Routledge
Taylor & Francis Group

Applied Developmental Science

ISSN: 1088-8691 (Print) 1532-480X (Online) Journal homepage: https://www.tandfonline.com/loi/hads20

Defining the Boundaries of Early Adolescence: A User's Guide to Assessing Pubertal Status and Pubertal Timing in Research With Adolescents

Lorah D. Dorn , Ronald E. Dahl , Hermi Rojahn Woodward & Frank Biro

To cite this article: Lorah D. Dorn , Ronald E. Dahl , Hermi Rojahn Woodward & Frank Biro (2006) Defining the Boundaries of Early Adolescence: A User's Guide to Assessing Pubertal Status and Pubertal Timing in Research With Adolescents, Applied Developmental Science, 10:1, 30-56, DOI: 10.1207/s1532480xads1001_3

To link to this article: https://doi.org/10.1207/s1532480xads1001_3

Applied Developmental Science
2006, Vol. 10, No. 1, 30–56

Defining the Boundaries of Early Adolescence: A User's Guide to Assessing Pubertal Status and Pubertal Timing in Research With Adolescents

Lorah D. Dorn, PhD
Cincinnati Children's Hospital Medical Center and the University of Cincinnati College of Medicine, Division of Adolescent Medicine

Ronald E. Dahl, MD and Hermi Rojahn Woodward
University of Pittsburgh School of Medicine, Western Psychiatric Institute and Clinic

Frank Biro, MD
Cincinnati Children's Hospital Medical Center and the University of Cincinnati College of Medicine, Division of Adolescent Medicine

This article addresses pragmatic issues regarding the assessment of puberty in research on adolescent health and development. Because pubertal processes have a major effect on physical, psychological, and social development, we posit that the assessment of pubertal status is at least as important as the specification of age for characterizing adolescent participants in research studies. Yet, a sampling of recent literature shows that the majority of publications addressing health and developmental issues in adolescence lack any measure of puberty. A more comprehensive review of 447 articles reporting to have assessed puberty reveals considerable inconsistencies in methods, definitions, and conceptualizations of puberty and its stages, which highlights the need for better standardization. This article provides an in-depth review of existing methods to assess pubertal status and timing and enumerates the relative merits and shortcomings of several approaches. Conceptual and practical guidelines are provided for selecting specific measures to assess puberty with an emphasis placed on the need for these choices to be driven by the specific goals of the research.

Consider the peer-review fate of a grant application or manuscript addressing an area of adolescent health or development without specifying the age of the sample. Whether the research is focused on cognitive development, self-esteem, athletic injuries, the management of diabetes, sleep changes, or the treatment of depression, it is safe to assume that the study would be considered unacceptable if it did not specify the age of the sample and give at least some consideration of the effects of age on the outcomes of interest. Yet, age is not a reliable indicator of stage of physical development during the period when pubertal processes are underway. There are several sources of maturational heterogeneity in physical and psychosocial development during adolescence that are not reflected by age. Some 13-year-olds can appear adult-like in size, sexual maturation, and behavior whereas others may show a more child-like immaturity in some or all of these domains.

This is not simply a matter of physical development. Pubertal development can influence numerous aspects of physiology, behavior, drug metabolism, motivation, emotion, and some aspects of cognitive development—in ways that are relevant to medical and psychiatric conditions (Dahl, 2004; Davison & Susman, 2001; Hein, 1994; Orr & Ingersoll, 1995). For some psychiatric disorders, such as depression and panic disorders in girls, it has been shown that pubertal processes have direct effects on risk (Angold, Costello, Erkanli, & Worthman, 1999; Angold, Costello, & Worthman, 1998; Angold & Rutter, 1992; Angold & Worthman, 1993; Hayward et al., 1992). Finally, timing and velocity of pubertal processes can significantly affect an individual's social experience and relationships with others. For over 50 years the literature has reported on the influence of early versus late puberty on such outcomes as achievement, social–emotional functioning, or behaviors (Dubas, Graber, & Petersen, 1991; Ge, Conger, & Elder, 2001; Graber, Lewinsohn, Seeley, & Brooks-Gunn, 1997; M. C. Jones, & Bayley, 1950; Mussen & Jones, 1957; Siegel, Yancey, Aneshensel, & Schuler, 1999;

This work was supported by a grant from the John D. and Catherine T. MacArthur Foundation Research Network on Psychopathology and Development (David J. Kupfer, Chair).

Correspondence should be addressed to Lorah D. Dorn, Ph.D., Professor of Pediatrics, Director of Research, Division of Adolescent Medicine, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue–ML 4000, Cincinnati, OH 45229. E-mail: lorah.dorn@cchmc.org.

Simmons, Blyth, Van Cleave, & Bush, 1979; Tschann et al., 1994; Wichstrom, 2000).

Timing of puberty is also reported to be related to numerous physiological disorders such as breast cancer risk (Apter & Vihko, 1985b; Key, 1999), polycystic ovary syndrome and insulin resistance (Dunaif, 1999; Ibáñez et al., 1993), or risk for lower bone density (Ito et al., 1995). In different individuals pubertal changes occur at different ages. That is, timing across individuals is different at a specific age. Although age is correlated with stage, the correlation is not perfect and therefore cannot be used as the sole descriptor. Age is said to be only marking the progression of time (Wohlwill, 1973). Thus, the ability to understand developmental processes, behaviors, and the disorders and diseases that occur during adolescence will remain severely limited if most of this research focuses solely on age.

A review of the literature (see subsequent section) reveals that the majority of studies on adolescent health and psychosocial development do not measure pubertal development and, of those that do, there is little consistency in the methods and conceptualization of pubertal processes and their role in adolescent development. This lack of consistency is illustrated by the widely discrepant or undefined use of terminology across studies: For example, some studies dichotomize their samples into "prepubertal" and "postpubertal" without specifying the criteria for these categories. Others have used a split in which participants at Tanner Stage 1 and 2 are called "prepubertal" whereas those at Tanner Stage 3 and above are "pubertal" or "postpubertal"—again with no clear-cut or explicit rationale for doing so. Alternatively, Tanner Stage 1 is the only prepubertal stage according to many studies. Yet, others use terms suggesting that puberty occurs in an almost event-like fashion. "Onset of puberty" is used to delineate the "prepubertal" versus "postpubertal" status of participants or, more rarely, participants are referred to as "prepubertal," "peripubertal," or "postpubertal," as if it could be determined that participants either have not yet experienced the event of puberty, are experiencing it right now, or have experienced it in the past. Needless to say, the reference point chosen to define these categories is usually missing. Thus, relevant conclusions about the impact that puberty or pubertal timing may have on health and development are difficult to discern. Moreover, a general lack of conceptual clarity and inconsistencies in the use of terms for pubertal maturation continue to limit our ability to compare results across studies and in turn hamper overall progress in the field of adolescent health and development.

The importance of including measures of pubertal maturation is relevant from middle childhood through adolescence. For example, the increase in adrenal androgens (the earliest part of puberty; see subsequent section on Physical and Neuroendocrine Changes in Puberty) often begins by 6 to 9 years of age and the onset of some external pubertal changes is evident in 10% to 30% of 9-year-old girls (Herman-Giddens et al., 1997). It may be best for studies of puberty to start at an age as young as 6. Other maturational changes, such as improvements in self-regulatory skills and judgment (that can affect risk behaviors among youth relevant to accidents, substance use, addiction, and other health problems), appear to be changing well into the early 20s.

In subsequent sections of this article, we present the literature review and discuss the lack of conceptual clarity and standardization of terms. This is followed by a summary of the major conceptual issues relevant to the consideration of the choice of methods for the assessment of pubertal status or timing. The goal of this article is not to advocate for one "best" measure or to simply provide a menu of available measures and their merits. Instead, the purpose is to describe the key conceptual issues that should be considered when assessing specific domains of pubertal maturation within adolescent studies and to describe ways that can guide decisions on the use of particular measures and instruments.

In other words, we do not believe there is a one-size-fits-all answer to the frequently posed question, "What is the best way to measure puberty?" Instead, what is often required is a clarification of the research questions being addressed by a study and the subsequent components of maturation that are of greatest relevance and interest.

Literature Review for Evaluating Methods of Measuring Puberty

To evaluate our impression that much of the existing literature lacks attention and consistency in assessing pubertal processes and measures we conducted a literature search using the terms *pubertal status, pubertal stage,* and *timing of puberty* in MEDLINE database from 1966 through January 2004 and PSYCINFO beginning in 1967 in English. We then selected articles that used puberty as a key variable in the study where puberty was used in the primary or secondary analyses and not just to describe characteristics of the sample.

Four hundred and forty-seven empirical articles met the previous criteria and were available for viewing. Table 1 presents the method of pubertal assessment used as well as the setting of these studies. "Clinic setting" indicates that the study most likely would have required a clinical setting to determine some of the measures (e.g., bone density) or that a patient group from a clinic (e.g., from a diabetes clinic) was seen. Thus, we assumed clinical facilities would be available to conduct a physical exam. "Nonclinic setting" indicates that the study was

Table 1. *Frequencies of method of measuring puberty in journal articles reviewed via Medline and PSYCINFO through January 2004*

Setting	PE	PDS	SR	SR-Various	SR & PE	Age Menarche	"According to Tanner"	Misc.	DK	Total
Clinic setting	182	2	17	9	4	2	72	6	21	315
Nonclinic setting	31	33	9	24	4	10	7	1	6	125
Unsure	3	4	0	0	0	0	0	0	0	7
Total	216	39	26	33	8	12	79	7	27	447

Note: Clinic setting had access to facility to conduct physical examination; Nonclinic setting presumably had no ready access facility to conduct physical examination such as in school; PE = Physical Examination; PDS = Pubertal Development Scale; SR = Self-report using pictures or drawings; SR-Various = Self-report various methods including surveys or unspecified methods; SR & PE = Self-assessment and physical examination; "According to Tanner" = not specified if PE or SR, unclear; Misc. = Miscellaneous methods such as hormones, ultrasound, etc.; DK = Don't Know, no method specified but referred to pubertal assessment.

conducted in a facility that was not used primarily for health care. Specifically in such settings it would have been more difficult (but not impossible) to conduct a physical examination (e.g., school).

The process of the literature review revealed an unexpected degree of difficulty determining the method used to measure pubertal status. For example, although Table 1 shows that 216 (48.3%) conducted physical examinations to determine pubertal stage, it was often very difficult to definitively conclude that a physical was actually conducted. Rarely did an article state something as clear as "pubertal stage was determined by physical examination using Tanner criteria." The methods section and procedures often implied that pubertal stage was determined from a medical record but without specifying whether a physical examination was conducted or by whom. For example, some statements indicated that "participants were seen in the clinic by our research assistants and pediatricians where pubertal stage was determined." Even among the studies that clearly stated that physical examinations were done, very few described anything about the training of the examiners or stated interrater agreement among raters in the study. Thus, the figure of 48.3% is the very optimistic estimate and the quality of the physical determination of the pubertal staging is uncertain.

In 79 (17.67%) articles it stated, "pubertal development was assessed according to Tanner," but no details were provided as to whether this Tanner staging was determined by physical exam or by self-report. An additional 7 studies (see column labeled "Misc.") utilized some other objective measures to assess pubertal development such as measurement of pubertal hormones or a combination of hormones and ultrasound of the ovaries or testes.

A number of articles utilized self-assessment measures of puberty. The most widely used measure for self-report of pubertal development is the Petersen Pubertal Development Scale (PDS; Petersen, Crockett, Richards, & Boxer, 1988) that was utilized in 39 (8.7%) articles. Eight articles used both a self-report measure and physical exam. There were 26 articles that employed self-assessment of pubertal development using pictures or drawings of the stages of puberty. An additional 33 (7.38%) articles used self-assessment surveys or some unspecified methods of self-report. In some cases, these were one-item measures of pubertal development or measures for which no reliability and validity data were reported. It was interesting to note that of the 98 (21.9%) articles that used only self-report by any method, 28 were conducted in a clinic setting where it would seem that a physical exam could have been used to determine stage of pubertal development. An additional self-report measure of puberty is menarcheal age. Twelve studies used self-report of age at menarche as the primary measure, again without much explanation regarding how the data were collected.

Finally, 27 (6.04%) articles that reported pubertal assessment did not provide any information explaining how it was measured. Yet, these articles reported conclusions in their discussion and abstract about the relationship between pubertal development and/or pubertal timing and other variables studied. Thus, in reality, 23.7% (17.67% and 6.04% cited previously) of the articles really did not have enough information to determine if puberty was determined by physical exam or self-report.

In summary, the variability and lack of specifying methods for pubertal assessment presents significant challenges to the interpretability and replicability for a high proportion of these studies. These concerns apply to articles published in medical journals and psychological/ behavioral journals. This literature review highlights the crucial need for improved standards in the assessment of pubertal status and in reporting of more detailed methodology in the literature.

Conceptual Issues Regarding Puberty

A review of the literature and perusal of popular press articles reveals a lack of conceptual clarity re-

garding puberty. Many conceptual issues need to be addressed to make meaningful progress toward a better understanding of the role of pubertal processes in adolescent health and behavior. Given the central focus on methods to assess pubertal status and timing, it is outside the scope of this article to address these conceptual issues in significant depth, but the issues are highlighted briefly.

Puberty Versus Adolescence

Ambiguities about the distinction between puberty and adolescence are evident at many levels ranging from the popular press to the scientific literature. We, like others in field, believe it is crucial to consider the fundamental distinctions between the two.

Adolescence

In some frameworks, adolescence is considered to encompass the 2nd decade of life. In others, adolescence is considered to extend into the early 20s when the transition into adulthood presumably occurs. Among the useful definitions of adolescence, we prefer the following: *Adolescence* is the interval between childhood and the assumption of adult roles and responsibilities, a broad interval of maturation that encompasses physical, mental, and emotional development, as well as coincident cognitive changes and change in social roles.

Puberty

Puberty encompasses a more specific set of processes involved in physical and reproductive maturation. Most but not all of these processes occur within the general context of adolescence. Most pubertal processes are progressing at peak velocity in early to mid-adolescence but some aspects such as adrenarche can begin by 6 to 8 years of age—a point that many people would not regard as adolescence.

Puberty Is Not an Event or Unitary Process

There is increasing recognition that puberty represents a complex suite of interrelated changes that span several domains of growth and development that proceed over a long interval of time. In addition to the three separate neuroendocrine axes, adrenarche, gonadarche, and growth axis, which are discussed in the following sections, other dimensions of pubertal maturation include changes in social experience, perceptions by peers, self-perception, and puberty-specific influences on brain development as well as the changes in physical maturation, size, growth rate, and metabolic changes. Each of these domains has its own developmental course. There is a wide range of individual differences with regard to the timing, overall velocity, and relative synchrony of these processes. Thus, a crude categorization into prepubertal or postpubertal has little or no value without specifying which "event" or set of maturational changes of pubertal development are being used as the reference point. The central issue—conceptually and methodologically—is to specify the particular aspects of pubertal development that are of interest to a specific study.

Puberty and Its Underlying Neuroendocrinology

Physical and Neuroendocrine Changes in Puberty

Puberty includes two independent but overlapping processes that are controlled by different mechanisms referred to as adrenarche and gonadarche (Grumbach, 2002). Many consider gonadarche to reflect actual puberty. A recent article noted that some girls do appear to have pubic hair rather than breast development as the initial manifestation of puberty (F. M. Biro et al., 2003). In this study onset of puberty was the same age, regardless of their pathway, and height velocity was advanced beyond prepubertal values. Puberty also includes changes in the growth axis and reproductive maturity/competence. In this section a brief overview is provided of the endocrinology of adrenarche, gonadarche, and growth. Figure 1 delineates specific measures of puberty in girls and portrays which phase(s) of pubertal development they may represent. The figure for boys should be approximately 1 to 2 years later (figure not shown). It is important to recognize that there are significant individual differences in growth and development, and these curves represent what may be an average. Figure 1 also illustrates a number of unknowns that still exist about pubertal development.

Adrenarche, or the "awakening of the adrenal glands," is the earliest phase of puberty and is still poorly understood (Grumbach & Styne, 1998, 2003). At the early part of adrenarche, adrenal androgens begin to rise. These include dehydroepiandrosterone (DHEA), its sulfate (DHEAS), and androstendione (Forest, 1989; Grumbach & Styne, 1992; Sizonenko, Paunier, & Carmignac, 1976). These are relatively weak androgens compared to testosterone, which is an androgen primarily from the gonad. Adrenarche begins between about ages 6 to 9 in girls and approximately a year later in boys (Cutler et al., 1990; Hubert & Carson, 1990; Parker, 1991; Parker, Sack, Fisher, & Odell, 1978) although more recent reviews cite potentially earlier ages of adrenarche (Grumbach & Styne, 2003). However, maturation of the adrenal axis contin-

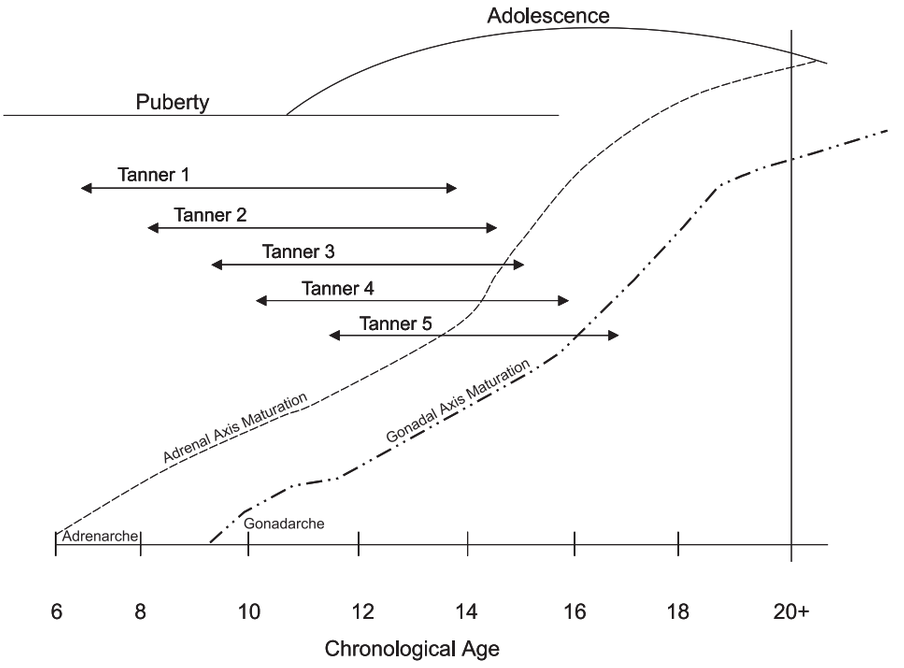

Figure 1. Representation of specific measures of puberty in girls within the phases of development (i.e., adrenarche and gonadarche). Average ages are represented but wide individual differences are possible.

ues well into the 20s. From an endocrine perspective, adrenarche is considered the earliest phase of puberty. Normal children in adrenarche still are referred to as prepubertal.

The early part of adrenarche is usually characterized by rising levels of hormones without any visible external physical signs of puberty like breast, genital, or pubic hair development. Thus, from some perspectives the early stages of adrenarche may be considered prepubertal. In general, it is only later in puberty (e.g., in gonadarche) that the concentrations of adrenal androgens are high enough to bring about the physical changes of pubic hair and body odor. Adrenal androgens continue to rise during the second phase of puberty. In fact, they continue to rise into the 3rd decade of life. The reproductive axis hormones are quiescent during early adrenarche. Low levels of steroids testosterone (T) and estradiol (E_2) exist but in small quantities and not as a result of gonadal output. In prepubertal boys and girls the contribution of T and E_2 appears to be primarily via the adrenal gland (Grumbach & Styne, 1998, 2003). The mechanism for the initiation of adrenarche remains unknown although some evidence suggests Adrenocorticotropic hormone may contribute (Weber, Clark, Honour, & Savage, 1997). Alternatively, 3 β hydroxysteroid dehydrogenase (Gell et al., 1996; Gell et al., 1998) or nutritional status (Remer & Manz, 2001) may contribute as well. Palmert and colleagues (2001) concluded that adrenarche may be more of a gradual process that begins in early childhood. Understanding that there is an early phase of puberty, called adrenarche, when hormones begin to increase, may mean that studies examining the pubertal process should begin at earlier ages than previously noted. Specifically, if hormones are thought to contribute to behavioral changes in adolescence, then longitudinal studies of early adolescence should begin enrolling participants as young as 6 years of age. Similarly, studies examining health- or disease-related issues may also want to consider the potential importance of this earlier phase of puberty.

Gonadarche represents a second phase of puberty and is characterized by the maturation of primary sexual characteristics (ovaries and testes) as well as full development of secondary sexual characteristics (pubic hair, breast, and genital development). The hallmark of gonadarche is reactivation of the hypothalamic-pituitary gonadal (HPG) axis (Knobil, 1988; Medhamurthy, Gay, & Plant, 1990; T. Plant, 1996; T. M. Plant, 1986, 2002; Terasawa, Bridson, Nass, Noonan, & Dierschke, 1984). The initial activation of the HPG axis was during the fetal and neonatal period. For gonadarche the mechanism also remains speculative (Grumbach & Styne, 2003; T. Plant, 1996; Suter, Pohl, & Plant, 1998), but we do know much more about gonadarche compared to adrenarche. This knowledge is based primarily on nonhuman primate models. With this reactivation of the HPG axis, lutenizing hormone (LH) and follicle stimulating hormone (FSH) are secreted from the pituitary in a pulsatile fashion and gonadal steroid secretion increases.

With gonadarche there is reproductive maturation and full development of the primary sexual characteristics (ovaries and testes) and secondary sexual characteristics (pubic hair and both genital and breast development). Again, it is the further increase in adrenal androgens that brings about the appearance of pubic hair, at least in girls. The gonadal steroids (T, E_2) are responsible primarily for more complex breast and genital growth. The secondary sex characteristics can be organized into stages (e.g., Tanner Stages) to reflect the developmental trajectory of puberty (see section on pubertal staging following). Within these reproductive changes, menarche occurs relatively late in gonadarche, and spermarche occurs in early to mid puberty in boys (H. Kulin, Frontera, Demers, Bartholomew, & Lloyd, 1989; Nielsen et al., 1986; Schaefer & Etemadi, 1990).

Gonadarche usually begins at approximately 9 or 10 in girls and approximately 1 year later in boys (Grumbach & Styne, 1992, 2003). Historical data on pubertal timing of girls (and studies by anthropologists in hunter–gatherer and other traditional societies) indicate that these ages represent a significant advancement in pubertal timing over the past 100 to 150 years. There is some evidence that puberty may be continuing to move to earlier ages in girls over the past few decades, however, this is still controversial in the field of endocrinology. Herman-Giddens and colleagues (1997) reported that, on average, Caucasian girls begin puberty at about age 9, but African American girls may begin showing signs, on average, as early as age 7. Menarche, however, was relatively unchanged from reports in the past decades. Although this study was the first to include such a large number of girls (over 17,000), it is important to note that this report is quite controversial based on a number of methodological issues (Emans & Biro, 1998) as well as the concluded early age of puberty (Reiter, 2001; Rosenfield et al., 2000). In addition, a recent study of referrals to a pediatric endocrine clinic suggested that puberty in these younger ages may not represent "normal" in all cases (Midyett, Moore, & Jacobson, 2003). The concern of many is that significant pathology may be missed in girls if puberty at these earlier ages is assumed to be normal. If pathology is evident, then perhaps interventions could occur (see Dorn & Rotenstein, 2004, for an example).

In a comparable large-scale study of boys using the National Health and Nutrition Examination Survey data, Herman-Giddens, Wang, and Koch (2001) reported on a secondary analysis that puberty was also earlier in boys than that reported in previous cohorts. The median age of pubic hair development was 12.0, 11.2, and 12.3 years in Caucasian, African American, and Mexican American boys, respectively (Herman-Giddens et al., 2001). Papadimitriou, Stephanou, Papantzimas, Glynos, and Philippidis (2002) reported that in 1,266 Greek boys, the mean age of genital development was at age 11, and pubic hair followed at a mean age of 11.5. These numbers were the same as those from a study in 1968.

Menarche, the onset of menses in girls, and *spermarche*, the onset of nocturnal emission in boys, are the only clear and unambiguous demarcated "events" of puberty that occur during gonadarche. Despite that fact, menarche is *not* a good proxy for marking the "onset" of puberty in girls because it occurs relatively late during the phase of gonadarche, that is well into the pubertal process. Both menarche and spermarche will be discussed in later sections.

A third set of neuroendocrine changes at puberty is related to changes in linear growth, body size, and composition (Reiter & Rosenfeld, 2003). The pulses of growth hormone (GH) become distributed more uniformly throughout the 24-hr day as contrasted with prepubertal children who have GH levels that increase primarily in the early part of the sleep period corresponding to Stage 4 of slow-wave sleep. There are increased amounts of GH in the 24-hr period and increased growth in puberty. Girls begin their growth spurt at an earlier age than boys, at approximately age 12, whereas boys typically have their growth spurt at age 14 (Marshall & Tanner, 1969, 1970). In fact, in girls the earliest outward manifestation of the beginning of puberty is an increase in growth rate (Tanner & Whitehouse, 1976; also reviewed by Parent et al., 2003). This fact is important to know if measures of height are being used to reflect pubertal development. For example, to capture part of the pubertal process using measures of linear growth, it may be beneficial to enroll girls at a younger age than boys. Similarly, linear growth in boys may continue into the late teens and even 20s. To further capture increased growth velocity, multiple measures in height are necessary each year (Parent et al., 2003).

In a Research Study, What Measure Should Be Used to Reflect Pubertal Development?

This question has no single answer. It is a bit analogous to asking a developmental psychologist for the best measure of cognitive development or asking a physician for the best measure of obesity. Of course, the answer to all of these is some version of "it depends." The first step for the investigator is to focus the questions—to determine *which* aspects of puberty (e.g., adrenarche or gonadarche) are most likely to have the greatest relevance to the goals of the particular research study. Selecting measures of puberty for a specific study requires a consideration of the processes of puberty and their concomitant physical, physiologic, neurobehavioral, and social changes. For example, in a study in which the youngest age of enrollment

is age 6 or 7, measures that reflect adrenarche may need to be considered. Specifically, hormone changes in adrenal androgens will reflect the process of adrenarche. Alternatively, if pubertal staging is used, many of the 6- to 9-year-olds will be considered prepubertal by Tanner criteria (i.e., Tanner Stage 1) even though adrenal androgens may have begun to make a significant rise. Different conclusions might be drawn when using one or the other method because they will yield different groups of children deemed to be prepubertal or pubertal. Thus, it is also very important to consider pubertal measures in the context of the dependent measures of the study.

Equally important is the need to specify the primary research questions to determine what part (or parts) of the pubertal process are most relevant to the study. For example, if pubertal staging is done by a physical examination using Tanner criteria and both breast and genital and pubic hair stages are determined, which measure should one use in the analyses? Consider the problem when a research question focuses on the influence of pubertal stage on parent or teacher report of aggressive behavior problems. In this case, the rationale may be that a measure of pubertal development that is more visible (e.g., breast development) should be used. Thus, the social signaling value of puberty may be paramount to understanding the question. Alternatively, if the research question focuses on whether pubertal development has an impact on an adolescent girl's report of her body image, then perhaps pubertal timing relative to peers or self-report of her pubertal stage may be the best measure to select. Establishing the adolescent's perception of her own development may have more of an impact on her body image than does her actual pubertal stage as indexed by physical exam (see Dorn, Ponirakis, & Susman, 2003, for empirical evidence). Striegel-Moore and colleagues (2001) showed that body mass index may mediate this relationship.

Measures of Pubertal Status or Pubertal Development

Objective Measures

In the following sections we discuss various measures that are defined in the literature as ways of measuring pubertal development. We refer to them as measures of pubertal status or pubertal development. Within each section we will describe the phase of puberty that the measure is most useful for and how it can be used to describe the pubertal process.

1. Pubertal staging by physical examination using Tanner criteria. Pubertal staging by physical examination using Tanner criteria (Tanner, 1962) has traditionally been considered the standard against which all other methods are measured. Tanner staging involves the categorization of pubertal changes (breast development in girls, genital development in boys, and pubic hair in both; Marshall & Tanner, 1969, 1970). There are 5 stages of development for both boys and girls that were adapted from Reynolds and Wines (1951). Stage 1 is indicative of the absence of any external signs of gonadal activation is underway. Stage 5 is used to designate full maturation (i.e., the presence of all necessary physical signs indicating that full physical maturity has been reached). Other sources have modified the staging for boys and define pubertal stage (PS)1 as absence of pubic hair and a testicular volume less than 3 cc, PSA2a as lack of pubic hair but a testicular volume of 3cc or greater, PSA2b is Tanner pubic hair Stage 2, and Stages 3, 4, and 5 are based on pubic hair only and parallel the usual Tanner ratings (F. M. Biro, Lucky, Huster, & Morrison, 1990, 1995). Typically, the Tanner pictures and sketches can also be found in textbooks of adolescent medicine (Hofman & Greydanus, 1989; Neinstein, 2002), pediatric endocrinology (Brook & Hindmarsh, 2001a), and psychology of adolescence textbooks (Steinberg, 1996).

There are important limitations to consider regarding Tanner staging. James Tanner originally conducted the study in England with fewer than 200 boys and 200 girls. More important, the sample collected by Marshall and Tanner was not representative of the population and included only Caucasian boys and girls. These photographs have not been standardized for differences in pubertal development of other races. Moreover, this set of black and white photos made decades ago continues to be referred to as the standard for what is used in both the clinical and research arenas for pubertal staging. These are significant limitations. Yet, to our knowledge there is no other source more frequently used as the standard than Tanner, although others have referenced additional photographs (van Wieringen, Roede, & Wit, 1985). It would be an important contribution for research and clinical care for a study to undertake updating these photos so as to represent diversity by race and ethnicity.

Tanner also contributed "norms" for examining the sequence of pubertal events in both girls and boys (Tanner, 1962). This source provides gender specific averages for the different pubertal changes (e.g., height spurt, breast development, menarche, axillary hair) that are shown in modified bar charts. For example, the figures show that most girls experience breast development as the first external change reflecting secondary sexual characteristics in puberty and menarche is nearly the last, whereas in boys testicular changes occur first. By knowing the sequence of events that contribute to the pubertal process, one can gain an understanding of the importance of evaluating which dimension(s) of puberty should be measured. That is, if a cross-sectional study is conducted in 8- to 12-year-old girls, age at menarche would not be a use-

ful measure for describing puberty because menarche occurs late in the pubertal process with a mean age of 12.0 to 12.6 years. If a study uses age at menarche as a measure of puberty, it is a misnomer to refer to those who are premenarcheal as prepubertal (see later section on menarche) as a great deal of external physical development and hormonal changes of puberty occur before menarche. Again, caution should be urged with these figures described by Tanner, as these norms were determined from a small number of participants in one race several decades ago.

Issues of measurement for pubertal staging by physical exam. Pubertal staging by physical examination is not appropriate for all research questions. If the physical exam is chosen as the appropriate method for the research question, however, there are several key issues concerning pubertal staging and an effective examination that should be addressed. First, in our consultation with both investigators and funding agencies, we have found a great deal of reluctance for some to even consider using this measure of puberty in research studies with adolescents. The pervasive myth is that adolescents would never consent to undergoing this examination. Our clinical and research experience has led us to strongly believe otherwise (see point 5 following on gender of examiner). We believe the key aspect in the adolescent's willingness to participate lies both with how comfortable the investigator is in explaining the research study and the included physical exam, as well as how experienced and comfortable the health care provider is in actually performing the examination. If one or both of these aspects is missing, then the study may fail to obtain pubertal staging by physical examination. It is important to note that examining the breast and genitalia is considered good clinical care and standard of practice in the physical examination of a child and adolescent. This has been emphasized by numerous textbooks and organizations including the American Academy of Pediatrics (2003; Burns, Brady, Dunn, & Starr, 2000; Hoekelman, Nelson, Weitzman, & Wilson, 2001). A specific example is *Bright Futures,* a guideline for health supervision that is used around the nation (Green, 1994). Specifically, the newborn and 9 month visits indicate evaluation of the genitalia and checking for the dissent of the testes. In middle childhood (age 10), early adolescence (age 11–14), and midadolescence (age 15–17), Tanner staging is indicated. In the latter two age groups palpation of the testes is suggested along with examination of the breasts in females and males (for gynecomastia). In a guide for parents for caring for teenagers, it indicates that during a physical examination an adolescent should expect an inspection of the external genitalia, palpation of the testes, and palpation of breasts in both genders (Bashe, 2003). In adolescent physical examinations in a primary care setting, pubertal stage should be recorded in the medical record. Unfortunately, not all practitioners adhere to these guidelines. We found evidence in one study of 56 boys age 10 to 14 years (Dorn, Susman, Nottelmann, Inoff-Germain, & Chrousos, 1990) during which examination for pubertal staging 3 of 56 boys were found to have undescended testes—a condition that had obviously been long-standing and never discovered by routine medical care (despite the fact that all 3 boys received regular health check-ups from their health care providers).

A second issue for conducting a physical examination lies with the training of examiners. One cannot assume that every physician or nurse practitioner has been trained to do pubertal staging accurately enough for research purposes. Pediatric residencies now require an adolescent medicine rotation that exposes residents to pubertal staging. However, the opportunity for extensive experience with pubertal staging both in medical and nurse practitioner programs may vary. Simple exposure to pubertal staging does not assure rater reliability. Thus, training in actual staging for a research study must be aggressively undertaken and interrater agreement checks are desirable throughout the study. This is illustrated by two recent studies. Hergenroeder, Hill, Wong, Sangi-Haghpeykar, and Taylor (1999) reported that agreement between physicians for breast stage was $\kappa = .5$ and for pubic hair stage was .79; coefficients that are not ideal for research purposes. Second, Albert, Hunsberger, and Biro (1997) reported misclassification when Tanner staging a longitudinal cohort of 9-year-old girls who participated in the National Heart, Lung, and Blood Institute (NHLBI) Growth and Health Study (NGHS). Forty-five percent of the 1,155 girls had an apparent "decrease" in maturation during a follow-up visit.

In addition, to conduct a successful physical exam for research purposes, the clinician must be experienced and comfortable in seeing adolescents. Methods of further training could include accompanying an experienced adolescent medicine physician or nurse practitioner to conduct multiple exams together and obtain interrater agreement prior to initiating the research study. As an alternative, experiences with pediatric endocrinology clinicians could lead to simultaneous examination of children and adolescents with various endocrine disorders (e.g., precocious puberty, premature adrenarche) or disorders of growth (e.g., constitutional delay) that may affect pubertal development. Still other studies have provided training sessions for reliability of performing pubertal staging. Even with such training, interrater agreement should be conducted throughout the study. That is, for a certain percentage of cases, two people independently conduct the physical examination and compare results. On our interrater agreement cases we have given adolescents a choice in having two examinations, and most

accept this addition. Our explanation of describing the importance of knowing that examiners consistently agree on the stage is understood by the adolescent when it is explained as "testing" the examiner. Most of the studies reviewed did not discuss training of the clinical examiner or interrater agreement for pubertal staging.

A third key issue lies with the context in which the examination can be done. It is helpful to conduct the examination in an appropriate setting such as a room designed and equipped for a physical examination with appropriate privacy. Also, appropriate gowns and drapes should be available, although we found that many adolescents prefer just to loosen clothing for the exam as it is not necessary to totally disrobe. In the NGHS (NHLBI, 1992), t-shirts of a known premeasured weight were used to maintain privacy and allow for assessment of maturation, skin folds, and body circumferences. Participants then kept the t-shirt as day clothing or a sleep shirt. Examinations can be done in clinical areas, and many schools have the appropriate facilities, particularly schools that have school-based clinics.

Fourth, the number of participants in the study may make direct examination difficult logistically because of time constraint or privacy. However, even with hundreds of adolescents a physical exam can be done with appropriate personnel. For example, in brief time periods, sports screening physicals are conducted effectively within several sessions set aside for this purpose. If one wants to conduct research in several middle schools or high schools, perhaps schools could be selected that already have school-based health clinics. Many of these clinics are staffed by nurse practitioners who already conduct physical exams. Experience of these clinicians varies across school systems so this option should be explored carefully. Training for research-grade pubertal assessments is still desirable.

A fifth consideration for conducting exams for pubertal staging includes the issue of gender of the rater. Some research studies provide an examiner of the same sex whereas others do not. Although the myth seems to be that adolescents prefer the same sex examiner, the literature does not always support this. For example, in a study of over 5,000 adolescents in Grades 5 through 12, 65% of males had no preference for the gender of their health care provider and 23% preferred a male (Kapphahn, Wilson, & Klein, 1999). For girls, 48% had no preference and 50% preferred a female provider. In a study of 108 (male and female) early adolescents seen three times across 1 year (Dorn et al., 1990), all but 1 participant allowed a female examiner to conduct the physical. The one exception was a female who refused to have the examination at her second visit. In another study with 6- to 9-year-olds, all 29 girls and boys consented to the physical exam by a female examiner (Dorn, Hitt, & Rotenstein, 1999). In the NGHS (NHLBI, 1992) and our study on children and adolescents with depression and anxiety (Ryan et al., 1992; Ryan et al., 1994), we always used same gender examiners. The first author's current study has thus far enrolled 120 girls who are 11 to 17 years of age (Dorn, 5/10/03–4/30/08). The study uses trained male and female examiners and no participant has refused the examination or requested an examiner of a specific sex. Although in all of our studies the compliance with examinations has been excellent, we do not know if potentially eligible participants choose not to enroll once they know that a physical examination will be conducted. The preferences of adolescents may not indicate that the same gender examiner is necessary. However, based on observations within the legal system, some have used chaperones when an examiner of the opposite sex is used.

Sixth, the technique of the examination in both boys and girls must be considered. One issue that has received attention has been whether the exam should consist solely of visualization of breast and genitals or if palpation should be included. Although Tanner's original study was based on breast inspection (Marshall & Tanner, 1969), breast staging is more accurate in girls when palpation is done. This is particularly critical for those girls who are in early puberty and who are slightly overweight or obese (Bonat, Pathomvanich, Keil, Field, & Yanovski, 2002). It is impossible to distinguish between breast tissue and adipose tissue without palpation and even difficult for an experienced pediatric endocrinologist or adolescent medicine physician to distinguish between the two with recommended palpation. Without palpation, an overweight girl can be easily categorized as Tanner Stage 2 or 3 breast development because of excess adipose tissue although she should be categorized as a Tanner Breast Stage 1 (F. M. Biro, Falkner, Khoury, Morrison, & Lucky, 1992). This obviously increases the margin of error in the data. Bonat and colleagues (2002) also suggested that self-reports may be overestimated at Stages 1 to 4 when increased adiposity is present.

For boys, a more objective examination can be conducted through palpation of the testes (see testicular volume determination following). This is favored over the subjective determination of genital stage in boys through visualization of length and width of the penis and scrotum (F. M. Biro et al., 1995; Largo & Prader, 1983). Again, it should be emphasized that a physical examination for adolescents in a primary care setting should include both visualization and palpation of breast and genitalia as standard medical practice (Bashe, 2003; Burns et al., 2000; Green, 1994). Our experience has shown that most adolescents do not have difficulty with this exam when the practitioner is experienced and comfortable conducting the examination.

With respect to the assessment of pubic hair, it is important to note that visualization of early pubic hair may require a close examination with an excellent light source. Distribution of hair may be sparse. In addition, one should distinguish between true pubic hair and the fine velus hair evident in prepubertal boys or girls. This velus hair may be more evident in African American than Caucasian youth.

Seventh, the cost–benefit ratio of conducting a physical examination deserves consideration. A trained health care professional can be costly but creative ways of employment are an option (e.g., hire health care professionals who can serve in multiple roles, hire on a per exam basis, collaborate with a fellowship program to provide training/clinical hours, etc.).

Finally, once the physical exam has been conducted, the decision needs to be made regarding the most appropriate measure for the statistical analysis. We refer readers back to an earlier section on the importance of dissecting the research question and how best to determine the appropriate measure of puberty. It is important to emphasize that studies that take the average of two processes (e.g., breast + pubic hair stage) may not reflect the best measure for any study. This method combines the phases of adrenarche and gonadarche and emphasizes different hormonal contributors especially in early puberty (F. M. Biro et al., 2003). Brook and Hindmarsh (2001b) stated that "under no circumstances should these ratings be lumped together for an overall stage of puberty" (p. 121). Such factors may have a significant impact on the conclusion of the study.

A frequently raised question centers around the incorporation of pubertal measures like the physical examination into studies. As a result of the *perceived* difficulty of using the physical exam, we discuss successful approaches to carrying out pubertal staging.

We have emphasized the importance of collaboration with experts in the field, expertise and comfort of the examiner, and appropriate context/environment for conducting the exam for pubertal staging. In addition, we have found several other strategies that are useful. For example, it is helpful during a physical exam to go from a least invasive aspect of the exam to the most invasive. In some studies we first measured the height and weight of each child or adolescent and then plotted the measurements on a growth chart. Both the meaning of percentiles and a person's development were then discussed. Next, we reviewed the gender appropriate figures of all the changes of puberty (e.g., the range for breast development, height spurt, menarche) with the adolescent. This provided a helpful introduction to the Tanner stage pictures and the adolescent's task of self-rating maturation. Finally, after explaining what was to be done in the exam, the physical exam was then conducted. Another option that has been useful in studies is to replace a full exam for pubertal staging with a "mini" physical exam that includes the Tanner staging. For example, eyes and ears are examined along with auscultation of the chest, palpation of the abdomen, followed by pubertal staging, and the like. It is also advantageous to consider providing a free physical exam that can be used for sports or camp readiness. If a study decides to include this in its recruitment strategy, it should be noted that a more thorough history and physical exam may be warranted depending on the purpose (e.g., history questions to rule out risk of sudden cardiac death, a more complete physical exam for neuromuscular status, etc.).

2. Pubertal staging by areolar development. Another method of pubertal staging by physical examination of girls is measurement of areolar diameter. This method is considered by some investigators to be a more accurate way of categorizing girls into pubertal stages because adipose tissue does not interfere with the measurement itself (Aygun, Akarsu, Guvenc, & Kocabay, 1998; F. M. Biro et al., 1992; Daniel & Paulshock, 1979; Morrison et al., 1994; Rohn, 1987; Sprecher et al., 1997). This approach is much less subjective; however, there is little normative data of areolar measurements as fewer studies have utilized such methodology. A current chapter does cite norms (Grumbach, 2002). Currently, this measure can be useful for girls in adrenarche and gonadarche similar to the usefulness of pubertal staging by physical exam using Tanner criteria (see previous section). This method may be particularly useful in longitudinal studies.

Issues of measurement for areolar development. The previous section of issues related to conducting a physical examination also applies to assessment of areolar development.

3. Pubertal staging by testicular volume. In clinical practice in both pediatric endocrinology and adolescent medicine, and in several research studies, measurement of testicular volume is used to categorize pubertal development. The most frequent method of measurement is the use of a Prader orchidometer (www.accuratesurgical.com) that consists of a set of either 12 or 14 elipsoidal beads that represent testes volume ranging from 1 to 25 cc or 1 to 35 cc, respectively. Calipers to measure length and width of the testes have also been used. F. M. Biro and colleagues (1995) used a method to define *Pubertal Stage 2a* (PS2a) as pubic hair absent but testicular volume greater than 3cc; this stage persisted, on the average, 6 months prior to the appearance of pubic hair. There are "rules of thumb" used clinically that state that a prepubertal testis is less than 3 cc. One source reported in Daniel and colleagues (Daniel, 1982) and cited by Neinstein (1996) showed that the left testis is on aver-

age 4.8 cc (*SD* = 2.8) at Tanner Stage 1, 6.4 cc (*SD* = 3.2 cc) at Tanner Stage 2, 14.6 cc (*SD* = 6.5) at Tanner Stage 3, 19.8 cc (*SD* = 6.2) at Tanner Stage 4, and 28.3 (*SD* = 8.5) at Tanner Stage 5. The right testis is slightly larger at all stages. A clinical reference (Genentech, 1997) reported that testicular volume at Tanner Stage 1 is ≤ 3 cc, Tanner Stage 2 is > 3 to 6cc, Tanner Stage 3 is > 6 to 10 cc, Tanner Stage 4 is > 10 to 15 cc, and Tanner Stage 5 is >15 cc. However, the methodology and sample for this information were not reported on which norms of testicular development were based. More important, most pediatric endocrinologists feel that testicular volume is the best way to determine genital stage in boys rather than merely by judging developmental changes in the scrotum or length and width of the penis from visualization only.

Testicular volume is a more precise measure of pubertal onset and may be particularly important if one wants to determine prepubertal versus peripubertal development in a study. Testicular enlargement is initiated by reactivation of the HPG axis in which sensitivity of the target tissues (e.g., the testes) is enhanced. In turn, T concentrations increase beyond prepubertal levels. Because an increase in testicular volume is generally the first visible change of puberty in boys, testicular volume is the best way to determine the change from prepubertal (Stage 1; within the phase of adrenarche) to pubertal at Stage 2 and above (when gonadarche begins). This increase in volume is usually not visible without palpation, specifically by a healthcare provider. Thus, self-report or parent-report of pubertal stage may underestimate pubertal development at this early stage. Testicular volume in boys is a particularly good choice for research if one is interested in knowing when pubertal processes begin.

Issues of measurement for testicular volume. It is not uncommon for testicular volume to be measured in boys undergoing a physical examination in a clinical practice, in an adolescent medicine clinic, or in endocrine clinics. In primary care clinics or offices an orchidometer is generally not used. Testicular volume also has successfully been measured in studies of puberty in boys age 6 and older. In our own studies no boys refused the examination using the Prader Orchidometer (Dorn, Hitt et al., 1999; Dorn et al., 1990) and there was less than a 5% refusal rate in another study (F. M. Biro et al., 1995). The same method of describing normative pubertal changes in adolescents before the physical exam may also contribute to the success of the adolescent accepting measurement of testicular volume. It also takes a skilled clinician who can talk to the adolescent during the exam to explain procedures, allay concerns, as well as carry on a conversation unrelated to the examination. In addition, training is required, albeit minimal, for those clinicians who have conducted examinations in adolescent boys. Interrater agreement must also be conducted periodically for research purposes just as described in the pubertal stage section. Finally, the cost of the orchidometer is approximately $45 to $100.

4. Hormone concentrations. It was not until the 1980s that the first studies conducted of normal adolescent development included hormone concentrations and an emphasis on a biopsychosocial perspective of pubertal maturation (Brooks-Gunn & Graber, 1994; Brooks-Gunn & Warren, 1989; Halpern, Udry, Campbell, & Suchindran, 1993; Nottelmann, Susman, Dorn et al., 1987; Nottelmann, Susman, Inoff-Germain et al., 1987; Susman, Inoff-Germain, Nottelmann, & Loriaux, 1987; Susman et al., 1985; Susman, Nottelmann, Inoff-Germain, & Dorn, 1987; Udry, Billy, Morris, Groff, & Raj, 1985; Udry & Talbert, 1988). These studies measured serum concentrations of adrenal androgens, gonadal steroids, and gonadotropins. The youngest participant in these studies was age 9; therefore the earliest part of adrenarche was not represented. Most of these studies simultaneously measured pubertal stage by physical exam. It is important to note a widespread belief that an individual cannot be categorized into a pubertal stage based solely on the concentration of a hormone. Although there are tables indicating ranges of hormones by stage (McAnarney, 1992), overlap exists that consequently limits the usefulness of hormones as a method for determining stage. Since the early studies of adolescents using serum hormones, new methodologies for examining hormone concentrations have included the measurement of hormones by blood spot, urine, and saliva. Advantages and disadvantages of these methods will be elaborated on in the following paragraphs.

Issues of measurement of hormone concentrations. The advent of radioimmunoassay in the 1960s provided an important step for examining hormone concentrations. Since that time new technology has developed that has further advanced the field. This includes the enzyme-linked immunosorbent assay, or ELISA, method that does not require any radioactive material, thus providing a simpler method of analysis. Although many hormones can now be easily measured, there are significant factors that must be carefully considered when hormones are included in a study of adolescent development. If such factors are not addressed in the protocol, significant errors in measurement can occur.

First, as discussed previously, it is necessary to consider the validity of hormone concentrations reflecting pubertal development based on the research questions of the study. This issue becomes particularly important due to skepticism surrounding a concentration of a hormone accurately reflecting a certain pubertal stage as a result of tremendous variability and overlap in hor-

mone concentrations within and between pubertal stages in both boys and girls.

Second, it is important to consider the normal physiology/endocrinology of a specific hormone. For example, when one is measuring T or E_2 in the serum, it is necessary to note that these hormones are located "downstream" from the hypothalamic releasing factors (e.g., gonadotropin releasing hormone, GnRH; the gonadotropins, LH, FSH). That is, significant hormone or releasing hormone increases are beginning prior to the secretion of T and E_2 from the gonad. These upstream hormone processes may be more relevant to some types of behavior than T or E_2, yet T and E_2 are *assumed* to contribute to behavior. Methods for directly measuring GnRH are not possible for most studies, whereas measures of T and E_2 can be obtained in clinical studies from peripheral blood and in some cases saliva. Other physiologic factors should be considered such as the temporal domain for optimal sampling. For example, some hormones have an ultradian rhythm (minute to minute fluctuation) or a circadian rhythm (varies by time of day), whereas others have monthly rhythms such as those hormones of the menstrual cycle (Cauter, 2001). Thus, study designs should consider the issue of rhythms and have participants come in at a consistent time of day and at a consistent time of the menstrual cycle for females. Alternatively, for those hormones with an ultradian rhythm, multiple samples must be collected so as to avoid obtaining the peak or the nadir of the hormone if only one sample is obtained. Some hormones such as T and cortisol have more than one rhythm that should be accounted for in the study design. Thus, the rhythmic properties of each individual hormone should be addressed. Some studies that were unable to conduct their research procedures at a consistent time of day have utilized the method of covarying "time of day" in the statistical analyses. Although including the covariate is feasible, tighter methodological constraints of enrolling at a consistent time of day may offer some advantages.

In addition, enrolling adolescent females during a certain phase of the menstrual cycle is most desirable. This task is easier to accomplish in studies of girls toward the end of puberty rather than at the beginning due to the normative irregularity of early menstrual cycles. Menstrual cycles are very irregular for approximately 2 to 3 years following menarche (Flug, Largo, & Prader, 1984; Widholm & Kantero, 1971) and an individual girl's normal cycle link may not be established for as many as 6 years postmenarche (Flug et al., 1984; Widholm & Kantero, 1971). Shortly after menarche, it is often difficult to predict when the next menstrual cycle will occur and for how long, making it difficult to schedule follow-up study visits. Also, the first adolescent menstrual cycles are generally anovulatory within the first 2 to 3 years following menarche (Venturoli et al., 1987; Venturoli et al., 1986). Thus, the hormonal characteristics of menstrual cycles of girls who have just begun menstruating are different from those of girls who have been menstruating for several years (Apter & Vihko, 1985a; R. Vihko & Apter, 1984). To address this issue on a practical level, we have attempted to bring adolescents in during the follicular phase of their menstrual cycle (Day 5–9) in some of our studies. However, if 45 days have lapsed between cycles, we bring them in for a study visit.

A third factor that is important when conducting research using hormones is the assay methods. For example, if one is enrolling prepubertal (i.e., pregonadarche) girls or early pubertal boys, the current serum E_2 assays are generally not sensitive enough to measure such prepubertal concentrations. That is, the level of detection of the assay is higher than the actual hormone concentration. However, a newer methodology in serum exists that can differentiate individual differences in Tanner Stage 1 (prepubertal) boys and girls (Klein, Martha, Blizzard, Herbst, & Rogol, 1996). Unfortunately, this methodology is less widely available. In addition, there is large variance within and across laboratories if one compares the measured concentrations of a hormone (Halpern & Udry, 1992; Schwartz, Granger, Susman, Gunnar, & Laird, 1998). Because there is variability across each assay run (interassay coefficient of variation), it is important to consider how samples are batched for the analyses. For example, if a longitudinal study is being conducted with three visits across a year, there are several questions to be addressed: (a) Should one analyze all three time points of one participant in one batch? (b) If a study includes a patient (or treatment) group and a healthy comparison group, should those samples be paired and run together? Each question deserves careful consideration by an investigator and with each dependent on individual studies and the research goals.

Fourth, it is important to consider hormone concentration differences due to gender, age, and race. All of these factors may have an impact on the outcome variable. Other important sources of variance may include activities such as fasting, smoking, exercise, or sleep. Medications and environmental factors may also influence some hormone concentrations as illustrated by hormone differences collected in the laboratory versus home setting (Dorn et al., unpublished observation).

Finally, it is crucial to consider the substance (e.g. blood, saliva, urine) that is used to measure the hormone. Saliva can now be used to analyze some but not all hormone concentrations and not all laboratories have the capability to measure hormone concentrations in this medium. Therefore, it is beneficial to investigate the experience of the laboratory. Laboratories frequently used for salivary assays include Salimetrics Inc. (salimetrics.com), the laboratory used by Kirschbaum and Hellhammer (1992), and numerous other laboratories do analyze salivary assays. Salivary

assay techniques are more highly perfected for cortisol, T, and some adrenal androgens. Others report that the saliva E_2 is a valid assay. To date, our review of the article by Shirtcliff and colleagues (2000) questions the assay methodology for saliva E_2 as not being sensitive enough to measure these low concentrations in prepubertal and early peripubertal girls and boys (see the following review). Measuring hormones in saliva may be easier to collect and process and also less invasive. The saliva collection method can also be taken to the field, but issues of collection remain a concern. For example, brushing teeth, eating, or drinking certain substances may interfere with the assay. See the chapters by Dahl, Dorn, and Ryan (1999) and Schwartz and colleagues (1998) for further information. Also, the type of collection device is important as illustrated by Dabbs, Jurkovic, and Frady (1991) in which the cotton used to collect the saliva interfered with the T assay.

Analyses of hormone concentrations by blood spot offer an alternative measure. With this technique the finger is pricked and blood is used to fill areas on filter paper. This technique was perfected by Worthman and Stallings (1994, 1997) and has been useful in the field because there are no issues of refrigerated storage for the sample. To date, hormones that have been analyzed using the blood spot technique include cortisol, gonadotropins, progesterone, and E_2.

Two recent articles have been published that have examined reliability and validity issues of serum hormones with blood spots and saliva. The first article by Shirtcliff and colleagues (2000) reported that correlations of serum and blood spot assays for E_2 were high in 17 boys ($r = .73$, $p < .001$) and 18 girls ($r = .96$, $p < .001$) whereas the correlations of serum and saliva were noticeably lower in 14 adult females ($r = .60$, $p < .013$) and not significant for adult males ($r = -.07$). Blood spot and saliva E_2 correlations were nonsignificant for boys ($r = -.18$) and relatively strong for girls ($r = .72$, $p \leq .002$). More important, the sensitivity of the E_2 salivary and blood spot assay was reported as adequate for the majority of prepubertal and postpubertal boys and girls (Shirtcliff et al., 2000). Unfortunately, it was not evident from the methods section how prepubertal was determined. Boys in the study ranged from 8 to 9 years. Although at this age it is *likely* that they were prepubertal, one cannot determine from the study that this was the case. Further, girls were described as "nonmenstruating" and between the ages of 10.78 and 12.27 years and equated with a prepubertal categorization. So as a result it is very likely that a number of these girls *were* pubertal, and, as we have illustrated before, premenarcheal does not equal prepubertal. More information is needed to determine if the assay is truly sensitive at prepubertal concentrations particularly in adolescent boys in which the concentration of the hormone was undetectable in 22.2% of the samples. A second important point of this article included the examination of hormone–behavior relations. Creatively using computer simulated behavioral variables, blood spot E_2 replaced serum and only underestimated the correlation by 3.45%. However, when saliva was substituted for serum, it was found to underestimate the correlation of E_2 and behavior by nearly 38% (Shirtcliff et al., 2000). The authors noted the importance of determining advantages and disadvantages of each sampling methodology.

A second excellent article rigorously examined sensitivity and reliability to obtain concentrations of testosterone, progesterone, and E_2 in adults using the blood spot methodology (Shirtcliff, Reavis, Overman, & Granger, 2001). The correlation between serum and blood spot was high; however, the variance was particularly low for E_2 and progesterone in men. Thus, there was a greater limitation in using this methodology in men.

Urinary gonadotropins such as LH and FSH have been used in studies of puberty since the 1960s (Fitschen & Clayton, 1965; H. E. Kulin, Rifkind, Ross, & Odell, 1967). In fact, these studies were able to measure prepubertal concentrations of gonadotropins. More recently ultrasensitive time-resolved immunofluorometric assays, or IFMAs, have been developed and showed a high correlation of serum and urine concentrations ($r = .72$) in 65 children age 0 to 15 years (Demir, Alfthan, Stenman, & Voutilainen, 1994). Further, the age-related trends closely paralleled serum levels (Demir, Dunkel, Stenman, & Voutilainen, 1995). Demir and colleagues reported a low sensitivity of the assays, but there was a decrease in levels when the samples were frozen, particularly when samples were repeatedly frozen and thawed. Thus, specific storage concerns must be followed. Urine collections offer a noninvasive method to examine hormone concentrations. Collection can be challenging in young children and adolescents but it is thought that 24 hr collections are no longer needed as a first morning-voided sample reflects 24 hr output (Girard, Baumann, & Ruch, 1980).

Next, the issue of binding proteins for hormones (e.g., sex hormone binding globulin, SHBG) must be considered. These proteins are carrier proteins that are the mechanism of hormone transport to the target tissues (e.g., breast) and can be analyzed in a blood sample. A full discussion of binding proteins is beyond the scope of this article; however, some studies have measured binding proteins and have used ratios of the binding protein to the hormone concentration. Binding proteins can be altered with medication (e.g., oral contraceptives), disease states, or normal physiological changes of development (e.g., pregnancy). Thus, analyzing binding proteins may be important in certain studies.

Finally, the cost of an assay is a consideration. Costs vary tremendously for each hormone as well as across laboratories. Research laboratories generally have lower rates than a commercial laboratory and costs can be as low as $5 per sample or as high as $100 or more per sample.

5. Gonadal ultrasound. Ultrasound technology can be used to examine the ovary and determine ovarian volume in adolescents, but there is no agreement for norms of these measures nor has the relationship been clarified between ovarian volume and other markers of puberty. A recent study included 139 girls with a mean age of 6 years (range 1 to 13 years) that reported significant differences in ovarian volume of prepubertal versus peripubertal girls (Herter, Golendziner, Flores, Becker, & Spritzer, 2002). Ovarian volume was positively correlated with chronological age and pubertal stage but, to our knowledge, could not be categorized into specific pubertal stages based on ovarian volume. More precise measures of ovarian volume can be obtained in a more invasive manner with a transvaginal probe, but this most likely is not feasible (or necessary in most studies) in pubertal age adolescents. This measurement is also costly. Thus, the use of ovarian volume by ultrasound would not be useful for most studies as a marker of pubertal development.

Ultrasound has also been utilized to determine testicular volume. The studies reported in the literature were generally not specific about adolescents but compared methods of determining volume in male adults. Fuse, Takahara, Ishii, Sumiya, and Shimazaki (1990) utilized three methods of measurement for testicular volume (e.g., orchidometer, calipers) compared to ultrasound and found that ultrasound was the best method. Further, they reported that measurement by calipers was not accurate. In another study, ultrasound was also favored to assess volume differences (Diamond et al., 2000). Alternatively, Carlsen and colleagues (2000) reported a significant difference by 3.6 to 4.3 ml in testicular measurement with an orchidometer compared to ultrasound with the latter underestimating size. In the one study with male adolescents, five methods of determining testicular volume were all highly correlated but not all equally accurate (Chipkevitch, Nishimura, Tu, & Galea-Rojas, 1996).

Issues of measurement using gonadal ultrasound. The primary issue of measurement lies with accessibility to the equipment and having an experienced radiologist to interpret the results. The high cost of the procedure will vary by the radiologist fees at each institution and cost is weighed against the lack of norms across puberty and hence the difficulty in interpretation.

6. Age at spermarche. There have been some investigations into the onset of spermatozoa production (spermarche or nocturnal emissions) in adolescent males. H. Kulin and colleagues (1989) found that the median age of spermarche (determined by morning urine samples) in a longitudinal study of boys age 10 through 17 was 14 years and that LH and FSH were also at adult male levels at that time. A similar age of spermarche was reported in 129 boys within the same age range (Schaefer & Etemadi, 1990). Self-report measures concur with this age range in nearly 87,000 Chinese boys from a national health study where the median age at spermarche ranged from 14.3 to 14.7 years depending on social class and geographical setting (urban vs. rural; Ji, 2001). Earlier, Nielsen and colleagues (1986) reported on a longitudinal sample of 40 boys slightly younger (age 8.6 to 11.7 years) at study initiation. They reported that spermarche occurred at a median age of 13.4 years but the age ranged from 11.7 to 15.3 years. It is important to note that there was great variability in the developmental markers that coincided with spermarche. The median testicular volume at the time of spermarche was 11.5 ml (range 4.7 to 19.6 ml) and pubic hair stage was 2.5. It would be difficult to determine if one had sperm production based on age or testicular volume with the wide range of those variables. In addition, spermarche (or lack thereof) reflects a dichotomous variable of puberty that can only be equated with being prepubertal (no sperm production) or pubertal (sperm production present). Thus, there are no more finely tuned stages of development. Spermarche appears to be a relatively early event in the pubertal process, but there is great variability. Age at spermarche may be important for some research questions, and it may have implications for health education. We know little about the significance of the event to boys as noted by the few reports on the psychological aspect of spermarche in adolescents (Adegoke, 1993; Downs & Fuller, 1991; Gaddis & Brooks-Gunn, 1985). Age at spermarche has also been determined subjectively by parent or self-report.

Issues of measurement for age at spermarche. To our knowledge there have been no known studies comparing the actual laboratory findings of spermarche (e.g., from morning urine samples) with self-report or parent report of spermarche. In addition, a laboratory cost would be involved.

Subjective Measures of Pubertal Development

1. Age at menarche by self- and parent-report. Age at menarche is frequently used as a variable to define pubertal status in adolescent girls. Because having one's first period is clearly a timed event, this measure is often and erroneously used to determine whether a

girl is prepubertal or postpubertal. The actual time this event occurred can be obtained by self- or parent-report using either a questionnaire or an interview methodology. Typically, a girl or her mother are asked, "How old were you (or your daughter) when you (she) got your (her) first period?" The age (e.g., 12, 13) is then recorded. Gynecological age can also be computed by subtracting the month and year of menarche from the date of the interview. For example, the variable of gynecologic age then used in analyses would be 2.5 years if an adolescent had reached menarche 2 years and 6 months prior to the interview date.

Issues of measurement for age at menarche. There are at least three important factors to consider when using the age at menarche variable in studies of pubertal development. First, menarche is a late event in the pubertal process. The majority of girls (nearly two-thirds) are Tanner Stage 4 when they get their first period and fewer are in Tanner Stage 2 (5%), Tanner Stage 3 (25%), or Tanner Stage 5 (10%; Neinstein, 1996). If a measure showing more finely tuned categories of the pubertal process or a distribution of pubertal participants is desired, age at menarche may not be the best measure. Investigators typically create a dichotomous variable of premenarcheal versus postmenarcheal, with some studies mistakenly equating premenarcheal versus postmenarcheal with prepubertal versus postpubertal. Many pubertal changes are well underway before menarche occurs (e.g., breast and pubic hair, growth spurt, axillary hair, gonadal and adrenal androgen changes; Tanner, 1962). This method of age at menarche represents assessment of the phase of gonadarche. Other concerns regarding this method appear in the Timing of Puberty section.

Second, reliability issues of self-report of age at menarche raise some concern. Although across the last 40 years studies have reported correlation coefficients for reporting age at menarche across many years (e.g., .60–.81; Bergsten-Brucefors, 1976; Casey et al., 1991; Damon & Bajema, 1974; Livson & McNeill, 1962; Must et al., 2002), there is tremendous variability in reporting. Again, the correlations may be high but, if accuracy is needed, this measure may not be useful. This may be of even greater concern for longitudinal studies of adolescents in which the age of menarche is obtained at each time of measurement. Dorn and colleagues (Dorn, Nottelmann, et al., 1999) reported that within an individual, self-report of age at menarche may vary by as much as 18 months across two methods of reporting (interview and questionnaire) at three times of measurement within 1 year. In this same article, methods were cited to enhance reliability and validity of self-report age at menarche. Although the sample was small, the issue deserves further investigation.

Third, statistical considerations should be noted in analysis of age at menarche data. For example Herman-Giddens, Kaplowitz, and Wasserman (2004) reported on how the average age of menarche varied in the same dataset depending on the statistical methodology utilized (see also Chumlea et al., 2003; T. Wu, Mendola, & Buck, 2002).

2. Pubertal stage using self- and parent-report. Defining pubertal stage by self- or parent-report has been used as a proxy for pubertal development by physical examination. This evolved from concerns and fears of both adolescent refusal of the physical exam as well as out of fears of investigators, school systems, or funding agencies that examinations could not be completed. There are two primary methods reported in the literature for assessing pubertal stage by self- and parent-report. First, a number of studies measured pubertal stage by having the adolescent and parent examine photographs or line drawings of Tanner stages (Bonat et al., 2002; Brooks-Gunn, Warren, Rosso, & Garguilo, 1987; Dills et al., 1995; Dorn et al., 1990; Duke, Litt, & Gross, 1980; Fewtrell, Cole, Bishop, & Lucas, 2000; Hammer et al., 1991; Morris & Udry, 1980; Neinstein, 1982; Taylor et al., 2001). Studies varied as to whether or not an explanation was given to the participants regarding the differences for each stage or if they were merely given the materials in a larger packet of measures. Not all research studies are explicit in telling whether they have done the former or the latter. The accuracy of self- and parent-report varies when compared to the physical examination conducted by a health care professional. Furthermore, it depends if accuracy is defined by "highly correlated" or by agreement coefficients (e.g., Kappa or percentage agreement; see Table 2 for review).

Some investigators reported comparisons of self-report with that of a physical examination in which Kappa coefficients were generally low and correlations were moderate to high (Dorn et al., 1990; Schlossberger, Turner, & Irwin, 1992). Low Kappa coefficients of self-report and physical exam, in a study of 107 girls age 8 to 17 years, revealed .34 and .37 for pubic hair and breast development, respectively (Hergenroeder et al., 1999). Similarly low coefficients (κ = .32–.51 for areolar and .36–.55 for pubic hair) were noted in a sample of Black and White girls from the NHLBI NGHS, with the investigators suggesting that self-assessment should only be used when crude ratings of pubertal development (i.e., prepubertal vs. postpubertal or less fine-tuned ratings) are needed (Y. Wu, Schreiber, Klementowicz, Biro, & Wright, 2001). Taylor and colleagues (2001) also reported low Kappa coefficients for breast or genital development (0.48) and a moderate coefficient for pubic hair (0.68). Alternatively, higher Kappa coefficients of .81 to .91 were reported by Duke and colleagues (1980) in 69 girls and boys. A recent study in overweight children by Bonat and colleagues (2002) reported that breast stage was overestimated by

Table 2. *Summary of Studies Comparing Agreement of Pubertal Stage by Physical Examination Versus Self- and/or Parent-Report*

Study	N	Status	Age Range	Measures	Results
Bonat et al., 2002	244 (135 girls; 109 boys)	41% obese	5–12 yo	1. PE Tanner 2. Self-report by Tanner drawings & text explanation	Kendall Rank Correlations with PE self breast: Obese (.37) non (.54) self hair: Obese (.64) non (.66) self hair boys: Obese (.45) non (.35)
Brooks-Gunn et al., 1987	151 Caucasian girls	Healthy	*n* = 38–11 yo *n* = 50–12 yo *n* - 63–13 yo	1. PE Tanner 2. Self-report: Tanner, line draw 3. Mother: Tanner, line draw	Correlations with PE self breast: *r* = .52–.68 self hair: *r* = .58–.74 mom rating breast: *r* = .69–.82 hair: *r* = .57–.83 total PDS: *r* = .54–.56 (*r*'s vary by age) Using % agreement highest was at: self breast stg 3 (62%) self hair stg 4 (55%) mom breast stg 3 (68%) mom hair stg 1 (86%)
Boas, Falsetti, et al., 1995	61 boys (34 c CF; 27 Healthy controls, race ?)	34 CF 24 controls	12–19 yo *M* = 14.3–14.9	1. PE Tanner 2. Self-report: Tanner, picture	Kappa Coefficients 1. pubic hair: CF κ = .802 (.946 weighted) controls κ = .732 (.905 weighted) 2. genitalia: CF κ = .489 (.840 weighted) controls κ = .345 (.657 weighted)
Carskadon & Acebo, 1993	698 (323 boys; 375 girls; race ?)	Healthy	5th & 6th grades	1. PDS Interview (adaptation): self parent teacher	self vs. parents item corr. ranged from .13 (growth sp) to .97 (menarche); mean PDS ranged from .23–.80 self vs. teacher ranged from .10–.44 parents from .10–.42 girls > boys in most cases
Dorn, Susman, et al., 1990	90 (46 boys;44 girls; primarily Caucasian)	Healthy	boys: 10–15 yo *M* = 12.7 + 1.3 girls: 9–15 yo *M* = 11.95 + 1.6	1. PE Tanner 2. Self-report: Tanner, Picture 3. Parent-report: Tanner	Correlations (*r*) of PE with self: .84 (genital), .77 (hair) .88 (breast), .91 (hair) κ = .33–.50 Correlations (*r*) of PE with parent: .75 (genital), .76 (hair), .86 (breast),.87 (hair) κ = .13–.55
Duke, Litt, & Gross, 1980	66 (43 girls; 23 boys; primarily Caucasian)	Healthy	girls: 9–17 yo boys: 11–18 yo	1. PE Tanner 2. Self-report: Tanner, picture	κ = .81 breast κ = .91 female pubic hair stage κ = .88 combined male pubic hair/genitalia

Continued

Table 2. *Continued*

Study	N	Status	Age Range	Measures	Results
Hardoff & Tamir, 1993	143 (Israel)	80 severely disadvantaged LD but normal intelligence, 70% boys 63 mainstream urban school (c) .52% boys	7th & 8th grade	1. PE Tanner; hair only 2. self-report: Tanner, picture)	c–group: k = .55 LD group: k = .43, $p < .001$ (not split by gender)
Hergenroeder et al., 1999	107 girls (48 Caucasian, 43 African American, 11 Hispanic, 5 Asian)	Healthy	8–17 yo	1. MD examiner for PE 2. Self-Assessment (no drawings & descriptions)	(Breast/Pubic Hair) k = .35/.44 Caucasion k = .42/.26 African American k = .09/.30 Hispanic
Hick & Katzman, 1999	40 girls	Anorexia nervosa	8–18 yo	1. PE by 2 pediatricians with adolescent medicine expertise 2. Standardized figure drawings	% agreement of MD & self-assess 30% breast 50% pubic hair
Morris & Udry, 1980	44 (47 girls; 48 boys; mixed ethnic)	Healthy	12–16 yo	1. PE Tanner 2. Self-report: Tanner, line draw 3. Questionnaire	Correlations $r = .59$ (genital), $r = .63$ (hair), $r = .18$ (test. vol.) $r = .63$ (breast), $r = .81$ (hair)
Neinstein, 1982	44 (22 girls; 22 boys; mixed ethnic)	Healthy	11–18 yo (70% Tanner 4–5)	1. PE Tanner 2. Self-report: Tanner, Picture	% agreement: 1. breast 87% 2. female pubic hair 86% 3. male genitalia 72% 4. male pubic hair 69% Correlations: $r = .87$ (breast), $r = .69$ (hair) $r = .73$ (genital), $r = .86$ (hair)
Schlossberger et al., 1992	83 (46 boys; 37 girls; race ?)	Healthy	11–14 yo $M = 12.4$–12.7	1. PE Tanner 2. Self-report: Tanner, line draw a) S_1 @ school b) S_2 @ clinic	$\kappa = .35$, $p < .0001$ boys pubic hair @ S_1 $\kappa = .66$, $p < .0001$ @ S_2 male genitalia $\kappa = .06$, $p < .49$ @ S_1 $\kappa = .18$, $p < .04$ @ S_2 breast development $\kappa = .43$, $p < .0001$ @ S_2 Pubic hair female $\kappa = .42$, $p < .0001$ @ S_1 $\kappa = .64$, $p < .0001$ @ S_2
Taylor et al., 2001	103 (62 boys,; 41 girls)	?	12–16 yo	1. PE Tanner by pediatric endo 2. Questionnaire with Tanner line drawings	$\kappa = .68$ pubic hair $\kappa = .48$ breast or genital
Wu et al., 2001	621girls (Caucasian & African American)	Healthy	9–10 yo	1. PE by trained RN (Areola stage) 2. Self-Assessment by Tanner pictures	$\kappa = .42$–.51 by age group, Caucasian $\kappa = .32$– .42 by age group, African American

Note: c = control group; yo = years old; PE = physical exam; CF = cystic fibrosis; PDS = Pubertal Development Scale by Petersen et al., 1988; LD = learning disabled; stg = stage; MD = Medical doctor. All studies appear in Medline or PSYCINFO.

38% of obese girls compared to 25% overestimation in nonobese girls, whereas pubic hair stages were generally the same as that of the practitioner in both groups. However, boys in both groups significantly overestimated pubic hair stage. It is interesting to note that many studies that use self-report measures of puberty provide their rationale by only citing studies showing high correlations and ignoring the studies that report low correlations and low Kappa coefficients. Table 2 provides a summary of studies that have compared agreement self- and parent-report of pubertal status with the actual physical exam ratings by a clinician.

The most widely used method of pubertal staging by self- or parent-report (without pictures or drawings) is the PDS (Petersen et al., 1988). Petersen's original desire for her longitudinal study in the late 1970s was to conduct physical examinations on the adolescents, whereas an alternative was to show photographs or line drawings of the pubertal stages to the participants. Both methods were met by resistance and were in turn rejected by the two school systems involved (personal communication, A. Petersen, 2003). It was at this point she developed the PDS. The PDS includes questions regarding growth in height, body hair, skin changes, breast changes, and additional questions for boys that focus on facial hair and voice changes. Each question has four possible responses. The measure can be completed by the adolescent and the parent. Alpha coefficients ranged from .68 to .83 across a longitudinal study (Petersen et al., 1988). We found only one report examining the validity for the PDS with a physical exam. Brooks-Gunn and colleagues (1987) reported that the correlation of the physical exam with the PDS in 11-, 12-, and 13-year-old girls was between .61 and .67. In addition to Petersen and colleagues (1988), numerous investigators have used the PDS (or a modification) in their research (Dick, Rose, Pulkkinen, & Kaprio, 2001; Ge, Conger, & Elder, 2001; Ge, Conger, & Elder, 2001; Graber et al., 1997; Keel, Klump, Leon, & Fulkerson, 1998; Klump, McGue, & Iacono, 2003; Martin et al., 2002; Robertson et al., 1992).

Issues of measurement for self- and parent-report. One issue in particular that has limited the use of self-report with the photographs or line drawings in school systems is the objection of distributing such pictures to adolescents. When taken out of context or put into the wrong hands, the intent of these pictures can be misconstrued. With society becoming increasingly litigious, it is prudent for school systems and investigators to be vigilant in this matter. However, it is unfortunate that the intention of this research is often misunderstood. Susman, Dorn, and Schiefelbein (2003) articulated that

> in North America, the study of puberty and reproduction is confused with sexual activity and the breaking of religious and social norms. Puberty and emerging sexuality take on negative societal connotations and thereby become shunned by the family and educational institutions. Adolescent sexuality is rarely viewed in the broader perspective as involving a social and cultural component. (p. 315)

Keeping this caveat in mind, it also would be prudent for investigators to think carefully about mailing these materials in a packet of questionnaires. In our studies, we have encountered no difficulty in showing the pictures in our lab to parents and adolescents after a description of normal pubertal growth is given.

The PDS has provided a significant contribution to the literature on adolescent development; however, it also has limitations. After examination of the PDS, it is apparent that many of the changes are events/processes that occur later in puberty (e.g., menarche, facial hair). Thus, this measure is limited in its ability to capture early changes of pubertal development during gonadarche and virtually provides no insight into changes in adrenarche. Further, it combines measures of puberty reflecting the gonadal and adrenal axis, which may or may not be appropriate, depending on the research question. Nevertheless, the PDS is a useful measure for some studies knowing that it reflects self- or parent perception of puberty rather than corresponding directly to actual pubertal stage by physical exam. It is important to note, however, that self-report of puberty may be very appropriate for some research questions. For example, in studies of eating disorders, one's self-perception of puberty may reflect more accurately constructs such as body image or self-esteem. In studies using self-report measures it is helpful to document for the reader that self-perception of pubertal stage or timing is being investigated rather than actual pubertal stage or timing. Further, the limitations of this method should be discussed in articles that utilize self-report. In our review of the literature, the discussion of the limitation of self-perceptions was not always apparent. We emphasize that such studies would be more accurate if self-perceptions are indicated as the methodology. For example, it would be helpful if titles of articles and abstracts say "perceived pubertal stage" or "self-report pubertal stage" rather than "pubertal stage"

Measures of Pubertal Timing

The previous measures of puberty reflect pubertal status or a more static measure of puberty determined at one point in time. Alternatively, the same measures (as well as others) have been used as measures of pubertal timing to reflect either a ranking of pubertal status (i.e., early, on time, or late) or sometimes as a continuum reflecting "earlier" to "later." Thus, timing is generally determined with respect to an anchor. For example, early puberty can be determined with respect to

national norms or with respect to the cohort within the study. In addition, the measure can be determined from a repeated measures strategy (e.g., Tanner stage at multiple time points) or from one time of measurement. Each method has its advantages and disadvantages. The following measures have been reported in the literature to reflect pubertal timing.

1. Bone age. Bone age has been used as a measure of pubertal status, but it has been more frequently used as a measure of pubertal timing. Interestingly, the early studies that examined timing of puberty (H. E. Jones, 1938; M. C. Jones, & Bayley, N., 1950) used bone. To determine bone age, an x-ray of the wrist is taken and then compared to norms using the Gruelich and Pyle (1959) method.

Issues of measurement of bone age. Comparing bone age to chronological age is a useful method to determine timing. For example, a bone age could be read at 8 years, 10 months when the chronological age is 10 years, 2 months. Bone age is then interpreted in standard deviations from the norm. Studies utilizing bone age benefit most when having a radiologist experienced in reading a bone age in children and who can remain throughout the entire study. To our knowledge, since the Jones study in the 1940s, no psychosocial studies of healthy adolescents have used bone age as a measure of timing of puberty. Although the exposure to radiation is minimal, exposing children and adolescents to radiation *without* a clinical indication is generally not favored by Institutional Review Boards (IRBs). Approval has been gained in some instances, particularly when the studies are examining a patient group with a clinical disorder or potential disorder.

2. Age at peak height velocity (PHV). Determining PHV requires longitudinal measurements of growth. A mathematical formula is applied to serial height measurements to obtain PHV (Bock et al., 1973; R. C. Hauspie, 1989; Largo, Gasser, Prader, Stuetzle, & Huber, 1978; Thissen, Bock, Wainer, & Roche, 1976). This requires careful measurement of height in pubertal age adolescents at least annually (preferably every 6 months) due to great variability of growth both within and between ages and gender. Investigators have used PHV in a variety of studies with adolescents (Bailey, Martin, McKay, Whiting, & Mirwald, 2000; Bernaards, Kemper, Twisk, van Mechelen, & Snel, 2001; F. Biro et al., 2001; R. Hauspie, Bielicki, & Koniarek, 1991; Iuliano-Burns, Mirwald, & Bailey, 2001; Mirwald, Baxter-Jones, Bailey, & Beunen, 2002; Petersen & Crockett, 1985; Petersen, Sarigiani, & Kennedy, 1991; Westin-Lindgen, 1982).

3. Categorization of pubertal status measures. Frequently, measures of pubertal status (e.g., breast or genital stage, age at menarche) have been used as a measure of timing by standardizing the scores and then categorizing the scores as early, on time, or late. Age at PHV has also been categorized in this way. For example, pubertal stage scores of the study sample may be compared to published norms of stages by age. Alternatively, scores of the study sample may be "normed" using the within gender study sample as the reference group. Within these distributions categories are formed by trichotomizing the distribution (Petersen & Crockett, 1985) or by defining cutoff points such as ±1 or more standard deviations.

Issues of measurement for pubertal timing as a categorical variable. Selecting the cutoffs for the timing categories can also be difficult. There needs to be a theoretical rationale and an empirical definition of "off time." For example, ±2.5 or 3 *SD* (e.g., for height or bone age) is generally considered *clinically* early or late. In other words, it is considered enough off time to warrant evaluation by a clinician. However, if one trichotomizes the data, this may not truly reflect clinical definitions of off time puberty. Alternatively, if one uses this ±2.5 *SD* cutoff, the cell sizes of the early or late category may be too small to successfully conduct statistical analyses if the overall sample size is small. For example, if a study has 100 participants and one uses a 2.5 *SD* cutoff, only a handful of participants may be in the off-time categories. Several studies have used a 20/60/20 trichotomization for research purposes (Brooks-Gunn & Warren, 1989). Ge, Conger, and Elder (2001) used a 30/40/30 distribution. Clinicians may use more stringent definitions but from a clinical perspective having 2.5% of outliers is relevant. Having 20% to 40% as outliers, however, may be more socially relevant. Examples of issues for measurement of puberty and timing were also addressed by Graber, Petersen, and Brooks-Gunn (1996).

Another critical issue is whether pubertal timing can be categorized and determined from a measurement at one point in time or if pubertal timing needs to be determined from a longitudinal study. It is important to note that an individual could be categorized as an early maturer during an initial visit but then 1 year later he or she could be categorized as on time with respect to peer development. Thus, the timing category may be inconsistent across a longitudinal study.

4. Using pubertal status to reflect pubertal timing as a continuous measure. Several investigations have determined a continuous measure of pubertal timing by using the residuals from regressing pubertal stage on chronological age (Dorn et al., 2003; Ellis & Garber, 2000; Ellis, McFadyen-Ketchum, Dodge, Pettit, & Bates, 1999). To our knowledge, this method has only been implemented in studies using psychological variables rather than physical-related variables (e.g., bone density, height) as the dependent

variable. Many other studies have used other measures of puberty as a continuous variable reflecting pubertal timing (Berzonsky & Lombardo, 1983; Ge, Brody, Conger, Simons, & Murry, 2002; Siegel et al., 1999; Susman et al., 1985; Tobin-Richards, Boxer, & Petersen, 1983).

Issues of measurement. The method of using residuals (cited previously) has its advantages and disadvantages. Advantages include its use with smaller samples, the range is not truncated as with categorical variables, and the available option of concluding that one is earlier rather than early; or later rather than late. However, a main disadvantage to this method is the lack of available norms. This makes it difficult to compare the actual similarities across studies.

5. Self- or parent-report of pubertal timing. The combined score on the PDS has been used to determine pubertal timing by providing some general type of categorization. In addition, the PDS includes a useful question of pubertal timing that asks the adolescent and parent to state if they feel the adolescent's pubertal timing is early, on time, or late with respect to peers. Again, this offers a subjective view of timing and is only one item, but it provides a self-perceived categorization that can be easily used in data analysis. This measure of timing may be very appropriate for some research questions that do not need an objective measure. For example, objective measures of pubertal timing would be more accurate for studies examining brain structure and function changes across puberty or studies examining the effect of pubertal timing on drug metabolism. Alternatively, a self-rating of pubertal timing may be more relevant with variables such as self-esteem or self-competence. A risk with these variables, though, could be low self-esteem causing an adolescent's view of development to be outside the normal range, thus not reflecting the validation against the true stage.

Necessity of Measuring Pubertal Stage in Adolescent Research

Pubertal development is a normal physiological process that may contribute to the expression of various physical and psychological processes encountered during the life span. Eliminating this important physiological process from research questions may lead us to make erroneous conclusions about health or developmental outcomes. In some psychosocial studies, researchers make conclusions about a phenomenon without accounting for puberty. For example, psychosocial studies focusing on temperament, learning and motivation, and aggression may have different findings if one factored pubertal status or timing of puberty into the equation. Alternatively, studies of physiological processes (e.g., drug metabolism, weight gain, bone accrual) may benefit from measures of puberty. Thus, without including appropriate measures of puberty, investigators may not be providing the field of adolescent health and development with the proper information to address health and developmental issues. Without including accurate or appropriate measures of puberty in investigations, researchers neglect potential gains that could be made in knowledge of physical and mental health problems in adolescents or adults. See Susman and colleagues (2003) for impact of puberty on adolescent and adult health.

To date, medical studies have had some success in obtaining some measures of puberty for studies that focus on a health outcome. Measures of pubertal status or timing have been used to quantify risks for disorders. For example, early timing of puberty has been noted as a risk for breast cancer (Apter, Reinila, & Vihko, 1989; Apter & Vihko, 1983; Rockhill, Moorman, & Newman, 1998; R. K. Vihko & Apter, 1986), osteoporosis (Blum et al., 2001; Lysen & Walker, 1997; Van der Sluis & De Muinck Keizer-Schrama, 2001), and testicular cancer (Weir, Kreiger, & Marrett, 1998). These studies have all used retrospective reports of age at menarche or age of male pubertal changes that all have limitations. Alternatively, it is interesting to note that many studies examining psychosocial factors that are conducted in, or with access to, health care settings are reluctant to use stronger objective measures of puberty in their studies. It seems that this is often due primarily to the discomfort of the investigator as well as a result of the reluctance of some IRBs.

Conclusions

Puberty represents a complex suite of physiological processes that has a major impact on physical, psychological, and social development during adolescence. Great interindividual variability in timing, velocity, and relative synchrony of pubertal processes often render age an insufficient descriptor of developmental stage in adolescent populations—particularly in early adolescence. We propose the inclusion of reliable and valid measures of pubertal development in studies as critical for meaningful progress in the field of adolescent health and development. Failing to consider the role of puberty will limit understanding of the physical and psychological changes that occur in adolescent populations and thus diminish our ability to advance the health and development of adolescents. In some cases, the lack of attention to these domains may lead to erroneous conclusions. Researchers must acknowledge the limitations of current methodologies and seek better approaches to addressing these concerns of the field.

The complexity of pubertal processes, the paucity of clear conceptualizations of their role in adolescent

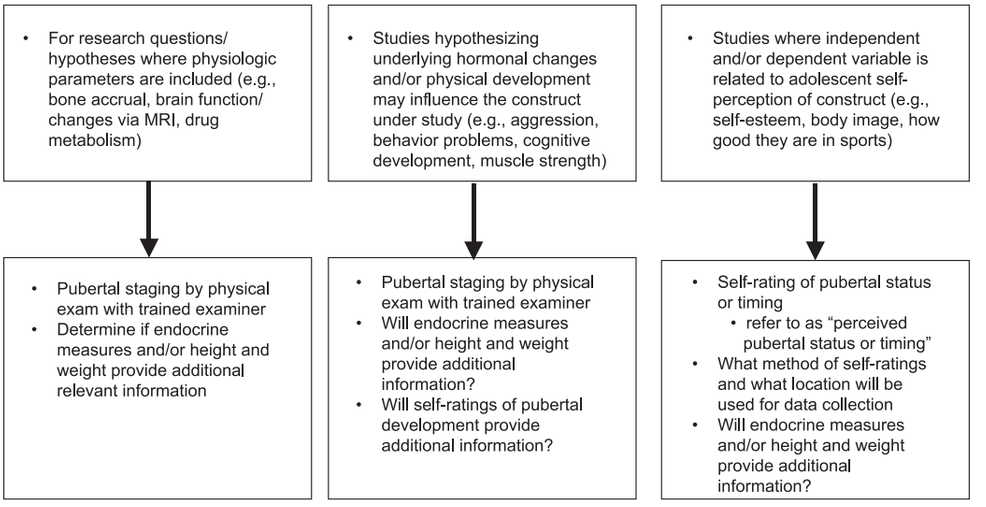

Figure 2. Strategy for determining pubertal measures by research question/hypothesis.

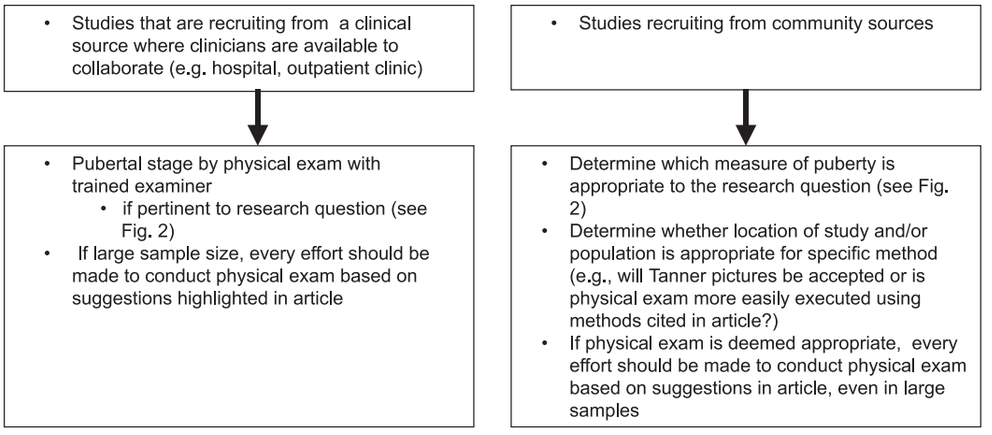

Figure 3. Sources of recruitment and selection of puberty measures.

health and development, and the absence of standards for the use of different methods and puberty-related terminology clearly demonstrate researcher's reluctance with using pubertal status and timing in their studies. However, we believe the role of puberty in adolescent health and development is too central to be omitted. Meaningful progress on key questions cannot be made without considering puberty. The situation we presently face in this regard is similar to that in earlier decades when researchers ignored the inclusion of women, minorities, or children and adolescents due to potential complications previously nonexistent in studies solely conducted with adult White males. Today, such omissions are no longer acceptable and must be justified from an ethical and scientific perspective whenever federal funding is sought. We propose that a similar perspective needs to be adopted when considering puberty in research on adolescent health and development.

The selection of methods to measure pubertal status and timing needs to occur in the context of the overall framework of a specific study and the key questions it seeks to address. In the selection process, the first step is to determine which phase or aspect of puberty may be most relevant to the central questions of the research. Stated otherwise, it could be argued that any of the measures of puberty discussed here are appropriate to use as long as they address the domains relevant to the research questions driving the study. The second step is to define and refine the research question to make sure that it can indeed be addressed by using the measure of choice. The third step may require incorporation of an appropriate collaborator with expertise in the area of puberty.

As we have described earlier, there is no "one-size-fits-all" answer to the question, "What is the best measure of puberty?" However, at this time, pu-

bertal staging by physical exam using Tanner criteria remains the "gold standard" in this field and is recommended for all studies where puberty by physical exam reflects the intent of the question and where the physical exam is feasible. Earlier in this article we described how self-report may be appropriate for some questions. Figure 2 and Figure 3 provide recommendations of strategies for determining what measures of puberty to use in research studies with adolescents.

Further we encourage investigators to provide their rationale for selecting a specific method or methods in assessing pubertal status and timing. When reporting the findings from research on adolescent health and development or in grant applications, clarity would be enhanced if any terms relating to puberty in the paper or grant were clearly defined. If participants are categorized, for example, as being prepubertal or postpubertal, the methodology and criteria of derivation should be included. To further advance developmental science, editors and reviewers for journals that cover adolescent health and development are encouraged to develop reporting standards of pubertal development or pubertal timing for their journals to increase comparability of findings across studies. Likewise, it would be advantageous for peer review committees to adopt similar standards for research applications, determine that appropriate puberty-related collaborators are on the study, and also promote the inclusion of pubertal measures in studies of adolescent health and development. A combined effort to best define and incorporate pubertal status and timing in research will largely account for advancement in this field.

References

Adegoke, A. A. (1993). The experience of spermarche (the age of onset of sperm emission) among selected adolescent boys in Nigeria. *Journal of Youth and Adolescence, 22*(2), 201–209.

Albert, P. S., Hunsberger, S. A., & Biro, F. M. (1997). Modeling repeated measures with monotonic ordinal responses and misclassification, with applications to studying maturation. *Journal of the American Statistical Association, 92*(440), 1304–1311.

American Academy of Pediatrics. (2003). *Clinical practice guidelines.* Elk Grove Village, IL: American Academy of Padiatrics.

Angold, A., Costello, E. J., Erkanli, A., & Worthman, C. M. (1999). Pubertal changes in hormone levels and depression in girls. *Psychological Medicine, 29*(5), 1043–1053.

Angold, A., Costello, E. J., & Worthman, C. M. (1998). Puberty and depression: The roles of age, pubertal status and pubertal timing. *Psychological Medicine, 28*(1), 51–61.

Angold, A., & Rutter, M. (1992). Effects of age and pubertal status on depression in a large clinical sample. *Development and Psychopathology, 4*, 5–28.

Angold, A., & Worthman, C. W. (1993). Puberty onset of gender differences in rates of depression: A developmental, epidemiologic and neuroendocrine perspective. *Journal of Affective Disorders, 29*, 145–158.

Apter, D., Reinila, M., & Vihko, R. (1989). Some endocrine characteristics of early menarche, a risk factor for breast cancer, are preserved into adulthood. *International Journal of Cancer, 44*(5), 783–787.

Apter, D., & Vihko, R. (1983). Early menarche, a risk factor for breast cancer, indicates early onset of ovulatory cycles. *Journal of Clinical Endocrinology and Metabolism, 57*(1), 82–86.

Apter, D., & Vihko, R. (1985a). Hormonal patterns of the first menstrual cycles. In C. Venturoli, C. Flamigni, & J. R. Givens (Eds.), *Adolescence in females* (pp. 215–238). Chicago: Yearbook Medical.

Apter, D., & Vihko, R. (1985b). Premenarcheal endocrine changes in relation to age at menarche. *Clinical Endocrinology, 22*, 753–760.

Aygun, A. D., Akarsu, S., Guvenc, H., & Kocabay, K. (1998). Nipple and areola diameter in Turkish pubertal girls. *Journal of Adolescent Health, 23*, 55–57.

Bailey, D. A., Martin, A. D., McKay, H. A., Whiting, S., & Mirwald, R. L. (2000). Calcium accretion in girls and boys during puberty: A longitudinal analysis. *Journal of Bone & Mineral Research, 15*, 2245–2250.

Bashe, P. G., & Greydanus, D. E. (Ed.). (2003). *American Academy of Pediatrics: The complete and authoritative guide for caring for your teenager.* New York: Bantum.

Bergsten-Brucefors, A. (1976). A note on the accuracy of recalled age at menarche. *Annals of Human Biology, 3,* 71–73.

Bernaards, C. M., Kemper, H. C., Twisk, J. W., van Mechelen, W., & Snel, J. (2001). Smoking behaviour and biological maturation in males and females: A 20-year longitudinal study. Analysis of data from the Amsterdam Growth and Health Longitudinal Study. *Annals of Human Biology, 28*(6), 634–648.

Berzonsky, M. D., & Lombardo, J. P. (1983). Pubertal timing and identity crisis: A preliminary investigation. *Journal of Early Adolescence, 3*, 239–246.

Biro, F., McMahon, R. P., Striegel-Moore, R., Crawford, P. B., Obarzanek, E., Morrison, J., et al. (2001). Impact of timing of pubertal maturation on growth in Black and White female adolescents: The National Heart, Lung, and Blood Institute Growth and Health Study. *The Journal of Pediatrics, 138*(5), 636–643.

Biro, F. M., Falkner, F., Khoury, P. R., Morrison, J. A., & Lucky, A. W. (1992). Areolar and breast staging in adolescent girls. *Adolescent and Pediatric Gynecology, 5*, 271–272.

Biro, F. M., Lucky, A. W., Huster, G. A., & Morrison, J. A. (1990). Hormonal studies and physical maturation in adolescent gynecomastia. *Journal of Pediatrics, 116*(3), 450–455.

Biro, F. M., Lucky, A. W., Huster, G. A., & Morrison, J. A. (1995). Pubertal staging in boys. *The Journal of Pediatrics, 127*(1), 100–102.

Biro, F. M., Lucky, A. W., Simbartl, L. A., Barton, B. A., Daniels, S. R., Striegel-Moore, R. H., et al. (2003). Pubertal maturation in girls and the relationship to anthropometric changes: Pathways through puberty. *Journal of Pediatrics, 142*(6), 643–646.

Blum, M., Harris, S. S., Must, A., Phillips, S. M., Rand, W. M., & Dawson-Hughes, B. (2001). Weight and body mass index at menarche are associated with premenopausal bone mass. *Osteoporosis International, 12*, 588–594.

Boas, S. R., Falsetti, D., Murphy, T. D., & Orenstein, D. M. (1995). Validity of self-assessment of sexual maturation in adolescent male patients with cystic fibrosis. *Journal of Adolescent Health, 17,* 42–45.

Bock, R. D., Wainer, H., Petersen, A., Thissen, D., Murray, J., & Roche, A. F. (1973). A parameterization for individual human growth curves. *Human Biology, 45*, 63–80.

Bonat, S., Pathomvanich, A., Keil, M., Field, A., & Yanovski, J. (2002). Self-assessment of pubertal stage in overweight children. *Pediatrics, 110*(4), 743–747.

Brook, C. G. D., & Hindmarsh, P. C. (Eds.). (2001a). *Clinical pediatric endocrinology* (4th ed). Oxford, England: Blackwell Science.

Brook, C. G. D., & Hindmarsh, P. C. (2001b). Growth assessment: Purpose and interpretation. In C. G. D. Brook & P. C. Hindmarsh (Eds.), *Clinical pediatric endocrinology* (Vol. 4th edition, pp. 115–123). Oxford, England: Blackwell Science.

Brooks-Gunn, J., & Graber, J. A. (1994). Studying links between hormones and negative affect: Models and measures. *Journal of Research on Adolescence, 4*(4), 469–486.

Brooks-Gunn, J., & Warren, M. P. (1989). Biological and social contributions to negative affect in young adolescent girls. *Child Development, 60*, 40–55.

Brooks-Gunn, J., Warren, M. P., Rosso, J., & Garguilo, J. (1987). Validity of self-report measures of girls' pubertal status. *Child Development, 58*(3), 829–841.

Burns, C. E., Brady, M. A., Dunn, A. M., & Starr, N. B. (2000). *Pediatric primary care: A handbook for nurse practitioners* (2nd ed.). Philadelphia: Saunders.

Carlsen, E., Andersen, A., Buchreitz, L., Jorgensen, N., Magnus, O., Matulevicuus, V., et al. (2000). Inter-observer variation in the results of the clinical andrological examination including estimation of testicular size. *International Journal of Andrology, 23*(4), 248–253.

Carskadin, M. A., & Acebo, C. (1993). A self-administered rating scale for pubertal development. *Journal of Adolescent Health, 14*, 190–195.

Casey, V. A., Dwyer, J., Colman, K. A., Krall, E. A., Gardner, J., & Valadian, I. (1991). Accuracy of recall by middle-aged participants in a longitudinal study of their body size and indices of maturation earlier in life. *Annals of Human Biology, 18*, 155–166.

Cauter, E. V. (2001). Endocrine rhythms. In K. L. Becker (Ed.), *Principles and practice of endocrinology and metabolism* (3rd ed.). Philadelphia: Lippincott Williams and Wilkins.

Chipkevitch, E., Nishimura, R. T., Tu, D. G., & Galea-Rojas, M. (1996). Clinical measurement of testicular volume in adolescents: Comparison of the reliability of 5 methods. *Journal of Urology, 156*(6), 2050–2053.

Chumlea, W. C., Schubert, C. M., Roche, A. F., Kulin, H. E., Lee, P. A., Himes, J. H., et al. (2003). Age at menarche and racial comparisons in U.S. girls. *Pediatrics, 111*(1), 110–113.

Cutler, G. B., Jr., Schiebinger, R. J., Albertson, B. D., Cassorla, F. G., Chrousos, G. P., Comite, F., et al. (1990). Adrenarche (human and animal). In M. M. Grumbach, P. C. Sizonenko, & M. L. Aubert (Eds.), *Control of the onset of puberty* (pp. 506–533). Baltimore: Williams & Wilkins.

Dabbs, J. M., Jr., Jurkovic, G. J., & Frady, R. L. (1991). Salivary testosterone and cortisol among late adolescent male offenders. *Journal of Abnormal Child Psychiatry, 19*(4), 469–478.

Dahl, R. E. (2004). Adolescent brain development: Vulnerabilities and opportunities. In R. E. Dahl & L. P. Spear (Eds.), *Annals of the New York Academy of Sciences* (Vol. 1021).New York, NY.

Dahl, R. E., Dorn, L. D., & Ryan, N. D. (1999). Sleep and neuroendocrine issues. In D. Shaffer, J. Richters, & C. P. Lucas (Eds.), *Assessment in child and adolescent psychopathology.* New York: Guilford.

Damon, A., & Bajema, C. J. (1974). Age at menarche: Accuracy of recall after thirty-nine years. *Human Biology, 46*, 381–384.

Daniel, W. A. (1982). Testicular volumes of adolescents. *Journal of Pediatrics, 101*(6), 1010–1012.

Daniel, W. A., & Paulshock, B. Z. (1979, May). A physician's guide to sexual maturity rating. *Patient Care, 13,* 122.

Davison, K. K., & Susman, E. J. (2001). Are hormone levels and cognitive ability related during early adolescence? *International Society for the Study of Behavioral Development, 25*(5), 416–428.

Demir, A., Alfthan, H., Stenman, U., & Voutilainen, R. (1994). Clinically useful method for detecting gonadotropins in children: Assessment of luteinizing hormone and follicle-stimulating hormone from urine as an alternative to serum by ultrasensitive time-resolved immunoflurometric assays. *Pediatric Research, 36*(2), 221–226.

Demir, A., Dunkel, L., Stenman, U., & Voutilainen, R. (1995). Age-related course of urinary gonadotropins in children. *Journal of Clinical Endocrinology and Metabolism, 80*(4), 1457–1460.

Diamond, D. A., Paltiel, H. J., DiCanzio, J., Zurakowski, D., Bauer, S. B., Atala, A., et al. (2000). Comparative assessment of pediatric testicular volume: Orchidometer versus ultrasound. *Journal of Urology, 164*(3 Pt 2), 1111–1114.

Dick, D. M., Rose, R. J., Pulkkinen, L., & Kaprio, J. (2001). Measuring puberty and understanding its impact: A longitudinal study of adolescent twins. *Journal of Youth and Adolescence, 30*(4), 385–399.

Dills, D. G., Allen, C., Palta, M., Zaccaro, D. J., Klein, R., & D'Alessio, D. (1995). Insulin-like growth factor-I is related to glycemic control in children and adolescents with newly diagnosed insulin-dependent diabetes. *Journal of Clinical Endocrinology & Metabolism, 80*(7), 2139–2143.

Dorn, L. D. (5/10/03–4/30/08). *Smoking and Metabolic Complications in Adolescent Girls: Funded by NIDA*. Federal Grant # R01DA016402–01.

Dorn, L. D., Hitt, S., & Rotenstein, D. (1999). Biopsychological and cognitive differences in children with premature vs. on-time adrenarche. *Archives of Pediatric and Adolescent Medicine, 153*(2), 137–146.

Dorn, L. D., Nottelmann, E. D., Susman, E. J., Inoff-Germain, G., Cutler, G. B. J., & Chrousos, G. P. (1999). Variability in hormone concentrations and self-reported menstrual histories in young adolescents: Menarche as an integral part of a developmental process. *Journal of Youth and Adolescence, 28*(3), 283–304.

Dorn, L. D., Ponirakis, A., & Susman, E. J. (2003). Pubertal timing and adolescent adjustment and behavior: Conclusions vary by rater. *Journal of Youth and Adolescence, 32*, 157–167.

Dorn, L. D., & Rotenstein, D. (2004). Early puberty in girls: The case of premature adrenarche. *Women's Health Issues, 14*(6), 177–183.

Dorn, L. D., Susman, E. J., Nottelmann, E. D., Inoff-Germain, G., & Chrousos, G. P. (1990). Perceptions of puberty: Adolescent, parent, and health care personnel. *Developmental Psychology, 26*(2), 322–329.

Downs, A. C., & Fuller, M. J. (1991). Recollections of spermarche: An exploratory investigation. *Current Psychology: Research & Reviews, 10*(1–2), 93–102.

Dubas, J. S., Graber, J. A., & Petersen, A. C. (1991). The effects of pubertal development on achievement during adolescence. *American Journal of Education, 99*, 444–460.

Duke, P. M., Litt, I. F., & Gross, R. T. (1980). Adolescents' self-assessment of sexual maturation. *Pediatrics, 66*(6), 918–920.

Dunaif, A. (1999). Insulin action in the polycystic ovary syndrome. *Endocrinology and Metabolism Clinics of North America, 28*(2), 341–359.

Ellis, B. J., & Garber, J. (2000). Psychosocial antecedents of variation in girls' pubertal timing: Maternal depression, stepfather presence, and marital and family stress. *Child Development, 71*(2), 485–501.

Ellis, B. J., McFadyen-Ketchum, S., Dodge, K. A., Pettit, G. S., & Bates, J. E. (1999). Quality of early family relationships and individual differences in the timing of pubertal maturation in girls: A longitudinal test of an evolutionary model. *Journal of Personality and Social Psychology, 77*(2), 387–401.

Emans, S. J., & Biro, F. (1998). Secondary sexual characteristics and menses in young girls. *Pediatrics, 101*(5), 949–950.

Fewtrell, M. S., Cole, T. J., Bishop, N. J., & Lucas, A. (2000). Neonatal factors predicting childhood height in preterm infants: Evidence for a persisting effect of early metabolic bone disease? *The Journal of Pediatrics, 137*(5), 668–673.

Fitschen, W., & Clayton, B. E. (1965). Urinary excretion of gonadotrophins with particular reference to children. *Archives of Disease in Childhood, 40*, 16–26.

Flug, D., Largo, R. H., & Prader, A. (1984). Menstrual patterns in adolescent Swiss girls: A longitudinal study. *Annals of Human Biology, 11*, 495–508.

Forest, M. G. (1989). Physiological changes in circulating androgens. *Pediatric and Adolescent Endocrinology, 19*, 104–129.

Fuse, H., Takahara, M., Ishii, H., Sumiya, H., & Shimazaki, J. (1990). Measurement of testicular volume by ultrasonography. *International Journal of Andrology, 13*(4), 267–272.

Gaddis, A., & Brooks-Gunn, J. (1985). The male experience of pubertal change. *Journal of Youth and Adolescence, 14*(1), 61–69.

Ge, X., Brody, G. H., Conger, R. D., Simons, R. L., & Murry, V. M. (2002). Contextual amplification of pubertal transition effects on deviant peer affiliation and externalizing behavior among African American children. *Developmental Psychology, 38*(1), 42–54.

Ge, X., Conger, R. D., & Elder, G. H. Jr. (2001). Pubertal transition, stressful life events and the emergence of gender differences in adolescent depressive symptoms. *Developmental Psychology, 37*, 404–417.

Ge, X., Conger, R. D., & Elder, G. H. Jr. (2001). The relation between puberty and psychological distress in adolescent boys. *Journal of Research on Adolescence, 11*, 49–70.

Gell, J. S., Atkins, B., Margraf, L., Mason, J. I., Sasano, H., Rainey, W. E., et al. (1996). Adrenarche is associated with decreased 3β-hydroxysteroid dehydrogenase expression in the adrenal reticularis. *Endocrine Research, 22*(4), 723–728.

Gell, J. S., Carr, B. R., Sasano, H., Atkins, B., Margraf, L., Mason, J. I., et al. (1998). Adrenarche results from development of a 3beta-hydroxysteroid dehydrogenase-deficient adrenal reticularis. *Journal of Clinical Endocrinology and Metabolism, 83*(10), 3695–3701.

Genentech. (1997). *Assessment of Pubertal Stages.* G70577–R1#LF0050, USA.

Girard, J., Baumann, J., & Ruch, W. (1980). Diagnostic possibilities in delayed puberty. In E. Cacciari & A. Prader (Eds.), *Pathophysiology of puberty* (pp. 223–237). New York: Academic.

Graber, J. A., Lewinsohn, P. M., Seeley, J. R., & Brooks-Gunn, J. (1997). Is psychopathology associated with the timing of pubertal development? *Journal of the American Academy of Child and Adolescent Psychiatry, 36*(12), 1768–1776.

Graber, J. A., Petersen, A. C., & Brooks-Gunn, J. (1996). Pubertal processes: Methods, measures, and models. In J. A. Graber, J. Brooks-Gunn, & A. C. Petersen (Eds.), *Transitions through adolescence* (pp. 23–53). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Green, M. (1994). *Bright futures: Guidelines for health supervision of infants, children, and adolescents.* Arlington, VA: National Center for Education in Maternal Child Health.

Gruelich, W. W., & Pyle, S. I. (1959). *Radiographic atlas of skeletal development of the hand and wrist.* Stanford and London: Oxford University Press.

Grumbach, M. M. (2002). The neuroendocrinology of human puberty revisited. *Hormone Research, 57*(Suppl 2), 2–14.

Grumbach, M. M., & Styne, D. M. (1992). Puberty: Ontogeny, neuroendocrinology, physiology, and disorders. In J. D. Wilson & P. W. Foster (Eds.), *Williams textbook of endocrinology* (pp. 1139–1231). Philadelphia: Saunders.

Grumbach, M. M., & Styne, D. M. (1998). Puberty: Ontogeny, neuroendocrinology, physiology and disorders. In J. D. Wilson, D. W. Foster, H. M. Kronenberg, & P. R. Larsen (Eds.), *Williams textbook of endocrinology* (9th ed., pp. 1509–1625). Philadelphia: Saunders.

Grumbach, M. M., & Styne, D. M. (2003). Puberty: Ontogeny, Neuroendocrinology, Physiology, and Disorders. In P. R. Larsen, H. M. Kronenberg, S. Melmed, & K. S. Polonsky (Eds.), *Williams textbook of endocrinology* (10th ed., pp. 1115–1286). New York: Elsevier.

Halpern, C. T., & Udry, J. R. (1992). Variation in adolescent hormone measures and implications for behavioral research. *Journal of Research on Adolescence, 2*(2), 103–122.

Halpern, C. T., Udry, J. R., Campbell, B., & Suchindran, C. (1993). Relationships between aggression and pubertal increases in testosterone: Panel analysis of adolescent males. *Social Biology, 40*(1–2), 8–23.

Hammer, L. D., Wilson, D. M., Litt, I. F., Killen, J. D., Hayward, C., Miner, B., et al. (1991). Impact of pubertal development on body fat distribution among White, Hispanic, and Asian female adolescents. *Journal of Pediatrics, 118*(6), 975–980.

Hardoff, D. & Tamir A. (1993). Self-assessment of pubertal maturation in socially disadvantaged learning-disabled adolescents. *Journal of Adolescent Health, 14*, 398–400.

Hauspie, R., Bielicki, T., & Koniarek, J. (1991). Skeletal maturity at onset of the adolescent growth spurt and at peak velocity for growth in height: A threshold effect? *Annals of Human Biology, 18*(1), 23–29.

Hauspie, R. C. (1989). Mathematical models for the study of individual growth patterns. *Revue d'epidemiologie et de sante publique, 37*, 461–476.

Hayward, C., Killen, J. D., Hammer, L. D., Litt, I. F., Wilson, D. M., Simmonds, et al. (1992). Pubertal stage and panic attack history in sixth- and seventh-grade girls. *American Journal of Psychiatry, 149*(9), 1239–1243.

Hein, K. (1994). Introduction to the special issue on pharmacokinetics and pharmacodynamics in adolescents. *Journal of Adolescent Health, 15*(8), 609–611.

Hergenroeder, A. C., Hill, R. B., Wong, W. W., Sangi-Haghpeykar, H., & Taylor, W. (1999). Validity of self-assessment of pubertal maturation in African American and European American adolescents. *Journal of Adolescent Health, 24*(3), 201–205.

Herman-Giddens, M. E., Kaplowitz, P. B., & Wasserman, R. (2004). Navigating the recent articles on girls' puberty in pediatrics: What do we know and where do we go from here? *Pediatrics, 113*(4), 911–917.

Herman-Giddens, M. E., Slora, E. J., Wasserman, R. C., Bourdony, C. J., Bhapkar, M. V., Koch, G. G., et al. (1997). Secondary sex characteristics and menses in young girls seen in office practice: A study from the pediatric research in office settings network. *Pediatrics, 99*(4), 505–512.

Herman-Giddens, M. E., Wang, L., & Koch, G. (2001). Secondary sexual characteristics in boys. *Archives of Pediatric and Adolescent Medicine, 155*, 1022–1028.

Herter, L. D., Golendziner, E., Flores, J. A. M., Becker, E., Jr., & Spritzer, P. M. (2002). Ovarian and uterine sonography in healthy girls between 1 and 13 years old: Correlation of findings with age and pubertal status. *American Journal of Roentgenology, 178*(6), 1531–1536.

Hick, K. M., & Katzman, D. K. (1999). Self-assessment of sexual maturation in adolescent females with anorexia nervosa. *Journal of Adolescent Health, 24*, 206–211.

Hoekelman, R. A., Nelson, N. M., Weitzman, M. L., & Wilson, M. H. (2001). *Primary pediatric care* (4th ed.). St. Louis, MO: Mosby.

Hofman, A. D., & Greydanus, D. E. (1989). *Adolescent medicine* (2nd ed.). East Norwalk, CT: Appleton & Lang.

Hubert, G. D., & Carson, S. A. (1990). Adrenarche. *Adolescent and Pediatric Gynecology, 3*, 3–14.

Ibáñez, L., Potau, N., Virdis, R., Zampolli, M., Terzi, C., Fussinyé, M., et al. (1993). Postpubertal outcome in girls diagnosed of premature pubarche during childhood: Increased frequency of functional ovarian hyperandrogenism. *Journal of Clinical Endocrinology and Metabolism, 76*(6), 1599–1603.

Ito, M., Yamada, M., Hayashi, K., Ohki, M., Uetani, M., & Nakamura, T. (1995). Relation of early menarche to high bone mineral density. *Calcified Tissue International, 57*(1), 11–14.

Iuliano-Burns, S., Mirwald, R. L., & Bailey, D. A. (2001). Timing and magnitude of peak height velocity and peak tissue velocities for early, average, and late maturing boys and girls. *American Journal of Human Biology, 13*, 1–8.

Ji, C. Y. (2001). Age at spermarche and comparison of growth and performance of pre- and post-spermarcheal Chinese boys. *American Journal of Human Biology, 13*, 35–43.

Jones, H. E. (1938). The California Adolescent Growth Study. *Journal of Educational Research, 31*, 561–567.

Jones, M. C., & Bayley, N. (1950). Physical maturing among boys as related to behavior. *Journal of Educational Psychology, 41*, 129–148.

Kapphahn, C. J., Wilson, K. M., & Klein, J. D. (1999). Adolescent girls' and boys' preferences for provider gender and confidentiality in their health care. *Journal of Adolescent Health, 25*(2), 131–142.

Keel, P. K., Klump, K. L., Leon, G. R., & Fulkerson, J. A. (1998). Disordered eating in adolescent males from a school-based sample. *International Journal of Eating Disorders, 23*(2), 125–132.

Key, T. J. (1999). Serum oestradiol and breast cancer risk. *Endocrine-Related Cancer, 6*, 175–180.

Kirschbaum, C., & Hellhammer, D. H. (1992). Methodological aspects of salivary cortisol measurement. In C. Kirschbaum, G. F. Read, & D. H. Hellhammer (Eds.), *Assessment of hormones and drugs in saliva in biobehavioral research* (pp. 19–32). Seattle, WA: Hogrefe & Huber.

Klein, K. O., Martha, P. M., Jr., Blizzard, R. M., Herbst, T., & Rogol, A. D. (1996). Longitudinal assessment of hormonal and physical alterations during normal puberty in boys II. Estrogen levels as determined by an ultrasensitive bioassay. *Journal of Clinical Endocrinology and Metabolism, 81*(9), 3203–3207.

Klump, K. L., McGue, M., & Iacono, W. G. (2003). Differential heritability of eating attitudes and behaviors in prepubertal versus pubertal twins. *International Journal of Eating Disorders, 33*, 287–292.

Knobil, E. (1988). The hypothalamic gonadotropic hormone releasing hormone (GnRH) pulse generator in the rhesus monkey and its neuroendocrine control. *Human Reproduction, 3*(1), 29–31.

Kulin, H., Frontera, M. A., Demers, L. M., Bartholomew, M. J., & Lloyd, T. A. (1989). The onset of sperm production in pubertal boys. Relationship to gonadotropin excretion. *American Journal of Diseases of Children, 143*(2), 190–193.

Kulin, H. E., Rifkind, A. B., Ross, G. T., & Odell, W. O. (1967). Total gonadotropin activity in the urine of prepubertal children. *Journal of Clinical Endocrinology, 27*, 1123–1128.

Largo, R. H., Gasser, T., Prader, A., Stuetzle, W., & Huber, P. J. (1978). Analysis of the adolescent growth spurt using smoothing spline functions. *Annals of Human Biology, 5*(5), 421–434.

Largo, R. H., & Prader, A. (1983). Pubertal development in Swiss boys. *Helvetica Paediatrica Acta, 38*(3), 211–228.

Livson, N., & McNeill, D. (1962). The accuracy of recalled menarche. *Human Biology, 34*, 218–221.

Lysen, V. C., & Walker, R. (1997). Osteoporosis risk factors in eighth grade students. *Journal of School Health, 67*(8), 317–321.

Marshall, W. A., & Tanner, J. M. (1969). Variations in pattern of pubertal changes in girls. *Archives of Disease in Childhood, 44*, 291–303.

Marshall, W. A., & Tanner, J. (1970). Variations in the pattern of pubertal change in boys. *Archives of Disease in Childhood, 45*, 13–23.

Martin, C. A., Kelly, T. H., Rayens, M. K., Brogli, B. R., Brenzel, A., Smith, W. J., et al. (2002). Sensation seeking, puberty, and nicotine, alcohol, and marijuana use in adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*(12), 1495–1502.

McAnarney, E. R. (1992). *Textbook of adolescent medicine*. Philadelphia: Saunders.

Medhamurthy, R., Gay, V. L., & Plant, T. M. (1990). The prepubertal hiatus in gonadotropin secretion in the male rhesus monkey (Macaca mulatta) does not appear to involve endogenous opiod peptide restrain of hypothalamic gonadotropin-releasing hormone release. *Endocrinology, 126*, 1036–1042.

Midyett, L. K., Moore, W. V., & Jacobson, J. D. (2003). Are pubertal changes in girls before age 8 benign? *Pediatrics, 111*(1), 47–51.

Mirwald, R. L., Baxter-Jones, A. D. G., Bailey, D. A., & Beunen, G. P. (2002). An assessment of maturity from anthropometric measurements. *Medicine & Science in Sports & Exercise, 34*(4), 689–694.

Morris, N. M., & Udry, J. R. (1980). Validation of a self-administered instrument to assess stage of adolescent development. *Journal of Youth and Adolescence, 9*(3), 271–280.

Morrison, J. A., Barton, B., Biro, F. M., Sprecher, D. L., Falkner, F., & Obarzanek, E. (1994). Sexual maturation and obesity in 9- and 10-year-old Black and White girls: The National Heart, Lung, and Blood Institute Growth and Health Study. *The Journal of Pediatrics, 124*(6), 889–895.

Mussen, P. H., & Jones, M. C. (1957). Self-conceptions, motivations, and interpersonal attitudes of late-and early-maturing boys. *Child Development, 28*, 243–256.

Must, A., Phillips, S. M., Naumova, E. N., Blum, M., Harris, S., Dawson-Hughes, B., et al. (2002). Recall of early menstrual history and menarcheal body size: After 30 years, how well do women remember? *American Journal of Epidemiology, 155*(7), 672–679.

National Heart, Lung, and Blood Institute. (1992). Obesity and cardiovascular disease risk factors in Black and White girls: The NHLBI Growth and Health Study. *American Journal of Public Health, 82*(12), 1613–1620.

Neinstein, L. S. (1982). Adolescent self-assessment of sexual maturation. *Clinical Pediatrics, 21*(8), 482–484.

Neinstein, L. S. (1996). *Adolescent health care: A practical guide* (3rd ed.). Baltimore: Williams & Wilkins.

Neinstein, L. S. (2002). *Adolescent health care: A practical guide* (4th ed.). Baltimore: Williams & Wilkins.

Nielsen, C. T., Skakkebæk, N. E., Richardson, D. W., Darling, J. A., Hunter, W. M., Jorgensen, M., et al. (1986). Onset of the release of spermatazoa (spermarche) in boys in relation to age, testicular growth, pubic hair, and height. *Journal of Clinical Endocrinology and Metabolism, 62*(3), 532–535.

Nottelmann, E. D., Susman, E. J., Dorn, L. D., Inoff-Germain, G. E., Loriaux, D. L., Cutler, G. B., Jr., et al. (1987). Developmental processes in early adolescence: Relations among chronological age, pubertal stage, height, weight, and serum levels of gonadotropins, sex steroids, and adrenal androgens. *Journal of Adolescent Health Care, 8*, 246–260.

Nottelmann, E. D., Susman, E. J., Inoff-Germain, G. E., Cutler, G. B., Jr., Loriaux, D. L., & Chrousos, G. P. (1987). Developmental processes in early adolescence: Relations between adolescent adjustment problems and chronologic age, pubertal stage, and puberty-related hormone levels. *Journal of Pediatrics, 110*, 473–480.

Orr, D. P., & Ingersoll, G. (1995). The contribution of level of cognitive complexity and pubertal timing to behavioral risk in young adolescents. *Pediatrics, 95*(4), 528–533.

Palmert, M. R., Hayden, D. L., Mansfield, M. J., Crigler, J. F. J., Crowley, W. F. J., Chandler, D. W., et al. (2001). The longitudinal study of adrenal maturation during gonadal suppression: Evidence that adrenarche is a gradual process. *Journal of Clinical Endocrinology and Metabolism, 86*, 4536–4542.

Papadimitrou, A., Stephanou, N., Papantzimas, K., Glynos, G., & Philippidis, P. (2002). Sexual maturation of Greek boys. *Annals of Human Biology, 29*(1), 105–108.

Parent, A. S., Teilmann, G., Juul, A., Skakkebaek, N. E., Toppari, J., & Bourguignon, J. P. (2003). The timing of normal puberty and the age limits of sexual precocity: Variations around the world, secular trends, and changes after migration. *Endocrine Reviews, 24*(5), 668–693.

Parker, L. N. (1991). Adrenarche. *Endocrinology and Metabolism Clinics of North America, 20*(1), 71–83.

Parker, L. N., Sack, J., Fisher, D. A., & Odell, W. D. (1978). Adrenarche: Prolactin, gonadotropins, adrenal androgens, and cortisol. *Journal of Clinical Endocrinology and Metabolism, 46*(3), 396–401.

Petersen, A. C., & Crockett, L. (1985). Pubertal timing and grade effects on adjustment. *Journal of Youth and Adolescence, 14*(3), 191–206.

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report of pubertal status: Reliability, validity, and initial norms. *Journal of Youth & Adolescence, 17*(2), 117–133.

Petersen, A. C., Sarigiani, P. A., & Kennedy, R. E. (1991). Adolescent depression: Why more girls? *Journal of Youth and Adolescence, 20*(2), 247–271.

Plant, T. (1996). Environmental factors and puberty in non-human primates. *Acta Paediatric Supplement, 417*, 89 - 91.

Plant, T. M. (1986). Gonadal regulation of hypothalamic gonadotropin-releasing hormone release in primates. *Endocrine Reviews, 7*, 75–88.

Plant, T. M. (2002). Neurophysiology of puberty. *Journal of Adolescent Health, 31*, 185–191.

Reiter, E. O. (2001). Have the onset and tempo of puberty changed? *Archives of Pediatric and Adolescent Medicine, 155*, 988–989.

Reiter, E. O., & Rosenfeld, R. G. (2003). Normal and aberrant growth. In P. R. Larsen, H. M. Kronenberg, S. Melmed, & K. S. Polonsky (Eds.), *Williams textbook of endocrinology* (10th ed., pp. 1003–1114). New York: Elsevier.

Remer, T., & Manz, F. (2001). The mid growth spurt in healthy children is not caused by adrenarche. *Journal of Clinical Endocrinology and Metabolism, 86*(9), 4183–4186.

Reynolds, E. L., & Wines, J. V. (1951). Physical changes associated with adolescence in boys. *AMA American Journal of Diseases in Children, 82*(5), 529–547.

Robertson, E. B., Skinner, M. L., Love, M. M., Elder, G. H., Conger, R. D., Dubas, J. S., et al. (1992). The Pubertal Development Scale: A rural and suburban comparison. *Journal of Early Adolescence, 12*(2), 174–186.

Rockhill, B., Moorman, P. G., & Newman, B. (1998). Age at menarche, time to regular cycling, and breast cancer (North Carolina, United States). *Cancer Causes and Control, 9*(4), 447–453.

Rohn, R. D. (1987). Nipple (papilla) development in puberty: Longitudinal observations in girls. *Pediatrics, 79*, 745–747.

Rosenfield, R. L., Bachrach, L. K., Chernausek, S. D., Gertner, J. M., Gottschalk, M., Hardin, D. S., et al. (2000). Current age of onset of puberty. *Pediatrics, 106*(3), 622.

Ryan, N. D., Birmaher, B., Perel., J. M., Dahl, R. E., Meyer, V., Al-Shabbout, M., et al. (1992). Neuroendocrine response to L-5-Hydroxytryptophan challenge in prepubertal major depression. *Archives of General Psychiatry, 49*, 843–851.

Ryan, N. D., Dahl, R. E., Birmaher, B., Williamson, D. E., Iyengar, S., Nelson, B., et al. (1994). Stimulatory tests of growth hormone secretion in prepubertal major depression: Depressed versus normal children. *Journal of the American Academy of Child and Adolescent Psychiatry, 33*(6), 824–833.

Schaefer, F., & Etemadi, N. (1990). Assessment of gonadal maturation by evaluation of spermaturia. *Archives of Disease in Childhood, 65*(11), 1205–1207.

Schlossberger, N. M., Turner, R. A., & Irwin, C. E. (1992). Validity of self-report of pubertal maturation in early adolescents. *Journal of Adolescent Health, 13*(2), 109–113.

Schwartz, E. B., Granger, D. A., Susman, E. J., Gunnar, M. R., & Laird, B. (1998). Assessing salivary cortisol in studies of child development. *Child Development, 69*(6), 1503–1513.

Shirtcliff, E. A., Granger, D. A., Schwartz, E. B., Curran, M. J., Booth, A., & Overman, W. H. (2000). Assessing estradiol in biobehavioral studies using saliva and blood spots: Simple radioimmunoassay protocols, reliability, and comparative validity. *Hormones and Behavior, 38*, 137–147.

Shirtcliff, E. A., Reavis, R., Overman, W. H., & Granger, D. A. (2001). Measurement of gonadal hormones in dried blood spots versus serum: Verification of menstrual cycle phase. *Hormones and Behavior, 39*, 258–266.

Siegel, J. M., Yancey, A., Aneshensel, C. S., & Schuler, R. (1999). Body image, perceived pubertal timing and adolescent mental health. *Journal of Adolescent Health, 25*, 155–165.

Simmons, R., Blyth, A., Van Cleave, E., & Bush, D. (1979). Entry into early adolescence: The impact of school structure, puberty, and early dating on self-esteem. *American Sociological Review, 44*(6), 948–967.

Sizonenko, P., Paunier, L., & Carmignac, D. (1976). Hormonal changes during puberty. IV. Longitudinal study of adrenal androgen secretions. *Hormone Research, 7*(4–5), 288–302.

Sprecher, D. L., Morrison, J. A., Simbartl, L. A., Schreiber, G. B., Sabry, Z. I., Biro, F. M., et al. (1997). Lipoprotein and apolipoprotein differences in Black and White girls. The National Heart, Lung, and Blood Institute Growth and Health Study. *Archives of Pediatric and Adolescent Medicine, 151*(1), 84–90.

Steinberg, L. (1996). *Adolescence* (4th ed.). New York: McGraw-Hill.

Striegel-Moore, R. H., McMahon, R. P., Biro, F. M., Schreiber, G., Crawford, P. B., & Voorhees, C. (2001). Exploring the relationship between timing of menarche and eating disorder symptoms in Black and White adolescent girls. *International Journal of Eating Disorders, 30*, 421–433.

Susman, E. J., Dorn, L. D., & Schiefelbein, V. L. (2003). Puberty, sexuality, and health. In R. M. Lerner, M. A. Easterbrooks, & J. Mistry (Eds.), *Handbook of psychology* (Vol. 6, pp. 295–324). Hoboken, NJ: Wiley.

Susman, E. J., Inoff-Germain, G., Nottelmann, E. D., & Loriaux, D. L. (1987). Hormones, emotional dispositions, and aggressive attributes in young adolescents. *Child Development, 58*, 1114–1134.

Susman, E. J., Nottelmann, E. D., Inoff, G. E., Dorn, L. D., Cutler, G. B., Loriaux, D. L., et al. (1985). The relationship of relative hormone levels and physical development and social-emotional behavior in young adolescents. *Journal of Youth and Adolescence, 14*, 245–264.

Susman, E. J., Nottelmann, E. D., Inoff-Germain, G., & Dorn, L. D. (1987). Hormonal influences on aspects of psychological development during adolescence. *Journal of Adolescent Health Care, 8*(6), 492–504.

Suter, K. J., Pohl, C. R., & Plant, T. M. (1998). The pattern and tempo of the pubertal reaugmentation of open-loop pulsatile gonadotropin-releasing hormone release assessed indirectly in the male Rhesus monkey (Macaca mulatta). *Endocrinology, 139*(6), 2774–2783.

Tanner, J. M. (1962). *Growth at adolescence* (2nd ed.). Oxford, England: Blackwell.

Tanner, J. M., & Whitehouse, R. H. (1976). Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. *Archives of Disease in Childhood, 51*(3), 170–179.

Taylor, S. J., Whincup, P. H., Hindmarsh, P. C., Lampe, F., Odoki, K., & Cook, D. G. (2001). Performance of a new pubertal self-as-

sessment questionnaire: A preliminary study. *Paediatric and Perinatal Epidemiology, 15*(1), 88–94.

Terasawa, E., Bridson, W. E., Nass, T. E., Noonan, J. J., & Dierschke, D. J. (1984). Developmental changes in the luteinizing hormone secretory pattern in peripubertal female rhesus monkeys: Comparison between gonadally intact and ovariectomized animals. *Endocrinology, 115*, 2233–2240.

Thissen, D., Bock, R. D., Wainer, H., & Roche, A. F. (1976). Individual growth in stature: A comparison of four growth studies in the USA. *Annals of Human Biology, 3*, 529–542.

Tobin-Richards, M. H., Boxer, A. M., & Petersen, A. C. (1983). The psychological significance of pubertal change: Sex differences in perceptions of self during early adolescence. In J. Brooks-Gunn & A. C. Petersen (Eds.), *Girls at puberty: Biological and psychosocial perspectives* (pp. 127–154). New York: Plenum.

Tschann, J. M., Adler, N. E., Irwin, C. E., Jr., Millstein, S. G., Turner, R. A., & Kegeles, S. M. (1994). Initiation of substance use in early adolescence: The roles of pubertal timing and emotional distress. *Health Psychology, 13*, 326–333.

Udry, J. R., Billy, J. O. G., Morris, N. M., Groff, T. R., & Raj, M. H. (1985). Serum androgenic hormones motivate sexual behavior in adolescent boys. *Fertility and Sterility, 43*(1), 90–94.

Udry, J. R., & Talbert, L. M. (1988). Sex hormone effects on personality at puberty. *Journal of Personality and Social Psychology, 54*(2), 291–295.

Van der Sluis, I. M., & De Muinck Keizer-Schrama, S. M. P. F. (2001). Osteoporosis in childhood: Bone density of children in health and disease. *Journal of Pediatric Endocrinology & Metabolism, 14*, 817–832.

van Wieringen, J. C., Roede, M. J., & Wit, J. M. (1985). Growth diagrams for patient care. *Tijdschrift voor kindergeneeskunde, 53*(4), 147–152.

Venturoli, S., Porcu, E., Fabbri, R., Magrini, O., Paradisi, R., Pallotti, G., et al. (1987). Postmenarcheal evolution of endocrine pattern and ovarian aspects in adolescents with menstrual irregularities. *Fertility and Sterility, 48*, 78–85.

Venturoli, S., Porcu, E., Fabbri, R., Paradisi, R., Ruggeri, S., Bolelli, G., et al. (1986). Menstrual irregularities in adolescents: Hormonal pattern and ovarian morphology. *Hormone Research, 24*, 269–279.

Vihko, R., & Apter, D. (1984). Endocrine characteristics of adolescent menstrual cycles: Impact of menarche. *Journal of Steroid Biochemistry, 20*(1), 231–236.

Vihko, R. K., & Apter, D. L. (1986). The epidemiology and endocrinology of the menarche in relation to breast cancer. *Cancer Surveys, 5*, 561–571.

Weber, A., Clark , A., Honour, J. L., & Savage, M. O. (1997). Diminished adrenal androgen secretion in familial glucocorticoid deficiency implicates a significant role for ACTH in the induction of adrenarche. *Clinical Endocrinology, 46*, 431–437.

Weir, H. K., Kreiger, N., & Marrett, L. D. (1998). Age at puberty and risk of testicular germ cell cancer. *Cancer Causes and Control, 9*(3), 253–258.

Westin-Lindgen, G. (1982). Achievement and mental ability of physically late and early maturing school children related to their social background. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 23*(4), 407–420.

Wichstrom, L. (2000). Predictors of adolescent suicide attempts: A nationally representative longitudinal study of Norwegian adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(5), 603–610.

Widholm, O., & Kantero, R. L. (1971). A statistical analysis of the menstrual patterns of 8,000 Finnish girls and their mothers. *Acta Obstetricia et Gynecologica Scandinavica, 14*, 1–36.

Wohlwill, J. F. (1973). The age variable in psychological research. In *The study of behavioral development*. New York: Academic.

Worthman, C. M., & Stallings, J. F. (1994). Measurement of gonadotropins in dried blood spots. *Clinical Chemistry, 40*(3), 448–453.

Worthman, C. M., & Stallings, J. F. (1997). Hormone measures in finger-prick blood spot samples: New field methods for reproductive endocrinology. *American Journal of Physical Anthropology, 104*, 1–21.

Wu, T., Mendola, P., & Buck, G. M. (2002). Ethnic differences in the presence of secondary sex characteristics and menarche among U.S. girls: The Third National Health and Nutrition Examination Survey, 1988–1994. *Pediatrics, 110*, 752–757.

Wu, Y., Schreiber, G. B., Klementowicz, V., Biro, F. M., & Wright, D. (2001). Racial differences in accuracy of self-assessment of sexual maturation among young Black and White girls. *Journal of Adolescent Health, 28*, 197–203.

Received May 5, 2004
Reviewed September 6, 2005
Accepted September 7, 2005