poses no problem in linear models, in which the coefficients are equivalent. However, in logistic and other nonlinear models, the coefficients are not equivalent, and in general for positive coefficients, $\beta_{SS} \geq \beta_{PA}$.

CHALLENGES IN APPLICATIONS

Sources of heterogeneity

Hierarchical models require the assumption that we can isolate unexplained heterogeneity in the random-effects components, and that we understand the structure of these components of the model. Further, we must have a reasonable idea of how the random effects contribute to variation. Models such as those described in the section on hierarchical models and Bayesian approaches are dependent on our specifying the random effects—and their distributions—correctly. In some settings, such as the successive pregnancies problem or the parental-exposures problem described in section on clustering outcomes, it is not immediately obvious how the random effects are structured; it may be that successive pregnancies from the same mother, but with different fathers, have a different random-effects structure than those from the same father, for example. Marginal models, although changing the target of inference from the subject-specific to the marginal, are more robust to these misspecifications because they do not rely on specification of the random effects structure.

Censoring (left, right, interval) and truncation

Censoring and truncation arise in time-to-event data when outcomes occur but are not observed (measured) as a part of the study. This may occur when the event or study outcome has occurred before individuals are enrolled in a study or after the study stops. Censoring occurs when an event is assumed to occur, but its precise timing is unknown. Right-censoring occurs when an event occurs after follow-up has stopped, so that we only know that the event occurred after the end of follow-up. In a survival setting, for example, we may know that on the last day of follow-up a subject is alive, so that we only know that death occurred after that date. If T is the time of the event, and C is the censoring time, we observe only the minimum value of C and T. Left-censoring occurs when we know that an event happened, but only that it occurred before enrollment or follow-up in a study. This is typical in the study of pregnancy, as pregnant women are often enrolled some time during pregnancy but not precisely before or at conception.

A rather unique censoring challenge in reproductive and perinatal epidemiology is the concept of interval censoring, which combines left- and right-censoring. We know that an event occurred within a time interval, but not the precise time within that interval. A simple example of interval censoring is the typical use of gestational age (GA). On birth certificates in the United States and Canada, GA is typically reported in completed weeks (Mathews et al. 2002). So, for example, we know that an infant born at 36 completed weeks was born between 36 weeks 0 days and 36 weeks 6 days, but we do not know the exact GA. Because of the shape of the GA distribution, we can infer that the actual GA distribution for all births with the same recorded GA must be skewed to higher values for preterm births and to lower values for post-term births.

The limiting case of censoring is current-status data, which is defined as occurring when, at the time we sample subjects, we only know that an event has occurred or not; that is, we only know the subject's current status. Current status data essentially combines left- and right-censoring because we know that, for subjects who have not yet had the event, it is right-censored, but for those who have had the event, it is left-censored (because we do not know precisely the timing at which the event occurred).

Left-, right-, or interval censoring can be handled in a variety of straightforward ways. If censoring is independent of the event, it is ignorable and can be dealt with easily (Klein and Moeschberger 2003). In effect, the available data form a random sample of the full data and can be used to represent the whole population. If censoring and the event of interest are not independent of each other, the subjects who are censored are not a random sample from the full dataset. In this case, the censoring mechanism must be modeled and accounted for. A simple way to do this is inverse probability weighting (Hernán et al. 2000). If one can model the censoring mechanism correctly, such that the probability of loss to follow-up can be estimated, then subjects can be weighted by the inverse probability of loss to follow-up. This weighting will account for the censoring by assigning extra weight to subjects with complete follow-up corresponding roughly to the number of subjects with similar characteristics who were lost to follow-up. Interval censoring

can be handled similarly, but with left- and right-censoring handled simultaneously (Klein and Moeschberger 2003).

Truncation occurs when time windows of observation are cut short or truncated, such that the endpoint may or may not occur during the time window of observation. Left truncation occurs when follow-up starts after a signal event (not necessarily the endpoint) has occurred, such that it is not known whether the endpoint occurred prior to the start of follow-up. This differs from left-censoring because we may not observe individuals at all if their data are truncated, whereas left-censoring allows us to know that an event has occurred, just not when it occurred. For example, if follow-up starts after a surgical procedure or after a prescription medication starts, subjects must survive to the end of the procedure in order to be followed-up; deaths prior to surgery may be truncated (Suissa 2008). Time-to-pregnancy studies typically enroll women or couples upon discontinuation of contraception. However, it is possible that contraception discontinuation occurs prior to enrollment, so that an unknown period of time exists when the couple is at risk for pregnancy (not using contraception) but not necessarily enrolled in the study. This situation may arise when a couple discontinues contraception, engages in sexual intercourse without any form of contraception, but does not enroll in the study for several days or weeks. This time is left truncated, because pregnancies and losses that occur during this period will not appear in the study; if a couple becomes pregnant successfully, they will simply not enroll, whereas pregnancies with early losses will not be recorded and the couple's fecundity will not be estimated correctly. Turnbull (1974, 1976) proposes methods for correcting for truncation. Truncation is a common phenomenon in reproductive and perinatal epidemiology, and represents an important area of research in the field.

CONCLUSION

Whatever the research question under study, it is imperative to consider a number of methodologic issues when designing reproductive or perinatal epidemiologic research. In keeping with the overarching epidemiologic method, investigators need to first define the referent and study population very carefully; as outlined above, sometimes these populations can be difficult to clearly define. In so doing, it should become apparent if issues such as "successive pregnancies" are a consideration, as may be encountered when enrolling women without regard to gravity or parity. The unit of analysis is another critical decision, particularly when multiple treatment cycles are a consideration or when anticipating the use of hierarchical data. This consideration is imperative for defining data collection and in building a data management system capable of supporting a hierarchical structure. Assuming the study design and cohort are appropriate for the research question, the investigator needs to then design an analytic plan that is sensitive to these, and other, unique challenges that underlie the field of reproductive and perinatal epidemiology. Collectively, these steps will help to ensure the correct analysis and interpretation of the study results within the context of study design, as well as inherent strengths and limitations.

REFERENCES

Ananth, C.V., Platt, R.W., and Savitz, D.A. 2005. Regression models for clustered binary responses: Implications of ignoring the intracluster correlation in an analysis of perinatal mortality in twin gestations. *Ann Epidemiol* 15(4): 293–301.

Anway, M.D. and Skinner, M.K. 2006. Epigenetic transgenerational actions of endocrine disruptors. *Endocrinology* 147(6 Suppl): S43–S49.

Bakketeig, L.S. and Hoffman, H.J. 1983. The tendency to repeat gestational age and birth weight in successive births, related to perinatal survival. *Acta Obstet Gynecol Scand* 62(5): 385–92.

Barker, D.J. 2002. Fetal programming of coronary heart disease. *Trends Endocrinol Metab* 13(9): 364–68.

Bonde, J.P., Hjollund, N.H., Jensen, T.K., Ernst, E., Kolstad, H., Henriksen, T.B., et al. 1998. A follow-up study of environmental and biologic determinants of fertility among 430 Danish first-pregnancy planners: Design and methods. *Reprod Toxicol* 12(1): 19–27.

Breslow, N. 2004. Whither PQL? In *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, D.Y. Lin and P. Heagerty, eds. New York: Springer.

Breslow, N.E. and Clayton, D.G. 1993. Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 88: 9–25.

Buck, G.M., Lynch, C.D., Stanford, J.B., Sweeney, A.M., Schieve, L.A., Rockett, J.C., et al. 2004. Prospective pregnancy study designs for assessing reproductive and developmental toxicants. *Environ Health Perspect* 112(1): 79–86.

Buck Louis, G., Dukic, V., Heagerty, P.J., Louis, T.A., Lynch, C.D., Ryan, L.M., et al. 2006. Analysis of repeated pregnancy outcomes. *Stat Methods Med Res* (15): 103–26.

Buck Louis, G.M., Schisterman, E.F., Dukic, V.M., and Schieve, L.A. 2005. Research hurdles complicating the analysis of infertility treatment and child health. *Hum Reprod* 20(1): 12–18.

Burton, P., Gurrin, L., and Sly, P. 1998. Extending the simple linear regression model to account for correlated responses: An introduction to generalized estimating equations and multi-level mixed modelling. *Stat Med* 17(11): 1261–91.

Carlin, J.B., Gurrin, L.C., Sterne, J.A., Morley, R., and Dwyer, T. 2005. Regression models for twin studies: A critical review. *Int J Epidemiol* 34(5): 1089–99.

Casella, G. and Berger, R.L. 2001. *Statistical inference*. Pacific Grove, CA: Duxbury.

Chapin, R.E., Robbins, W.A., Schieve, L.A., Sweeney, A.M., Tabacova, S.A., and Tomashek, K.M. 2004. Off to a good start: The influence of pre- and periconceptional exposures, parental fertility, and nutrition on children's health. *Environ Health Perspect* 112(1): 69–78.

de Mouzon, J., Spira, A., and Schwartz, D. 1988. A prospective study of the relation between smoking and fertility. *Int J Epidemiol* 17(2): 378–84.

Dunson, D.B., Bigelow, J.L., and Colombo, B. 2005. Reduced fertilization rates in older men when cervical mucus is suboptimal. *Obstet Gynecol* 105(4): 788–93.

Dunson, D.B. and Stanford, J.B. 2005. Bayesian inferences on predictors of conception probabilities. *Biometrics* 61(1): 126–33.

France, J.T., Graham, F.M., Gosling, L., and Hair, P.I. 1984. A prospective study of the preselection of the sex of offspring by timing intercourse relative to ovulation. *Fertil Steril* 41(6): 894–900.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 2003. *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall.

Goldstein, H., Browne, W., and Rasbash, J. 2002. Multilevel modelling of medical data. *Stat Med* 21(21): 3291–315.

Greenland, S. 2000. Principles of multilevel modeling. *Int J Epidemiol* 29: 158–67.

Hamilton, B.E. and Ventura, S.J. 2006. Fertility and abortion rates in the United States, 1960–2002. *Int J Androl* 29(1): 34–45.

Hassan, M.A. and Killick, S.R. 2004. Negative lifestyle is associated with a significant reduction in fecundity. *Fertil Steril* 81(2): 384–92.

Hernán, M.A., Brumback, B., and Robins, J.M. 2000. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 11(5): 561–70.

Hernán, M.A., Hernandez-Diaz, S., Werler, M.M., and Mitchell, A.A. 2002. Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology. *Am J Epidemiol* 155(2): 176–84.

Howards, P.P., Schisterman, E.F., and Heagerty, P.J. 2007. Potential confounding by exposure history and prior outcomes. *Epidemiology* 18: 544–41.

Hutcheon, J.A. and Platt, R.W. 2008. The impact of past pregnancy experience on subsequent perinatal outcomes. *Paediatr Perinat Epidemiol* 22: 400–408.

Hutcheon, J.A., Platt, R.W., Meltzer. S.J., and Egeland, G.M. 2006. Is birth weight modified during pregnancy? Using sibling differences to understand the impact of blood glucose, obesity, and maternal weight gain in gestational diabetes. *Am J Obstet Gynecol* 195(2): 488–94.

Jensen, T.K., Sobotka, T., Hansen, M.A., Pedersen, A.T., Lutz, W., and Skakkebaek, N.E. 2008. Declining trends in conception rates in recent birth cohorts of native Danish women: A possible role of deteriorating male reproductive health. *Int J Androl* 31(2): 81–92.

Klein, J.P. and Moeschberger, M.L. 2003. *Survival analysis: Techniques for censored and truncated data*. New York: Springer.

Liang, K.Y. and Zeger, S.L. 1993. Regression analysis for correlated data. *Annu Rev Public Health* 14: 43–68.

Louis, G.M., Cooney, M.A., Lynch, C.D., and Handal, A. 2008. Periconception window: Advising the pregnancy-planning couple. *Fertil Steril* 89(2 Suppl): e119–21.

Mathews, T.J., Menacker, F., and MacDorman, M.F. 2002. Infant mortality statistics from the 2000 period linked birth/infant death data set. *Natl Vital Stat Rep* 50(12): 1–28.

Neuhaus, J.M. and Kalbfleisch, J.D. 1998. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 54(2): 638–45.

Neuhaus, J.M., Kalbfleisch, J.D., and Hauck, W.W. 1991. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int Stat Rev* 59(1): 25–35.

Olsen, J. 1994. Options in making use of pregnancy history in planning and analysing studies of reproductive failure. *J Epidemiol Community Health* 48(2): 171–74.

Robins, J.M. 1998. *Marginal structural models*. 1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science.

Robins, J.M., Hernán, M.A., and Brumback, B. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11(5): 550–60.

Savitz, D.A., Dole, N., and Herring, A.H. 2006. Methodologic issues in the design and analysis of epidemiologic studies of pregnancy outcome. *Stat Methods Med Res* 15(2): 93–102.

Singer, J.D. and Willett, J.B. 2003. *Applied longitudinal data analysis: Modeling change and event occurrence*. London: Oxford University Press.

Suissa, S. 2008. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol* 167(4): 492–99.

Turnbull, B.W. 1974. Nonparametric estimation of a survivorship function with doubly censored data. *JAMA* 69: 169–73.

Turnbull, B.W. 1976. The empirical distribution function with arbitrarily grouped, censored and truncated data. *J R Stat Soc Series B Stat Methodol* 38: 290–95.

Verbeke, G. and Molenberghs, G. 2001. *Linear mixed models for longitudinal data*. New York: Springer.

Weinberg, C.R. 1993. Toward a clearer definition of confounding. *Am J Epidemiol* 137(1): 1–8.

Weinberg, C.R. 1995. Should we adjust for pregnancy history when the exposure effect is transient? [comment]. *Epidemiology* 6(3): 335–37.

Zeger, S.L., Liang, K.-Y., and Albert, P.S. 1988. Models for longitudinal data: A generalized estimating equation approach. Biometrics 44: 1049–60.

Zegers-Hochschild, F., Adamson, G.D., de Mouzon, J., Ishihara, O., Mansour, R., Nygren, K., et al. 2009. The International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary on ART terminology, 2009. *Hum Reprod* 24(11): 2683–87.

Zelen, M. 2004. Forward and backward recurrence times and length biased sampling: Age specific models. *Lifetime Data Anal* 10(4): 325–34.

Zinaman, M.J., Clegg, E.D., Brown, C.C., O'Connor, J., and Selevan, S.G. 1996. Estimates of human fertility and pregnancy loss. *Fertil Steril* 65(3): 503–09.

17

Novel Study Designs and Their Application in Reproductive and Perinatal Epidemiology

ENRIQUE F. SCHISTERMAN

The ultimate goal of an epidemiologist is to uncover the determinants of health and disease. Frequently, he or she is left to rely upon observational designs that necessitate caution in interpreting results, particularly with regard to causality. In the absence of experimental designs, an analogy of a time machine can be used to discuss the basic assumptions required for inferring causality (i.e., exchangeability, consistency, positivity, and proper model specification) when utilizing observational designs (Hernán and Robins 2010). In this analogy, study participants are observed for exposure over the course of their lives, travel back in time using the time machine to when their exposures are changed, and are then observed for a "second life." Researchers observe both lives to determine whether disease develops with the only difference being the presence of the exposure in one life and absence in the other. This paradigm allows the researcher to observe what happens to individuals in the presence and absence of the exposure with all else being equal. If occurrence of the outcome is identical in the two lives, despite differences in exposure, the researcher would conclude that the exposure did not cause the outcome. However, if the outcome varied by exposure, the researcher may conclude causality.

Obviously, because epidemiologists do not have a time machine study design available, there is a need to formalize assumptions underlying the use of observational methods when making inferences regarding causality. One crucial assumption is called *exchangeability*. Technically, exchangeability implies that the distribution of unobserved outcomes under unobserved treatment level (e.g., for the truly unexposed, this refers to the outcome they would have experienced had they been exposed) is the same as that of the observed outcomes under actual treatment (e.g., for the truly exposed, this refers to the outcomes they actually experienced). Under the counterfactual treatment, the individuals actually treated would experience disease like those not actually treated had they received treatment; individuals having received treatment are *substitutes* for individuals having received treatment. More basically, this assumption boils down to that of no unmeasured

confounding and no informative censoring, or ignorability of the treatment assignment and measurement of the outcome (Maldonado and Greenland 2002).

Reproductive and perinatal epidemiologists are faced with many unique study design challenges. Using the analogy of the time machine, and considering the assumptions necessary for determining causality, we explore the use of alternative designs that keep this framework in mind. This chapter reviews a description of select novel designs with relevancy for reproductive and perinatal epidemiology, viz, case-cohort study, case-crossover study, case-time-control study, a hybrid design for studies of biomarkers, case-only studies of gene–environment interactions, the case-parent triad design, a hybrid design for genetic studies, and time-to-pregnancy and current duration approaches for assessing fecundability. Each of these designs attempts to emulate findings that would be generated if study participants could in fact live two lives, with the exposure of interest being the only difference. The purpose of this chapter is not to review the basic principles and methods of study designs but, rather, to focus on the adaptations of the traditional case-control and cohort designs in reproductive and perinatal epidemiology. Several excellent textbooks address study designs more globally (Savitz 2003; Rothman et al. 2008). Appropriate settings for the application of each design to reproductive and perinatal epidemiology are considered.

NEW STUDY DESIGNS: ADAPTATIONS OF THE COHORT DESIGN

Case-cohort design

The case-cohort design is an approach for sampling from an established cohort and is an alternative to the well-known nested case-control study. As proposed by Prentice and Pyke (1979), the main feature of this study design is the ability to estimate the relative risk (as opposed to the odds ratio) for multiple outcomes without having to sample different controls for each specific outcome. In reproductive or perinatal epidemiology, we are frequently interested in assessing a spectrum of outcomes, such as length of gestation, gestational diabetes, and birth defects. In the case-cohort study design, a random sample of the cohort is selected at baseline and designated as the subcohort. Outcomes have not yet occurred at baseline; therefore, after a period of time, outcomes (event or censoring, administrative or otherwise) are observed for everyone in the cohort. For any particular outcome, the analysis is restricted to all cases and the subcohort for comparison purposes. Figure 17.1 illustrates the case-cohort design.

A common scenario that illustrates how this design might be used is one in which data and blood samples have been collected for a full cohort and there is interest in a biomarker of exposure; biochemical analyses may be restricted to cases and compared with a randomly selected subcohort to estimate the relative risk. A strength of this approach is that the source population is clearly specified and that each cohort member has an identical likelihood of being selected regardless of person-time experience contributed or disease status, even if the cohort

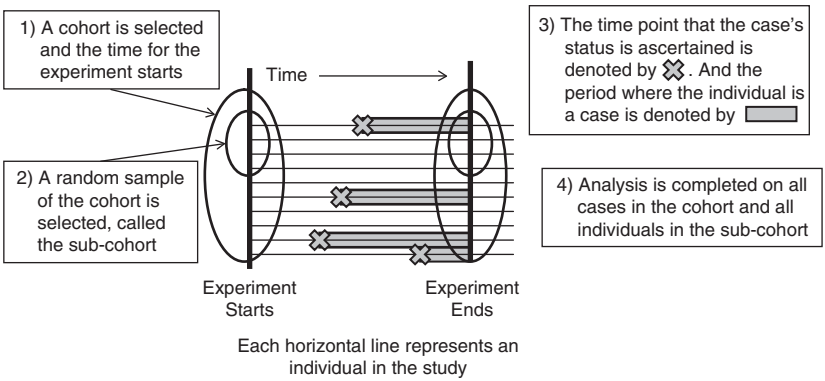


Figure 17.1 Case-cohort study.

is stratified in some manner (Rothman et al. 2008). Thus, the subcohort may be used to estimate the exposure distribution in the source population. This design is a logical way to conduct a study when the effect measure of interest is the ratio of incidence proportions (risk ratio) rather than the rate ratio, as is common in perinatal studies (Rothman et al. 2008). As such, this design permits estimation of the risk ratio without the need for the rare-disease assumption. It is also a cost-efficient design option, as the number of laboratory or other types of measurements need only be measured for cases and members of the subcohort, and not for the entire population. Unlike the case-control design that samples from the cohort as outcomes occur (i.e., incidence density sampling), the case-cohort design samples from the cohort at baseline to create the subcohort.

A key disadvantage of the case-cohort design is the overlap of membership between the subcohort and the cases. Since selection into the subcohort disregards disease status, a proportion of the subcohort will be cases and, thereby, contribute exposure (and covariate) information to both the case group and the subcohort. If the overlap is large, then larger subcohort sizes are needed to adjust for the controls that become cases. Compared to the ordinary case-control design, the case-cohort design might require sampling a larger number of controls to obtain the same statistical precision (Prentice and Pyke 1979).

Overlapping membership between the subcohort and cases also affects the data analysis. As noted previously, the random selection of the subcohort means that it will include cases selected at a proportion equal to the average cumulative incidence proportion. For example, in a cohort of women followed from conception to (preterm) delivery, approximately 12% of the subcohort may be cases while contributing also to the subcohort. This necessitates the use of appropriate data analysis techniques for addressing the nonindependence of risk sets. Conditional logistic regression is a readily available analytical technique that is appropriate for case-cohort designs, given that it approximates the partial likelihood (Wacholder and Boivin 1987).

Epidemiologists might be interested in the efficiency of the case-cohort design when assessing the association between genotype and a spectrum of pregnancy

outcomes, such as gravid health conditions (e.g., preeclampsia, gestational diabetes) or infant outcomes (e.g., length of gestation, birth size, and/or birth defects). Another example where this design would be of interest may be in studying the effects of insulin in pregnancy on multiple outcomes, such as preeclampsia, preterm delivery, and gestational diabetes.

Case-crossover design

The case-crossover design was originally proposed as an efficient study design for assessing acute effects (Maclure 1991). In this design, individuals with a particular outcome (cases) contribute information about exposures at the time(s) before onset of the outcome, as well as at time(s) other than preceding onset (controls). In essence, each person serves as his or her own control, thus eliminating the need for a separate comparison group (Rothman et al. 2008). The case exposure window is defined as the "at-risk" period preceding the acute study outcome and corresponds to the time that exposure might reasonably affect outcomes. The control windows are periods of comparable time that exclude the case window, as illustrated in Figure 17.2.

As Figure 17.2 illustrates, a higher proportion of exposure during the case versus control window would suggest an association between the exposure and outcome.

One important assumption is that neither the exposure nor confounder(s) change systematically over time. In addition to the concern for time-trends in exposure, important assumptions for use of the case-crossover design include: (a) timing of the exposure does not affect reporting; (b) the exposure is intermittent; (c) induction periods are short; and (d) study outcomes or effects are transient or acute rather than chronic in nature. Various analytic methods are appropriate for case-crossover designs, including conditional logistic regression, since it is similar to a matched analysis.

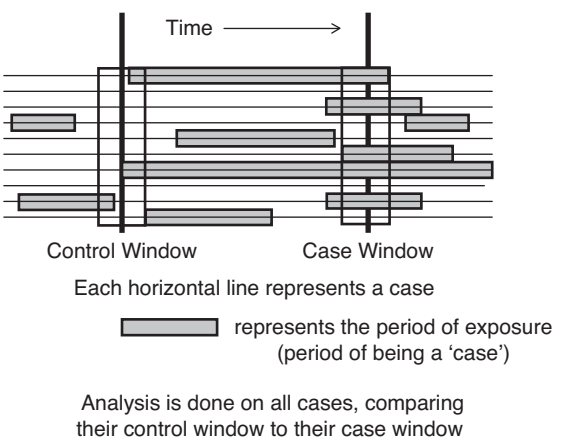


Figure 17.2 Case-crossover study.

The original application of this design was in the cardiovascular epidemiologic literature as an effort to identify triggers of myocardial infarction (Maclure 1991). Its utility for reproductive or perinatal epidemiology is considerable, given the diurnal patterns in reproductive and other hormones and the purported seasonality in natality or adverse birth outcomes (e.g., sudden infant death syndrome). Other applications might be the study of time-to-pregnancy (TTP) in attempting to identify determinants of pregnancy relative to earlier menstrual cycles without conception. For the case-crossover design, an important assumption is that the exposure must have some variability over time, thereby allowing for comparisons. Examples of a case-crossover design in reproductive and perinatal epidemiology include the assessments of daily air particulate matter or frequency of sexual intercourse as exposures and preterm delivery as the outcome. Differences in air particulate or sexual intercourse during the case window can be compared with the control windows to assess differences using a matched odds ratio. Since the woman serves as her own control, selection and recall bias are minimal.

An example of a creative use of the case-crossover design is a study assessing the use of folic acid and cardiovascular birth defects (Hernandez-Diaz et al. 2003). Investigators assessed maternal use of folic acid in a control window that was defined as the 2 months preceding the last menstrual period, followed by a washout period of 1 month; the case window was the following 2 months. The authors were able to vary the length and timing of the control window to assess the robustness of their findings with other designs (Hernandez-Diaz et al. 2003). As described, the case-crossover design has been applied to study several reproductive outcomes, and should be considered when interest is in the effects of acute exposures that are very common in reproductive and perinatal epidemiology.

Case-time-control design

Although similar to the case-crossover design in that subjects are observed during two or more time periods, the case-time-control design allows for consideration of exposures with a longer duration, rather than only those with acute or transient effects, and the design includes both cases and controls. Similar to the case-crossover design, the case-time-control design requires that each study participant be observed twice, once in a case window and at least once in a control window. The difference between the case-crossover and case-time-control designs is that, in the case-time-control design cases (subjects that have the outcome during the case window) and controls (subjects that do not have the outcome during the case window) are observed; in the case-crossover design, only cases are observed. For each case subject, exposure frequencies from the case window and the control windows are then compared through a matched odds ratio, as for the case-crossover design. Figure 17.3 illustrates the case-time-control design.

Among control study participants, the frequencies of exposure during the case windows are used as an estimate of the exposure distribution in the underlying cohort during two different periods of time and, therefore, the matched odds ratio between windows measures the time trend in exposure (Hernandez-Diaz et al. 2003). To minimize bias stemming from time trends in the exposure, two

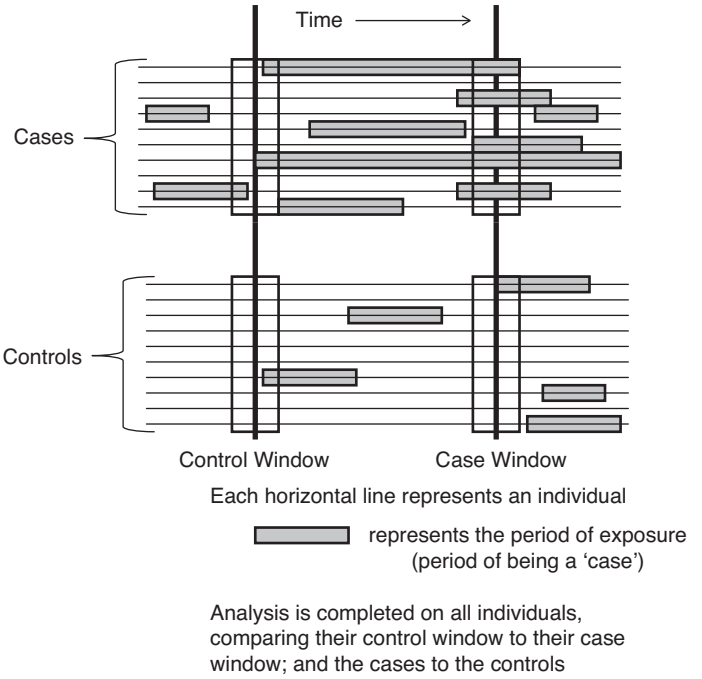


Figure 17.3 Case-time-control study.

assumptions must be met: (1) the case-crossover odds ratio is the product of the two component odds ratios—an odds ratio due to the causal effect of the exposure on the outcome, and an odds ratio due to the time trend in exposure prevalence; and (2) the time trend odds ratio is similar for cases and controls. Therefore, the case-time-control odds ratio is the case-crossover odds ratio (from the cases) divided by the time trend odds ratio (from the controls).

The advantage of the case-time-control design is that it reduces between-person confounding and bias arising from time trends. The reduction in bias due to time trends is only valid under the assumption that the exposure trends in cases are equivalent to the exposure trends in controls. In the absence of bias from time trends, a case-control design would be just as efficient as the case-time-control design, since both include controls. Unlike the case-crossover design, which is a case-only study type, the case-time-control design must consider control selection and any associated potential biases. Information bias, however, should be less of a concern for this type of design in comparison with the case-crossover design if exposure ascertainment is comparable for cases and controls. The case-time-control design can estimate relative risks in the absence of confounding for exposure of interest (Suissa 1995). To avoid effects of time-trends that will induce bias in case-crossover designs, one should consider the case-time-control design. This design has been applied in the birth defects literature without much gain in efficiency or reduction of bias; however, other areas of reproduction may benefit from application of this design (Hernandez-Diaz et al. 2003).

Hybrid design for studies involving biomarkers

Biomarkers of exposure and health outcomes are a main component of epidemiologic research, in part, to help assure the true classification of study participants with regard to exposure, susceptibility, or outcome. However, cost is often a limiting factor in determining the feasibility of conducting biomarker research, particularly the number and extent of laboratory analysis. This issue has prompted the development of hybrid study designs that utilize the pooling of biospecimens based upon a priori criteria. The basic premise of pooled designs is the physical combining, or "pooling" of a small amount of biological samples (e.g., urines or blood components) from a number of individuals for analysis. Assays are then performed to quantify the biomarker of exposures or outcomes in pooled samples and not in the individual samples. Pooling allows for fewer assays than would be required for evaluation of individual samples and is, thereby, an efficient approach for estimating means. However, an obvious downside is the total loss of all individual-level information. The hybrid design is a variation of basic pooling approaches that creates a sample of both pooled and unpooled samples to efficiently estimate the unknown parameters of the biomarker's distribution, those of central tendency (e.g., mean) as well as shape (e.g., variance). Figure 17.4 illustrates the hybrid design.

Key advantages of the hybrid and pooled designs are the reduction in cost and improved efficiency by reducing measurement error (i.e., pooling error, random measurement error, and limit of detection [LOD]). Measurements on both pooled and unpooled samples reduce measurement error without requiring replications, as in a validation study, which would be needed in unpooled designs. A common complexity arises when some participants have levels below some lower detection threshold, such as the LOD. Fortunately, methods to address these issues have been developed (Schisterman and Vexler 2008). Under these circumstances, biomarker

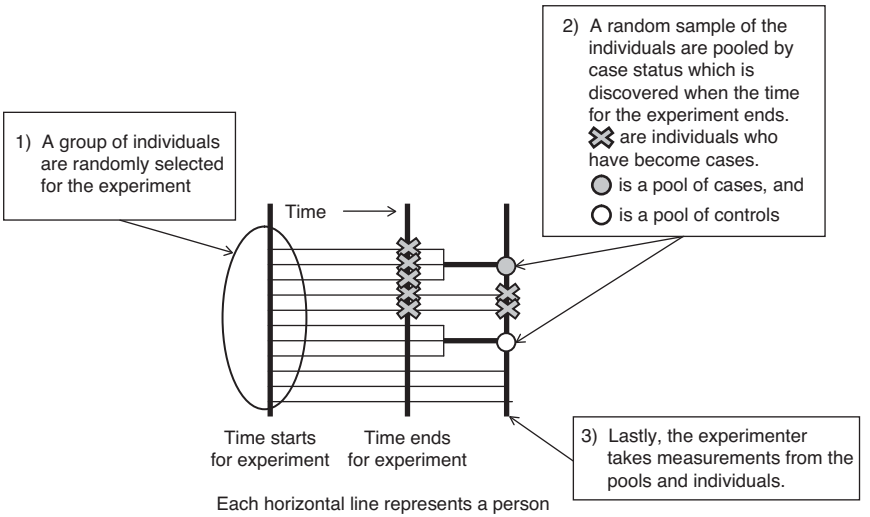


Figure 17.4 Hybrid design for studies with biomarkers.

values at or above the LOD are measured and reported, but values below the LOD may be unobservable. Maximum likelihood estimators to estimate means and variances are easily available using the hybrid design (Mumford et al. 2006). For example, if one is interested in the effect of polychlorinated biphenyl compounds (PCBs) on spontaneous abortion, one might consider this design, given that measurements of PCBs are very expensive, subject to measurement error and the LOD issue. Therefore, using this design, one can disentangle all of these limitations and unmask small and not obvious effects. The application of this design extends to other areas of reproductive and perinatal epidemiology where measurement of biospecimens is expensive and subject to measurement error and LOD.

Case-only studies of gene–environment interactions

Conventional study designs aimed at assessing interactions require large random samples of cases and controls for sufficient statistical power. This potential limitation of conventional study designs for gene–environment interactions has particular impact for studies of rare perinatal outcomes, such as birth defects. In addition, fiscal considerations that arise related to the expense of measuring genetic and environmental factors in large case-control studies have prompted the development of novel study designs for answering questions about gene–environment interactions. One such example is the case-only design (Piegorsch et al. 1994). As the name implies, only cases are required, thus eliminating the expenses and complexities related to sampling controls. Figure 17.5 illustrates the case-only design.

Gene–environment interactions are assessed by cross-tabulation of genetic (G) and environmental (E) exposures that classify cases as either G+/E+, G+/E–, G–/E+, or G–/E–. The cross-product ratio depends on both the interaction and the population association of G and E. Only if the genetic and environmental risk factors are independent in the population is the cross-product ratio equal to the risk ratio. This is a strong underlying assumption of the case-only design, viz, if a particular genotype affects the distribution of the environmental factor (e.g., a case-only study of the interaction between gene G and smoking status would be invalid due to failure to meet this assumption if gene G is involved in the development

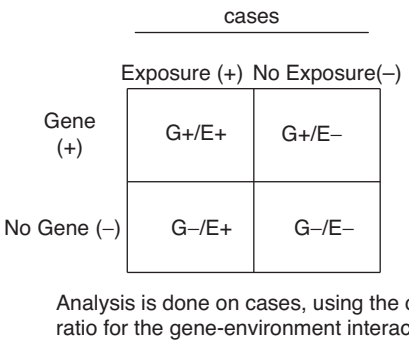


Figure 17.5 Case-only studies of gene-environment interactions.

of a smoking addiction). If the risk of disease is small at all levels of the study variables, it is approximately equal to the odds ratio for the gene–environment interaction that is computed from case-control data. The case-only design has been demonstrated to be an efficient design for estimating gene–environment interactions when important assumptions can be met about the exposure, particularly independence of the genetic and environment factors in the population and the rare-disease assumption.

Piegorsch and colleagues (1994) have demonstrated that this design offers greater precision in estimating interactions (i.e., smaller standard errors) in comparison to the case-control design. However, caution is necessary with regards to assigning a causal relation between the genotype evaluated and the outcomes assessed, as the at-risk genotype may be a marker for other genes or nongenetic causal factors. Noncausal gene variants may correlate with causal gene variants due to admixture. Accordingly, as with a regular case-control study investigating gene–environment interactions, associations may arise due to linkage disequilibrium between the genetic marker and the true susceptibility allele(s) at neighboring loci.

The case-only design offers much promise for perinatal epidemiology, particularly for the study of gene–environment interactions of rare outcomes, such as birth defects. For example, by examining infants with cleft palates (cases only), it has been found that the gene for transforming growth factor (TGFα) combined with maternal smoking during pregnancy significantly increased the risk of cleft palate (Piegorsch et al. 1994; Umbach and Weinberg 1997). Although population-based sampling frameworks are available for studying many birth defects, choice of controls is exceedingly difficult as many affected pregnancies are terminated (only birth prevalence of birth defects may be available for study). In addition, there is the strong possibility of recall bias when sampling mothers of unaffected babies (Rockenbaurer et al. 2001). Due to the many limitations and assumptions that accompany the case-only design, it cannot substitute for a well-conducted case-control study. However, this design does offer an important epidemiologic tool for assessing gene–environment interactions.

Case-parent triads

The case-parent triad design has been proposed as a variant of the case-cohort design in studying associations between genetic factors and disease (Ahsan et al. 2002; Khoury and Flanders 1996; Laird and Lange 2008). This design also allows for assessment of gene–gene interactions, gene–environment interactions, and maternal effects (Umbach and Weinberg 2000; Weinberg et al. 1998; Wilcox et al. 1998). The case-parent triad design compares the genotype of the cases with the genotype of a fictitious control formed by the nontransmitted allele from each parent of the case subject (Piegorsch et al. 1994; Khoury and Flanders 1996). Cases for which information from both genetic parents is not available are therefore excluded, although methods have been developed to account for their missing data (Bergemann and Huang 2009). Relative risks can be estimated directly from the frequency distribution of the case-parent triads.

Compared with the case-control design, the case-parent triad design has both advantages and disadvantages. Although case-control studies are vulnerable to population stratification or confounding arising from the characteristics of the population structure, case-parent triad designs overcome this bias. The case-parent triad design minimizes population stratification bias by using controls of family members of the case, thus matching with respect to genetic ancestry. In addition, case-parent triad designs can produce valid estimates of relative risks with selected cases, given the inherent matching of parents with cases. The stratification on parental mating type absorbs variations in recruitment rates that lead to over-representation of certain parental mating types (e.g., because of cultural factors). Case-parent triad designs can also work well in settings where cases are highly selected (e.g., drawn from a clinic or support-group setting). Cases and parents of cases may be more motivated to participate in research than are parents of unaffected children, especially in genetic-based diseases. Last, similar to the case-only design, the case-parent triad approach does not allow assessment of the independent effects of environmental exposure, but on the other hand, it does allow for assessment of the effect of the genotype. These advantages make the case-parent triad design a method of choice for preliminary studies of candidate alleles related to conditions of pregnancy or early life (Wilcox et al. 1998), and the design holds promise for application to other areas of reproductive and perinatal epidemiology.

Hybrid design for genetic studies

To bring together the advantages of both family-based and population-based approaches, recent papers have proposed hybrid designs that blend elements of the family- and population-based study designs (Nagelkerke et al. 2004; Epstein et al. 2005; Weinberg and Umbach 2005). The proposed hybrid blends case-parent and control-parent triads, and genotypes case-parent triads limited to the parents of control offspring (case-parent triad/control parents design) (Weinberg and Umbach 2005). This hybrid design greatly improves power for the evaluation of offspring- and maternally-mediated genetic effects. It allows testing of potential biases arising from population stratification and the use of case-parent triads to address any observed bias. In addition, the assumption of population parental mating symmetry can be tested, and maternal genetic effects can be validly estimated even if mating symmetry is rejected. The case-parent triad/control-parents hybrid design calls for both parents of controls to be genotyped, which can prove challenging, given the reported difficulty in recruiting fathers in some studies.

Recently, a new hybrid design called the control-mothers dyad has been developed (Vermeulen et al. 2009). Similar to the previous designs, information on exposures is collected for both the case offspring and the unrelated control offspring. In the previous hybrid design that entails genotyping control parents, one can directly estimate the mating-type frequencies in the source population. In the control-mothers dyad, genotyping information is collected from control mothers and control children, who act as surrogates for their fathers' genotyping data. This recent alternative hybrid design has greater power than the case-parents or the case-mother/control-mother

designs, but at the expense of reduced power when using hybrid designs inclusive of control parents. This alternative hybrid design also retains the ability to test for bias due to population stratification and, thereby, to examine whether case-parents and control data can be safely combined. By combining the advantages of both family- and population-based approaches, these hybrid designs are a valuable tool for reproductive and perinatal epidemiologists interested in evaluation of offspring- and maternally mediated genetic effects.

Assessing fecundability–time-to-pregnancy and current-duration designs

Perhaps one of the most exciting, though challenging, aspects of studying human reproduction and development is the importance of timing of various events. As a result, reproductive epidemiologists must utilize study designs that can address timing and the correlated data structure that is biologically consistent with human reproduction and development. In addition, past reproductive history is informative about future reproductive outcomes, thus necessitating use of modeling techniques capable of addressing the correlated data structure as outlined in Chapter 16 (Platt et al. 2004; Buck Louis et al. 2005; Ananth 2007).

Reproductive epidemiologists recognize the need for study designs that may accommodate different units of analysis within a data structure. For example, TTP studies may require day-specific information on lifestyle, cycle-level information about menstruation, woman-level information such as age, or couple-level information such as frequency of intercourse. The ability to capture acute exposures (e.g., intercourse during the fertile window) together with the need for capturing acute effects, such as the rise in luteinizing hormone as a proxy of ovulation, is paramount for answering critical data gaps regarding most aspects of human reproduction and development. Examples of circumstances in which these study designs are appropriate include the current-duration approach (Keiding et al. 2002), prospective TTP (Buck et al. 2004), and retrospective TTP designs (Joffe et al. 2005); most studies utilize prospective cohort designs with longitudinal capture of data and often in varying levels ranging from day- or cycle-level to couple- or population-level.

Advantages and disadvantages of the retrospective and prospective TTP designs have been extensively discussed in the literature (Weinberg et al. 1994; Joffe et al. 2005; Bonde et al. 2006; Key et al. 2008). Essentially, the main disadvantage of the prospective TTP study lies in the sampling frame being based on couples planning pregnancy, or wanting to become pregnant. The retrospective TTP sampling frame is usually based on couples who became pregnant, thereby excluding infertile couples. A newer study design aimed at addressing the period of time when women are at risk of pregnancy is the current-duration approach (Keiding et al. 2002; Slama et al. 2002). This design utilizes a cross-sectional sampling framework of women/couples at risk for pregnancy, irrespective of their pregnancy intentions, and then queries them about the length of time they have been at risk for pregnancy. Figure 17.6 illustrates the TTP using a current-duration approach.

The current-duration approach produces a distribution of time when couples have been at risk or the duration of the current time period at risk. An adaptation

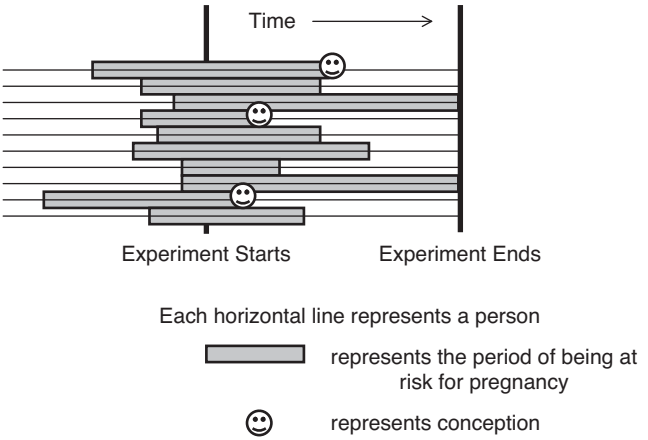


Figure 17.6 Time-to-pregnancy studies.

of this design is a prospective component allowing for the follow-up of couples for TTP after initial contact, and reporting is based on the case-cohort assumptions. Major advantages of the current-duration approach are related to the inclusion of couples who do not conceive and the absence of an a priori requirement for couples to define when they first attempted to become pregnant (Slama et al. 2006). It is important to note that, unlike the classical prospective TTP study, in which only couples that are activity trying to become pregnant are recruited, in this design, couples are recruited during the period they are "at risk" of pregnancy, given the absence of contraceptive usage while being sexually active. Sampling options for the current-duration approach include random selection of a population via various venues, then screening for eligibility criteria such as sexually active, not biologically or surgically sterile, or other required criteria.

The current-duration approach cannot distinguish between attempts to conceive that end in a pregnancy and attempts to conceive that are terminated for other reasons, nor does it systematically capture spontaneous pregnancies (i.e., those without a clear time duration of attempting to conceive). Furthermore, it has not yet been empirically demonstrated how practical it is to distinguish long current durations from sterile or behaviorally nonfecund couples (Scheike and Keiding 2006).

Recently, a population-based current-duration study was undertaken in France by Slama and colleagues (2006). Sixty-nine (<1% of the 7,699 dialed telephone numbers) women at risk for pregnancy were interviewed. Using a parametric approach relying on a Pareto distribution, the estimated percentages of couples not pregnant after 12 and 24 months despite the absence of contraception were 34% and 16%, respectively. The authors noted that the current-duration approach yielded a higher eligibility rate than incident cohort designs that rely on couples initiating attempts to become pregnant. Although a more expensive and intense data collection process is required to apply this design, it may overcome some of the potential limitations of the other previously utilized TTP studies, especially those retrospective in design. More research is needed to find ways to reduce cost

while efficiently reaching the target population of couples at risk for becoming pregnant.

CONCLUSION

Reproductive and perinatal epidemiology faces many unique study design challenges to adequately address bias and measurement error when evaluating etiologic or predictive questions. This chapter has provided an introduction and broad overview of novel study design options, including hybrid designs for a more specific analysis focusing on the genetic determinants of adverse outcomes. Using the analogy of the time machine for observational study designs, each of the novel designs reviewed attempted to emulate findings that would be generated if study participants could in fact live two lives, with the exposure of interest being the only difference. Despite the promise of newer observational study designs, many unresolved issues remain which support the need for even more design options.

REFERENCES

Ahsan, H., Hodge, S.E., Heiman, G.A., Begg, M.D., and Susser, E.S. 2002. Relative risk for genetic associations: The case-parent triad as a variant of case-cohort design. *Int J Epidemiol* 31:669–78.

Ananth, C.V. 2007. Menstrual versus clinical estimate of gestational age dating in the United States: Temporal trends and variability in indices of perinatal outcomes. *Paediatr Perinat Epidemiol* 21(Suppl 2):22–30.

Bergemann, T.L. and Huang, Z. 2009. A new method to account for missing data in case-parent triad studies. *Hum Hered* 68:268–77.

Bonde, J.P., Hjollund, N.H., Jensen, T.K., Ernst, E., Kolstad, H., Henriksen, T.B., et al. 1998. A follow-up study of environmental and biologic determinants of fertility among 430 Danish first-pregnancy planners: Design and methods. *Reprod Toxicol* 12:19–27.

Bonde, J.P., Joffe, M., Sallmén, M., Kristensen, P., Olsen, J., Roeleveld, N., and Wilcox, A. 2006. Validity issues relating to time-to-pregnancy studies of fertility. *Epidemiology* 17:347–49.

Buck Louis, G.M., Schisterman, E.F., Dukic, V.M., and Schieve, L.A. 2005. Research hurdles complicating the analysis of infertility treatment and child health. *Hum Reprod* 20:12–18.

Buck, G.M., Lynch, C.D., Stanford, J.B., Sweeney, A.M., Schieve, L.A., Rockett, J.C., et al. 2004. Prospective pregnancy study designs for assessing reproductive and developmental toxicants. *Environ Health Perspect* 112:79–86.

Epstein, M.P., Veal, C.D., Trembath, R.C., Barker, J.N., Li, C., and Satten, G.A. 2005. Genetic association analysis using data from triads and unrelated subjects. *Am J Hum Genet* 76:592–608.

Hernán, M.A. and Robins, J.M. 2010. *Causal inference*, 1st ed. Boca Raton, FL: CRC.

Hernandez-Diaz, S., Hernán, M.A., Meyer, K., Werler, M.M., and Mitchell, A.A. 2003. Case-crossover and case-time-control designs in birth defects epidemiology. *Am J Epidemiol* 158:385–91.

Joffe, M., Key, J., Best, N., Keiding, N., Scheike, T., and Jensen, T.K. 2005. Studying time to pregnancy by use of a retrospective design. *Am J Epidemiol* 162:115–24.

Keiding, N., Kvist, K., Hartvig, H., Tvede, M., and Juul, S. 2002. Estimating time to pregnancy from current durations in a cross-sectional sample. *Biostatistics* 3:565–78.

Key, J., Best, N., Joffe, M., et al. 2009. Methodological issues in analyzing time trends in biological fertility. Protection bias. *Am J Epidemiol* 169: 285–93.

Khoury, M.J. and Flanders, W.D. 1996. Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: Case-control studies with no controls! *Am J Epidemiol* 144:207–13.

Laird, N.M. and Lange, C. 2008. Family-based methods for linkage and association analysis. *Adv Genet* 60:219–52.

Maclure, M. 1991. The case-crossover design: A method for studying transient effects on the risk of acute events. *Am J Epidemiol* 133:144–53.

Maldonado, G. and Greenland, S. 2002. Estimating causal effects. *Int J Epidemiol* 31:422–29.

Mumford, S.L., Schisterman, E.F., Vexler, A., and Liu, A. 2006. Pooling biospecimens and limits of detection: Effects on ROC curve analysis. *Biostatistics* 7:585–98.

Nagelkerke, N.J., Hoebee, B., Teunis, P., and Kimman, T.G. 2004. Combining the transmission disequilibrium test and case-control methodology using generalized logistic regression. *Eur J Hum Genet* 12:964–70.

Piegorsch, W.W., Weinberg, C.R., and Taylor, J.A. 1994. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Stat Med* 13:153–62.

Platt, R.W., Joseph, K.S., Ananth, C.V., Grondines, J., Abrahamowicz, M., and Kramer, M.S. 2004. A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death. *Am J Epidemiol* 160:199–206.

Prentice, R.L., and Pyke, R. 1979. Logistic disease incidence models and case-control studies. *Biometrika* 66(3):403–11.

Rockenbaurer, M., Olsen, J., Czeizel, A.E., Pedersen, L., S rensen, H.T., and EuroMAP Group. 2001. Recall bias in a case-control surveillance system on the use of medicine during pregnancy. *Epidemiology* 12:461–66.

Rothman, K.J., Greenland, S., and Lash, T.L. 2008. *Modern epidemiology*, 3rd ed. Philadelphia: Lippincott Williams and Wilkins.

Savitz, D.A. 2003. *Interpreting epidemiologic evidence: Strategies for study design and analysis*. New York: Oxford University Press.

Scheike, T.H. and Keiding, N. 2006. Design and analysis of time-to-pregnancy. *Stat Methods Med Res* 15:127–40.

Schisterman, E.F., Vexler, A., and Liu, A. 2005. To pool or not to pool: That is the question? *Epidemiology*.

Schisterman E.F. and Vexler, A. 2008. To pool or not to pool, from whether to when: Applications of pooling to biospecimens subject to a limit of detection. *Paediatr Perinat Epidemiol* 22:486–96.

Slama, R., Eustache, F., Ducot, B., et al. 2002. Time to pregnancy and semen parameters: A cross-sectional study among fertile couples from four European cities. *Hum Reprod* 17:503–15.

Slama, R., Ducot, B., Carstensen, L., Lorente, C., de La Rochebrochard, E., Leridon, H., et al. 2006. Feasibility of the current-duration approach to studying human fecundity. *Epidemiology* 17:440–49.

Suissa, S. 1995. The case-time-control design. *Epidemiology* 6:248–53.

Umbach, D.M. and Weinberg, C.R. 1997. Designing and analysing case-control studies to exploit independence of genotype and exposure. *Stat Med* 16(15):1731–43.

Umbach, D.M. and Weinberg, C.R. 2000. The use of case-parent triads to study joint effects of genotype and exposure. *Am J Hum Genet* 66:251–61.

Vermeulen, S.H., Shi, M., Weinberg, C.R., et al. 2009. A hybrid design: Case-parent triads supplemented by control-mother dyads. *Genet Epidemiol* 33:136–44.

Wacholder, S. and Boivin, J.F. 1987. External comparison with the case-cohort design. *Am J Epidemiol* 126:1198–1209.

Weinberg, C.R., Baird, D.D., and Wilcox, A.J. 1994. Sources of bias in studies of time to pregnancy. *Stat Med* 13:671–81.

Weinberg, C.R., Wilcox, A.J., and Lie, R.T. 1998. A log-linear approach to case-parent-triad data: Assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am J Hum Genet* 62:969–78.

Weinberg, C.R. and Umbach, D.M. 2005. A hybrid design for studying genetic influences on risk of diseases with onset early in life. *Am J Hum Genet* 77:627–36.

Wilcox, A.J., Weinberg, C.R., and Lie, R.T. 1998. Distinguishing the effects of maternal and offspring genes through studies of "case-parent triads." *Am J Epidemiol* 148:893–901.

Index

Note: Page numbers followed by "*f*" and "*t*" denote figures and tables, respectively.