Published in final edited form as:
Dev Psychol. 2013 June ; 49(6): 1187–1193. doi:10.1037/a0029424.

Longitudinal Reliability of Self-Reported Age at Menarche in Adolescent Girls: Variability Across Time and Setting

L. D. Dorn, Ph.D.[1,2], **L. M. Sontag-Padilla, Ph.D**[3], **S. Pabst, M.Ed**[1], **A. Tissot, Ph.D**[1,2], and **E.J. Susman, Ph.D**[4]

[1]Cincinnati Children's Hospital Medical Center, Cincinnati OH

[2]University of Cincinnati College of Medicine, Cincinnati, OH

[3]The RAND Corporation, Pittsburgh, PA

[4]The Pennsylvania State University, State College, PA

Abstract

Objective—Age at menarche is critical in research and clinical settings, yet there is a dearth of studies examining its reliability in adolescents. We examined age at menarche during adolescence, specifically, 1) average method reliability across three years, 2) test-retest reliability between time points and methods, 3) intra-individual variability of reports, and 4) whether intraindividual variability differed by setting or individual characteristics.

Methods—Girls (n = 253) were enrolled in a cross-sequential study in age cohorts (11, 13, 15, and 17 years). Age at menarche was assessed using three annual, in-person clinician interviews followed by nine quarterly phone interviews conducted by research assistants.

Results—Reliability of age at menarche across time was moderate and varied by method. In-person interviews showed greater reliability [intraclass correlation coefficient (ICC) = .77)] versus phone interviews (ICC = .64). Test-retest reliability in reports did not decrease across time. However, average differences in reported age varied as much as 2.3 years (SD = 2.2 years), with approximately 9% demonstrating differences greater than 4.5 years. Pubertal timing category (i.e., early, late) changed for 22.7% if categorized at the final versus the first report of age at menarche.

Conclusions—Reliability was moderate, but average differences in reported age are notable and concerning. Using in-person clinician interviews may enhance reliability. Researchers and clinicians should be cognizant of the implications of using different methods measuring age at menarche when interpreting research findings.

Keywords

Age at menarche; adolescents; consistency; accuracy; pubertal timing

Age at menarche is collected during routine adolescent or adult health care visits, but it is a key variable in social-behavioral as well as clinical research studies. For example, studies focusing on the secular trend of early maturation use age at menarche to describe pubertal differences across cohorts (Chumlea et al., 2003), whereas biobehavioral studies use age at

Corresponding Author: Lorah D. Dorn, Ph.D/Cincinnati Children's Hospital Medical Center/3333 Burnet Ave./MLC 4000/Cincinnati, OH 45229/Phone: 513-636-7204/Fax: 513-636-1129/ lorah.dorn@cchmc.org.

Financial Disclosure: This research was supported in part by Grant Number R01 DA16402, National Institute of Drug Abuse, NIH, PI: Dorn and by the National Center for Research Resources and the National Center for Advancing Translational Sciences, NIH, through Grant 8 UL1 TR000077-04.

menarche to categorize pubertal timing (e.g., “early”, “on-time”, “late”) (Deardorff, Gonzales, Christopher, Roosa, & Millsap, 2005; Marklein, Negriff, & Dorn, 2009; Mendle, Turkheimer, & Emery, 2007; Obeidallah, Brennan, Brooks-Gunn, & Earls, 2004). Alternatively, clinical studies use age at menarche to describe risk associated with certain diseases as illustrated by literature showing girls with earlier menarche may be at greater risk for cardiovascular disease (Remsberg et al., 2005) or adult breast cancer (Apter, Reinila, & Vihko, 1989; Hamilton & Mack, 2003). On an individual level, age at menarche can provide critical input into clinical decision making, whereas findings from research studies about risks associated with timing of menarche can be used to direct patient care or inform community-wide prevention.

With the application of age of menarche for numerous purposes and settings, it is critical to note the dearth of research describing the reliability of age at menarche across time or whether reliability varies across methods of assessment. Age of menarche is a marker that must rely on self-report, or in some cases parent report, since identifying the actual onset objectively would require tracking efforts prior to the event that would be labor intensive and virtually prohibitive. Thus, in the strictest sense, no real “gold standard” exists for determining age at menarche. Using correlations, studies of age at menarche report reliability across as much as 40 years range from .60 to .83(Artaria & Henneberg, 2000; Bergsten-Brucefors, 1976; Damon & Bajema, 1974; Hediger & Stine, 1987; Koo & Rohan, 1997; Koprowski, Coates, & Bernstein, 2001; Livson & McNeill, 1962; Must et al., 2002). However, most of these studies rely on as few as two assessments, a decade or more apart. Correlational studies do not fully address consistency in reports; in turn, they may underestimate the degree of fluctuation in self-reported age at menarche across time. For example, kappa coefficients were reported as low to moderate (e.g., $k = .35$) when comparisons included an adolescent self-report and then self-report at age 48 (Cooper et al., 2006). Further, there may be some instances in which repeated self-report varies by several months. The resulting error from this methodological influence may be more crucial for certain research questions than others. For example, reports varying by a year may be less important for a study examining an outcome 30–40 years later compared with a longitudinal study categorizing pubertal-age girls into timing categories where the event is more proximal. Such variability was noted by a study of 9–13 year-olds where they were asked their age at menarche by clinician interview and then by questionnaire at baseline, 6- and 12-months later. Reports varied by as much as 18 months across one year (Dorn et al., 1999). As such, different conclusions about timing of age at menarche and its effect on an outcome (e.g., substance use, depression) will be drawn in the absence of a reliable method of assessing age at menarche.

To address the gap in the adolescent literature, our aims were to use two methods of assessment (in-person, phone) to determine: 1) the average method reliability of reporting age at menarche across the three years, 2) test-retest reliability of reports between different time points and methods, 3) intraindividual variability of reports, and 4) whether intraindividual variability of reports differed by method or individual characteristics (i.e., race, pubertal timing and gynecological age). The analytic goal was descriptive rather than predictive. Thus, specific hypotheses were not made.

Method

Design and Participants

This study was part of a longitudinal study that focused on the impact of depressive symptoms and smoking on reproductive and bone health in pubertal-age girls (Dorn et al., 2008). Two hundred sixty-two girls were enrolled by age cohort (11, 13, 15, and 17 years) in a cross-sequential design (Miyazaki & Raudenbush, 2000). Between December 2003 and

October 2007, girls were recruited from an adolescent medicine clinic in a large Midwestern children's hospital and the surrounding community.

Exclusion criteria for the parent study included: 1) pregnancy or breast feeding within 6 months, 2) primary amenorrhea (> 16 years), 3) secondary amenorrhea (< 6 cycles/year), 4) body mass index (BMI) for age < 1st percentile or weight > 300 pounds, 5) medication/ medical disorder influencing bone health, and 6) psychological disorders impairing comprehension or compliance. Additionally, for purposes of this paper, girls who were premenarcheal at the final assessment (n = 6) or girls who had fewer than two assessment points (n = 3) by the final assessment were excluded. Thus, 253 girls were included for this analysis.

Three annual visits were conducted in a Clinical Translational Research Center and telephone interviews were conducted quarterly following annual visits (e.g., 3, 6, 9 months). The focus for the current analyses was on age at menarche. The study was approved by the Institutional Review Board of the medical center. Parents provided written informed consent and adolescents provided assent.

Measures and Procedures

Primary measures for this report focus on age at menarche assessed via two methods. First, a trained clinician (i.e., pediatric nurse practitioner, adolescent medicine physician or fellow) conducting the physical exam interviewed each adolescent in private using standard questions (e.g., How old were you when you started your first period?) to determine age at menarche (year and month). Additional standard questions were used as necessary to probe for a more accurate date (e.g., What grade in school were you? What season was it? Was it near a holiday or in summer? etc.) (Dorn, Dahl, Woodward, & Biro, 2006; Dorn et al., 1999). The second methodology included similar questions but in a year (e.g., 2001) and month format (e.g., In which year and month did you start your first period?) and asked via phone by a trained research assistant. Age was then determined using probes if necessary. Research staff was blind to earlier reports of age at menarche.

Other variables in the analyses included race (white versus other) and family socioeconomic status (SES) (Hollingshead, 1975) with possible scores ranging from 8 (low) to 66 (high). Both were obtained by parent report. Pubertal timing was categorized using standard criteria (e.g., + 1 SD) for timing groups (early, on-time, and late) based on the sample distribution of first reported age at menarche within white and black racial groups in this study. Those within + 1 SD (between 11.53 and 13.75 years for white girls, and between10.56 and 13.33 years for black and other girls) were considered "on-time". Gynecological age was calculated as age at baseline minus age-at-menarche in years and months.

Statistical Analyses

To estimate reliability of self-reported age at menarche across the study, we calculated the intra-class correlation coefficient (ICC) across all assessments, and by methodology (in-person versus phone interviews). The ICC reflects the proportion of the total outcome variation that lies between people (or occasions) (Singer, 2003). Variance components used to calculate the ICC were estimated using SPSS 18.0 linear mixture modeling; maximum likelihood estimation was used to handle missing data. Bivariate correlations were used to describe test-retest reliability of reports across time. Intrainidividual variability of reports were calculated for each participant as difference scores between maximum and minimum reported age at menarche across the study, across in-person interviews only, and across phone interviews only. Finally, we estimated differences in variability of reports (i.e.,

difference scores) by race, pubertal timing and gynecological age using a combination of univariate analysis of variance (ANOVA) and hierarchical linear regression.

Results

Descriptive Statistics

On average, the 253 girls included in this study were 15.1 years old (*SD* =2.1 years), and the majority were white (61.7%) or black (33.2%) with some mixed race/other (5.1%). Race was dichotomized as white versus non-white. At baseline, most participants were in later puberty (Tanner breast stage I = 1.5%, II = 1.9%, III = 10.3%, IV = 14.9%, and V = 71.4%), and 80.2% were postmenarcheal. Family SES ranged from 14–66 (mean = 37.4, *SD* = 13.6).

Average method reliability of self-reported age at menarche

The ICC across all 12 assessments (i.e., both in-person interviews and phone interviews) was 0.64, whereas the ICC for in-person interviews only (3 time points) and phone interviews only (9 time points) was 0.77 and .64, respectively. The two methods were significantly different ($p < .05$; using 2-tailed *r* to *z* transformation) suggesting that across time girls were more reliable during in-person interviews than during phone interviews.

Change in test-retest reliability in reports across time and method

Bivariate correlations across all assessment points were examined to determine whether test-retest reliability of age at menarche decreased across time (Table 1). Correlations across the study corroborate findings from the ICC, such that greater test-retest reliability exists for in-person clinician interviews (0.74 – 0.81) compared to phone interviews (0.55 – 0.76). However, results do not provide evidence that test-retest reliability in reports decreased across the course of the study. Although correlations vary within a single year and across measurement methods, correlations between age at menarche reports do not become progressively lower across time.

Average intraindividual variability of self-reported age at menarche

Mean age of menarche as determined by first report in the study (e.g., Time 1 for girls who were postmenarcheal at Time 1; later times for girls who were premenarcheal at Time 1) was 12.4 years (SD = 1.25) and did not differ by age cohort (Table 2). Using difference scores (maximum age reported – minimum age), reports of age at menarche across all assessments (see “average difference: in-person & phone” in Table 2) varied on average by 2.3 years (SD = 2.2), with 8.8% of girls demonstrating differences greater than 4.53 years (1 SD above mean). When examined separately by method (in-person or phone interview), reports were fairly consistent across annual visits with an average difference score of 0.58 years (7 months) and 16.4% of girls demonstrating differences greater than 1 year. In contrast, reports were highly inconsistent for phone interviews (mean difference score = 2.2 years), with 66.3% demonstrating differences greater than 1 year.

Intraindividual variability of age at menarche differ by race, pubertal timing, and gynecological age

Difference scores of age at menarche, as well as reliability in self-reports, were examined by race, pubertal timing, and gynecological age. Controlling for chronological age, white participants reported a later age at menarche (compared to non-whites (Table 3). Although there was no difference ($p > .05$) by race between mean age at menarche reported during the in-person interviews versus phone interviews, there were significant differences by race in intraindividual variability (i.e., the average difference score) of age at menarche. Specifically, after controlling for family SES, difference scores for in-person interviews

(maximum reported age at menarche - minimum reported age at menarche) were significantly ($p < .05$) higher for non-whites (0.7 years, SD = .08) compared to whites (0.5 years, SD = .06). Findings suggest that non-whites were less reliable or consistent in their reports across the in-person interviews compared to whites. This difference by race did not emerge across all assessments (in-person and phone combined) or phone interviews only. Additionally, lower SES was associated with higher difference scores during the phone visits ($p < .05$).

Next, pubertal timing differences in reliability of age at menarche, as well as difference scores, were examined. After controlling for chronological age, no pubertal timing differences were observed for age at menarche difference scores across all methodologies (in-person, phone). Important to note, in this study, pubertal timing groups were calculated based on *first* report of age at menarche, with most girls reporting age at menarche at the Year 1 annual visit. However, these categorizations are likely to vary depending on which assessment point is used. Hence, for purposes of comparison, pubertal timing groups were also estimated based on the Year 3 annual visit reports. Cut-off ranges varied slightly between Year 3 and Year 1 annual reports. Overall, cutoffs were younger for Year 1 ("on-time" categorization ranged between 11.74 and 13.88 years for white girls, and between 10.96 and 13.58 years for black and other girls) compared to Year 3 ("on-time" categorization ranged between 11.53 and 13.75 years for white girls, and 10.56 and 13.33 years for black and other girls). As a result of these differences, 46 (22.7%) of the 203 girls who reported age at menarche at the Year 3 annual visit fell within *different* timing groups from the original categorization; 14 of these girls (6.9% of the 203 girls) were categorized as "early" after originally falling within the "on-time" group. (Table 4). Change in timing categories indicated that actual counts differed from expected counts (X^2 (4) = 125.3, p < . 001).

The relationship between gynecological age and difference scores for reporting age at menarche were examined via hierarchical linear regression controlling for chronological age. Fifty girls were pre-menarcheal at baseline; for exploratory purposes, these girls were assigned a negative gynecological age representing age until menarche. Controlling for age at the Year 1 annual visit, gynecological age was positively associated with difference scores across all assessments ($\beta = .34$, $p < .01$), clinician interviews ($\beta = .31$, $p < .01$), and phone interviews ($\beta = .25$, $p < .05$). However, when premenarcheal girls at the baseline visit were removed from the analyses, effects were no longer significant.

Discussion

To our knowledge, this is the first report to examine reliability and intraindividual variability of self-reported age at menarche in adolescents that were repeatedly sampled across three years. Additionally, two methodologies were utilized; in-person clinician interviews and phone interviews by research assistants. Earlier reports on age at menarche included primarily adult women utilizing two points in time to look at reliability where correlation coefficients ranged from .60–.83 (Artaria & Henneberg, 2000; Bergsten-Brucefors, 1976; Cooper et al., 2006; Damon & Bajema, 1974; Hediger & Stine, 1987; Livson & McNeill, 1962; Must et al., 2002). However, in the study at hand, reliability coefficients across all 12 time points and both methodologies (i.e., in-person and phone interviews combined) was moderate (Cohen & Cohen, 1983) (ICC = 0.64), as was the reliability for phone interviews (ICC = 0.64). However, reliability for the in-person interviews was identified as strong (ICC = 0.77). Similarly we found that test-retest reliability was greater for in-person interviews compared to phone interviews suggesting investigators may want to consider utilizing in-person interviews by trained clinicians in order to obtain the most reliable reports of age of menarche. Our methodology also included probes (with both in-person and phone

interviews) that might enhance the recollection of the event. Although not tested by this research, this methodology may have contributed to obtaining higher test-retest reliability and greater method reliability for in-person interviews. Alternatively, future studies could test whether differences lie with method of delivery (in-person vs. phone) or experience (clinician vs. research assistant).

It also was surprising to see such large discrepancy across methodology *and* time, particularly since menarche was such a recent event for most girls. On average, differences in reports varied by 2.3 years, with nearly 9% of girls demonstrating differences over 5 years. Again discrepancies were smaller across in-person interviews (average 7 months) compared to an average 2.2 years across the phone interviews, with nearly 67% of the girls being "off" by more than one year.

Why do girls have such difficulty remembering an event that many clinicians and parents consider "memorable"? Indeed, for some girls menarche was memorable and they reported the actual date and location where it occurred (e.g., October 13 at my sixth grade best friend's birthday). Other girls may have been reluctant to talk about an event they view as private or as something that they want to forget; a response style reflecting "pubertal amnesia" (Petersen & Taylor, 1980). Alternatively, for some girls the "event" may be less clear cut. Premenarcheal girls may experience light spotting on 1–2 occasions and then have a lag of several months before a full menstrual period (i.e., true menarche) occurs. In such instances, they are confused about what constitutes their first period. Further, anecdotal reports by study personnel suggested that many girls were not knowledgeable about other facts related to reproductive maturity such as menstrual cycles, day in cycle, regularity and the like. This lack of knowledge may complicate responding to questions in an interview. Thus, when assessing age of menarche, clinicians and researchers should ensure that a girl understands exactly what question is being asked of her and what does or does not constitute a menstrual period.

Other important findings emerged from these data. For example, it is puzzling that reliability of self-reporting during in-person interviews was lower for non-white compared to white girls, whereas race was not a significant factor in the reliability of phone interviews. This difference might reflect the impact of racial congruence when interviewing patients, as the vast majority of the clinicians were white and their race would only have been identifiable during in-person interviews. However, the impact of racial congruence on clinician-patient interaction is not consistent within the literature. The differences in reliability may also reflect ethnic differences in the comfort of discussing the menstrual cycle or salience of menarche within specific race or ethnic groups. Additionally, SES differences in reliability of self-reports were evident for phone interviews, such that girls with lower SES had higher difference scores indicating less consistency in reporting. Anecdotally, many participants used cell phones for the phone interviews. It may be that lower SES girls were not able to devote full attention to the interview due to concerns of limited cell phone minutes.

Another important finding is that pubertal timing categorization (early, on-time, late) changed for 22.7% of the sample depending upon whether timing was categorized from their report at the last visit compared to the first. Other researchers have reported a similar change in category (15–20%) (Smolak, Krieg, Hayward, Shisslak, & Taylor, 2007). This difference has tremendous implications for those conducting and interpreting research. If large discrepancies in categorization occur, the conclusions made about the influence of timing of puberty on an outcome (i.e., substance use, early pregnancy, breast cancer) also may change.

Exploring recency effects on the variability of self-reported age at menarche, we found that greater gynecological age was associated with less consistency in reports of age at

menarche. However, this finding did not hold once girls who were premenarcheal at the first visit were removed from analyses. Such an effect (or lack thereof) suggests that beyond an initial recency effect to the onset of menarche, girls may be "consistently inconsistent" in reporting. Reasons for inconsistency should be explored in future research.

In spite of the strengths of this study, limitations exist. First, determining accuracy of age at menarche is difficult since a "gold standard" for determining such is not readily available. A multi-informant approach may use parent report to corroborate the child's report, but this approach still lacks an objective marker and also reflects a subjective report since the amount of communication with parents about menarche may vary widely. Some clinicians suggest maternal involvement may be helpful in gynecological care of adolescents (Rosenthal, Cohen, Burklow, & Hillard, 1996). Empirically, Caspi and Moffit report a moderately sized correlation of .66 between maternal report of her daughter's menarcheal age with daughter's self-report (Caspi & Moffitt, 1991). Second, the impact of the slight inadvertent variation in questioning regarding age at menarche (i.e., age in years vs. what year) may have influenced the results. However, in the latter we then would translate when they said age 13 to the year (i.e., 2001) on the form.

In sum, age at menarche is a variable that is crucial in both clinical and research settings. This study with adolescents provides evidence that self-reported age at menarche has moderate to strong reliability, but its reliability is not perfect. Methods are available (i.e., in-person clinician interview with probes) that can effectively be used and may increase reliability and decrease discrepancy across time or method. Capitalizing on methods to improve the reliability of reports is essential since self-report of menarche is commonly used in clinical and research settings. Consumers of the literature and/or those designing studies should consider the impact of methodology and discrepancies associated with each method as they interpret published findings, make clinical decisions, and design new studies.

References

Apter D, Reinila M, Vihko R. Some endocrine characteristics of early menarche, a risk factor for breast cancer, are preserved into adulthood. International Journal of Cancer. 1989; 44(5):783–787.10.1002/ijc.2910440506

Artaria MD, Henneberg M. Why did they lie? Socio-economic bias in reporting menarcheal age. Annals of Human Biology. 2000; 27(6):561–569.10.1080/03014460050178669 [PubMed: 11110222]

Bergsten-Brucefors A. A note on the accuracy of recalled age at menarche. Annals of Human Biology. 1976; 3(1):71–73.10.1080/03014467600001151 [PubMed: 945036]

Caspi A, Moffitt TE. Individual differences are accentuated during periods of social change: The sample case of girls at puberty. Journal of Personality and Social Psychology. 1991; 61(1):157–168.10.1037/0022-3514.61.1.157 [PubMed: 1890586]

Chumlea WC, Schubert CM, Roche AF, Kulin HE, Lee PA, Himes JH, Sun SS. Age at menarche and racial comparisons in US girls. Pediatrics. 2003; 111(1):110–113.10.1542/peds.111.1.110 [PubMed: 12509562]

Cohen, J.; Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum; 1983.

Cooper R, Blell M, Hardy R, Black S, Pollard TM, Wadsworth MEJ, Kuh D. Validity of age at menarche self-reported in adulthood. Journal of Epidemiology and Community Health. 2006; 60(11):993–997.10.1136/jech.2005.043182 [PubMed: 17053289]

Damon A, Bajema CJ. Age at menarche: Accuracy of recall after thirty-nine years. Human Biology. 1974; 46(3):381–384. [PubMed: 4426599]

Deardorff J, Gonzales NA, Christopher FS, Roosa MW, Millsap RE. Early puberty and adolescent pregnancy: the influence of alcohol use. Pediatrics. 2005; 116:1451–1456.10.1542/peds.2005-0542 [PubMed: 16322170]

Dorn LD, Dahl RE, Woodward HR, Biro F. Defining the boundaries of early adolescence: A user's guide to assessing pubertal status and pubertal timing in research with adolescents. Applied Developmental Science. 2006; 10(1):30–56.10.1207/s1532480xads1001_3

Dorn LD, Nottelmann ED, Susman EJ, Inoff-Germain G, Cutler GB, Chrousos GP. Variability in hormone concentrations and self-reported menstrual histories in young adolescents: Menarche as an integral part of a developmental process. Journal of Youth and Adolescence. 1999; 28(3):283–304.10.1023/A:1021680726753

Dorn LD, Susman EJ, Pabst S, Huang B, Kalkwarf H, Grimes S. Association of depressive symptoms and anxiety with bone mass and density in ever-smoking and never-smoking adolescent girls [Journal Article]. Archives of Pediatrics & Adolescent Medicine. 2008; 162(12):1181–1188.10.1001/archpedi.162.12.1181 [PubMed: 19047547]

Hamilton AS, Mack TM. Puberty and Genetic Susceptibility to Breast Cancer in a Case-Control Study in Twins. The New England Journal of Medicine. 2003; 348:2313–2322. [PubMed: 12788995]

Hediger ML, Stine R. Age at menarche based on recall information. Annals of Human Biology. 1987; 14(2):133–142.10.1080/03014468700006862 [PubMed: 3606026]

Hollingshead, AB. A Four-Factor Classification of Social Status. New Haven, CT: Yale University CT Press; 1975.

Koo MM, Rohan TE. Accuracy of short-term recall of age at menarche. Annals of Human Biology. 1997; 24(1):61–64.10.1080/03014469700004782 [PubMed: 9022907]

Koprowski C, Coates RJ, Bernstein L. Ability of young women to recall past body size and age at menarche. Obesity Research. 2001; 9(8):478–485.10.1038/oby.2001.62 [PubMed: 11500528]

Livson N, McNeill D. The accuracy of recalled menarche. Human Biology. 1962; 34:218–221. [PubMed: 13931081]

Marklein E, Negriff S, Dorn LD. Pubertal timing, friend smoking, and substance use in adolescent girls. Prevention Science. 2009; 10(2):141–150.10.1007/s11121-008-0120-y [PubMed: 19067163]

Mendle J, Turkheimer E, Emery RE. Detrimental psychological outcomes associated with early pubertal timing in adolescent girls. Developmental Review. 2007; 27(2):151–171.10.1016/j.dr.2006.11.001 [PubMed: 20740062]

Miyazaki Y, Raudenbush SW. Tests for linkage of multiple cohorts in an accelerated longitudinal design. Psychological Methods. 2000; 5(1):44–63.10.1037/1082-989X.5.1.44 [PubMed: 10937322]

Must A, Phillips SM, Naumova EN, Blum M, Harris S, Dawson-Hughes B, Rand WM. Recall of early menstrual history and menarcheal body size: After 30 years, how well do women remember? American Journal of Epidemiology. 2002; 155(7):672–679.10.1093/aje/155.7.672 [PubMed: 11914195]

Obeidallah D, Brennan RT, Brooks-Gunn J, Earls F. Links between pubertal timing and neighborhood contexts: Implications for girls' violent behavior. Journal of the American Academy of Child & Adolescent Psychiatry. 2004; 43(12):1460–1468.10.1097/01.chi.0000142667.52062.1e [PubMed: 15564815]

Petersen, AC.; Taylor, B. The biological approach to adolescence: Biological change and psychosocial adaptation. In: Adelson, J., editor. Handbook of the Psychology of Adolescence. New York: Wiley; 1980. p. 115-155.

Remsberg K, Demerath E, Schubert C, Chumlea W, Sun S, Siervogel R. Early menarche and the development of cardiovascular disease risk factors in adolescent girls: The Fels longitudinal study. Journal of Clinical Endocrinology and Metabolism. 2005; 90(5):2718–2724.10.1210/jc.2004-1991 [PubMed: 15728207]

Rosenthal SL, Cohen SS, Burklow KA, Hillard PA. Family involvement in the gynecologic care of adolescents. Journal of Pediatric and Adolescent Gynecology. 1996; 9(2):59–65. [PubMed: 8795778]

Singer, JD.; Willett, JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. New York: Oxford University Press; 2003.

Smolak L, Krieg DB, Hayward C, Shisslak CM, Taylor CB. The reliability of self-reported menarcheal timing. The Journal of Early Adolescence. 2007; 27(3):386–397.10.1177/0272431607302007

Table 1

Bivariate correlations of self-reported age at menarche for annual in-person interviews and quarterly phone interviews

	Year 1				Year 2				Year 3			
	Annual	**3 mo**	**6 mo**	**9 mo**	**Annual**	**3 mo**	**6 mo**	**9 mo**	**Annual**	**3 mo**	**6 mo**	**9 mo**
Year 1												
Annual (n = 209)	--											
3 month (n = 212)	.74	--										
6 month (n = 211)	.74	.68	--									
9 month (n = 218)	.67	.63	.72	--								
Year 2												
Annual (n = 192)	**.78**	.72	.78	.75	--							
3 month (n = 211)	.63	.67	.68	.64	.69	--						
6 month (n = 208)	.67	.68	.74	.76	.74	.74	--					
9 month (n = 206)	.61	.59	.70	.64	.76	.61	.66	--				
Year 3												
Annual (n = 203)	**.74**	.64	.70	.63	**.81**	.67	.66	.64	--			
3 month (n = 203)	.63	.67	.66	.70	.66	.65	.65	.59	.73	--		
6 month (n = 194)	.61	.55	.61	.63	.63	.61	.65	.60	.65	.69	--	
9 month (n = 194)	.60	.64	.62	.63	.64	.61	.60	.63	.68	.75	.62	--

Note. All correlations are significant at $p < .01$. Annual represents in-person interviews, whereas 3 month, 6 month, and 9 month represent phone interviews. **Bolded** correlations are between annual in person visits. n at each assessment represents non-missing reports of age at menarche (i.e., does not include girls who were premenarcheal or missed the visit); 253 girls reported age at menarche at least twice across the course of the study.

Table 2

Descriptive statistics and difference scores * of self-reported age at menarche in a longitudinal study of 253 adolescent girls

	Total ***M (SD)***	**Age 11** ***M (SD)***	**Age 13** ***M (SD)***	**Age 15** ***M (SD)***	**Age 17** ***M (SD)***
n	253	43	52	87	71
Age at Menarche First Report	12.40 (1.25)	12.20 (0.94)[a]	12.44 (1.11)[a]	12.37 (1.22)[a]	12.53 (1.52)[a]
In-person Interview Mean Age at Menarche	12.43 (1.17)	12.15 (.89)[a]	12.53 (1.07)[a]	12.40 (1.14)[a]	12.55 (1.38)[a]
Phone Interview Mean Age at Menarche	12.19 (1.30)	12.10 (1.02)[a]	12.05 (1.36)[a]	12.12 (1.25)[a]	12.42 (1.46)[a]
In-Person Interview Mean – Phone Interview Mean	0.52 (0.50)	0.29 (0.37)[a,b]	0.55 (0.53)[a]	0.61 (0.55)[a,b]	0.51 (0.44)[b]
Average difference (years): In-person & phone *	2.33 (2.20)	1.39 (2.40)[a]	2.57 (2.55)[b]	2.60 (2.23)[b]	2.39 (1.57)[b]
Average difference (years): In-person interview only *	0.58 (0.74)	0.28 (0.42)[a]	0.62 (0.63)[a]	0.66 (0.72)[a]	0.61 (0.94)[a]
Average difference (years): Phone interview only *	2.16 (1.58)	1.31 (2.41)[a]	2.33 (2.57)[b]	2.45 (2.28)[b]	2.16 (1.55)[b]

Note.

* Difference scores represent discrepancy estimates calculated as the difference between the maximum reported age at menarche and minimum reported age at menarche across all assessments (in-person and phone), across in-person interviews only, and across phone interviews only. Means in same row not sharing superscripts (a, b, c) differ at $p < .001$ in the Bonferroni significant difference comparison. Annual in-person interviews were conducted three times and phone interviews were conducted quarterly (i.e, 3, 6, and 9 months; 9 total) after each annual visit.

Table 3

Age at menarche (years) and difference scores * by racial group and pubertal timing in 253 adolescent girls

	Racial Group[1]		Pubertal Timing Group[2]		
	White *M (SD)*	**Non-White** *M (SD)*	**Early** *M (SD)*	**On-Time** *M (SD)*	**Late** *M (SD)*
n	156	97	38	174	41
Age at Menarche First Report	12.63 (0.10)[a]	12.05 (0.13)[b]	10.54 (0.12)[a]	12.39 (0.06)[b]	14.18 (0.12)[c]
In-person Interview Mean Age at Menarche	12.64 (0.10)[a]	12.12 (0.12)[b]	10.77 (0.12)[a]	12.44 (0.06)[b]	14.00 (0.12)[c]
Phone Interview Mean Age at Menarche	12.41 (0.10)[a]	11.86(0.13)[b]	10.69 (0.16)[a]	12.18 (0.08)[b]	13.65 (0.16)[c]
In-Person Interview Mean – Phone Interview Mean	0.52 (0.04)[a]	0.52 (0.05)[a]	0.63 (0.08)[a]	0.51 (0.04)[a]	0.44 (0.08)[a]
Average difference (years): In-person & phone *	2.12 (0.18)[a]	2.66 (0.23)[a]	3.06 (0.35)[a]	2.20 (0.17)[a]	2.20 (0.35)[a]
Average difference (years): In-person interview only *	0.50 (0.06)[a]	0.70 (0.08)[b]	0.74 (0.12)[a]	0.54 (0.06)[a]	0.59 (0.12)[a]
Average difference (years): Phone interview only *	1.96 (0.18)[a]	2.46 (0.23)[a]	2.75 (0.36)[a]	2.03 (0.17)[a]	2.11 (0.35)[a]

Note.

* Difference scores represent discrepancy estimates calculated as the difference between the maximum reported age at menarche and minimum reported age at menarche across all assessments (in-person and phone), across in-person interviews only, and across phone interviews only.

1 Means are adjusted at sample mean of age = 15.1 years and SES = 37.43.

2 Means are adjusted at sample mean of age = 15.1 years. Means in same row not sharing superscripts (a, b, c) differ at $p < .01$ in the Bonferroni significant difference comparison. Comparisons between racial groups and pubertal timing groups were estimated separately. Annual clinic visits were conducted three times and phone visits were conducted quarterly at 3, 6, and 9 months (9 total) after each annual visit.

Table 4

Change in pubertal timing categories based on Year 1 reports versus Year 3 reports

	Pubertal Timing Year 3 Counts			
Pubertal Timing Year 1 Counts	**Early**	**On-Time**	**Late**	**Total**
Early	18 (8.8%)	7 (3.4%)	0 (0%)	25 (12.3%)
On-Time	14 (6.9%)	119 (58.6%)	9 (4.4%)	142 (70.0%)
Late	0 (0%)	16 (7.9%)	20 (9.9%)	36 (17.7%)
Total	32 (15.7%)	142 (70.0%)	29 (14.3%)	203 (100%)

Note. 22.7% (N = 46) demonstrate change in timing categories; chi-square comparison test indicates actual counts differ from expected counts: $\chi^2(4) = 125.3$, $p < .001$