

# Examination of US Puberty-Timing Data from 1940 to 1994 for Secular Trends: Panel Findings

Susan Y. Euling, PhD<sup>a</sup>, Marcia E. Herman-Giddens, PA, DrPH<sup>b</sup>, Peter A. Lee, MD, PhD<sup>c</sup>, Sherry G. Selevan, PhD<sup>a</sup>, Anders Juul, MD, DMSc<sup>d</sup>, Thorkild I. A. Sørensen, MD, DMSc<sup>e</sup>, Leo Dunkel, PhD<sup>f</sup>, John H. Himes, PhD, MPH<sup>g</sup>, Grete Teilmann, MD, PhD<sup>d</sup>, Shanna H. Swan, PhD<sup>h</sup>

<sup>a</sup>National Center for Environmental Assessment, Office of Research and Development, US Environmental Protection Agency, Washington, DC; <sup>b</sup>Department of Maternal and Child Health, School of Public Health, University of North Carolina, Chapel Hill, North Carolina; <sup>c</sup>Department of Pediatrics, Penn State College of Medicine, Milton S. Hershey Medical Center, Hershey, Pennsylvania; <sup>d</sup>University Department of Growth and Reproduction, Rigshospitalet, Blegdamsvej, Copenhagen, Denmark; <sup>e</sup>Danish Epidemiology Science Centre, Institute of Preventive Medicine, Copenhagen University Hospital, Copenhagen, Denmark; <sup>f</sup>Division of Endocrinology Diabetes and Metabolic Diseases, Hospital for Children and Adolescents, Biomedicum, University of Helsinki; <sup>g</sup>Division of Epidemiology, University of Minnesota School of Public Health, Minneapolis, Minnesota; <sup>h</sup>Department of Family & Community Medicine, University of Missouri, Columbia, Missouri

The authors have indicated they have no financial relationships relevant to this article to disclose.

## ABSTRACT

Whether children, especially girls, are entering and progressing through puberty earlier today than in the mid-1900s has been debated. Secular trend analysis, based on available data, is limited by data comparability among studies in different populations, in different periods of time, and using different methods. As a result, conclusions from data comparisons have not been consistent. An expert panel was asked to evaluate the weight of evidence for whether the data, collected from 1940 to 1994, are sufficient to suggest or establish a secular trend in the timing of puberty markers in US boys or girls. A majority of the panelists agreed that data are sufficient to suggest a trend toward an earlier breast development onset and menarche in girls but not for other female pubertal markers. A minority of panelists concluded that the current data on girls' puberty timing for any marker are insufficient. Almost all panelists concluded, on the basis of few studies and reliability issues of some male puberty markers, that current data for boys are insufficient to evaluate secular trends in male pubertal development. The panel agreed that altered puberty timing should be considered an adverse effect, although the magnitude of change considered adverse was not assessed. The panel recommended (1) additional analyses of existing puberty-timing data to examine secular trends and trends in the temporal sequence of pubertal events; (2) the development of biomarkers for pubertal timing and methods to discriminate fat versus breast tissue, and (3) establishment of cohorts to examine pubertal markers longitudinally within the same individuals.

**C**HANGING TRENDS in the timing of pubertal development may identify public health concerns. Before the mid-1900s, secular trend analyses of age at female puberty relied on studies of age at menarche among specific populations, and these studies suggest that the average age of menarche declined in the United States between the late 1800s and the mid-1900s.<sup>1,2</sup> This decline was attributed to improvements in general health, nutrition, and other living conditions during this time frame. From the mid-1900s to the present, some studies have reported that the age at menarche and breast and pubic hair development have declined in US girls<sup>3–5</sup>; however, individual study methods and study comparisons have been challenged, and different studies have drawn different conclusions, leading to an unresolved debate regarding a secular trend in female puberty timing during this time frame. Although there are fewer published studies of puberty timing for US boys, some studies have reported an earlier puberty onset in boys as well.<sup>6,7</sup>

An expert panel was asked to resolve the debate by performing a thorough weight-of-the-evidence evaluation of the data during and after “The Role of Environmental Factors on the Onset and Progression of Puberty” Workshop (see introduction article<sup>8</sup> for details). This is the first effort, to our knowledge, to develop consensus on whether there has been a secular trend in the timing of puberty in the

[www.pediatrics.org/cgi/doi/10.1542/peds.2007-1813D](http://www.pediatrics.org/cgi/doi/10.1542/peds.2007-1813D)

doi:10.1542/peds.2007-1813D

The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the US Environmental Protection Agency. Mention of trade names of commercial products does not constitute endorsement or recommendation for use.

Dr Dunkel's current affiliation is Department of Pediatrics, Kuopio University Hospital, Kuopio, Finland. Dr Swan's current affiliation is Department of Obstetrics and Gynecology, University of Rochester School of Medicine and Dentistry, Rochester, New York.

### Key Words

puberty, secular trends, puberty, menarche, Tanner stage, precocious puberty, growth and development

### Abbreviations

Bn—breast development Tanner stage *n*  
PHn—pubic hair development Tanner stage *n*  
PROS—Pediatric Research in Office Settings  
NHANES—National Health and Nutrition Examination Survey  
HPA—hypothalamic-pituitary-adrenal  
Gn—genital development Tanner stage *n*  
HPG—hypothalamic-pituitary-gonadal  
SES—socioeconomic status  
NHES—National Health Examination Survey  
HHANES—Hispanic Health and Nutrition Examination Survey

Accepted for publication Sep 5, 2007

Address correspondence to Susan Y. Euling, PhD, US Environmental Protection Agency, National Center for Environmental Assessment, Office of Research and Development, Mail Code 8623P, US Environmental Protection Agency, 1200 Pennsylvania Ave, NW, Washington, DC 20460. E-mail: [euling.susan@epa.gov](mailto:euling.susan@epa.gov)

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275). Copyright © 2008 by the American Academy of Pediatrics

United States from the mid-1900s to the present. This article presents the panel conclusions on the question, "Are data sufficient to suggest or establish a secular trend in the timing of puberty onset and/or progression (as measured by puberty markers) in boys and girls from US data over the time period 1940 to 1994." This time frame was selected because (1) it is at the center of the debate about a secular trend in puberty timing and (2) many studies that were performed during this interval assessed breast and pubic hair staging as well as menarche.

### PUBERTY-TIMING MARKERS

Human puberty is a complex temporal sequence of biological events leading to the maturation of secondary sex characteristics, accelerated linear growth, and attainment of reproductive capacity. During normal puberty, the sequence of events requires gonadal and adrenal activation. Breast development, pubic hair development, and age at menarche have been most frequently assessed in female puberty-timing studies in humans. The male puberty markers included in this discussion are genital and pubic hair development. Breast, genital, and pubic hair development have been assessed using a progressive scale, usually the Tanner staging. Tanner staging is based on a scale of 5 progressive stages described and depicted in photographs by Marshall and Tanner in which stage 1 is prepubertal, stage 2 is onset, and stage 5 is adult.<sup>9-12</sup> Earlier studies that assessed breast and pubic hair staging used similar definitions.<sup>13,14</sup> They are based on visual observation alone, either from photographs or in person by a trained professional or a child (ie, self-assessment).

The first sign of puberty in girls is either the initiation of breast development, designated as Tanner stage 2 (B2; also called breast budding), or pubic hair development (PH2).<sup>10,15</sup> Their relative order of appearance is subject to interindividual and interracial variability.<sup>5,16</sup> Black girls often begin pubic hair development before or close to the time of breast development onset, whereas white girls typically begin breast development before pubic hair development.<sup>5</sup> B2 can be difficult to assess using visual inspection alone, either in person or from photographs. Discerning breast from fat tissue is a key concern, especially in overweight girls with excessive subcutaneous fat in the chest. Palpation under the areolar area by a trained professional can distinguish breast (glandular) from fat tissue, improving data quality, and is used in clinical diagnosis. The Lee<sup>17</sup> study and a subset of the Pediatric Research in Office Settings (PROS) study were the only U.S. studies to use this technique.<sup>18,19</sup> Comparing data from the US National Health and Nutrition Examination Surveys (NHANES), NHANES III<sup>20</sup> and NHANES 1999–2000, Jolliffe<sup>21</sup> found an increased prevalence and extent of overweight US children aged 2 to 19 years between 1971 and 2000. With an increased percentage of overweight children, the use of palpation is particularly useful in identifying B2. Visual inspection remains the only method to stage breast development progression (Tanner stages 3–5).<sup>15</sup>

PH1 to PH5 in boys and girls have been evaluated by visual inspection by a clinician or self-assessment. Male pubic hair development occurs in response to dihy-

drotosterone, produced by the testes, binding to androgen receptors and activating androgen-dependent gene transcription in genital skin regions. Female pubic hair development is thought to occur in response to androgens produced from the activation of the hypothalamic-pituitary-adrenal (HPA) axis. Racial differences in patterns and progressive stages of pubic hair development increase the variability in pubic hair staging measures that use a scale (Tanner staging) developed for white individuals.

Age at menarche can be determined by asking a girl or her parent her "current status" (also called "status quo"), whether she has had her first menses by the time of assessment and her birth date, and/or by asking postmenarcheal females (or their mothers) to "recall" their age at first menses. Recall may be less valid and decreases in accuracy with greater time elapsed between menarche and asking for the date.<sup>22,23</sup> Height and weight were measured in a number of studies, some longitudinal. In girls, the onset of the pubertal growth spurt usually precedes or is concomitant with the onset of breast development, at least in white girls.<sup>10</sup>

In boys, the first external signs of puberty are an increase in testicular volume above 3 mL<sup>24,25</sup> and Tanner's genital growth stage 2 (G2).<sup>11</sup> The increase in testicular volume is attributed primarily to the growth of testicular seminiferous tubules stimulated by follicle-stimulating hormone. A volume of >3 mL is considered the most reliable and valid male pubertal onset marker developed to date.<sup>26-28</sup> Testicular volume can be measured by an examiner, using an orchidometer (first developed by Prader<sup>29</sup>) or a ruler or by ultrasonography. The most common method is using an orchidometer, either by comparing the testis size by palpation to the orchidometer beads or by using a punched-out (ie, card with holes) orchidometer. Measurements of testicular volume using either orchidometer method are well correlated to those using ultrasound.<sup>30,31</sup> Although testicular volume is often measured in the clinical setting, it has not been used in human studies of puberty timing in the US population. Instead, most studies used visual assessment of genital development using Tanner staging. Stage 2 corresponds to the first visible signs of testicular and scrotal growth and changes in the scrotal skin.<sup>9,11</sup> Additional male pubertal markers include axillary hair development, age at first ejaculation, acne, and voice break, all relatively late male pubertal events. The timing of these markers is subject to higher interindividual variability, and the relationship to the timing of hypothalamic-pituitary-gonadal (HPG) axis and HPA axis activation is not as well understood. Age at spermarche can be determined by the presence of spermatozoa in the first morning urine void (spermaturia) but is not considered a reliable marker because of a high number of false-negative results.<sup>32,33</sup> In boys, the onset of the pubertal growth spurt is a relatively late pubertal event, usually corresponding to the time of late G3 or early G4.<sup>11</sup>

Hormone and receptor measurements would be very useful for determining the timing of the activation of the HPG and HPA axes underlying the physical manifestations of puberty. Endogenous hormone levels and other

biochemical assays to evaluate the physiologic pubertal status are costly for puberty studies with a large participant number, and, although the technology is improving, they have validity issues such as detection limits and interindividual hormone level variability<sup>34</sup> (reviewed by Rockett et al<sup>35</sup>). This discussion focuses on the physically manifested puberty markers because they were assessed in the studies from 1940 to 1994.

### ISSUES IN SECULAR TREND ANALYSIS FOR PUBERTAL TIMING STUDIES

A secular trend is a change in the distribution of an outcome in a population during a specified time frame, usually “a long period of time, generally years or decades.”<sup>36</sup> Secular trend analysis of health measures is frequently used to track public health concerns or initiatives. If an adverse trend is detected, then studies are often performed to understand better the possible causes or determinants, and interventions may result. Secular trend analysis is, by its nature, historic, and projections to the future are inherently uncertain. Confidence in predicting future trends is greater when the past trend is stable or represents a steady change, but even smooth trends can change. Understanding the secular trend and

the trend of its determinants should improve prediction of future trends.

The ability to compare adequately 2 studies, performed at different points in time, is affected by the similarities in the study designs. A number of study design choices, including general study design (eg, longitudinal), population characteristics (eg, race), age(s) of measurement, puberty measures assessed, and method(s) of puberty measurement affect the quality of the data comparisons. Study differences among selected studies for the age of menarche are illustrated in Table 1. Some reports are based on the same study (eg, NHANES III, Bogalusa Heart Study) but with a different subset of participants, age definition, or analytic approach. Although no 2 studies performed at different times can replicate definitions and analyses precisely, some studies are judged to be similar enough to allow for comparison. Comparability is somewhat limited by the methodologic choices of the historic studies. For example, newer studies may have used more reliable methods (eg, palpation for breast development onset) that are less similar to the methods used in an older study (eg, photographs for breast development onset).

Most of the published studies on puberty timing were

**TABLE 1** Examples of Study Design Choices That Can Affect Data Comparability Among Selected US Studies of Menarche

Study	Study Design Type	Questionnaire Data Used in Analysis (Respondent)	Calculation/Analysis	Age Increment Reported	Age Range of Population, y <sup>a</sup>
Fels Institute Study; Reynolds and Wines <sup>13</sup> (1948)	Longitudinal	“Menstrual record” (NR)	Mean	Midpoint between visits (ie, half-year)	8–18
Guidance Study; Nicolson and Hanley <sup>14</sup> (1953)	Longitudinal	NR (NR)	Mean	Midpoint between visits	8–17
NHES I, II, and III; MacMahon <sup>63</sup> (1973), Lee et al <sup>173</sup> (1980)	CS; nationally representative, longitudinal	I: recall (self); II and III: current status and recall (parent); “menstrual history” (self)	I: mean; II and III: median/fitted logistic curve; mean	I: year; II and III: midpoint of the year (added 0.5 years); half-year	I: 18–79; II and III: 11–14; 8.6–17.8
Bogalusa Heart Study; Foster et al <sup>61</sup> (1977)	CS (1 cohort) <sup>b</sup>	Recall (self)	Median	Year	5–14
Pediatric Research in Office Settings (PROS) Network; Herman-Giddens et al <sup>5</sup> (1997)	CS; well-child or problem visit at volunteer physicians office	Current status (self)	Means/probit analysis	Year	3–12
Bogalusa Heart Study; Wattigney et al <sup>93</sup> (1999)	Mixed longitudinal/CS (2 nonoverlapping CS cohorts) <sup>b</sup>	Limited to postmenarcheal girls; recall (self)	Mean and analysis of variance to assess differences between 2 cohorts	Early (8–11 years); intermediate (12–13 years); late (14–17 years)	8–17
Bogalusa Heart Study; Freedman et al <sup>4</sup> (2002)	Mixed longitudinal/CS (7 overlapping CS cohorts) <sup>b</sup>	Current status (self)	Median/logistic regression	Midpoint of year	Third grade (~8 to 17)
NHANES III; Wu et al <sup>47</sup> (2002)	CS; nationally representative	Current status (self); recall (self)	Current status data: mean/probit analysis; recall data: mean/failure time model	Current status: year; postmenarcheal girls: midpoint of year (added 0.5 years)	10–16
NHANES III; Chumlea et al <sup>46</sup> (2003)	CS; nationally representative	Current status (parent for 8- to 9-year-olds; self for 10- to 19-year-olds)	Median/probit analysis	Midpoint of a 3-mo age group	8–19

NR indicates not reported; CS, cross-sectional.

<sup>a</sup> Year ranges described in whole numbers indicate the beginning of the first age to the end of the second age (eg, 3–12 is 3.0 years to 12.99 years of age).

<sup>b</sup> The Bogalusa Heart Study used an overlapping panel design. There were 7 CS examinations of children (or cohorts) in a specific age range, which were timed closely together so that some of the participants were examined multiple times (ie, longitudinally). Freedman et al<sup>4</sup> analyzed data from all 7 CS cohorts; Wattigney et al<sup>93</sup> analyzed data from 2 nonoverlapping CS cohorts; Foster et al<sup>61</sup> analyzed data from 1 CS cohort.

either cross-sectional or longitudinal. In a cross-sectional study, all measurements are made at a single point in time; pubertal development is assessed only once per individual. For Tanner stage (breast, pubic hair, and genital development), current developmental stage and current age are recorded. For menarche, both current status (menstruating: yes or no) and age are obtained, and recalled age at menarche can be asked of those who report menstruating. The age at menarche may have greater measurement error than for Tanner staging for reports that provided menarche in whole year of age. Conversely, determining menarche by the current status method is more reliable compared with Tanner staging measurements, which are somewhat subjective. In a longitudinal study, each boy or girl is examined at several points in time, with age and status recorded at each examination. Both cross-sectional and longitudinal approaches give information about the distribution of age at onset. In a cross-sectional study, onset for any individual is known to have occurred either before or after the current age at the time of assessment. In a longitudinal study, onset is known to have occurred before the first age of assessment, between the ages at 2 different assessments, or after the age at the last assessment.<sup>10</sup> From cross-sectional study data, mean age of onset of successive stages can be estimated; however, this approach, based on single observations of an individual, does not provide the highest quality information on individual variability in the progression through the stages. Longitudinal studies can observe progression through Tanner stages within the same individuals. A longitudinal study provides information on the time each individual spends in intermediate stages, with the quality of the information improving as the number of examinations increases and the time between examinations decreases.

Timing of breast, genital, and pubic hair stages in a population may be described by several characteristics: (1) the average age of onset or attainment of the milestone; (2) the proportion of children who have attained the milestone by a given age; and (3) the average age of all children in a specific stage. Estimates of the first 2 characteristics may be obtained using censored data (from "current status") techniques. The common approach in puberty studies is to smooth the distribution by assuming that it is either normal or logistic (using probit or logistic regression, respectively). Both distributions are determined by 2 parameters, a mean and a variance; estimates that are based on the 2 assumptions will be similar, because the shapes of the normal and logistic distributions differ little.<sup>37</sup> For longitudinal studies, techniques similar to these standard methods can be used to study progression, producing estimates of the distribution of duration of time within a stage.

The age range of the children studied is another important factor. To get the best estimates for age at attainment, the ages at which children are examined should span the range during which most children attain the milestone. If few of the oldest children included have attained the milestone at the end of the study (or, alternatively, most of the youngest included have attained

it), then estimates of mean age of attainment will be less precise. A related problem is the possibility of biased estimates of mean age in a stage in a study with a restricted age range. Children should not be included or excluded in analyses on the basis of attainment of the milestone. For example, in a longitudinal study, elimination of children who have not attained the milestone will bias the mean age at attainment downward. Because children change rapidly in the pubertal period, information improves with more precise age measurements. Creating broad groupings by current age (eg, quarter year, whole years) is less desirable than use of exact age. In a longitudinal study, reporting the age of first observation of a stage as the "age of attainment" overestimates the true age of attainment. Finally, rounding of recalled age at attainment, which can occur if age is reported in whole year only or date is reported as a calendar year only, most likely decreases the accuracy of the estimate. It is important to cope with these problems by applying statistical tools that account for the structure of the data (eg, left truncation, right censoring, interval censoring). Such tools are available (eg, life time analysis or analysis of time to event), and models that incorporate continuous traits that vary over time may be applied. Rather than simply estimating possibly biased observed mean ages, such analyses may provide age distributions and allow for estimation of median or quartiles of ages.

Differences in methods to measure puberty markers can affect findings and, thus, comparability of findings between studies. For breast stage, studies may use palpation and/or observation to determine breast stage in girls. In addition, the rater (eg, self, trained professional) and his or her training as well as the setting where the measurement takes place can affect comparability. Intrarater and interrater variability is important information for comparing across studies. Self-assessment is based on showing drawings or photographs to a child and having him or her mark the stage most like himself or herself.<sup>38,39</sup> The quality of self-assessment compared with that of a trained examiner has been examined in a number of studies in varying populations, using a variety of instruments in different settings (reviewed by Rockett et al<sup>35</sup>). Self-assessment in boys and girls for Tanner staging tended to overestimate stage during the early puberty stages and underestimate in the later stages compared with the staging by a physician.<sup>39</sup> In a study of visual self-assessment by children aged 6 to 12 years, assessment of breast development by girls and pubic hair development by boys was often different from that of the trained examiner, and obese girls were more likely to overestimate breast stage than nonobese girls.<sup>40</sup>

#### **POPULATION VARIABLES THAT AFFECT PUBERTY TIMING**

Studies of higher quality examine a defined group of people with known characteristics (eg, gender, age, socioeconomic status [SES], geographic location) during the time period of interest. Information about genetic and environmental factors that affect puberty timing improves the quality of a study of puberty timing.



Knowledge of genetic factors that could affect puberty timing would be useful. Polymorphisms in genes that regulate puberty timing have been identified, but these genetic markers have not been available until recently.<sup>41–44</sup> Racial and ethnic differences reflect a combination of genetic, social, and environmental factors that are different. Race/ethnicity, defined as racial and/or ethnic subpopulations (eg, Hispanic black, non-Hispanic black), was not often recorded in the historic studies but has been recorded in more recent studies. Analyses of the US PROS and NHANES III data revealed racial differences in the timing of female puberty; black girls were younger than white girls at the same stage of breast development, pubic hair development,<sup>5,45</sup> and menarche.<sup>5,46</sup> Ages for Mexican American girls were between those for the 2 other groups.<sup>45–47</sup> For boys, the onset of genital and pubic hair development occurs earlier in black than white boys.<sup>6,7</sup>

Information on environmental factors including previous nutrition, body fat or size, and race are important in studies of puberty timing. Early severe protein-energy malnutrition presenting as marasmus (severe wasting and growth retardation as a result of insufficient intake) but not kwashiorkor (edematous protein-deficiency malnutrition) delays timing of puberty, at least in girls,<sup>48–50</sup> and chronic malnutrition during childhood through to puberty retards pubertal development in boys and girls.<sup>51</sup> Among chronically undernourished children, age at menarche was delayed by ~2 years among the girls and sexual maturity measures were delayed by ~3 years in boys.

Higher subcutaneous fat levels and BMI at prepubertal ages (5–9 years) are associated with increased likelihood of relatively early (<11 years) menarche.<sup>4</sup> A dose-response association between a higher BMI and increased relative hazards for earlier menarche in girls (8–13 years at baseline) has been observed.<sup>52</sup> The relationship of body fat and puberty timing is reviewed by Kaplowitz.<sup>53</sup>

Some studies found that children who had a low birth weight (defined as <2500 g) or were small for gestational age (typically defined as birth weight <10th percentile for gestational age and gender) are more likely to have advanced timing of puberty (girls and boys<sup>54</sup>; girls only<sup>55</sup>); however, other results are mixed. For example, 1 study found an association between small size at birth and early puberty in girls but not boys,<sup>56</sup> and another found no differences between pubertal timing in children who were born at very low birth weight (<1251 g) and control subjects of both genders.<sup>57</sup> An additional issue is that definitions for small birth size and puberty timing outcomes have not been consistent across studies. One hypothesis is that children who are born small and subsequently go through very rapid catchup growth may start puberty earlier.<sup>58–60</sup>

## EXAMINATION OF US PUBERTY-TIMING DATA (1940–1994) FOR SECULAR TRENDS

### Girls' Data

The earliest published US data on female puberty timing that included breast and pubic hair development as well

as menarche were collected in the 1930s and 1940s (Tables 2 and 3). These 2 small, longitudinal studies included mainly high-SES, white populations.<sup>13,14</sup> Although the 2 studies used different Tanner stage definitions, they reported similar mean ages of breast development onset and menarche, but pubic hair development onset ages were more divergent. Studying a small, largely white population in the early 1970s, Lee<sup>17</sup> reported older ages of breast development onset and menarche than were observed in either the 2 earlier studies or the Bogalusa Heart Study, a study of a similar time period.<sup>61</sup> Analyzing the Bogalusa Heart Study sequential cohort data, Freedman et al<sup>4</sup> found that the mean age of menarche in white girls declined 0.2 years and in black girls declined 0.8 years from 1973–1974 to 1992–1994.

The first pubertal timing data on a representative US sample were from the National Health Examination Survey (NHES) cycles I to III conducted by the Centers for Disease Control (later the Centers for Disease Control and Prevention) from 1963 to 1970 (Tables 2 and 3).<sup>62,63</sup> For NHES cycles II and III, the mean age at menarche was calculated.<sup>63</sup> Tanner stage data were collected only in NHES III, and Harlan et al<sup>62</sup> reported the percentage of children of specific ages who were in each stage. Comparisons with other studies were difficult because mean age at attainment for Tanner stages was not calculated. From 1982 through 1984, pubertal data on menarche, breast development, and pubic hair development were collected in the Hispanic Health and Nutrition Examination Survey (HHANES). Villarreal et al<sup>64</sup> analyzed the Mexican American data only from HHANES and reported mean ages of B2 and PH2. Because neither NHES III nor HHANES collected pubic hair and breast development stage data on children who were younger than 12 and 10 years, respectively, the results are less useful for examining the age of onset (Tanner stage 2) than studies that assessed earlier ages (eg, 37% of 10-year-old Mexican American girls in HHANES had already begun breast development<sup>64</sup>; in NHES III, few participants who were 12–17 years of age were in stages 1 or 2). Advantages of the NHES II and III, HHANES, and NHANES III data for assessing secular trends are the similar sampling strategies, which include oversampling of black and other racial/ethnic groups, and statistical weighting designed to provide a representative US sample. A possible disadvantage is the inclusion of multiple participants from the same household (ie, nonindependent participants who may be genetically related and have similar environmental exposures).

The PROS study was a cross-sectional study on the timing of US female pubic hair and breast development and menarche on a large, racially diverse population of 3- to 12-year-old girls conducted from 1992 to 1993 by Herman-Giddens et al.<sup>5</sup> It is the largest study ( $n = 17\,077$ ) using Tanner staging and menarche, with a much larger sample size than NHES II and III (combined,  $n = 3000$ ) or NHANES III data sets ( $n = 2000–2500$ ; Table 2). The youngest age assessed was 3 years, which is optimal for estimating the onset of breast and pubic hair development; however, the age truncation at 12

**TABLE 2 Description of Puberty Timing Studies for US Girls**

Study	Years Data Collected	Geographic Region; SES	Race/Ethnicity of Participants Included in Analysis	Study Type; Interval Assessed	No. of Participants	Markers Assessed (Method)	Age Range, y <sup>a</sup>
Fels Study; Reynolds and Wines <sup>13</sup> (1948)	1940s	Yellow Springs, OH; mid-upper SES	NR but known to be 100% white	L; 6 mo (birthday and half birthday)	49	Menarche, breast and pubic hair staging <sup>b</sup> (photos)	8–18
Guidance Study; Nicolson and Hanley <sup>14</sup> (1953)	1930s–1940s	Berkeley, CA; high educational level	NR but known to be 100% white	L; 6 mo	95	Menarche, breast and pubic hair staging <sup>c</sup> (photos)	8–17
NHES I, II, and III; MacMahon <sup>65</sup> (1973)	1960–1970	Nationally representative; mixed SES	5635 white; 1043 black; 32 other <sup>d</sup>	CS	I: 3581; II and III <sup>e</sup> : 2242	Menarche	I: 18–79; II and III <sup>e</sup> : 11–14
NHES cycle III; Harlan et al <sup>62</sup> (1980) <sup>f</sup>	1966–1970	Nationally representative; mixed SES	2688 white; 500 black	CS	7514	Breast and pubic hair Tanner staging (visual examination)	12–17
Lee <sup>17</sup> (1980)	1969–1974	Baltimore, MD; upper middle SES	94% white; 6% black and Asian	L; 6 mo	18	Menarche and breast (palpation) and pubic hair Tanner staging (visual examination)	8.6–17.8
Bogalusa Heart Study; Foster et al <sup>61</sup> (1977) <sup>9</sup>	1973–1974	Bogalusa, LA; low-average SES	1059 white <sup>h</sup> ; 621 black	CS	1680	Menarche, breast and pubic hair Tanner staging (visual examination)	5–14
Freedman et al <sup>4</sup> (2002)	1973–1994	Bogalusa, LA; low-average SES	61% white <sup>h</sup> ; 39% black	ML/CS	9158	Menarche	7–16.9
Wattigney et al <sup>93</sup> (1999) <sup>j</sup>	1978–1979; 1992–1994	Bogalusa, LA; low-average SES	78–79: 332 white <sup>h</sup> ; 238 black 92–94: 348 white <sup>h</sup> ; 305 black	ML/CS	78–79: 570 92–94: 653	Menarche	8–17
HHANES; Villarreal et al <sup>64</sup> (1989)	1982–1984	Mexican Americans from the Southwest	100% Mexican American <sup>i</sup>	CS	699	Breast and pubic hair Tanner staging (visual examination)	10–17
PROS; Herman-Giddens et al <sup>5</sup> (1997)	1992–1993	Volunteer physician practices in United States (34 states and Puerto Rico); NR	90.4% white; 9.6% black	CS	17 077	Menarche, breast and pubic hair Tanner staging (visual examination for all; palpation for subset <sup>k</sup> )	3–12
NHANES III; Wu et al <sup>47</sup> (2002)	1988–1994	Nationally representative; mixed SES	466 white; 589 black; 568 Mexican American (Tanner stages); 330 white; 419 black; 419 Mexican American (menarche)	CS	1623 (Tanner stages); 1168 (menarche)	Menarche, pubic and breast development Tanner staging (visual examination)	8–16 (Tanner stages); 10–16 (menarche)

**TABLE 2 Continued**

Study	Years Data Collected	Geographic Region; SES	Race/Ethnicity of Participants Included in Analysis	Study Type; Interval Assessed	No. of Participants	Markers Assessed (Method)	Age Range, y <sup>a</sup>
Sun et al <sup>45</sup> (2002)	1988–1994	Nationally representative; mixed SES	594 white; 788 black; 763 Mexican American <sup>l</sup>	CS	2145	Breast and pubic hair development; Tanner staging (visual examination)	8–18
Chumlea et al <sup>46</sup> (2003)	1988–1994	Nationally representative; mixed SES	710 white; 917 black; 883 Mexican American <sup>l</sup>	CS	2510	Menarche	8–19
NHES II/III; NHANES III; Anderson et al <sup>3</sup> (2003)	1963–1965 (NHES II); 1966–1970 (NHES III); 1988–1994 (NHANES)	Nationally representative; mixed SES	White, black, other; NHES II and III <sup>c</sup> ; 77.8% white; NHANES III: 86.0% white <sup>l</sup>	CS	3272 NHES II/III <sup>m</sup> and 1326 NHANES III	Menarche	10–15
NHES III; HHANES; NHANES III; Sun et al <sup>65</sup> (2005)	1966–1970 (NHES III); 1982–1984 (HHANES); 1988–1994 (NHANES III)	Nationally representative; mixed SES (NHES and HHANES); Hispanic (NHANES)	NHES III: white (83.9%); black; HHANES: 100% Mexican American; NHANES III: white (23.9%); black; Mexican American <sup>l</sup>	CS	3130 NHES III; 712 HHANES; 1218 NHANES III	Breast and pubic hair development; Tanner staging (visual examination)	12–17 (NHES III–NHANES III comparison); 10–17 (HHANES–NHANES III comparison)

L Indicates longitudinal; CS, cross-sectional; ML/CS, mixed longitudinal and cross-sectional components or overlapping panel design; NR, not reported by authors.

<sup>a</sup> Years described in whole numbers indicate the range from the beginning of the first age to the end of the second age (eg, 3–12 is 3.0 years to 12.99 years of age).

<sup>b</sup> Reynolds and Wines<sup>13</sup> used a 5-stage rating system for breast and pubic hair development that had similar definitions to Tanner staging.

<sup>c</sup> Nicolson and Hanley<sup>14</sup> used a 4-stage system for breast development. They defined stage 2 as onset (similar to Tanner stage 2), stage 3 as similar to Tanner stage 2), stage 3 as similar to Tanner stage 2), and stage 4 as similar to Tanner stage 3, and stage 5 as similar to Tanner stage 5. A 5-stage system similar to Tanner staging was used for pubic hair development.

<sup>d</sup> Hispanic white was included in white; Hispanic black was included in black.

<sup>e</sup> Data from NHES cycles II and III were combined, and the age range was limited to 11 to 14 years.

<sup>f</sup> Data from Hairan et al<sup>62</sup> are not included because neither means nor medians were reported.

<sup>g</sup> Foster et al<sup>61</sup> analyzed 1 CS data set.

<sup>h</sup> White was defined as nonblack.

<sup>i</sup> The analysis of Wattigney et al<sup>63</sup> grouped ages as 8 to 11 (“early”), 12 to 13 (“intermediate”), or 14 to 17 (“late”) years of age. Data from this study are not presented in Table 3 because of lack of comparisons with other studies.

<sup>j</sup> Mexican Americans from the Southwest United States from HHANES were included.

<sup>k</sup> Data for breast staging using palpation were reported by Kaplowitz and Oberfield.<sup>18</sup>

<sup>l</sup> Non-Hispanic white and non-Hispanic black.

<sup>m</sup> Anderson et al<sup>3</sup> combined data from NHES cycles II and III and limited the age range to 10 to 15 years.

**TABLE 3** Ages of Pubertal Events From Studies of US Girls

Study	Mean/Median <sup>a</sup> Age of Attainment of a Stage, y (SD, SE, or FL/CL) <sup>b</sup>							Menarche	
	B2	B3	B4	B5	PH2	PH3	PH4		PH5
Fels Study; Reynolds and Wines <sup>13</sup> (1948)	10.8 (1.1)	11.4 (NR)	12.2 (NR)	13.7 (NR)	11.0 (1.1)	11.9 (NR)	12.5 (NR)	13.9 (NR)	12.9 (1.4) white
Guidance Study; Nicolson and Hanley <sup>14</sup> (1953)	10.6 (1.2)	11.2 (1.1)	ND <sup>c</sup>	13.9 (0.9)	11.6 (0.9)	12.5 (1.0)	13.2 (0.9)	NR <sup>d</sup>	12.8 (1.1)
NHES I, II, and III; MacMahon <sup>6</sup> (1973)	ND	ND	ND	ND	ND	ND	ND	ND	12.73 (1.386) white; 12.85 (1.534) black; II and III: 12.80 (0.41) white; 12.52 (0.111) black
NHES II and III; Anderson et al <sup>7</sup> (2003)	ND	ND	ND	ND	ND	ND	ND	ND	12.75 (12.69–12.82) all; 12.80 (12.73–12.87) white; 12.48 (12.28–12.67) black
NHES III									
Sun et al <sup>8</sup> (2005)	ND	11.52 (9.98–12.15) white; NS for black <sup>e</sup>	12.78 (12.49–13.02) white; 11.46 (10.09–12.21) black	15.47 (15.26–15.69) white; 14.29 (13.80–14.71) black	ND	NS for white; NS for black <sup>e</sup>	12.74 (12.54–12.92) white; 11.85 (10.97–12.39) black	14.02 (13.58–14.39) white; 14.02 (13.58–14.39) black	ND
Lee <sup>17</sup> (1980)	11.2 (1.6)	12.4 (1.2)	13.1 (0.7)	14.5 (1.6)	11.9 (1.5)	12.7 (0.5)	13.4 (1.2)	14.6 (1.1)	13.3 (1.3) all
Bogalusa Heart Study; Foster et al <sup>8</sup> (1977)	10.37 (0.105) white; 10.22 (0.125) black	11.60 (0.09) white; 10.91 (0.12) black	NR <sup>g</sup>	NR <sup>g</sup>	10.86 (0.095) white; 10.13 (0.125) black	11.66 (0.09) white; 10.91 (0.12) black	NR <sup>h</sup>	NR <sup>h</sup>	1973–1974; 12.69 (0.085) white; 12.83 (0.11) black
HHANES; Villarreal et al <sup>6a</sup> (1989) <sup>i</sup>	10.95 (10.93–11.31)	12.16 (12.00–12.44)	13.86 (13.57–14.13)	15.10 (14.78–15.36)	11.32 (11.19–11.61)	12.44 (12.25–12.65)	14.07 (13.74–14.26)	15.50 (15.11–15.65)	ND
HHANES; Sun et al <sup>8</sup> (2005)	Mexican American ND	Mexican American 11.98 (11.52–12.4)	Mexican American 13.72 (13.46–13.98)	Mexican American 14.97 (14.70–15.25)	Mexican American ND	Mexican American 12.24 (11.67–12.76)	Mexican American 13.92 (13.68–14.16)	Mexican American 15.27 (15.01–15.54)	ND
Bogalusa Heart Study; Freedman et al <sup>8</sup> (2002)	ND	ND	ND	ND	ND	ND	Mexican American ND	Mexican American ND	1973–1974; 12.7 (12.6–12.9) white; 12.9 (12.7–13.1) black; 1992–1994: 12.5 (12.4–12.8) white; 12.1 (11.9–12.3) black
PROS; Herman-Giddens et al <sup>9</sup> (1997)	9.96 (1.82) white; 8.87 (1.93) black	11.30 (1.42) white; 10.19 (1.42) black	ND	ND	10.51 (1.67) white; 8.78 (2.00) black	11.53 (1.21) white; 10.35 (1.63) black	ND	ND	12.88 (1.20) white; 12.16 (1.21) black
NHANES III; Wu et al <sup>10</sup> (2002)	10.3 (10.0–10.5) white; 9.5 (9.3–9.8) black; 9.7 (9.4–9.9) Mexican American	ND	ND	ND	10.6 (10.4–10.9) white; 9.5 (9.2–9.8) black; 10.3 (10.1–10.6) Mexican American	ND	ND	ND	12.2 (12.0–12.4) black; 12.2 (12.0–12.5) Mexican American <sup>k</sup>
NHANES III; Sun et al <sup>8</sup> (2002) <sup>l,m</sup>	10.38 (10.11–10.65) white; 9.48 (9.14–9.76) black; 9.80 (0–11.78 <sup>n</sup> ) Mexican American	11.75 (11.49–12.02) white; 10.79 (10.50–11.08) black; 11.43 (8.64–14.50) Mexican American	13.29 (12.97–13.61) white; 12.24 (11.87–12.61) black; 13.07 (12.79–13.36) Mexican American	15.47 (15.04–15.94) white; 13.92 (13.57–14.29) black; 14.70 (14.37–15.04) Mexican American	10.57 (10.29–10.85) white; 9.43 (9.05–9.74) black; NS for Mexican American <sup>o</sup>	11.80 (11.54–12.07) white; 10.57 (10.30–10.83) black; 11.70 (11.14–12.27) Mexican American	13.00 (12.71–13.30) white; 11.90 (11.38–12.42) black; 13.19 (12.88–13.52) Mexican American	16.33 (15.86–16.88) white; 14.70 (14.32–15.11) black; 16.30 (15.90–16.76) Mexican American	ND



**TABLE 3 Continued**

Study	Mean/Median <sup>a</sup> Age of Attainment of a Stage, y (SD, SE, or FL/CL) <sup>b</sup>									
	B2	B3	B4	B5	PH2	PH3	PH4	PH5	Menarche	
NHANES III; Chumlea et al <sup>6</sup> (2003)	ND	ND	ND	ND	ND	ND	ND	ND	12.43 (12.33–12.53) all; 12.55 (12.31–12.79) white; 12.06 (11.84–12.28) black; 12.25 (12.04–12.46) Mexican-American	
NHANES III; Anderson et al <sup>7</sup> (2003)	ND	ND	ND	ND	ND	ND	ND	ND	12.54 (12.44–12.64) all; 12.60 (12.48–12.71) white; 12.14 (11.87–12.39) black	
NHANES III; Sun et al <sup>8</sup> (2005)	ND	11.82 (10.49–12.32) white; NS for black <sup>9</sup> ; NS for Mexican American <sup>9</sup>	12.90 (12.18–13.39) white; 11.46 (9.73–12.31) black; 13.04 (12.70–13.37) Mexican American <sup>9</sup>	15.16 (14.61–15.71) white; 13.54 (12.77–14.10) black; 14.74 (14.42–15.09) Mexican American	ND	NS for white <sup>9</sup> ; 11.48 (7.54–12.24) black; 11.65 (11.01–12.15) Mexican American	12.73 (12.08–13.16) white; NS for black <sup>9</sup> ; 13.15 (12.77–13.53) Mexican American	16.13 (15.55–16.92) white; 14.29 (13.62–14.84) black; 16.24 (15.83–16.73) Mexican American	ND	

ND, not determined; NS, not shown in this table; CL, 95% confidence limits; FL, fiducial limits.

<sup>a</sup> Means and medians are considered comparable values because the ages for various pubertal markers are typically normally distributed.

<sup>b</sup> See references for details.

<sup>c</sup> Nicolson and Hanley<sup>14</sup> used a 4-stage breast development scale in which B3 was equivalent to Tanner B3 and B4 was equivalent to Tanner B5. Thus, Tanner B4 was not determined.

<sup>d</sup> Not reported because PH5 was considered by the authors to be unreliable as a result of the interrater agreement score.

<sup>e</sup> Mean age data were restricted to women aged 18 to 34 years in cycle I; women were born 1926 to 1944 and thus were peripubertal ages (10–15 years) ~1936–1956.

<sup>f</sup> Data from cycles II and III were combined, and the analysis was limited to ages 11 to 14 years.

<sup>g</sup> Not shown in this table because the FLs were labeled “unstable” by the authors.

<sup>h</sup> Foster et al<sup>6</sup> only reported median ages for transition into B2, B3, PH2, and PH3 because of the “limited age range,” 5 to 14 years, of their population.

<sup>i</sup> Villarreal et al<sup>6</sup> calculated percentages of girls in a given stage by age, means for being in a stage, means (from probit analysis), and medians (from Spearman-Kärber analysis) for entry into a stage. Means for entry into a stage are presented here for comparison with the other studies.

<sup>j</sup> Freedman et al<sup>7</sup> reported data from 7 cohorts assessed between 1973 and 1994. Because of space limitations, only median ages are shown for the first (1973–1974) and last (1992–1994) cohorts studied.

<sup>k</sup> Wu et al<sup>8</sup> calculated mean menarcheal age separately for current status and recalled age. Mean ages from current status data are shown.

<sup>l</sup> Sun et al<sup>8</sup> calculated means for “age in a stage” and medians for “age at entry” into a stage. Medians for “age at entry” are reported here for comparison with other studies that looked at age of attainment.

<sup>m</sup> FLs were computed at 98.3% CLs to adjust for multiple comparisons between racial/ethnic groups.

<sup>n</sup> The left side FL of 0 is likely attributable to the small sample size for Mexican Americans.

<sup>o</sup> Not shown in this table because the authors did not report the FL because of the small sample size for Mexican Americans.

years excludes observation of puberty completion (Tanner stage 5 and menarche) for many girls. For example, 64.8% of the white girls and 37.9% of the black girls had not achieved menarche by the end of their 12th year (up to 12.99 years of age). The PROS study group is a convenience sample and not representative of a larger population; however, all of the earlier study populations (1930–1970), including Marshall and Tanner (1969),<sup>10</sup> on which US clinical standards for puberty timing were derived, were also convenience samples. The participants included girls who were examined during pediatric visits for either a (1) well-child visit at offices that assessed Tanner staging and menarcheal status routinely in the well-child visit (95.74%) or (2) a problem visit that required a complete examination (4.26%). The study by Herman-Giddens et al<sup>5</sup> reported younger median ages than any preceding study (Table 3). For example, the median B2 reported was ~0.6 to 1.2 years earlier and the median PH2 was ~0.5 to 1.4 years earlier than the values from Reynolds and Wines,<sup>13</sup> Nicolson and Hanley,<sup>14</sup> or Lee.<sup>17</sup>

Several articles presented analyses of puberty-timing data of children in 3 racial/ethnic groups (non-Hispanic white, non-Hispanic black, and Mexican American) from NHANES III, conducted between 1988 and 1994 (Table 3).<sup>45–47,65</sup> Comparing ages of attainment of pubertal stages between PROS and NHANES III, conducted during similar time intervals (1980–1990s), the median ages from NHANES III are slightly higher than values from PROS (Table 3); however, the different study designs complicate any comparison (Tables 1 and 2) and may explain the differences in findings. An alternative interpretation is that the NHANES III and PROS findings corroborate one another because both reported an earlier age of onset of breast and pubic hair development compared with the ages from data collected in the 1930s and 1940s. Additional comparisons of the NHANES III with data from the 1960s and 1970s were conducted. Comparing NHANES III (1988–1994) and NHES II/III data,<sup>63</sup> Chumlea et al<sup>46</sup> reported a decrease of 4 months in the median age of menarche (combining racial/ethnic groups) between 1963–1970 and 1988–1994; however, the authors considered the change too small to be meaningful, but statistical results were not presented. In another analysis that compared the NHANES III with reanalyzed NHES II/III data, Anderson et al<sup>3</sup> calculated a decrease of 2.5 months in the median age of menarche that the authors interpreted as a meaningful change; nonoverlapping confidence intervals for the median ages were for whites only, suggesting statistical significance. Furthermore, Herman-Giddens et al<sup>18</sup> argued that there has been a statistically significant decline in the median age of menarche from 1963–1970 to 1988–1994 because confidence intervals do not overlap for most calculations from NHANES III and NHES II/III data.

To assess whether Tanner stage age attainment has changed between the late 1960s and early 1990s, Sun et al<sup>65</sup> performed a reanalysis of NHES III, HHANES, and NHANES III data (reviewed at the workshop in unpublished form). For making better comparisons among the

data sets, the analysis was limited to stages 3 to 5 (because NHES III included children 12 to 17, considered too old to estimate stage 2 reliably) and Mexican American children from HHANES. The authors concluded that there is no evidence of an earlier puberty (as measured by median ages of Tanner stages 3–5) during the time spanning the 3 surveys (1960s through 1990s) for either non-Hispanic black or white girls but “some evidence” for Mexican American girls between 1982 and 1994. An advantage of this study is the consistent reanalysis of data from 3 national surveys. Disadvantages of the analysis are the exclusion of information on onset (Tanner stage 2), which many consider to be the key puberty-timing issue of concern, and decreased study power as a result of limiting the sample sizes for comparisons.

A subset of the female puberty-timing studies could assess whether the duration of the puberty, defined as the interval from onset (as measured by B2 or PH2, whichever occurs first) to completion (as measured by menarche). Because the earlier studies<sup>13,14,17</sup> (Table 3) all calculated approximately a 2.1- to 2.2-year interval for largely white populations, the duration for white girls only could be compared. Foster et al<sup>61</sup> and Wu et al<sup>47</sup> observed an increase of ~2.3 years in the duration of puberty for white girls; however, Herman-Giddens et al<sup>5</sup> observed a longer duration of ~3.1 year for white girls. One interpretation that explains the reported increased duration is an earlier puberty onset with a relatively unchanged or less-changed age at menarche. When comparing studies that allowed for racial comparisons, black girls were found to have a longer duration than white girls.<sup>5,47,61</sup>

### Boys' Data

Far fewer data are available on the ages and patterns of pubertal development in US boys than girls. This difference may indicate a lower interest in studying boys' puberty that may reflect less cultural awareness of male pubertal development compared with girls (eg, breast development, menstruation). In addition, the greater number of published studies assessing girls' puberty timing may reflect the concern about the relationship between early female puberty and later disease, which has been less investigated for boys<sup>66,67</sup> (reviewed by Golub et al<sup>68</sup>).

All of the US analyses of male puberty timing have assessed Tanner stages for genital and pubic hair development staging (Table 4), albeit with study design differences. Testicular volume was measured in only 1 US study, but Tanner genital development was not also assessed.<sup>25</sup> The history of US boys' studies is similar to the girls' because most of the US puberty studies collected data on both boys and girls (eg, Fels, Guidance, NHES III, HHANES, NHANES, Bogalusa). For example, the earliest data on the age of puberty in US boys were from small, mostly high-SES white populations.<sup>14,17,69</sup> The first boys' data representative of the US population were collected from 1963 to 1970 in NHES III.<sup>70</sup> The earliest ages included in NHES III or HHANES data are not optimal for estimating the mean ages of the onset (eg, 14.5% of 10-year-old Mexican American boys in HHANES already had begun genital development) com-

**TABLE 4 Description of Puberty Timing Studies for US Boys**

Study	Years Data Collected	Geographic Region; SES	Race/Ethnicity of Participants Included in Analysis	Study Type; Interval Assessed	N	Markers Assessed (Method)	Age Range, y <sup>a</sup>
Fels Study; Reynolds and Wines <sup>69</sup> (1951) <sup>b</sup>	1940s	Yellow Springs, OH; mid-upper SES	NR but known to be 100% white	L; 6 mo (birthday and half birthday)	59	Pubic hair and genital development	9–21
Guidance Study; Nicolson and Hanley <sup>70</sup> (1953) <sup>c</sup>	1930s–1940s	Berkeley, CA; high educational level	NR but known to be 100% white	L; 6 mo	92 <sup>d</sup>	Tanner staging <sup>b</sup> (photos) Sex stage <sup>d</sup> (photos)	8–19
NHES III; Harlan et al <sup>70</sup> (1979) <sup>e</sup>	1966–1970	Nationally representative; mixed SES	499 black and 3047 white	CS	3546	Pubic hair and genital development Tanner staging (visual examination)	12–17
Lee <sup>77</sup> (1980)	1969–1974	Baltimore, MD; upper middle SES	94% white; 6% black and Asian	L	36	Pubic hair and genital development Tanner staging (visual examination)	9.2–17.5
Bogalusa Heart Study; Foster et al <sup>61</sup> (1977) <sup>f</sup>	1973–1974	low-average SES	1153 white; 676 black <sup>g</sup>	CS <sup>5</sup>	1829	Pubic hair and genital development Tanner staging (visual examination)	5–14
HHANES; Villarreal et al <sup>64</sup> (1989)	1982–84	Mexican Americans of the Southwest	100% Mexican American <sup>h</sup>	CS	704	Pubic hair and genital development Tanner staging (visual examination)	10–17
NHANES III Sun et al <sup>65</sup> (2002)	1988–1994	Nationally representative; mixed SES	537 white; 798 black; 783 Mexican American <sup>i</sup>	CS	2118	Pubic hair and genital development Tanner staging (visual examination)	8–18
Biro et al <sup>25</sup> (1995)	NR	Greater Cincinnati, OH area; NR	~50% white; ~50% black (precise value NR)	L	515	Pubic hair and genital development Tanner staging (visual examination) and testicular volume <sup>l</sup>	10–15
NHANES III; Herman-Giddens et al <sup>6</sup> (2001)	1988–1994	Nationally representative; mixed SES	536 white; 797 black; 781 Mexican American <sup>i</sup>	CS	2114	Pubic hair and genital development Tanner staging (visual examination)	8–18
NHES III; and NHANES III; Karpati et al <sup>7</sup> (2002)	NHES II: 1963–1965; NHES III: 1966–1970; NHANES III: 1988–1994	Nationally representative; mixed SES	NHES III: white, black, other <sup>9</sup> ; NHANES III: white, black, Mexican American, other <sup>1</sup>	CS	3010 (NHES II); 3514 (NHES III); 2481 (NHANES III)	Pubic hair and genital development Tanner staging (visual examination)	NHES II: 8–11; NHES III: 12–17; NHANES III: 8–18
NHES III; HHANES III; Sun et al <sup>65</sup> (2005)	1966–1970 (NHES III); 1982–1984 (HHANES); 1988–1994 (NHANES III)	Nationally representative; mixed SES (NHES and HHANES); Hispanic (HHANES)	NHES III: white (86.4%); black <sup>7</sup> ; HHANES: 100% Mexican American; NHANES III: white (20.8%), black, Mexican American <sup>i</sup>	CS	3520 NHES III; 717 HHANES; 1246 NHANES III	Breast and pubic hair development Tanner staging (visual examination)	12–17 (NHES III–NHANES III comparison); 10–17 (HHANES–NHANES III comparison)

<sup>a</sup> Years, described in whole numbers, indicate the range from the beginning of the first age to the end of the second age (eg, 3–12 is 3.0 years to 12.99 years of age).

<sup>b</sup> For pubic hair development, Reynolds and Wines<sup>69</sup> used a 5-stage rating scale that had similar definitions to Tanner staging. For genital development, they used 2 separate scales, a 5-stage genital development scale that had some differences from Tanner GD stages, and a 7-stage penis growth rating scale. G2 was defined as "enlargement of scrotum, first reddening and texture change." G3 was defined as "first sculpturing" and enlargement of penis. G4 was defined as "pronounced sculpturing and darkening." G5 was defined as "essentially adult; reddish brown color, loose penile skin, loss of sharp sculpturing."

<sup>c</sup> Nicolson and Hanley<sup>70</sup> used a 5-stage "sex stage" system of Greulich et al<sup>68</sup> that combined assessment of pubic hair and genital development.

<sup>d</sup> The number depends on the stage or outcome assessed and varied between 80 and 92 (eg, n = 92 for sex stage II).

<sup>e</sup> Harlan et al<sup>70</sup> data are not included because neither means nor medians were reported.

<sup>f</sup> The Bogalusa Heart Study used an overlapping panel design of multiple cohorts. Foster et al<sup>61</sup> analyzed data from 1 CS cohort.

<sup>g</sup> Hispanic white was included in white; Hispanic black was included in black.

<sup>h</sup> Mexican Americans from the Southwest United States from HHANES were included.

<sup>i</sup> White is non-Hispanic white; black is non-Hispanic black.

<sup>l</sup> Biro et al<sup>25</sup> assessed testicular volume using a Prader orchidometer.

pared with the earliest ages assessed in the Bogalusa and NHANES III studies.

Comparison of the timing of genital development for white boys between the studies conducted in the 1940s to 1970s<sup>14,17,61,69</sup> and the 1980s to 1990s<sup>6,7,45</sup> suggests that the mean or median age of G2 declined by 1 to 2 years (Table 5). For the age of G3, the difference is less dramatic but is also earlier in 1988–1994. Comparing NHES III and NHANES III data, Sun et al<sup>65</sup> calculated a decrease in the median ages for G3 and G4 for non-Hispanic white boys but not for non-Hispanic black boys during this time span. Comparing HHANES and NHANES III data for Mexican Americans, Sun et al<sup>65</sup> reported a decrease in the ages of G3 and G4 as well. For the age of G4, differences are not consistently found when studies are compared (eg, G4 is much later according to Lee<sup>17</sup> than to Reynolds and Wines<sup>69</sup>). The age of G5 varies greatly, even among studies conducted within the same decade, calling rater or marker reliability into question (eg, G5 is earlier or later when comparing the 1940s studies with various 1970–1990 studies).

Comparison of the age of PH2 onset for white boys between the earlier studies<sup>17,61,69</sup> and NHANES III (1988–1994)<sup>6,7,45</sup> suggests that the mean or median age for pubic hair development onset has declined by ~0.2 to 0.5 years; however, Foster et al<sup>61</sup> reported a higher PH2 mean age for white boys than previous or subsequent studies (from the 1940s to 1950s and the 1988–1994 NHANES III; Table 5), suggesting that pubic hair staging is not a reliable marker. Sun et al<sup>65</sup> reported that PH3 and PH4 were earlier for white boys in NHANES III compared with NHES III. Mean ages for PH4 and PH5 exhibit variability among studies, even among those performed in the same decade. Looking at the percentages of 12-year-old white boys in PH2 or higher between NHES III and NHANES III data, similar percentages (73.5% for NHES III; 71.5% for NHANES III) were reported, suggesting no change in the timing of progression of pubic hair development overall between 1966–1970 and 1988–1994.

For white boys' data, comparisons of the duration of male puberty are limited to the studies of Reynolds and Wines,<sup>69</sup> Lee,<sup>17</sup> and NHANES III, from which the pubertal interval for white boys can be calculated as ~4.6 years (1940s), ~3.4 years (1960s/1970s), and ~5.5 to 5.6 years (1980s/990s). Excluding the study by Lee, which found a shorter pubertal interval than the earlier study (perhaps reflecting pubertal measure variability issues), the other studies suggest that the pubertal interval has increased between the 1940s and the 1990s. Given the caveats of low reliability of male puberty markers and the small number of studies, the reported increased pubertal duration may reflect an earlier age of onset (stage 2) with no significant change in the age of completion (stage 5), as with the girls' study findings.

## DISCUSSION

### Panel Consensus: Identified Areas of Agreement and Disagreement

The panel agreed that these data indicate racial/ethnic differences in the timing and order of pubertal events for

boys and girls and recommended that studies present results by race/ethnic group. Racial/ethnic distinctions may reflect genetic or environmental factors or a combination of both. For example, an individual identified as "non-Hispanic black" in NHANES III might have more in common, genetically and environmentally, with others identified differently. Even with the problems of defining race/ethnicity, the panel preferred and recommended that studies present results by race/ethnic group. Stratification by body size, particularly body fat, was also considered appropriate because body fat and growth have a profound impact on puberty timing.

All panel members questioned whether studies spanning the time frame 1940–1994 were similar enough to compare. Many of the existing studies differ in study size, participant characteristics (eg, race/ethnicity, SES), study group selection method, study type (eg, longitudinal), marker assessment methods, and statistical techniques.

### Girls' Data

#### Breast Development

##### Onset (B2)

The majority of panelists concluded that there are sufficient data to suggest (most) or establish (fewer) a secular trend toward an earlier age at onset of breast development on the basis of B2 as a fairly reliable marker of breast development onset and the opinion that an early onset of breast development, even if isolated, was a public health concern (Table 6). B2 is considered a more reliable external puberty-timing marker than pubic hair development because breast development requires estrogen action from the HPG axis. The basis for this opinion was a weight-of-evidence approach from comparing either the recent PROS or NHANES III data for white girls with the earlier studies that assessed B2<sup>13,14,17</sup> (Tables 2 and 3). Additional corroboration comes from comparing PROS and Bogalusa Heart Study B2 ages for black and white girls separately.<sup>61</sup> Comparisons of B2 ages for Hispanic girls were not possible because earlier studies before HHANES did not report data for Hispanic girls and the comparison by Sun et al<sup>65</sup> of HHANES and NHANES III did not report values for B2; however, some panelists thought that the data for Mexican American girls from HHANES<sup>64</sup> could be reliably compared with those from NHANES III.

A minority of panelists concluded that there are insufficient data to suggest a secular trend on the basis of the limited quantity and quality of the data, particularly the lack of studies that assessed girls young enough to determine reliably the timing of puberty onset markers, such as B2. As mentioned, girls 10 or 12 years and older were assessed in studies before the 1970s. There was concern about the lack of palpation to determine B2 as well as the difference in Tanner stage assessment method (photographs versus visual examination) across studies. The only study to use palpation in all breast assessments was the small study by Lee et al<sup>17</sup>; however, the PROS study performed palpation and visual assess-

**TABLE 5** Ages of Pubertal Events From Studies of US Boys

Study	Mean/Median <sup>a</sup> Age of Attainment of a Stage, y (SD, SE, or 95% CI) <sup>b</sup>							
	G2	G3	G4	G5	PH2	PH3	PH4	PH5
Reynolds and Wines <sup>69</sup> (1951)	11.5 (0.9)	NS <sup>c</sup>	13.4 (0.7)	17.3 (0.8)	12.2 (1.1)	13.3 (0.8)	13.9 (0.7)	16.1 (1.2)
Nicolson and Hanley <sup>14</sup> (1953) <sup>d</sup>	11.8 (1.0)	13.1 (1.0)	13.8 (1.0)	15.2 (1.0)	ND	ND	ND	ND
NIHES III; Karpatti et al <sup>7</sup> (2002)	NR	13.0 (13.0–13.1)	13.9 (13.8–14.0)	15.1 (14.8–15.3)	NR	13.3 (13.2–13.4)	14.0 (13.9–14.1)	15.3 (15.1–15.5)
NIHES III; Sun et al <sup>65</sup> (2005)	ND	13.05 (12.91–13.17) white; 12.74 (12.38–13.02) black	13.88 (13.72–14.04) white; 13.72 (13.47–13.94) black	15.07 (14.96–15.18) white; 14.76 (14.42–15.11) black	ND	13.29 (13.20–13.38) white; 13.28 (13.01–13.51) black	14.04 (13.96–14.11) white; 14.03 (13.80–14.24) black	15.39 (15.30–15.49) white; 15.20 (14.95–15.46) black
Lee <sup>7</sup> (1980)	11.9 (1.1)	13.2 (0.8)	14.3 (0.8)	15.1 (1.1)	12.3 (0.8)	13.9 (0.9)	14.7 (0.9)	15.3 (0.8)
Bogalusa; Foster et al <sup>61</sup> (1977) <sup>e</sup>	11.82 (0.11) white; 11.16 (0.185) black	NR	NR	NR	12.52 (0.085) white; 11.72 (0.125) black	NR	NR	NR
HHANES; Villarreal et al <sup>64</sup> (1989) <sup>f</sup>	12.39 (12.14–12.58)	13.50 (13.27–13.73)	14.59 (14.35–14.89) Mexican American	16.26 (15.94–16.54) Mexican American	12.81 (12.43–12.83) Mexican American	13.62 (13.36–13.84) Mexican American	14.64 (14.32–14.94) Mexican American	16.06 (15.71–16.33) Mexican American
HHANES; Sun et al <sup>65</sup> (2005)	ND	13.29 (12.90–13.70)	14.38 (14.19–14.59) Mexican American	16.08 (15.76–16.46) Mexican American	ND	13.41 (13.10–13.72) Mexican American	14.42 (14.17–14.69) Mexican American	15.86 (15.61–16.14) Mexican American
Biro et al <sup>25</sup> (1995) <sup>g</sup>	ND	ND	ND	ND	12.79 (NR)	13.74 (NR)	14.63 (NR)	15.19 (NR)
NHANES III; Herman-Giddens et al <sup>6</sup> (2001)	10.1 (9.6–10.6) white; 9.5 (8.9–10.0) black; 10.4 (9.6–11.1) Mexican American	12.4 (12.0–12.7) white; 11.8 (11.3–12.3) black; 12.5 (12.2–12.8) Mexican American	13.5 (13.2–13.8) white; 13.4 (13.1–13.6) black; 13.7 (13.4–14.1) Mexican American	15.9 (15.3–16.4) white; 14.9 (14.4–15.5) black; 15.7 (15.3–16.2) Mexican American	12.0 (11.7–12.3) white; 11.2 (10.9–11.4) black; 12.3 (12.1–12.5) Mexican American	12.6 (12.3–13.0) white; 12.5 (12.3–12.8) black; 13.1 (12.9–13.3) Mexican American	13.5 (13.2–13.8) white; 13.7 (13.5–13.9) black; 14.1 (13.8–14.4) Mexican American	15.7 (15.3–16.0) white; 15.4 (14.9–15.9) black; 15.8 (15.5–16.2) Mexican American
NHANES III; Karpatti et al <sup>7</sup> (2002)	10.1 (9.6–10.6) white; 9.3 (8.7–9.9) black; 10.4 (9.6–11.2) Mexican American	12.4 (12.1–12.7) white; 11.8 (11.3–12.3) black; 12.5 (12.2–12.8) Mexican American	13.5 (13.2–13.8) white; 13.4 (13.2–13.7) black; 13.8 (13.4–14.1) Mexican American	16.0 (15.4–16.6) white; 14.9 (14.3–15.5) black; 15.8 (15.4–16.3) Mexican American	12.0 (11.7–12.3) white; 11.1 (10.8–11.5) black; 12.3 (12.1–12.6) Mexican American	12.7 (12.3–13.0) white; 12.6 (12.3–12.8) black; 13.1 (12.9–13.3) Mexican American	13.6 (13.3–13.9) white; 13.7 (13.6–13.9) black; 14.0 (13.8–14.3) Mexican American	15.6 (15.2–16.0) white; 15.2 (14.8–15.7) black; 15.7 (15.4–16.1) Mexican American
NHANES III; Sun et al <sup>65</sup> (2002) <sup>h</sup>	10.03 (9.61–10.40) white; 9.64 (9.4–10.60) Mexican American	12.32 (12.00–12.67) white; 11.78 (11.50–12.08) black; 12.53 (12.29–12.79) Mexican American	13.52 (13.22–13.83) white; 13.40 (13.15–13.66) black; 13.77 (13.51–14.03) Mexican American	16.01 (15.57–16.50) white; 15.00 (14.70–15.32) black; 15.76 (15.39–16.14) Mexican American	11.98 (11.69–12.29) white; 11.16 (10.89–11.43) black; 12.30 (12.06–12.56) Mexican American	12.65 (12.37–12.95) white; 12.51 (12.26–12.77) black; 13.06 (12.79–13.36) Mexican American	13.73 (13.49–13.99) black; 14.08 (13.83–14.32) Mexican American	15.67 (15.30–16.05) white; 15.32 (14.99–15.67) black; 15.75 (15.46–16.03) Mexican American
NHANES III; Sun et al <sup>65</sup> (2005)	ND	12.44 (11.73–12.84) white; 12.37 (11.82–12.70) black; 12.56 (12.34–12.80) Mexican American	13.49 (13.12–13.81) white; 13.46 (13.10–13.75) black; 13.76 (13.52–14.01) Mexican American	15.83 (15.38–16.36) white; 14.86 (14.38–15.31) black; 15.60 (15.23–16.02) Mexican American	ND	12.55 (11.95–12.93) white; 12.69 (12.27–12.99) black; 13.08 (12.84–13.33) Mexican American	13.48 (13.12–13.79) white; 13.79 (13.45–14.09) black; 14.08 (13.85–14.31) Mexican American	15.59 (15.23–15.96) white; 15.21 (14.71–15.71) black; 15.79 (15.49–16.11) Mexican American

CI, confidence interval.

<sup>a</sup> Means and medians are considered comparable values because the ages for various pubertal markers are typically normally distributed.

<sup>b</sup> See references for details.

<sup>c</sup> Reynolds and Wines<sup>69</sup> definitions of G stages differ from Tanner G stages (see Table 4 footnote b). Their G3 mean age is not shown here because it was defined differently from the other studies.

<sup>d</sup> Nicolson and Hanley<sup>14</sup> used a 5-stage “sex stage” system with assessment of pubic hair and genital development; genital stage descriptions were similar to those of Tanner and are reported here as G2 to G5, but comparability with other studies that used the Tanner stage definitions is a concern.

<sup>e</sup> Villarreal et al<sup>64</sup> only presented median ages for transition into G2 and PH2 because of the “limited age range,” 5 to 14 years, of their population.

<sup>f</sup> Foster et al<sup>61</sup> calculated percentages of boys in a given stage by age, means (from probit analysis), and medians (from Spearman-Kärber analysis) for entry into a stage. Means for entry into a stage are presented in this table.

<sup>g</sup> Biro et al<sup>25</sup> defined a P51; “P51 was defined as the absence of pubic hair and a testicular volume of the larger testis <3 mL.” Their P52 was similar to Tanner’s PH2. Data for comparable Tanner PH stages are presented here. P51 data are not presented here. Data were not presented by race.

<sup>h</sup> Sun et al<sup>65</sup> calculated means for “age at entry” into a stage, and medians for “age at entry” into a stage. Medians are reported here for comparison with other studies that looked at age of attainment.



**TABLE 6 Panel Conclusions About Evidence of Secular Trends in the Timing of Puberty for 1940–1994 Data**

Marker	Sufficient Data:		Insufficient Data:	
	Yes Trend	No Trend	Yes Trend	No Trend
Girls' BD onset (B2)	++	+	–	–
Girls' BD progression (B3–B5)	+	++	–	–
Girls' PH onset (PH2)	+	++	–	–
Girls' PH progression (PH3–PH5)	–	+++	–	–
Girls' menarche	++	+	–	–
Boys' GD onset (G2)	–	+++	–	–
Boys' GD progression (G3–G5)	–	+++	–	–
Boys' PH onset (PH2)	+	++	–	–
Boys' PH progression (PH3–PH5)	+	++	–	–

BD indicates breast development; –, no one's opinion; +, minority opinion; ++, majority opinion; +++, unanimous opinion.

ment for 39% of the participants,<sup>19</sup> and comparisons of findings for those participants indicated no evidence of biased staging when visual assessment alone was performed (ie, mistaking fat tissue for breast tissue).<sup>19</sup>

Some studies did not describe race/ethnicity of the participants, include multiple racial/ethnic groups (eg, all white population), or adjust for race/ethnicity, yielding 1 summary result for all race/ethnic groups, which limited comparisons for separate racial/ethnic groups between the studies. For this reason, some panelists considered data only for white girls because the older studies assessed only or mainly white girls. The older studies are mainly a high-SES population, which presents another complication when comparing data with those of the more recent studies.

Some panelists did not consider B2 timing changes in isolation, without subsequent breast development progression, as a marker of puberty. This viewpoint is based on the fact that B2 can result without pubertal progression from exogenous estrogen exposure and, thus, does not necessarily indicate HPG axis activation.<sup>71,72</sup> However, if pubertal development continues to progress after onset, then this suggests that the breast development onset was attributable to HPG activation. For this reason, some panelists considered B2 to be a less reliable marker of breast development timing than B3 to B5.

#### *Progression (B3–B5)*

The majority of the panelists concluded that the data are insufficient to suggest or establish a secular trend toward an earlier achievement of B3 to B5 stages. The reasons for this opinion include a lack of study design comparability between the historical and more recent studies. Some panelists argued that the NHES II/III, HHANES, and NHANES III studies were the only set that could be reliably compared, and they concluded that the change in ages of B3 to B5 was not significant over this time period (1960s–1990s); however, NHES II/III and NHANES III studies do not span the complete interval of time (1940s–1990s) considered by the panel.

The minority concluded that there are sufficient data to suggest a secular trend toward an earlier B3 to B5. They viewed B3 to B5 as more reliable markers of breast development timing because progression indicates acti-

vation of the HPG and they are less prone to inaccuracies in staging than B2. The basis for earlier B3 to B5 came from comparing either the more recent NHANES III or PROS data for white girls with the earlier studies<sup>13,14,17</sup> (Table 3).

#### *Pubic Hair Development*

##### *Onset (PH2)*

The majority concluded that PH2 is highly variable (ie, interindividual and interracial/ethnicity variability of ages for PH2), and, therefore, the data are insufficient to suggest or establish a secular trend in the age of PH2. Some panelists viewed pubic hair development, in general, as less reliable and more variable than breast development. A concern about the appropriateness of Tanner staging of pubic hair development<sup>10</sup> across racial/ethnic groups, because it was developed on the basis of white children, was raised. As with breast development, the lack of studies with similar designs was considered a key issue that hindered conclusions about secular changes.

A minority concluded that the data are sufficient to suggest a secular trend toward an earlier onset of pubic hair development because they considered PH2 a more reliable marker than B2 (in the absence of palpation). The more recent studies suggest that the mean age of PH2 is ~6 months earlier than several decades ago. This conclusion comes from comparing the recent PROS and NHANES III data for white girls that showed an earlier PH2 to any of the earlier studies<sup>13,14,17</sup> (Table 3).

##### *Progression (PH3–PH5)*

The panel unanimously concluded that the data are insufficient to suggest or establish a secular trend in the timing of pubic hair stages PH3 through PH5 on the basis of (1) the view that PH3 to PH5 markers are relatively subjective (particularly discerning between 4 and 5) and variable (eg, interracial/ethnicity and interindividual); (2) data comparability issues; and (3) exclusion of reliance on the PROS PH3 to PH5 timing data, because the oldest girls were 12 years old and, thus, many were too young to assess these stages.

##### *Menarche*

The majority concluded that the data are sufficient to suggest (most) or establish (fewer) a secular trend toward an earlier age of menarche. Most panelists with this opinion included only the NHES II/III and NHANES III studies in their evaluation because these studies used similar methods and population selection criteria. These panelists were convinced by the finding that menarche has decreased by 2.5<sup>3</sup> to 4<sup>46</sup> months during ~25 years. Others considered the full spectrum of studies spanning 1940–1994 on which to base their conclusions.

A minority of the panelists concluded that the cross-sectional study data are insufficient to conclude that there has been a biologically meaningful change in the age of menarche. Some panelists agreed with the conclusions of Chumlea et al<sup>46</sup> that interpreted the 4-month decrease as not significant. In addition, some panelists

excluded the menarche data from the PROS study<sup>5</sup> from consideration because the age range (3–12 years) was not optimal for calculating the mean age at menarche. Some panelists were concerned about the impact of the different statistical methods used in the studies that compared the NHES II/III and NHANES III data.

### Boys' Data

All agreed that statistical power and data quality issues make robust conclusions about the boys' data difficult. The panel discussed but did not come to conclusions on data regarding the tempo of puberty and the pubertal growth spurt.

### Genital Development

The unanimous opinion was that the data for genital development (G2 and G3–G5) are insufficient to suggest or establish a trend toward an earlier puberty in boys, on the basis of the lack of data quantity, data quality, and marker reliability. Many boys who were included in NHES III and NHANES III were too old for adequate G2 or PH2 assessment. In addition, the lack of validity and reliability of genital staging, particularly for G2, was considered problematic. Panelists agreed that testicular volume assessment, not performed in the studies, was a more meaningful measure of puberty onset where a testicular volume of >3 mL is conclusive evidence of HPG activation. Data quality was a concern given the degree of disparity, particularly for genital development, between the NHANES III findings and NHES III findings.<sup>6</sup>

### Pubic Hair Development

An overwhelming majority found the pubic hair development (PH2 and PH3–PH5) timing data to be insufficient to suggest or establish a trend. Similar to the argument for insufficiency of genital development data, this opinion was based on the lack of data quantity, data comparability, and marker reliability.

A very small minority of panelists concluded that there is suggestive evidence of an earlier pubic hair onset (PH2) and progression (PH3–PH5) for Mexican American boys but not for white or black boys during this time frame. This opinion was based on the viewpoint that pubic hair development may be easier to stage than genital development and, thus, more reliable. Data supporting a trend toward an increase in linear height in boys from the 1963–1970 to 1988–1994<sup>7</sup> and from 1973 to 1992<sup>73</sup> were also considered corroborative of an earlier male pubertal growth spurt.

### Reasons for Differences of Opinion

Differences of opinion among the panelists were based on judgment calls that placed more or less value on key aspects of study comparability. Overall, some panelists placed more weight on nationally representative studies (NHES I–III, HHANES, and NHANES III) because of similar designs yet performed at different times. For example, the clinicians tended to place more value on these data perhaps because they look to nationally represen-

tative data on which to base normative puberty-timing clinical guidelines. In assessing the girls' data, some of the panelists placed more weight on the study by Herman-Giddens et al<sup>5</sup> because of its much larger sample size and, hence, statistical power and inclusion of younger ages for assessing Tanner staging. One important design choice was the Tanner stage progression definition (eg, some studies reported the mean age of being "in a stage," others reported the mean age of "transition to a stage," and some reported both). Different statistical methods also affect comparisons among studies; however, few panelists were experts in statistics and none in survey statistics. For this reason, detailed scrutiny of similar and appropriate statistical analyses, within and among the studies, was not a focus of the discussions.

### Panel Research Recommendations

#### *New Analyses of Existing Data*

If the original data from historical puberty studies are available, then parallel analyses of data from multiple studies could yield more meaningful results. Thus, when possible, reanalysis of some of the existing studies with similar statistical approaches is recommended. For example, multiple examinations of the same participants should not be treated as separate cross-sectional studies or included in a single analysis as though they were from separate children. For the studies of largest study populations, 1 approach to examining secular trends is to compare children at specific ages in studies that were conducted at different times and in which each child is only examined once.

When analyzing a probability-based study, such as 1 that uses stratified multistage sampling, analyses must incorporate sample weights to yield accurate results. The size of the population affects the power and, in turn, the study quality. It is important to use statistical methods that are appropriate for the type of data and question at hand, as well as to have comparable statistical methods.

The data from the existing studies could be examined for changes in temporal sequence and tempo of puberty for different racial/ethnic groups. The relative timing of pubertal events may be a critical factor in separating out different types of puberty-timing phenotypes. For example, discordance between Tanner stages (eg, simultaneous PH4 and B2) suggests a relatively high androgen-to-estrogen concentration.

#### *Priority New Studies*

The panel recommended longitudinal studies as the optimal design for examining puberty timing in future studies. Tanner stages are best assessed for an individual over time, because it is a relative measure. Pubertal marker data collection may also be imbedded within larger studies. Examples of longitudinal studies in which pubertal timing are or will be tracked are the National Children's Study<sup>74,75</sup> (<http://nationalchildrensstudy.gov>) of 100 000 children to be followed from in utero to 21 years of age, the Danish National Birth Cohort study<sup>76</sup> ([www.ssi.dk/sw3820.asp](http://www.ssi.dk/sw3820.asp)) of 100 000 women and their

offspring during the first 6 years of life, and the Norwegian Mother and Child Cohort Study ([www.fhi.no/artikler/?id=51488](http://www.fhi.no/artikler/?id=51488)). These and other studies will provide baseline information with which to compare past studies and future cohorts of children and test hypotheses regarding specific exposures and puberty timing.

A second useful study design is a short-term longitudinal study with sequential cohorts (eg, the Bogalusa Heart Study), sometimes referred to as a mixed cross-sectional/longitudinal study. The sequential cohort design is excellent for looking at trends within 1 study because design choices are maintained and secular trends could be detected sooner than with the longitudinal study design. For either of these study designs, collaboration among disciplines and between countries could enhance and encourage the use of similar study designs, which would allow for geographic and temporal comparisons.

For the purposes of secular trend analysis, any new study using different methods (eg, testicular volume assessment) is still limited by methods used in the earlier studies. Thus, new studies would benefit from including design aspects that are in common with the historical studies. For example, the most recent NHANES surveys (4 cycles spanning 1999–2006) have not collected Tanner stage data for boys and girls, losing comparability with the past NHES (I–III), HHANES, and NHANES III data. Recognizing that there are practical limitations of conducting large studies of puberty timing, the panel recommended a study design “wish list” for longitudinal studies:

- individual observations repeated every 6 months;
- Tanner staging;
- breast palpation and testicular volume measurement for puberty onset;
- hormonal measurements;
- adequately wide age ranges (eg, 6–18 years of age);
- geographic differences considered, including migration and adoption;
- SES noted;
- race/ethnicity noted and defined and data stratified when appropriate;
- body fat and weight covariates (eg, BMI);
- inclusion (or targeted studies) of other racial/ethnic groups (eg, Native Americans, Asians) in large enough numbers to analyze; and
- hypothesis testing.

Tanner staging method was considered critical to the collection of puberty-timing data, but inclusion of breast tissue palpation and testicular volume assessment was recommended to increase the reliability of estimating the ages of breast and genital development onset. Of note, testicular volume in early puberty is being assessed in a US boys’ puberty-timing PROS study.<sup>77</sup> Training standards for assessing Tanner stage were considered necessary (for girls<sup>15</sup>; for boys<sup>78</sup>), and some suggested

that only trained clinicians should be raters; however, the use of palpation by a trained examiner for breast and testicular development has practical (eg, financial, technical training) and ethical (eg, consent for breast palpation or testicular measurement of study participants) constraints.

Underlying the interest in performing a secular trend analysis is the concern about whether environmental factors alter the age at puberty. If so, then unless the environmental factors are distributed uniformly, they should result in geographic variation that is measurable in cross-sectional studies. This reasoning led to well-controlled, multicenter studies of semen quality conducted recently in the United States, Europe, and Japan that demonstrated significant geographic variation in semen quality.<sup>79–81</sup> These studies identified areas with unusually high and low semen quality and thus suggest etiologic studies to identify causes of these differences.<sup>82</sup> Such an approach may be useful in identifying factors that affect age at puberty. One testable hypothesis with new studies is whether the secular trend toward an earlier puberty (at least concluded for girls) may be attributable to subpopulation differences that may in turn reflect differential exposures to factors that affect puberty. If puberty timing varies by geography, then the next question is which factor(s) distinguishes 1 geographic region from another (eg, an exposure, genetics).

In support of the hypothesis of geographic variation in puberty-timing data, studies from several countries have found somewhat different mean ages of menarche (reviewed by Parent et al<sup>58</sup>). For example, European data collected in the 1990s showed mean ages of menarche that are older than those from the US data (eg, menarche mean ages ranged from 13.0 to 13.4 years in Denmark,<sup>83,84</sup> ~13.2 years in Norway<sup>85</sup> and Sweden,<sup>86</sup> and 12.6–12.9 years for white girls in the United States<sup>3,5,46,47</sup>). Urban–rural differences in the mean age of menarche have also been observed within a country<sup>87</sup>; however, it is not possible to conclude definitively that mean age at menarche is different among countries without an analysis of the study design comparability.

Racial patterns of differences in puberty timing may also differ by geographic areas. For example, reported racial differences from South African data on female puberty timing have a different pattern than those reported from US data. The median ages for breast development initiation (B2) and completion (B5) were similar for black and white girls, but black girls had a later PH2 and PH5 than white girls who lived in Johannesburg,<sup>88</sup> and the median age at menarche for black girls was 10 months later than for white girls. These differences in puberty timing by geography may be a reflection of SES, race/ethnicity, and/or environmental factors.

#### *Recommended Methods Development*

Puberty markers that are easier to measure, less invasive, less expensive, and more sensitive and reliable need to be explored, developed, and validated. New methods to enhance and/or replace Tanner staging are recommended. These fall into 2 categories: physical markers and physiologic markers.



### Physical Markers

Tanner staging was recognized as an excellent system for assessing pubertal onset and progression; however, a less invasive examination could decrease stress to the participant. Improvements that may achieve this goal include using a same-gender examiner and developing methods and markers. Methods and markers to be developed and validated include the following:

- behavioral markers of male and female puberty onset;
- muscle mass increase in boys;
- infrared photographs for Tanner stage analysis;
- computer imaging for testicular volume measurements;
- race/ethnic-specific Tanner staging (eg, pubic hair staging in Peruvian Indians is not accurate for estimating sexual maturation<sup>89</sup>; it was noted that PH4 for Asians is typically the last, adult-type stage); and
- male pubertal markers not currently extensively used (eg, spermarche, voice break).

### Physiologic Markers

The development of reliable biochemical markers would represent a major step forward for puberty-timing studies in that biochemical changes may identify precursor events and provide validation of later phenotypic puberty changes. Important biochemical markers include measurement of hormone and hormone receptor levels. Hormone level assays are available, but improved assay sensitivity and reliability for urine, blood, and saliva are needed for use in large-scale human studies. Assessment of changing hormone levels during a 24-hour period complicates measurements. For example, in prepubertal boys and girls, gonadotropins in serum exhibit large diurnal variations as a result of their pulsatile secretion; therefore, any single measurement is unreliable.<sup>34</sup> The uncertainty can be overcome by measuring urinary secretion of luteinizing hormone, which increases before physical signs of puberty are evident.<sup>90</sup> During early puberty in boys, testosterone measures are most reliable when blood sampling is done in the morning, when testosterone levels are expected to be highest.<sup>91</sup> There are a few available assays for steroid receptor levels in humans that are limited in their validity and still under additional development.<sup>92</sup> Recommended physiologic marker methods for development and validation include the following:

- improved sensitivity and reliability and reduced cost of hormone level measurements to define further the associations between hormone levels and pubertal status;
- steroid hormone receptor level measurements; and
- puberty markers that distinguish between central precocious puberty and variants of precocious puberty (eg, nonprogressive early B2).

### CONCLUSIONS

The optimum data for studying puberty timing would be the collection of a suite of puberty-timing measures that

include physical as well as physiologic markers. The combination of physical and physiologic measures of human puberty timing will help in the understanding of the progression from onset to completion and of inter-individual variability and in the characterization of biochemical precursors. For example, sensitive hormone assay data in conjunction with Tanner staging data could define a marker that precedes overt physical changes yet still allow for comparison with earlier studies of puberty timing.

### ACKNOWLEDGMENTS

Support for “The Role of Environmental Factors on the Onset and Progression of Puberty” workshop was provided by US Environmental Protection Agency cooperative agreement 830774, the National Institute of Environmental Health Sciences (this research was supported in part by the Intramural Research Program of the National Institutes of Health and National Institute of Environmental Health Sciences), and Serono Inc.

We gratefully acknowledge David Schlumper and the Serono Symposia International for organizing and holding “The Role of Environmental Factors on the Onset and Progression of Puberty” Expert Panel Workshop. We thank Tim Mealey of the Meridien Institute for skillful moderation of the workshop and Beth Gladen for educating us about puberty study design and analytical methods and providing detailed comments on the manuscript. We thank the other members of the Expert Panel for participation and contributions: David Abbott, Carlos Bourdony, Jean-Pierre Bourguignon, Germaine Buck Louis, Farid Chehab, Gwen Collman, Ralph Cooper, Paul Foster, Beth Gladen, Mari Golub, L. Earl Gray, Mel Grumbach, Paul Kaplowitz, Carole Kimmel, Robert Lustig, Michele Marcus, Sergio Ojeda, Ora Pescovitz, Tony Plant, Ed Reiter, Steve Schrader, Richard Sharpe, Wolfgang Sippell, Niels Skakkebaek, Ana Soto, Jorma Toppari, Rochelle Tyl, Martin Wabitsch, Patricia Whitte, and Selma Witchel.

### REFERENCES

1. Zacharias L, Wurtman RJ. Age at menarche: genetic and environmental influences. *N Engl J Med.* 1969;280(16):868–875
2. Wyshak G, Frisch RE. Evidence of a secular trend in age of menarche. *N Engl J Med.* 1982;306(17):1033–1035
3. Anderson SE, Dallal GE, Must A. Relative weight and race influence average age at menarche: results from two nationally representative surveys of US girls studied 25 years apart. *Pediatrics.* 2003;111(4 pt 1):844–850
4. Freedman DS, Kahn LK, Serdula MK, Dietz WH, Srinivasan SR, Berenson GS. Relation of age at menarche to race, time period, and anthropometric dimensions: the Bogalusa Heart Study. *Pediatrics.* 2002;110(4). Available at: [www.pediatrics.org/cgi/content/full/110/4/e43](http://www.pediatrics.org/cgi/content/full/110/4/e43)
5. Herman-Giddens ME, Slora EJ, Wasserman RC, et al. Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings Network. *Pediatrics.* 1997;99(4):505–512
6. Herman-Giddens ME, Wang L, Koch G. Secondary sexual characteristics in boys: estimates from the National Health and Nutrition Examination Survey III, 1988–1994. *Arch Pediatr Adolesc Med.* 2001;155(9):1022–1028

7. Karpati AM, Rubin CH, Kieszak SM, Marcus M, Troiano RP. Stature and pubertal stage assessment in American boys: the 1988–1994 Third National Health and Nutrition Examination Survey. *J Adolesc Health*. 2002;30(3):205–212
8. Euling SY, Selevan SG, Pescovitz OH, Skakkebaek NE. The role of environmental factors in the timing of puberty. *Pediatrics*. 2008;121(2 suppl):S167–S171
9. Tanner JM. *Growth at Adolescence*. 2nd ed. Oxford, England: Blackwell Scientific Publications; 1962
10. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child*. 1969;44(235):291–303
11. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child*. 1970;45(239):13–23
12. Van Wieringen JC, Wafelbakker F, Verbrugge HP, De Haas JH. Growth diagrams. In: Yen SSC, Jaffe RB, eds. *Reproductive Endocrinology: Physiology, Pathophysiology and Clinical Management*. 2nd ed. Philadelphia, PA: WB Saunders; 1978:1–68
13. Reynolds EL, Wines JV. Individual differences in physical changes associated with adolescence in girls. *Am J Dis Child*. 1948;75:329–350
14. Nicolson AB, Hanley C. Indices of physiological maturity: derivation and interrelationships. *Child Dev*. 1953;24(1):3–38
15. Herman-Giddens ME, Bourdony CJ. *Assessment of Sexual Maturity Stages in Girls*. Elk Grove Village, IL: Pediatric Research in Office Settings, American Academy of Pediatrics; 1995. Report MA0089
16. Biro FM, Lucky AW, Simbartl LA, et al. Pubertal maturation in girls and the relationship to anthropometric changes: pathways through puberty. *J Pediatr*. 2003;142(6):643–646
17. Lee PA. Normal ages of pubertal events among American males and females. *J Adolesc Health Care*. 1980;1(1):26–29
18. Herman-Giddens ME, Kaplowitz PB, Wasserman R. Navigating the recent articles on girls' puberty in Pediatrics: what do we know and where do we go from here? *Pediatrics*. 2004;113(4):911–917
19. Kaplowitz PB, Oberfield SE; Drug and Therapeutics and Executive Committees of the Lawson Wilkins Pediatric Endocrine Society. Reexamination of the age limit for defining when puberty is precocious in girls in the United States: implications for evaluation and treatment. *Pediatrics*. 1999;104(4 pt 1):936–941
20. US Department of Health and Human Services. *Third National Health and Nutrition Examination Survey, 1988–1994, NHANES III Laboratory Data File* [CD-ROM]. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention; 1996. Public use data file documentation No. 76200
21. Jolliffe D. Extent of overweight among US children and adolescents from 1971 to 2000. *Int J Obes Relat Metab Disord*. 2004;28(1):4–9
22. Must A, Phillips SM, Naumova EN, et al. Recall of early menstrual history and menarcheal body size: after 30 years, how well do women remember? *Am J Epidemiol*. 2002;155(7):672–679
23. Damon A, Damon ST, Reed RB, Valadian I. Age at menarche of mothers and daughters, with a note on accuracy of recall. *Hum Biol*. 1969;41(2):160–175
24. Zachmann M, Prader A, Kind HP, Hafliger H, Budliger H. Testicular volume during adolescence: cross-sectional and longitudinal studies. *Helv Paediatr Acta*. 1974;29(1):61–72
25. Biro FM, Lucky AW, Huster GA, Morrison JA. Pubertal staging in boys [published correction appears in *J Pediatr*. 1995;127(4):674]. *J Pediatr*. 1995;127(1):100–102
26. Rivkees SA, Hall DA, Boepple PA, Crawford JD. Accuracy and reproducibility of clinical measures of testicular volume. *J Pediatr*. 1987;110(6):914–917
27. Chipkevitch E, Nishimura RT, To DS, Galea-Rojas M. Clinical measurement of testicular volume in adolescents: comparison of the reliability of 5 methods. *J Urol*. 1996;156(6):2050–2053
28. Carlsen E, Anderson AG, Buchreitz L, et al. Inter-observer variation in the results of the clinical andrological examination including estimation of testicular size. *Int J Androl*. 2000;23(4):248–253
29. Prader A. Testicular size: assessment and clinical importance. *Triangle*. 1966;7(6):240–243
30. al Salim A, Murchison PJ, Rana A, Elton RA, Hargrave TB. Evaluation of testicular volume by three orchidometers compared with ultrasonographic measurements. *Br J Urol*. 1995;76(5):632–635
31. Shiraishi K, Takihara H, Kamiryo Y, Naito K. Usefulness and limitation of punched-out orchidometer in testicular volume measurement. *Asian J Androl*. 2005;7(1):77–80
32. Nielsen CT, Skakkebaek NE, Richardson DW, et al. Onset of the release of spermatozoa (spermarche) in boys in relation to age, testicular growth, pubic hair, and height. *J Clin Endocrinol Metab*. 1986;62(3):532–535
33. Jørgensen M, Keiding N, Skakkebaek NE. Estimation of spermarche from longitudinal spermaturia data. *Biometrics*. 1991;47(1):177–193
34. Wu FC, Butler GE, Kelnar CJ, Sellar RE. Patterns of pulsatile luteinizing hormone secretion before and during the onset of puberty in boys: a study using an immunoradiometric assay. *J Clin Endocrinol Metab*. 1990;70(3):629–637
35. Rockett JC, Lynch CD, Buck GM. Biomarkers for assessing reproductive development and health: part 1—pubertal development. *Environ Health Perspect*. 2004;112(1):105–112
36. Last JM, ed. *A Dictionary of Epidemiology*. 4th ed. New York, NY: Oxford University Press; 2001
37. Cox DR. *Analysis of Binary Data*. London, England: Methuen and Co Ltd; 1970
38. Wu Y, Schreiber GB, Klementowicz V, Biro F, Wright D. Racial differences in accuracy of self-assessment of sexual maturation among young black and white girls. *J Adolesc Health*. 2001;28(3):197–203
39. Schlossberger NM, Turner RA, Irwin CE Jr. Validity of self-report of pubertal maturation in early adolescents. *J Adolesc Health*. 1992;13(2):109–113
40. Bonat S, Pathomvanich A, Keil MF, Field AE, Yanovski JA. Self-assessment of pubertal stage in overweight children. *Pediatrics*. 2002;110(4):743–747
41. Ibáñez L, Ong KK, Mongan N, et al. Androgen receptor gene CAG repeat polymorphism in the development of ovarian hyperandrogenism. *J Clin Endocrinol Metab*. 2003;88(7):3333–3338
42. Kadlubar FF, Berkowitz GS, Delongchamp RR, et al. The CYP3A4\*1B variant is related to the onset of puberty, a known risk factor for the development of breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2003;12(4):327–331
43. Stavrou I, Zois C, Ioannidis JP, Tsatsoulis A. Association of polymorphisms of the oestrogen receptor alpha gene with the age of menarche. *Hum Reprod*. 2002;17(4):1101–1105
44. Xita N, Tsatsoulis A, Stavrou I, Georgiou I. Association of SHBG gene polymorphism with menarche. *Mol Hum Reprod*. 2005;11(6):459–462
45. Sun S, Schubert CM, Chumlea WC, et al. National estimates of the timing of sexual maturation and racial differences among US children. *Pediatrics*. 2002;110(5):911–919
46. Chumlea WC, Shubert CM, Roche AF, et al. Age at menarche and racial comparisons in US girls. *Pediatrics*. 2003;111(1):110–113
47. Wu T, Mendola P, Buck GM. Ethnic differences in the presence of secondary sex characteristics and menarche among US girls: the Third National Health and Nutrition Examination Survey, 1988–1994. *Pediatrics*. 2002;110(4):752–757



48. Galler JR, Ramsey FC, Salt P, Archer E. Long-term effects of early kwashiorkor compared with marasmus: I—physical growth and sexual maturation. *J Pediatr Gastroenterol Nutr.* 1987;6(6):841–846
49. Cameron N, Mitchell J, Meyer D, et al. Secondary sexual development of Cape coloured girls following kwashiorkor. *Ann Hum Biol.* 1988;15(1):65–75
50. Cameron N, Mitchell J, Meyer D, et al. Secondary sexual development of Cape coloured boys following kwashiorkor. *Ann Hum Biol.* 1990;17(3):217–228
51. Kulin HE, Bwibo N, Mutie D, Santner SJ. The effect of chronic childhood malnutrition on pubertal growth and development. *Am J Clin Nutr.* 1982;36(3):527–536
52. Koprowski C, Ross RK, Mack WJ, Henderson BE, Bernstein L. Diet, body size and menarche in a multiethnic cohort. *Br J Cancer.* 1999;79(11–12):1907–1911
53. Kaplowitz PB. The link between body fat and the timing of puberty. *Pediatrics.* 2008;121(2 suppl):S208–S217
54. Lazar L, Pollak U, Kalter-Leibovici O, Pertzalan A, Phillip M. Pubertal course of persistently short children born small for gestational age (SGA) compared with idiopathic short children born appropriate for gestational age (AGE). *Eur J Endocrinol.* 2003;149(5):425–432
55. Ibáñez L, Ferrer A, Marcos MV, Hierro FR, de Zegher F. Early puberty: rapid progression and reduced final height in girls with low birthweight. *Pediatrics.* 2000;106(5). Available at: [www.pediatrics.org/cgi/content/full/106/5/e72](http://www.pediatrics.org/cgi/content/full/106/5/e72)
56. Veening MA, van Weissenbruch MM, Roord JJ, de Delme-marre-van Waal HA. Pubertal development in children born small for gestational age. *J Pediatr Endocrinol Metab.* 2004;17(11):1497–1505
57. Powls A, Botting N, Cooke RWI, Pilling D, Marlow N. Growth impairment in very low birthweight children at 12 years: correlation with perinatal and outcome variables. *Dis Child Fetal Neonatal Ed.* 1996;75:F152–F157
58. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr Rev.* 2003;24(5):668–693
59. Krstevska-Konstantinova M, Charlier C, Craen M, et al. Sexual precocity after immigration from developing countries to Belgium: evidence of previous exposure to organochlorine pesticides. *Hum Reprod.* 2001;16(5):1020–1026
60. dos Santos Silva I, De Stavola BL, Mann V, Kuh D, Hardy R, Wadsworth MEJ. Prenatal factors, childhood growth trajectories and age at menarche. *Int J Epidemiol.* 2002;31(2):405–412
61. Foster TA, Voors AW, Webber LS, Frerichs, Berenson GS. Anthropometric and maturation measurements of children, ages 5 to 14 years, in a biracial community: the Bogalusa Heart Study. *Am J Clin Nutr.* 1977;30(4):582–591
62. Harlan WR, Harlan EA, Grillo GP. Secondary sex characteristics of girls 12 to 17 years of age: the U.S. Health Examination Survey. *J Pediatr.* 1980;96(6):1074–1078
63. MacMahon B. *Age at Menarche: United States.* Rockville, MD: National Center for Health Statistics; 1973. DHEW publication (HRA)74–1615
64. Villarreal SF, Martorell R, Mendoza F. Sexual maturation of Mexican-American adolescents. *Am J Hum Biol.* 1989;1(1):87–95
65. Sun SS, Schubert CM, Liang R, et al. Is sexual maturity occurring earlier among U.S. children? *J Adolesc Health.* 2005;37(5):345–355
66. Weir HK, Kreiger N, Marrett LD. Age at puberty and risk of testicular germ cell cancer (Ontario, Canada). *Cancer Causes Control.* 1998;9(3):253–258
67. United Kingdom Testicular Cancer Study Group. Aetiology of testicular cancer: association with congenital abnormalities, age at puberty, infertility, and exercise. *BMJ.* 1994;308(6941):1393–1399
68. Golub MS, Collman GW, Foster MD, et al. Public health implications of altered puberty timing. *Pediatrics.* 121;2(2 suppl):S218–S230
69. Reynolds EL, Wines JV. Physical changes associated with adolescence in boys. *Am J Dis Child.* 1951;82(5):529–547
70. Harlan WR, Grillo GP, Cornoni-Huntley J, Leaverton P. Secondary sex characteristics of boys 12 to 17 years of age: the U.S. Health Examination Survey. *J Pediatr.* 1979;95(2):293–297
71. Hertz R. Accidental ingestion of estrogens by children. *Pediatrics.* 1958;21(2):203–206
72. Tiwary CM. Premature sexual development in children following the use of estrogen- or placenta-containing hair products. *Clin Pediatr (Phila).* 1998;37(12):733–739
73. Freedman DS, Khan LK, Serdula MK, Srinivasan SR, Berenson GS. Secular trends in height among children during 2 decades: the Bogalusa Heart Study. *Arch Pediatr Adolesc Med.* 2000;154(2):155–161
74. Branum AM, Collman GW, Correa A, et al. The National Children's Study of Environmental Effects on Child Health and Development. *Environ Health Perspect.* 2003;111(4):642–646
75. Wang RY, Needham LL, Barr DB. Effects of environmental agents on the attainment of puberty: considerations when assessing exposure to environmental chemicals in the National Children's Study. *Environ Health Perspect.* 2005;113(8):1100–1107
76. Olsen J, Melbye M, Olsen SF, et al. The Danish National Birth Cohort: its background, structure and aim. *Scand J Public Health.* 2001;29(4):300–307
77. Herman-Giddens ME. Recent data on pubertal milestones in United States children: the secular trend toward earlier development. *Int J Androl.* 2006;29(1):241–246
78. Herman-Giddens ME, Bourdony CJ, Dowshen SA. *Assessment of Sexual Maturity Stages in Boys.* Elk Grove Village, IL: Pediatric Research in Office Settings, American Academy of Pediatrics; 2005
79. Jorgensen N, Carlsen E, Nermoen I, et al. East-West gradient in semen quality in the Nordic-Baltic area: a study of men from the general population in Denmark, Norway, Estonia and Finland. *Hum Reprod.* 2002;17(8):2199–2208
80. Punab M, Zilaitiene B, Jorgensen N, et al. Regional differences in semen qualities in the Baltic region. *Int J Androl.* 2002;25(4):243–252
81. Swan SH, Brazil C, Drobnis EZ, et al. Geographic differences in semen quality of fertile US males. *Environ Health Perspect.* 2003;111(4):414–420
82. Swan SH, Kruse R, Liu F, et al. Biomarkers of pesticide exposure in relation to semen quality. *Environ Health Perspect.* 2003;111(12):1478–1484
83. Helm P, Grolund L. A halt in the secular trend towards earlier menarche in Denmark. *Acta Obstet Gynecol Scand.* 1998;77(2):198–200
84. Juul A, Teilmann G, Scheike T, et al. Pubertal development in Danish children: comparison of recent European and US data. *Int J Androl.* 2006;29(1):247–255; discussion 286–290
85. Nafstad P, Stray-Pedersen B, Solvberg M, Tangen T. Menarche and menstruation problems among teenagers in Oslo 1993 [in Norwegian]. *Tidsskr Nor Laegeforen.* 1995;115(5):604–606
86. Lindgren GW, Degerfors IL, Fredriksson A, et al. Menarche 1990 in Stockholm schoolgirls. *Acta Paediatr Scand.* 1991;80(10):953–955
87. Rimpelä AH, Rimpela MK. Towards an equal distribution of health? Socioeconomic and regional differences of the secular

- trend of the age of menarche in Finland from 1979 to 1989. *Acta Paediatr.* 1993;82(1):87–90
88. Channing-Pearce SM, Solomon L. Pubertal development in black and white Johannesburg girls. *S Afr Med J.* 1987;71(1):22–24
89. Freyre EA, Llaza G, Rebaza-Gonzalez R, Sami D. Tanner's pubic hair staging is not applicable to all adolescents. *J Adolesc Health.* 2002;30(3):144–145
90. Demir A, Voutilainen R, Juul A, et al. Increase in first morning voided urinary luteinizing hormone levels precedes the physical onset of puberty. *J Clin Endocrinol Metab.* 1996;81(8):2963–2967
91. Wu FC, Brown DC, Butler GE, Stirling HF, Kelnar CJ. Early morning plasma testosterone is an accurate predictor of imminent pubertal development in prepubertal boys. *J Clin Endocrinol Metab.* 1993;76(1):26–31
92. Mudarris A, Peck EJ Jr. Human anti-estrogen receptor antibodies: assay, characterization, and age- and sex-related differences. *J Clin Endocrinol Metab.* 1987;64(2):246–254
93. Wattigney WA, Srinivasan SR, Chen W, Greenlund KJ, Berenson GS. Secular trend of earlier onset of menarche with increasing obesity in black and white girls: the Bogalusa Heart Study. *Ethn Dis.* 1999;9(2):181–189
94. Greulich WW, Dorfman RI, Catchpole HR, Solomon CI, Culotta CS. Somatic and endocrine studies of pubertal and adolescent boys. *Monogr Soc Res Child Dev.* 1942;7(3):1

**Examination of US Puberty-Timing Data from 1940 to 1994 for Secular Trends:  
Panel Findings**

Susan Y. Euling, Marcia E. Herman-Giddens, Peter A. Lee, Sherry G. Selevan, Anders Juul, Thorkild I. A. Sørensen, Leo Dunkel, John H. Himes, Grete Teilmann and Shanna H. Swan

*Pediatrics* 2008;121;S172  
DOI: 10.1542/peds.2007-1813D

<b>Updated Information &amp; Services</b>	including high resolution figures, can be found at: <a href="http://pediatrics.aappublications.org/content/121/Supplement_3/S172">http://pediatrics.aappublications.org/content/121/Supplement_3/S172</a>
<b>References</b>	This article cites 83 articles, 0 of which you can access for free at: <a href="http://pediatrics.aappublications.org/content/121/Supplement_3/S172#BIBL">http://pediatrics.aappublications.org/content/121/Supplement_3/S172#BIBL</a>
<b>Subspecialty Collections</b>	This article, along with others on similar topics, appears in the following collection(s): <b>Pediatrician Wellness</b> <a href="http://www.aappublications.org/cgi/collection/pediatrician_wellness">http://www.aappublications.org/cgi/collection/pediatrician_wellness</a>
<b>Permissions &amp; Licensing</b>	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: <a href="http://www.aappublications.org/site/misc/Permissions.xhtml">http://www.aappublications.org/site/misc/Permissions.xhtml</a>
<b>Reprints</b>	Information about ordering reprints can be found online: <a href="http://www.aappublications.org/site/misc/reprints.xhtml">http://www.aappublications.org/site/misc/reprints.xhtml</a>

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™



# PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

## **Examination of US Puberty-Timing Data from 1940 to 1994 for Secular Trends: Panel Findings**

Susan Y. Euling, Marcia E. Herman-Giddens, Peter A. Lee, Sherry G. Selevan, Anders Juul, Thorkild I. A. Sørensen, Leo Dunkel, John H. Himes, Grete Teilmann and Shanna H. Swan

*Pediatrics* 2008;121;S172  
DOI: 10.1542/peds.2007-1813D

The online version of this article, along with updated information and services, is located on the World Wide Web at:

[http://pediatrics.aappublications.org/content/121/Supplement\\_3/S172](http://pediatrics.aappublications.org/content/121/Supplement_3/S172)

Pediatrics is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. Pediatrics is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2008 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 1073-0397.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™

