American Journal of Epidemiology

Vol. 162, No. 2
Printed in U.S.A.
DOI: 10.1093/aje/kwi172

REVIEW

Studying Time to Pregnancy by Use of a Retrospective Design

Michael Joffe[1], Jane Key[1], Nicky Best[1], Niels Keiding[2], Thomas Scheike[2], and Tina Kold Jensen[3]

[1] Department of Epidemiology and Public Health, School of Medicine, Imperial College, London, United Kingdom.
[2] Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark.
[3] University of Southern Denmark, Odense, Denmark.

Received for publication October 28, 2004; accepted for publication March 2, 2005.

Biologic fertility can be measured using time to pregnancy (TTP). Retrospective designs, although lacking detailed timed information about behavior and exposure, are useful since they have a well-defined target population, often have good response rates, and are simpler and less expensive to conduct than prospective studies. This paper reviews retrospective TTP studies from a methodological viewpoint and shows how methodological problems can be avoided or minimized by appropriate study design, conduct, and analysis. Sensitivity analyses using data from four European retrospective TTP studies are presented to explore the issues. Although the identified biases tend to have small impacts, the effects are not systematic across studies, and sensitivity analyses are recommended routinely. Planning bias can be checked by comparing propensity to report contraceptive failures in different exposure groups. Medical intervention bias can be avoided by censoring and inclusion of unsuccessful pregnancy attempts. Truncation bias can be a serious problem if unrecognized, but it is avoidable with appropriate study design and/or analysis. Behavior change bias can be minimized by assessing the covariates at the beginning of unprotected intercourse. More complete inference is possible if the study design covers the whole population, not just those who achieve a pregnancy.

data collection; fertility; infertility; questionnaires; reproduction

Abbreviation: TTP, time to pregnancy.

Time to pregnancy (TTP) measures how long a couple takes to conceive. The TTP distribution in a population describes its degree of fertility. It is a functional measure, the final common path of a large number of biologic mechanisms in both sexes (1); its use is complementary to more mechanistic research on the biologic processes necessary for fertility and studies of specific medical conditions. Information about TTP is easy to obtain, and it has proven useful in descriptive epidemiology, for example, to identify time trends (2) and spatial variation (3) and to identify risk factors (4–7). Since the unit of study is the couple, covariates relating to both partners are required.

TTP can be studied by use of either a prospective or retrospective design. Each has its strengths and weaknesses, and their roles are complementary. Prospective studies (the detailed discussion of which is beyond the scope of this paper) recruit couples at the start of their attempt, that is, couples who are prepared to undertake regular testing and to be followed up. These studies are therefore able to obtain detailed, timed information on key biologic events, including ovulation, implantation, and exposures, and on certain covariates, such as the frequency/timing of intercourse (8). They are becoming more sophisticated, and their feasibility has been demonstrated (8–11). The drawbacks are lack of a clear sampling frame (except in occupational studies) (12) and need for highly motivated participants, which might cause response bias and planning bias (see below). These drawbacks may adversely impact external validity to

Correspondence to Dr. Michael Joffe, Department of Epidemiology and Public Health, School of Medicine, Imperial College, St. Mary's Campus, Norfolk Place, London W2 1PG, United Kingdom (e-mail: m.joffe@imperial.ac.uk).

a degree that cannot be empirically evaluated, so that the findings may be difficult to generalize to all couples. They are, however, considered acceptable if the primary focus of the study is on etiologic factors, such as environmental exposures, where within-cohort comparisons of exposures are performed. Information from nonparticipants may also be available, mitigating the response bias.

A key advantage of the retrospective design is that it is possible to achieve a sample that is representative of the target population, with the consequent benefit of high external validity, which is particularly important for descriptive studies, for example, on time trends or spatial comparisons. If studies are well designed, conducted, and analyzed, most of the theoretical problems can be avoided or minimized in practice. The main drawback is that it is not possible to obtain detailed time-specific information about behavior and risk factors.

A third possible design is the current duration approach (13–15). This uses a cross-sectional survey design and focuses on couples who are currently having unprotected intercourse, asking participants about the duration of their current attempt. Under certain assumptions, these data can be used to generate the TTP distribution (14). The sampling frame is clearly defined, recall bias is minimal, and biologic samples can also be collected (15). In addition, a combination design is possible with follow-up of the couples, based on principles from the case-cohort design (16). These very promising newer designs are discussed elsewhere (14).

This paper reviews retrospective TTP studies from a methodological viewpoint, not including technical statistical issues. It covers study design, data collection, and potential biases. TTP analysis has evolved a series of sensitivity analyses, designed to investigate the extent to which different definitions or assumptions could lead to bias. We illustrate a series of such analyses, using four data sets from previously analyzed European studies: the European Multicentre Study on Infertility and Subfecundity (3), the Office for National Statistics (ONS) Omnibus Survey (2), the Danish Twin Study (17), and the Odense Prenatal Study (18). Each of these was approved by the appropriate ethics committee.

DESIGNING A RETROSPECTIVE STUDY

Sampling frame

Retrospective TTP data should be derived from special surveys with a population basis or with a well-defined source population; volunteer samples should be avoided since they suffer from the same drawbacks as prospective studies. Clinical samples (e.g., women seeking infertility treatment) are subject to strong selection and self-selection forces that are difficult to assess and do not provide satisfactory data. There are three main sampling frames.

Pregnancy-based studies. A sample of pregnant or delivering women is asked about the TTP relating to the current pregnancy. Where prenatal care is not universal, recruitment at delivery is preferable to use of prenatal clinics. The advantage of this design is that it is easy to define and contact the women and, as they are asked about the current pregnancy, recall bias is minimal. The disadvantages are that sterile couples are excluded and that subfecund couples are underrepresented.

Cross-sectional population-based or occupationally based studies. Couples are selected randomly from the general population or an occupational group. They are asked about TTP for either all previous pregnancies or just one (e.g., the first), and it is possible to inquire about unsuccessful attempts and sterility (see below). As the focus is on previous exposures, this design approximates a retrospective cohort study. Its advantage is that information on the whole reproductive history can be obtained. The disadvantage is long recall with possible recall bias, including for exposure variables and covariates.

Population-based birth cohort study. Members of a previously defined birth cohort are questioned about their reproductive history up to the interview (only possible when the cohort is old enough). This is similar to cross-sectional population-based or occupationally based studies, but it also allows a longitudinal study of factors operating earlier in life.

Pregnancies and unsuccessful attempts

While the term ''time to pregnancy'' implies that data are collected on periods of unprotected intercourse only if they lead to conception, it is possible also to include unsuccessful attempts if the study design is not pregnancy based. This is preferable, as stronger inferences can then be made: Regression estimates are not conditional on achieved conception, as they are with pregnancy-based studies (19), and selection bias from including only fertile couples is avoided. The length of the unsuccessful attempt generally has to be at least 6 months for the couple to remember these attempts and to provide valid information about them.

In data collection, one decision is whether to include only one pregnancy (or unsuccessful attempt) per couple or to study all of them. If more than one event is included in the analysis, which is more efficient, it is necessary to use mixed (frailty) models (20, 21), as the TTP values for a couple's different pregnancies are likely correlated. However, there are currently no standard programs for the discrete multivariate survival analysis that would be required. Only data on the first pregnancy/attempt can be considered unbiased, as relatively fertile couples may tend to have more children, and adverse reproductive experience may influence behavior; these complex issues need to be considered if later pregnancies/attempts are studied.

Obtaining TTP information and its validity

Ascertainment of TTP and unsuccessful attempts requires only a few simple questions (see the suggested questionnaire in the Appendix). These questions have been found to be readily acceptable in a wide range of cultures, and refusal to answer them is rare (12). Validation studies show that replies give an accurate representation of the true TTP distribution (22–25), even with recall up to 20 years (25). This is not true for nonbirth outcomes, for example, miscarriages, which are therefore commonly excluded (1). Men can also provide valid information, generating the same distributions and analytical results as for women drawn

from the same population (2, 7), albeit with more item nonresponse (typically 10–15 percent compared with 5–10 percent for women) and digit preference.

However, individual-level misclassification is common, including digit preference, leading to nondifferential errors (24), requiring a larger number of respondents than would otherwise be needed. In practice, stable estimates of the TTP distribution can be obtained with approximately 200 values per exposure group or even fewer with ordered categories, such as successive 5-year periods (2).

POSSIBLE BIASES IN RETROSPECTIVE TIME TO PREGNANCY STUDIES AND HOW TO MINIMIZE THEM

Several methodological issues have been raised in relation to retrospective TTP studies, and their theoretical impact has been discussed (12, 26). We present here a series of analyses that explore the practical impact of these potential problems. They are sensitivity analyses, exploring to what extent findings change if inclusion criteria or details of the analytical method are altered. We use data from four existing data sets for these analyses. Such inclusion criteria could be whether or not to include unsuccessful attempts or pregnancies ending in miscarriages; an example of an analytical detail is the value chosen for censoring. In some cases, the alternate analysis is quite extreme and unrealistic, but if the findings change little this provides reassurance. It has long been recommended that such analyses be conducted routinely in TTP studies (1, 26), but they are generally not presented in substantive research reports.

Sterility and severe subfertility

TTP in its strict meaning relates only to periods of unprotected intercourse that end with conception, excluding sterile couples and systematically underrepresenting subfertile couples. It is therefore unsuitable for assessing clinical subfertility or infertility. By construction, the cumulative TTP distribution eventually reaches 100 percent (figure 1, circles). If unsuccessful attempts are added to the pregnancy-related TTP values, this generates what has been termed ''time of unprotected intercourse'' (3), which applies to all ''attempts,'' whether successful or not, for which the cumulative distribution approaches but does not reach 100 percent, even after many years (figure 1, triangles).

In figures 2, 3, and 4, we present sensitivity analyses for the effect of maternal age and smoking (where available) on TTP, using four major European data sets; the legend to figure 2 contains details on these studies. All analyses are based on the discrete-time analogue of the Cox model (27), and they show the estimated odds ratios with 95 percent confidence intervals. Analysis *a* shows the baseline analysis, using TTP for successful attempts only and excluding contraceptive failures. Analysis *b* illustrates the impact on the odds ratios for maternal age and smoking of adding data on infertile phases (unsuccessful attempts) to the TTP distribution, for those data sets that contain the relevant information. Inclusion of infertile phases makes the effect of smoking (which is no longer conditional on achieved conception) on fertility more clear-cut, in the one data set where this analysis is possible, but no consistent impact is seen on the effect of maternal age.

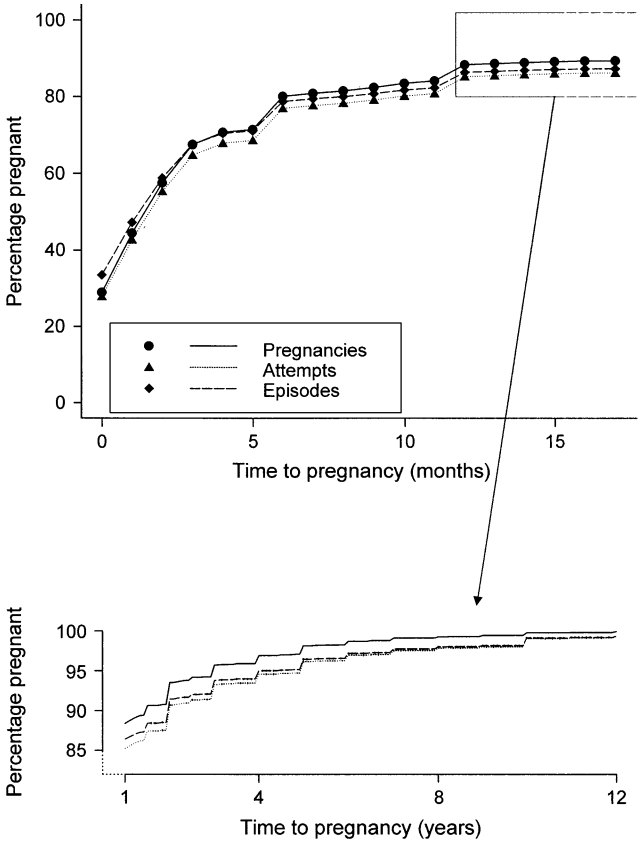

FIGURE 1. Population fecundity distributions by different inclusion criteria, with data taken from a Danish study conducted in 1998–1999 of twins born in 1931–1952 (Hum Reprod 2005;20:955–64) (17). The figure shows the cumulative distribution of time to pregnancy (circles), all attempts (time to pregnancy + unsuccessful attempts) (triangles), and all episodes (time to pregnancy + unsuccessful attempts + contraceptive failures) (diamonds). In the latter analysis, contraceptive failures are assigned the value of 0 for time to pregnancy. The inset shows the right-hand tail of the distribution.

''Accidental'' and unplanned pregnancies

At the other end of the continuum of fertility, the concept of TTP does not apply to conceptions that are ''accidental,'' in the sense of resulting from failure of a contraceptive method. These conceptions could result from inconsistent use and/or from failure of the method itself. Such a pregnancy does not have an eligible TTP value, and a screening question is therefore used, asking whether the pregnancy resulted from contraceptive failure. The dashed line with diamonds in figure 1 shows the typical distribution, which now includes all known episodes (''attempts'' plus ''accidents,'' the latter being included with the TTP value set at 0). If one exposure group (e.g., smokers) has a higher degree of

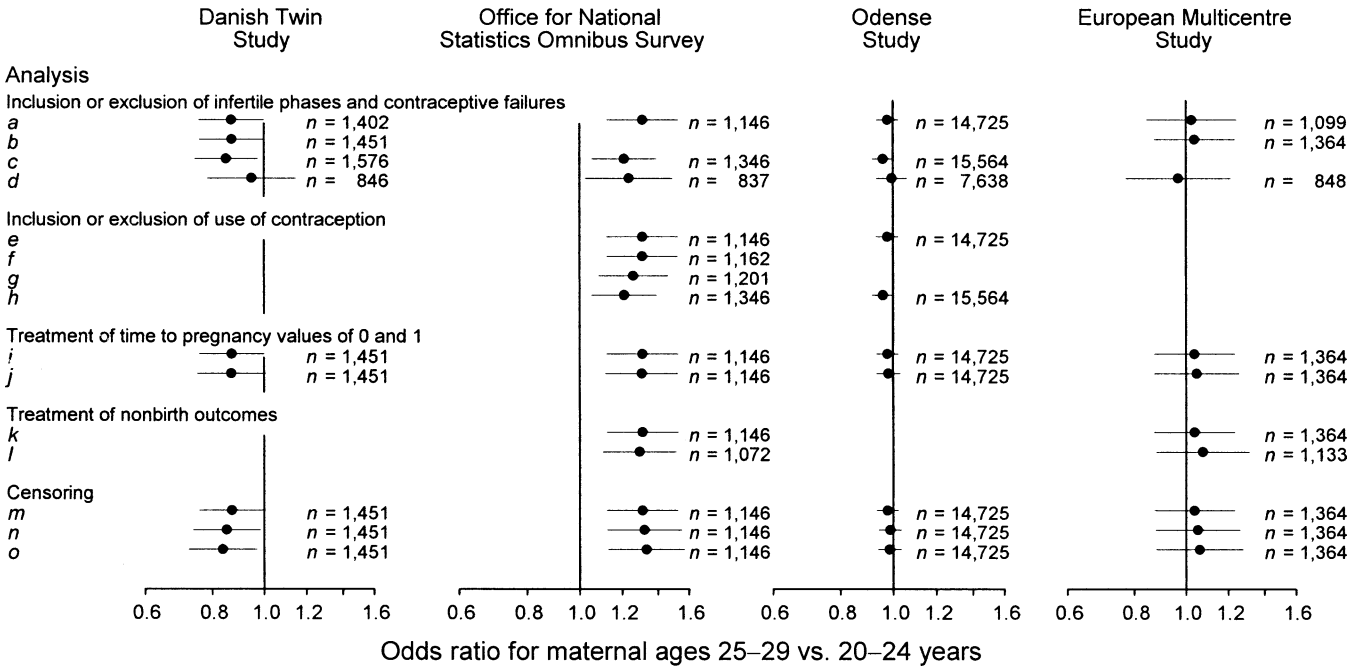

FIGURE 2. Illustrative examples of sensitivity analyses for studying fecundity, based on four European data sets, showing the effects on odds ratios for maternal age group 25–29 years compared with 20–24 years. The Danish Twin Study is a population-based survey conducted in 1998–1999 of twins born in 1931–1952 (Hum Reprod 2005;20:955–64) (17). The Omnibus Survey is a cross-sectional survey representative of the British population, conducted in 1996 by the Office for National Statistics (ONS) (Lancet 2000;355:1961–5) (2). The Odense Study is pregnancy based and hospital based but thought to be representative of the island of Fünen, from a routinely asked question in the prenatal clinic during 1972–1984 (Am J Epidemiol 2000;152:565–72) (18), and the European Multicentre Study is a population-based survey carried out in 1991–1993 in Denmark, Italy, Germany, Poland, and Spain (Eur J Public Health 1999;9:229–35) (3). Plots show odds ratios (symbols) and 95% confidence intervals (lines) for the effect on fertility of maternal age at the start of the attempt (relative to the baseline group aged 20–24 years) and smoking (relative to not smoking) at the start of the attempt (or during pregnancy, depending on how the question was asked). They are estimated from the following analyses: *a*, time to pregnancy (TTP) only; *b*, TTP + infertile phases (unsuccessful attempts); *c*, TTP + infertile phases + contraceptive failures (set at TTP = 0); *d*, TTP only, excluding values of 0 and 1; *e*, same as analysis *a*, which includes TTP values only for couples not using contraception; *f*, same as *e* + use of contraception "only sometimes"; *g*, same as *f* + use of contraception "most of the time"; *h*, same as *g* + use of contraception "all of the time"; *i*, same as analysis *b*, treating TTP values of 0 and 1 separately; *j*, same as *i*, but with TTP values of 0 and 1 grouped together; *k*, same as analysis *b*; *l*, same as analysis *k*, but for births only, i.e., excluding pregnancies that ended with a miscarriage and so on; *m*, same as analysis *b*, where censoring is at 14 months; *n*, same as analysis *m* but with censoring at 10 months; and *o*, same as analysis *m* but with censoring at 7 months. For the ONS Omnibus Survey and the Odense Study, because infertile phases were not available, analyses *i*, *k*, and *m* were the same as analysis *a* rather than analysis *b*. Maternal smoking is not presented for the Danish Twin Study, as this information was not available for infertile phases, which were important to include in the analysis. In the other studies, information on maternal smoking was ascertained at the starting time only in the European Multicentre Study; in the others, questions on smoking habit during pregnancy were asked. The recommended analysis is *l*.

risk taking and/or a lower degree of conscious planning, the members of this group are more likely to have "accidental" pregnancies that are excluded from the main TTP analysis, so that their eligible pregnancies apply only to a subset who are less prone to "accidents," very likely associated with lower fecundity, resulting in "planning bias" (26). In addition, some couples may change their attitude to the pregnancy when retrospective questioning is used, reporting it as planned even though it was a contraceptive failure. This has been termed "wantedness bias" (26), since reporting the pregnancy as an accident could be considered as saying the child was unwanted. This would cause bias if differential propensity to reinterpret one's recall were associated with one or more covariates.

To test for the presence of planning and wantedness bias, researchers must obtain full information about accidental pregnancies. Then, a parallel analysis can be performed with the same covariates but with the proportion of accidental pregnancies in each exposure group as the outcome variable. Table 1 shows the results of such an analysis for the four data sets, indicating that older women are less likely (except in the Odense Study) and that smokers are more likely to have a pregnancy that is declared a contraceptive failure. The proportion of accidental pregnancies also varies, even within Western Europe, from 5–9 percent in the Danish studies to 17.5 percent in the United Kingdom.

To assess the impact of the potential bias this may introduce in the main analysis, researchers should carry out sensitivity analyses, first including the accidental pregnancies in the TTP analysis with an assigned value of 0 (figures 2, 3, and 4, analysis *c*), and second, by repeating the TTP analysis excluding all TTP values of 0 and 1

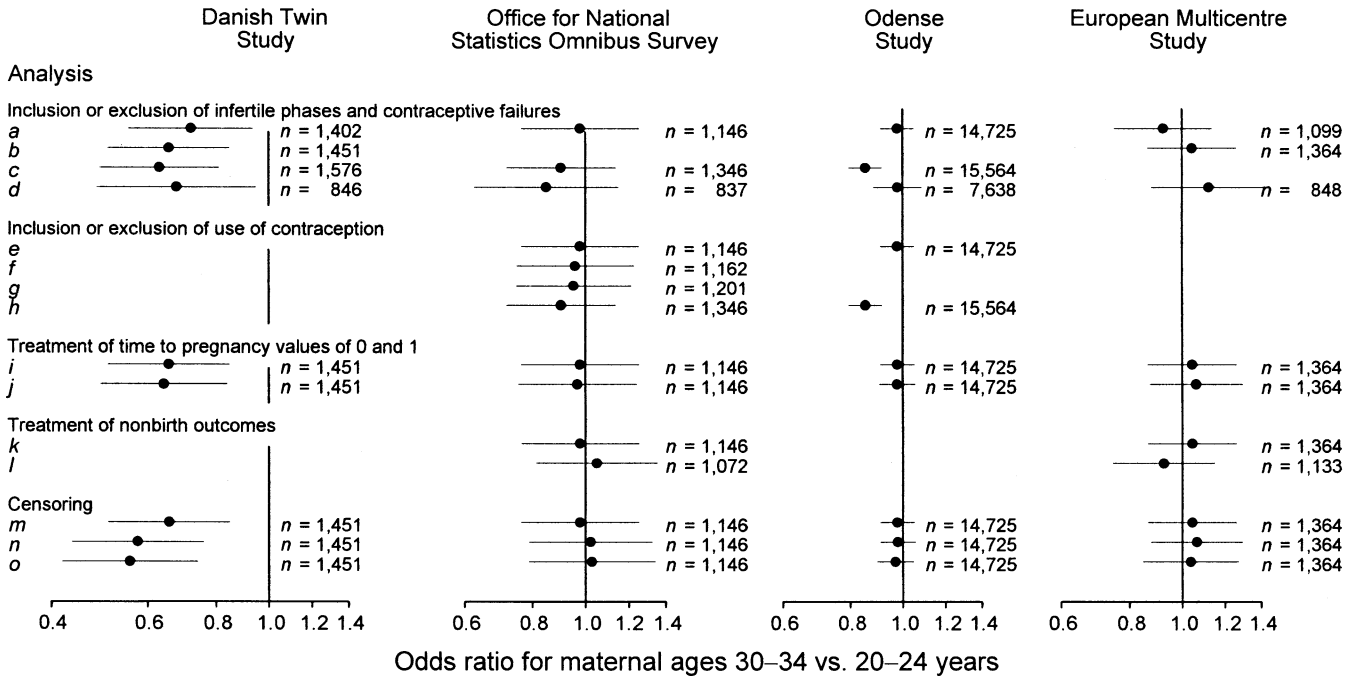

FIGURE 3. Illustrative examples of sensitivity analyses for studying fecundity, based on four European data sets between 1972 and 1999, showing the effects on odds ratios for the maternal age group 30–34 compared with 20–24 years. See the legend to figure 2 for a more detailed explanation of symbols, lines, and analyses.

(figures 2, 3, and 4, analysis *d*) (TTP values of 1 are excluded since some respondents may report immediate conceptions as a TTP of 1; see below). The latter analysis, which is admittedly highly artificial, reveals some sensitivity of the estimated effects of smoking and maternal age to the classification of accidental pregnancies in these data sets.

An additional sensitivity analysis is possible, if the questionnaire is designed appropriately, to test the robustness of the accidental/nonaccidental boundary: Respondents are asked whether they were using a contraceptive method around the time of conception and, if so, whether they were just using it "irregularly" or "only sometimes" (with additional possible categories of "most of the time" and "all of the time"). The principal analysis uses the pregnancies with reported contraceptive use as "no" (figures 2, 3, and 4, analysis *e*), with supplementary analyses having "no" plus "irregularly"/"only sometimes" (figures 2, 3, and 4, analysis *f*), and alternatively those as in figures 2, 3, and 4, analysis *f*, plus pregnancies where contraceptive use was reported as most or all of the time (figures 2, 3, and 4, analyses *g* and *h*). Irregular birth control use makes little difference to the effect estimates, and we recommend using the simpler dichotomous criterion, based on occurrence or not of a contraceptive failure, to simplify the questionnaire.

Reporting of immediate conceptions may be inconsistent (some respondents may use "1" to indicate the first cycle or month), and it has therefore been usual practice to group the values 0 and 1. This is precautionary and entails little loss of information. The comparison is shown in figures 2, 3, and 4, analyses *i* and *j*, which indicates that it is an issue of no importance.

Nonbirth pregnancy outcomes

A more important decision is whether to include all conceptions or only those ending in births (or for current pregnancies, those reaching at least 6 months' gestation). Figures 2, 3, and 4, analyses *k* and *l*, show this comparison, and the differences are generally small. However, as already mentioned, we favor exclusion of pregnancies ending in miscarriage from TTP analyses for validity reasons.

Pregnancy recognition bias

The available pregnancy tests make it possible to diagnose a pregnancy a few days after overdue menstruation. Some of these early recognized pregnancies may result in early miscarriages and should not be included since, if earliness of pregnancy recognition differs according to the hypothesis variable, bias may occur, although it is likely to be small (26). It has been suggested that participants could be asked when they first detected the pregnancy, as a way of recognizing and allowing for this possible source of bias (26).

Medical intervention bias

If the exposure of interest is associated with the probability that a couple would seek medical care for infertility,

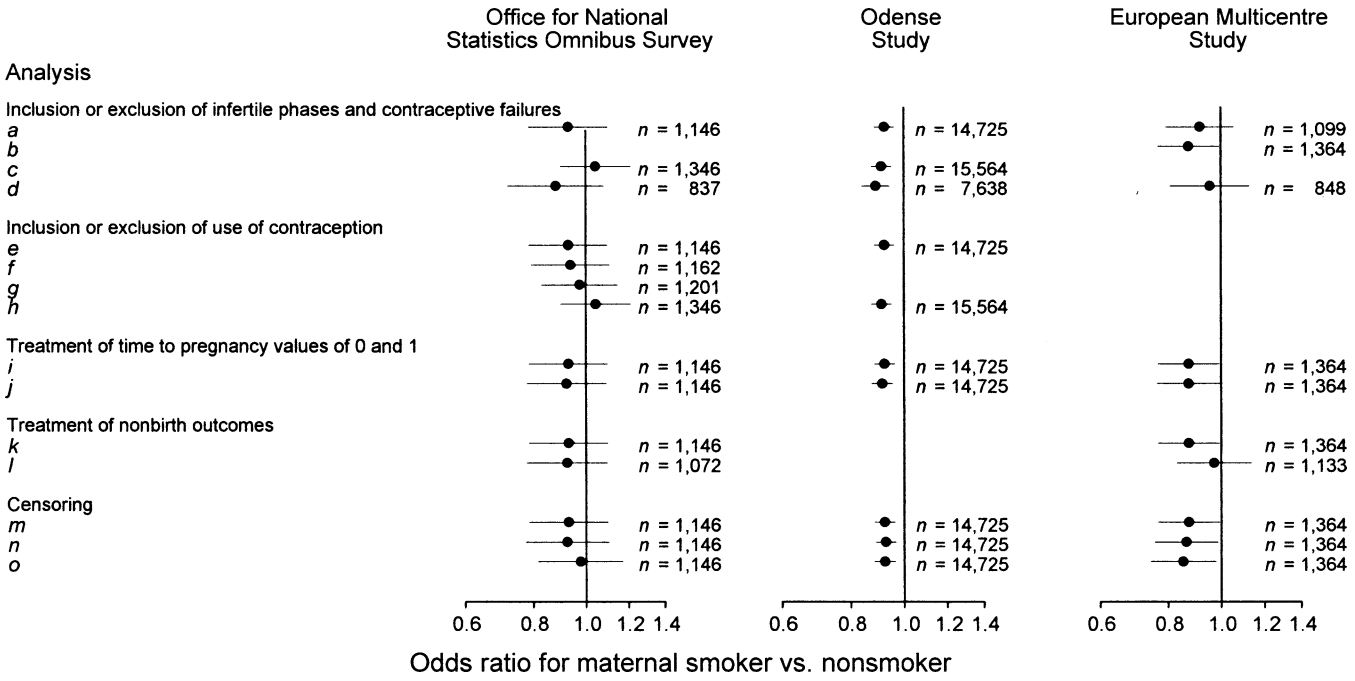

FIGURE 4. Illustrative examples of sensitivity analyses for studying fecundity, based on four European data sets between 1972 and 1999, showing the effects on odds ratios for smokers compared with nonsmokers. See the legend to figure 2 for a more detailed explanation of symbols, lines, and analyses.

there is a possibility of bias to the extent that the intervention is successful. As TTP is analyzed using survival analysis, this is readily dealt with using censoring, so that only the duration of the unsuccessful attempt up to the censoring time is included in the analysis. The ideal is to choose individual censoring times, when each couple sought medical assistance. However, as this adds to data collection, a simpler alternative is to censor all attempts at some reasonable value

TABLE 1. Risk of contraceptive failure in relation to maternal age and smoking, with data from four European data sets collected between 1972 and 1999*

Study	Contraceptive failures		No. of pregnancies not declared contraceptive failures	Contraceptive failure†					
				Maternal age 25–29 years		Maternal age 30–34 years		Maternal smoking	
	No.	%		Odds ratio	95% confidence interval	Odds ratio	95% confidence interval	Odds ratio	95% confidence interval
Danish Twin Study	125	8.6	1,451	0.67	0.44, 1.02	0.39	0.15, 1.03		
Odense Prenatal Study	839	5.7	14,725	1.22	1.03, 1.43	2.32	1.87, 2.88	1.18	1.03, 1.36
ONS Omnibus Survey	200	17.5	1,146	0.84	0.60, 1.17	0.60	0.33, 1.10	1.67	1.21, 2.31
European Multicentre Study‡	442	13.1	3,366	0.80	0.64, 1.00	0.86	0.59, 1.26	1.08	0.88, 1.31

* Additional background on the four European studies is as follows: the Danish Twin Study, a population-based survey conducted in 1998–1999 of twins born during 1931–1952 (Hum Reprod 2005;20:955–64) (17); the Odense Prenatal Study, a pregnancy- and hospital-based study thought to be representative of the island of Fünen, from a routinely asked question in the prenatal clinic in 1972–1984 (Am J Epidemiol 2000;152:565–72) (18); the Office for National Statistics (ONS) Omnibus Survey, a cross-sectional survey representative of the British population conducted in 1996 (Lancet 2000;355:1961–5) (2); and the European Multicentre Study, a population-based survey carried out in 1991–1993 in Denmark, Italy, Germany, Poland, and Spain (Eur J Public Health 1999;9:229–35) (3).

† Estimated using logistic regression, with the outcome variable indicating whether pregnancy was declared as a contraceptive failure. The referent category was a nonsmoking mother aged 20–24 years.

‡ In the European Multicentre Study, previous analyses (figures 2, 3, 4) used the "most recent" episode because this has information on infertile phases. In this analysis, we have used the first pregnancy, because information was available on contraceptive failures.

that will encompass the vast majority of couples, such as 12 months (we prefer to use 14 months, as this avoids the problem of digit preference in this data region). Although censoring means that some information is lost on the length of time taken to conceive after the censoring time, including for couples who do not receive medical help, this has only a slight effect on the analysis, as the statistical power of TTP analyses is concentrated in the first 6 months, with little information being added at the end of the curve (figure 1). As well as the standard censoring time, a sensitivity analysis can also be carried out with alternative values, for example, at 10 and at 7 months, as shown in figures 2, 3, and 4, analyses *m–o*. While the effect estimates are slightly affected in some analyses, this does not alter the interpretation for any of them; nevertheless, this is a simple check that can readily be carried out in all principal analyses.

Censoring does not deal with the whole of medical intervention bias if data are collected only on successful attempts. In this case, pregnancies that result from medical treatment are included in the data set, whereas without intervention they would not have been. This means that the effect of successful treatment is to add additional pregnancies to the data, and these have long TTP values; in other words, it makes the TTP distribution of successful attempts appear *less* fertile.

Truncation bias

It is important to be aware of truncation, as it distorts the data set and has potential to cause large biases. It occurs when TTP distributions and/or covariates are nonstationary over time, and it is especially important when studying trends or any exposure that varies systematically over time. There are various types of truncation that differ according to the pattern of sampling.

In a cross-sectional or birth cohort study that questions people about the TTP of previous births not including current attempts or pregnancies, suppose that data collection takes place in December 2005. Then, couples who started unprotected intercourse in December 2004 will be included in the sample only if they conceived within 3 months, assuming a gestational length of 9 months, and less fertile couples will be systematically excluded. The result is that the couples with the most recent starting times will appear to be more fertile, because of the exclusion of the less fertile couples (28). This is right truncation.

With pregnancy-based studies, both right and left truncation can occur. To illustrate this, we used the Odense Prenatal Clinic Study to examine the impact of truncation on the estimated effect of year of starting the attempt and of female age, smoking, and parity (figure 5). As expected, this shows a severe bias for time trend effects, which is ''corrected'' by use of a crude solution that excludes the period effects for which truncation occurs. Biases are much less severe, although not completely absent, for the effects of maternal age, smoking, and parity. This is to be expected since, provided there is no time trend in the exposure distribution, the effects of left and right truncation on the estimated covariate effects should cancel out. Crude adjustment for truncation reduces bias in some but not all of the estimates. For exposures with a trend, for example, certain occupational exposures, a similar but weaker bias to that shown for the trend itself would be expected. The ''time trend bias'' described in the earlier TTP literature (26) is a special case of truncation bias.

There are two ways of avoiding truncation bias in the design of the study, both of which apply to cross-sectional or birth cohort studies. One is to collect data from couples who are old enough to have finished their reproductive careers. Although this requires long-term recall, as we have seen, it has been done with apparent success in the Danish Twins Study. The other is to collect information on continuing attempts (see also current duration approach, above), so that the ascertainment of couples who are having unprotected intercourse no longer depends on the occurrence of conception to be included in the study. Alternatively, the bias can be handled in the analysis, by using either the crude exclusion method, as above, which reduces sample size, or statistical modeling (20).

Behavior change bias

Information about all variables should be assessed at the ''starting time,'' when unprotected intercourse began. If, instead, the assessment relates to some subsequent point, such as the time of conception, bias may occur: Couples are likely to adopt behavior perceived as healthier, with each cycle that elapses without conceiving; for example, they may quit smoking. As the later ''healthier'' (nonsmoking) value is then the assessed one, this systematic misclassification leads to an apparent association of nonsmoking with delayed conception, which could readily be interpreted as a protective effect of smoking (or other behavior perceived as harmful) (26). With assessment at the starting time, misclassification still occurs, as information on the nonsmoking cycles (postquitting) later in the attempt is lost, its effect being to underestimate the harmful effect of smoking. However, this is less serious than judging it to be protective. A further possibility is to collect cycle-specific information (3–6), but this is laborious.

If information about the variable is ascertained during pregnancy instead of during the attempt, the dilution effect would be still greater, especially as pregnancy is a time when many women change behavior. Consistent with this, figure 4 shows that the estimated effect for maternal smoking is greater in the European Multicentre Study, in which smoking was ascertained at the starting time, than in the Odense and Omnibus studies in which smoking was asked during pregnancy.

Knowledge and behavior as determinants of time to pregnancy

TTP is influenced by not only biologic factors but also knowledge and behavior. As far as possible, the methodology should be designed to trace the biologic events. Thus, the filter question, and therefore the criterion for eligibility for a TTP value, is based on behavior that affects biology (contraceptive practice). This is associated with, but distinct from, ''trying to conceive,'' a less specific behavioral

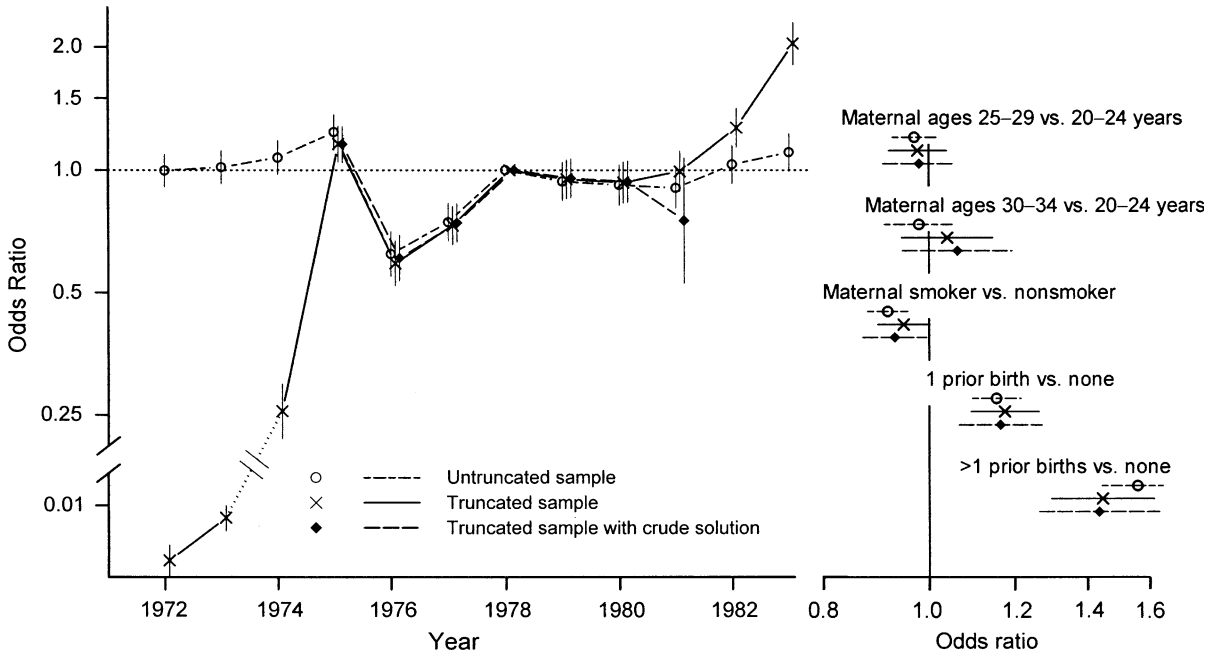

FIGURE 5. Effect of left and right truncation on estimates of time trends in fecundity and the effects of maternal age, smoking, and parity, using a pregnancy-based data set—the Odense Prenatal Clinic Study (Am J Epidemiol 2000;152:565–72) (18), which has data on pregnancies occurring from 1972 to 1987. We artificially created the following subsets: 1) a nontruncated sample based on *starting* times in the period 1972–1983; 2) a truncated sample based on *conception* times in 1975–1983, so that those with a starting time in 1972–1975 would be excluded if they conceived before 1975 (left truncation: they would be missed by prenatal recruitment as they conceived too soon for the study), and those with a starting time leading up to 1983 would be excluded unless they had conceived by 1983 (right truncation: they conceived after prenatal recruitment for this subset had finished); 3) to remedy this, truncation is crudely corrected for, by including only those with starting times from 1975 onward to deal with left truncation and excluding those with starting times after 1979, so that those who had conceived by 1983 were unaffected by right truncation. The left graph shows the odds ratios with 95% confidence intervals for time trends in the year of starting the attempt, estimated with each of the three data subsets. The right graph shows the odds ratios with 95% confidence intervals for the effects of female age, smoking, and parity estimated from each data subset.

measure, and with attitudinal measures, such as ''wanting'' to conceive or ''planning,'' which is a combination of both (29). Another reason why contraceptive failure is preferable as a filter is that it is clear-cut, whereas the relation between these other concepts is complex (29–33), women do not spontaneously use these terms to describe their own situations (32, 34), and reports are not stable (35). The implication is that pregnancies conceived when not using contraception are eligible for a TTP value, even if the couple reported that they were not actively ''trying to conceive.''

Nevertheless, interpretation of TTP studies must always take account of possible behavioral factors. For example, it is possible that the higher fertility observed among more educated couples (7), and in recent decades (2), is due to the spread through the population of better information on the timing of the fertile days in the woman's cycle. Information about this can easily be obtained from questionnaires (36).

Other behavioral factors may also be operating, and some of these could be important determinants and/or could cause bias. For example, persistence in trying to conceive could be differential within populations (e.g., in different age groups) and/or between populations, and this could generate important bias, especially in studies that do not supplement pregnancy-related TTP values with information on infertile phases (37).

More generally, we need better evidence on the extent to which TTP reflects biology or such behavioral factors. The importance depends on the subject of the analysis; for example, something like persistence is likely to be important in comparisons of age groups and of time trends, geographic/ethnic differences, and educational/social class differences but not in the case of a homogeneous population when the focus is on differential environmental exposures.

CONCLUSION

Retrospective TTP studies can be used to estimate biologic fertility in a population. Such studies have proved readily acceptable in a wide variety of populations. As with all retrospective studies, exposures and other covariates are difficult for participants to remember accurately, especially with longer duration of recall. The analyses presented here

suggest that, while retrospective TTP studies can also be susceptible to various other biases, these are not systematic across studies (either of the same or different design), and in many cases the impact on the effects of interest is small. Most potential biases and other problems can be effectively controlled if the studies are properly designed and conducted, and if careful sensitivity analysis along the lines presented here is carried out.

ACKNOWLEDGMENTS

This work was funded by Medical Research Council grant G9900635.

Maria He Huang-Wong organized the data sets and carried out essential quality control, and James Bennett assisted in producing the figures. The authors are grateful to Dr. Svend Juul for access to data from the European Multicentre Study on Infertility and Subfecundity and to the many people who helped to create the data sets that were used in this study.

Conflict of interest: none declared.

REFERENCES

1. Joffe M. Time to pregnancy: a measure of reproductive function in either sex. Occup Environ Med 1997;54:289–95.
2. Joffe M. Time trends in biological fertility in Britain. Lancet 2000;355:1961–5.
3. Karmaus W, Juul S. Infertility and subfecundity in population-based samples from Denmark, Germany, Italy, Poland and Spain. Eur J Public Health 1999;9:229–35.
4. Bolumar F, Olsen J, Boldsen J, et al. Smoking reduces fecundity: a European multicenter study on infertility and subfecundity. Am J Epidemiol 1996;143:578–87.
5. Bolumar F, Olsen J, Rebagliato M, et al. Caffeine intake and delayed conception: a European multicenter study on infertility and subfecundity. Am J Epidemiol 1997;145:324–34.
6. Olsen J, Bolumar F, Boldsen J, et al. Does moderate alcohol intake reduce fecundability? A European multicenter study on infertility and subfecundity. Alcohol Clin Exp Res 1997;21:206–12.
7. Joffe M, Li Z. Male and female factors in fertility. Am J Epidemiol 1994;140:921–9.
8. Buck GM, Lynch CD, Stanford JB, et al. Prospective pregnancy study designs for assessing reproductive and developmental toxicants. Environ Health Perspect 2004;112:79–86.
9. Chapin RE, Buck GM. Our once-in-a-lifetime opportunity. Environ Health Perspect 2004;112:67–8.
10. Tingen C, Stanford JB, Dunson DB. Methodologic and statistical approaches to studying human fertility and environmental exposure. Environ Health Perspect 2004;112:87–93.
11. Rockett JC, Buck GM, Lynch CD, et al. The value of home-based collection of biospecimens in reproductive epidemiology. Environ Health Perspect 2004;112:94–104.
12. Joffe M. Invited commentary: the potential for monitoring of fecundity and the remaining challenges. Am J Epidemiol 2003;157:89–93.
13. Weinberg CR, Gladen BC. The beta-geometric distribution applied to comparative fecundability studies. Biometrics 1986;42:547–60.
14. Keiding N, Kvist K, Hartvig H, et al. Estimating time to pregnancy from current durations in a cross-sectional sample. Biostatistics 2002;3:565–78.
15. Slama R, Ducot B, Keiding N, et al. Studying human fertility and environmental exposures. (Letter). Environ Health Perspect 2004;112:A604.
16. Olsen J, Andersen PK. We should monitor human fecundity, but how? A suggestion for a new method that may also be used to identify determinants of low fecundity. Epidemiology 1999;10:419–21.
17. Jensen TK, Joffe M, Scheike T, et al. Time trends in waiting time to pregnancy among Danish twins. Hum Reprod 2005;20:955–64.
18. Jensen TK, Scheike T, Keiding N, et al. Selection bias in determining the age dependence of waiting time to pregnancy. Am J Epidemiol 2000;152:565–72.
19. Olsen J, Juul S, Basso O. Measuring time to pregnancy: methodological issues to consider. Hum Reprod 1998;13:1751–3.
20. Scheike T, Jensen TK. A discrete survival model with random effects: an application to time to pregnancy. Biometrics 1997;53:318–29.
21. Scheike TH, Petersen JH, Martinussen T. Retrospective ascertainment of recurrent events: an application to time to pregnancy. J Am Stat Assoc 1999;94:713–25.
22. Baird DD, Weinberg CR, Rowland AS. Reporting errors in time-to-pregnancy data collected with a short questionnaire. Am J Epidemiol 1991;133:1282–90.
23. Zielhuis GA, Hulscher ME, Florack EI. Validity and reliability of a questionnaire on fecundability. Int J Epidemiol 1992;21:1151–6.
24. Joffe M, Villard L, Li Z, et al. Long-term recall of time-to-pregnancy. Fertil Steril 1993;60:99–104.
25. Joffe M, Villard L, Li Z, et al. A time to pregnancy questionnaire designed for long term recall: validity in Oxford, England. J Epidemiol Community Health 1995;49:314–19.
26. Weinberg CR, Baird DD, Wilcox AJ. Sources of bias in studies of time to pregnancy. Stat Med 1994;13:671–81.
27. Fahrmeir L, Tutz G. Multivariate statistical modelling based on generalized linear models. 2nd ed. New York, NY: Springer Verlag, 2001.
28. Jensen TK, Keiding N, Scheike T, et al. Declining human fertility? Fertil Steril 2000;73:421–3.
29. Klerman LV. The intendedness of pregnancy: a concept in transition. Matern Child Health J 2000;4:155–62.
30. Kaufmann RB, Morris L, Spitz AM. Comparison of two question sequences for assessing pregnancy intentions. Am J Epidemiol 1997;145:810–16.
31. Trussell J, Vaughan B, Stanford J. Are all contraceptive failures unintended pregnancies? Evidence from the 1995 National Survey of Family Growth. Fam Plann Perspect 1999;31:246–7, 260.
32. Stanford JB, Hobbs R, Jameson P, et al. Defining dimensions of pregnancy intendedness. Matern Child Health J 2000;4:183–9.
33. Sable MR, Libbus MK. Pregnancy intentions and pregnancy happiness: are they different? Matern Child Health J 2000;4:191–6.
34. Barrett G, Wellings K. What is a 'planned' pregnancy? Empirical data from a British study. Soc Sci Med 2002;55:545–57.
35. Joyce T, Kaestner R, Korenman S. The stability of pregnancy intentions and pregnancy-related maternal behaviors. Matern Child Health J 2000;4:171–8.

36. Demographic and health surveys. Calverton, MD: Macro International, Inc, 2005. (http://www.measuredhs.com). (Accessed February 14, 2005).
37. Basso O, Juul S, Olsen J. Time to pregnancy as a correlate of fecundity: differential persistence in trying to become pregnant as a source of bias. Int J Epidemiol 2000;29:856–60.

APPENDIX

Suggested Questionnaire

Filtering is not shown but is available from the authors.

Screening question (if needed). Have you ever been pregnant or tried to get pregnant?/Have you ever fathered a pregnancy or tried to do so? [Y/N]

A1. Have you and your husband/wife/partner ever tried to become pregnant over a period of at least a year without success? [Y/N]

Text. The following questions concern the *first* period in which you waited longer than a year to get pregnant.

A2. When did you first experience this? [From: month/year]

A3. Are you still trying? [Y/N]

A4. Did you get pregnant by the end of this period? [Y/N]

A5. How long did you try to get pregnant before you gave up? [months/years, or month/year of giving up]

A6. Why did you give up?

A7. At the beginning of this period, were you a regular smoker? [Y/N]

A8. At the beginning of this period, was your husband/wife/partner a regular smoker? [Y/N]

A9. When was s/he born? [month/year]

B1. Have you/has your wife/partner ever been pregnant? [Y/N]

B2. How many times have you been pregnant?/How many pregnancies have you fathered? (Do not include deliberate terminations of pregnancy.) [number of times]

B3. For each pregnancy, please enter the following information [in table format], starting with the first:

- month/year when it ended (or, if still pregnant, month/year when delivery is expected)
- was it a miscarriage? (>8 weeks; confirmed by a test)
- was the baby stillborn?
- was the baby born alive?
- other
- baby's sex (or twins/higher multiples)
- birth weight and gestational age

Text (if more than one pregnancy). I want to ask you now about the first of these pregnancies (ask about current pregnancy if it is the first).

B4. How long did it take you/your wife/partner to get pregnant? [months/years]

B5. Before this pregnancy, what was the last type of birth control that you or your husband/wife/partner used? [list of main possible methods]

B6. Did you/she/he stop this before the pregnancy started, or did the pregnancy result from a birth control failure? [stopped before/birth control failure/not clear]

B7. At the beginning of this period, were you a regular smoker? [Y/N]

B8. At the beginning of this period, was your husband/wife/partner a regular smoker? [Y/N]

B9. (if not already asked) What is her/his date of birth? [month/year]

C1. When were you born? [month/year]

Date of completion of questionnaire/interview [month/year]

Other questions could include:

- frequency of intercourse (if recall is short enough)
- when the pregnancy was first detected (if recall is short enough)
- one or more behavioral questions, e.g., on desired family size; knowledge of the woman's fertile days in the menstrual cycle
- socioeconomic status and educational class
- information on medical treatment and medical conditions relevant to fertility