

Comparison of Clinical, Maternal, and Self Pubertal Assessments: Implications for Health Studies

Mary Beth Terry, PhD,^{a,b} Mandy Goldberg, MPH,^a Sarah Schechter, BA,^c Lauren C. Houghton, PhD,^a Melissa L. White, MSW,^a Karen O'Toole, RN,^d Wendy K. Chung, MD, PhD,^{b,e} Mary B. Daly, MD, PhD,^f Theresa H.M. Keegan, PhD, MS,^g Irene L. Andrulis, PhD,^{h,i} Angela R. Bradbury, MD,^{j,k} Lisa Schwartz, PhD,^l Julia A. Knight, PhD,^{i,m} Esther M. John, PhD,^{n,o} Sandra S. Buys, MD^d

abstract

BACKGROUND: Most epidemiologic studies of puberty have only 1 source of pubertal development information (maternal, self or clinical). Interpretation of results across studies requires data on reliability and validity across sources.

METHODS: The LEGACY Girls Study, a 5-site prospective study of girls aged 6 to 13 years ($n = 1040$) collected information on breast and pubic hair development from mothers (for all daughters) and daughters (if ≥ 10 years) according to Tanner stage (T1–5) drawings. At 2 LEGACY sites, girls ($n = 282$) were also examined in the clinic by trained professionals. We assessed agreement (κ) and validity (sensitivity and specificity) with the clinical assessment (gold standard) for both the mothers' and daughters' assessment in the subcohort of 282. In the entire cohort, we examined the agreement between mothers and daughters.

RESULTS: Compared with clinical assessment, sensitivity of maternal assessment for breast development was 77.2 and specificity was 94.3. In girls aged ≥ 11 years, self-assessment had higher sensitivity and specificity than maternal report. Specificity for both mothers and self, but not sensitivity, was significantly lower for overweight girls. In the overall cohort, maternal and daughter agreement for breast development and pubic hair development (T2+ vs T1) were similar (0.66, [95% confidence interval 0.58–0.75] and 0.69 [95% confidence interval 0.61–0.77], respectively), but declined with age. Mothers were more likely to report a lower Tanner stage for both breast and pubic hair compared with self-assessments.

CONCLUSIONS: These differences in validity should be considered in studies measuring pubertal changes longitudinally when they do not have access to clinical assessments.



^aDepartment of Epidemiology, Columbia University Mailman School of Public Health, New York, New York; ^bHerbert Irving Comprehensive Cancer Center and Departments of ^cPediatrics and Medicine, Columbia University Medical Center, New York, New York; ^cColumbia University College of Physicians and Surgeons, New York, New York; ^dDepartment of Medicine, University of Utah Health Sciences Center, Huntsman Cancer Institute, Salt Lake City, Utah; ^eDepartment of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, Pennsylvania; ^fDivision of Hematology and Oncology, University of California (UC) Davis School of Medicine, and UC Davis Comprehensive Cancer Center, Sacramento, California; ^gDepartment of Molecular Genetics and ^hDivision of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; ⁱLunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Canada; ^jDepartments of ^kMedicine and Hematology/Oncology, ^lMedical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; ^mDepartment of Pediatrics, Division of Oncology, The Children's Hospital of Philadelphia and Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; ⁿCancer Prevention Institute of California, Fremont, California; and ^oDepartment of Health Research and Policy (Epidemiology), Stanford University School of Medicine, and Stanford Cancer Institute, Stanford, California USA

Dr Terry conceptualized the design of the overall parent study and the analyses presented in this study, directed the data analysis and interpretation, and drafted the initial manuscript; Drs John, Andrulis, Daly, Buys, and Bradbury conceptualized the design of the overall parent study and

WHAT'S KNOWN ON THIS SUBJECT: Mothers and girls underreport pubertal development relative to clinical measurements. Many epidemiologic studies base pubertal assessment on a single source (clinical, maternal, or self) and/or change sources over time as girls age and clinical and maternal assessments become more difficult.

WHAT THIS STUDY ADDS: Maternal breast Tanner assessments were more valid than self-assessments compared with clinical only before age 11 years. Among girls ≥ 11 years, self-assessments had higher sensitivity and specificity. These differences in validity should be considered in studies measuring pubertal changes longitudinally.

To cite: Terry MB, Goldberg M, Schechter S, et al. Comparison of Clinical, Maternal, and Self Pubertal Assessments: Implications for Health Studies. *Pediatrics*. 2016;138(1):e20154571

Early age at menarche is associated with increased risk of breast cancer, ovarian cancer, Type 2 diabetes, and other health conditions.¹⁻⁹ Earlier menarche has also been shown to be associated with higher rates of depression, anxiety, eating disorders, smoking, and substance abuse.¹⁰⁻¹⁵ Studies have also found that early menarche was associated with a 13% to 16% increased risk of all-cause mortality, even after adjusting for BMI.^{16,17}

Age at menarche has been used as a proxy for onset of pubertal development. Markers of puberty, including breast and pubic hair development, often begin several years before first menses.¹⁸ Age at menarche started to decline in the early 1900s but has remained fairly constant at 12 to 13 years for the past 60 years.¹⁹ Over the past generation, there has been a dramatic decline in the age at onset of breast development.^{20,21} Because the age of onset of breast development is decreasing without corresponding decreases in age at first menses, menarche is an inaccurate indicator of pubertal onset.

Independent of age at menarche, earlier age at breast development has been associated with a 20% increased breast cancer risk in a prospective cohort of 104 931 women.²² Importantly, the study confirmed that the number of years between onset of breast development and menarche, referred to as tempo, may affect risk over and above the age at attainment of any single pubertal milestone.²² Thus, the window between onset of breast development and first menses has become wider in most populations worldwide,^{23,24} suggesting a possible future increase in breast cancer incidence.^{21,25}

Pubertal onset, defined as the beginning of breast and/or pubic hair development, is often assessed by using Tanner staging,²⁶ which is routinely used in clinical evaluations.

Tanner stages range from T1 to T5, with T1 referring to prepubertal development and T5 indicating full development. T2 is the first appearance of either breast buds or pubic hair and is used to indicate the onset of puberty. Tanner stage is generally assessed by a clinician but can also be evaluated by self- or maternal report using drawings of Tanner stages with explanatory text.²⁷

Most studies of pubertal development use only a single source of Tanner staging. For example, of the large epidemiologic studies, 4 use clinical staging by a trained professional,^{20,21,28,29} 1 uses self-assessment,³⁰ and 1 uses maternal staging, self-staging, or a combined measure.³¹ To interpret data across studies, it is important to determine whether pubertal development information differs by source, which can only be examined within the few cohorts that collect pubertal data from multiple sources.³²⁻³⁴ It is also important to determine whether factors such as age, family history of breast cancer, BMI, and race/ethnicity affect measurements to assess whether there could be any differential bias in pubertal assessment by source of Tanner staging (eg, clinical, maternal, or self). We report results from reliability and validity analyses comparing maternal and self-assessment to clinical staging in a large study of girls' health and development.

METHODS

LEGACY Girls Study

The LEGACY Girls Study is a 5-site prospective study of pubertal development in girls ages 6 to 13 years at recruitment, half of whom have a family history of breast cancer (for details, see John et al³⁵). Classification of pubertal timing is based on the Growth and Development Questionnaire completed every 6 months by

mothers/guardians for girls of all ages and by girls aged ≥ 10 years. It includes questions on age at menarche and breast and pubic hair development using line drawings that show 5 stages of development, Tanner T1 through T5, for breast and pubic hair.²⁶ Because 97% of girls participated in LEGACY with their biological mother,³⁵ we will refer to the guardian as the mother from here on. The girls' self-assessment will be used for sensitivity analyses.

At 2 sites, we also collected clinical measures of breast development. Three clinical raters from New York and 1 from Utah were trained concurrently on the determination of Tanner breast stage using visual inspection along with palpation when necessary. Palpation was used in addition to visual assessment in a subset of girls, if they consented, to help the clinical raters distinguish between Tanner stages 1 and 2. Palpation was used in 32.2% of baseline clinical Tanner measures. The addition of palpation did not change the clinical Tanner rating in 92.1% of instances when palpation was used. The clinical raters did not evaluate pubic hair Tanner stage. Clinician interrater reliability for breast Tanner stage was almost perfect, with weighted κ scores ranging from 0.93 to 1.00 and κ for T2+ versus T1 ranging from 0.94 to 1.00 (based on 181 assessments with 2 clinical raters, see Supplemental Table 7).

Statistical Analysis

We calculated measures of validity by treating the clinical assessment as the gold standard: sensitivity (percentage correctly identifying the onset of breast development, T2+) and specificity (percentage correctly identifying prepubertal stage, T1), separately for mothers and daughters. We calculated concordance (overall agreement), κ (T1 vs T2+) and weighted κ (for T1-T5) for the first visit with clinical

TABLE 1 Population Characteristics of the LEGACY Girls Study Cohort

	Girls with Clinical Assessment (<i>n</i> = 282)	All Girls (<i>n</i> = 1040)
	Median (Range)	Median (Range)
Age, y	9.5 (6.0 to 15.1)	10.1 (5.2 to 16.6)
BMI, kg/m ²	16.5 (11.1 to 37.4)	17.0 (10.3 to 49.5)
BMI for age z score	0.03 (−4.9 to 2.9)	0.04 (−6.4 to 3.0)
Weight, kg	31.0 (17.7 to 92.7)	34.5 (15.4 to 144.8)
Weight for age z score	0.13 (−3.3 to 3.4)	0.27 (−3.4 to 3.9)
Height, cm	136.3 (107.0 to 181.8)	141.6 (107.0 to 181.8)
Height for age z score	0.24 (−2.4 to 3.7)	0.54 (−2.8 to 6.6)

staging available for New York and Utah girls. For both breast and pubic hair Tanner staging, we calculated κ between maternal and self-assessment for girls ages ≥ 10 years from all study sites. κ statistics were interpreted by strength of agreement as follows: <0.00 , poor; $0.00-0.20$, slight; $0.21-0.40$, fair; $0.41-0.60$, moderate; $0.61-0.80$, substantial; $0.81-1.00$, almost perfect.³⁶

We examined differences in validity and agreement by age, breast cancer family history in first- and second-degree relatives, BMI, race/ethnicity, and study site. We calculated percentiles and z scores for each girl's age based on age and gender using the Centers for Disease Control and Prevention SAS source code³⁷ and compared girls with a BMI <85 th percentile with those with a BMI ≥ 85 th percentile.³⁸ Race/ethnicity was mother-reported and categorized as non-Hispanic white, non-Hispanic black, Hispanic, Asian/Pacific Islander, or other for analyses using the full cohort. We combined girls identified as non-Hispanic black, Asian/Pacific Islander, or

other into 1 group for analyses using clinical staging because of small numbers in each category. We formally tested differences in sensitivity and specificity of maternal and self-assessment by each characteristic using a 2-sample test of proportions.

We used polytomous logistic regression to examine factors (ie, girl's age, family history status, BMI at visit [<85 th percentile vs ≥ 85 th percentile], race/ethnicity, and study site) associated with discordant clinical and maternal assessments of breast onset compared with the referent group of girls with concordant staging.

RESULTS

Clinical Versus Maternal Assessment of Breast Development

Girls with clinical assessments (*n* = 282) were slightly younger and smaller than girls in the overall cohort (*n* = 1040; Table 1). Of the clinical and maternal assessments, 73% were in agreement (Table 2). When there was disagreement,

mothers were more likely to underestimate than overestimate their daughter's breast stage. Unweighted κ for all 5 breast Tanner stages was 0.54 (95% confidence interval [CI] 0.47 to 0.62), and weighted κ was 0.72 (95% CI 0.67 to 0.78), indicating that discrepant assessments typically differed by only 1 stage. The κ for breast T2+ compared with T1 was 0.73 (95% CI 0.65 to 0.81), indicating substantial agreement between maternal and clinical assessments for the onset of breast development. Seventy-seven percent of mothers accurately identified when their daughters were T2+ (sensitivity), and 94.3% accurately identified when their daughters were T1 (specificity).

Validity of maternal report, when compared with the clinical assessment as the gold standard, differed significantly by age (sensitivity and specificity) and BMI (specificity; Table 3). Sensitivity of maternal report of T2+ was 56.0% for girls <10 years and 82.7% for those ≥ 10 years of age; specificity was 96.4% and 79.0%, respectively. Specificity was lower for mothers of overweight girls (≥ 85 th percentile) (73.7% vs 97.0%). When we examined discordances between maternal and clinical reports using polytomous logistic regression models, only daughter's BMI ≥ 85 th percentile was associated with maternal overestimation of breast onset (odds ratio [OR] = 6.0, 95% CI 1.5 to 23.1).

TABLE 2 Agreement Between Maternal Assessment of Breast Development Versus Clinical Assessment as Gold Standard

Maternal rating, <i>n</i> (%)	Clinical Rating, <i>N</i> (%)					Total
	1	2	3	4	5	
1	150 (53.2)	27 (9.6)	1 (0.4)	0 (0)	0 (0)	178
2	8 (2.8)	25 (8.9)	7 (2.5)	3 (1.1)	0 (0)	43
3	1 (0.4)	7 (2.5)	19 (6.7)	4 (1.4)	1 (0.3)	32
4	0 (0)	1 (0.4)	4 (1.4)	12 (4.3)	8 (2.8)	25
5	0 (0)	0 (0)	0 (0)	4 (1.4)	0 (0)	4
Total	159	60	31	23	9	282

TABLE 4 Agreement Between Maternal and Self-Assessment of Breast Development in Girls Aged ≥ 10 Years

Description	<i>n</i>	Tanner Staging—5 Categories		Tanner Stage T2+ vs T1
		κ (95% CI)	Weighted κ (95% CI)	κ (95% CI)
All girls, age ≥ 10	457	0.51 (0.45 to 0.56)	0.68 (0.64 to 0.72)	0.66 (0.58 to 0.75)
Breast cancer family history				
Positive	250	0.55 (0.47 to 0.63)	0.70 (0.64 to 0.76)	0.70 (0.58 to 0.82)
Negative	207	0.45 (0.36 to 0.54)	0.66 (0.59 to 0.72)	0.62 (0.50 to 0.74)
BMI percentile				
≥ 85 th	85	0.44 (0.31 to 0.57)	0.65 (0.55 to 0.74)	0.38 (0.05 to 0.72)
< 85 th	356	0.51 (0.45 to 0.58)	0.68 (0.63 to 0.73)	0.68 (0.59 to 0.77)
Race/ethnicity				
Non-Hispanic white	269	0.51 (0.43 to 0.58)	0.67 (0.61 to 0.73)	0.65 (0.55 to 0.76)
Hispanic	91	0.43 (0.29 to 0.56)	0.61 (0.50 to 0.71)	0.73 (0.51 to 0.95)
Non-Hispanic black	37	0.44 (0.24 to 0.65)	0.69 (0.57 to 0.81)	0.72 (0.36 to 1.00)
Asian	45	0.51 (0.34 to 0.69)	0.69 (0.58 to 0.81)	0.54 (0.28 to 0.80)
Other	15	0.80 (0.53 to 1.00)	0.80 (0.53 to 1.00)	1.00 (1.00 to 1.00)

TABLE 5 Agreement Between Maternal and Self-Assessment of Breast and Pubic Hair Development (5 Tanner Stages) by Age

Daughter's Age (y)	Breast Development			Pubic Hair Development		
	Weighted κ (95% CI)	Mean Tanner Stage		Weighted κ (95% CI)	Mean Tanner Stage	
		Maternal	Self		Maternal	Self
10	0.63 (0.53 to 0.73)	1.7	1.9	0.65 (0.55 to 0.75)	1.6	1.9
11	0.60 (0.49 to 0.71)	2.1	2.3	0.58 (0.47 to 0.68)	2.1	2.5
12	0.45 (0.33 to 0.58)	2.9	3.0	0.59 (0.48 to 0.70)	3.1	3.4
13	0.61 (0.49 to 0.73)	3.5	3.6	0.61 (0.44 to 0.77)	4.0	4.0
$\geq 14^a$	0.40 (0.14 to 0.67)	3.9	4.0	0.43 (0.09 to 0.77)	4.4	4.5

^a Ages 14, 15, and 16 were combined because only 34 girls were aged ≥ 14 y at recruitment.

TABLE 6 Agreement Between Maternal and Self-Assessment of Pubic Hair Development in Girls Aged ≥ 10 Years

Description	<i>n</i>	Tanner Staging—5 Categories		Tanner Stage T2+ vs. T1
		κ (95% CI)	Weighted κ (95% CI)	κ (95% CI)
All girls, aged ≥ 10	436	0.57 (0.51 to 0.62)	0.72 (0.68 to 0.76)	0.69 (0.61 to 0.77)
Breast cancer family history				
Positive	234	0.57 (0.50 to 0.65)	0.74 (0.69 to 0.80)	0.65 (0.52 to 0.77)
Negative	202	0.55 (0.47 to 0.64)	0.69 (0.63 to 0.76)	0.73 (0.62 to 0.84)
BMI percentile				
≥ 85 th	83	0.54 (0.41 to 0.67)	0.70 (0.59 to 0.80)	0.75 (0.55 to 0.94)
< 85 th	337	0.57 (0.51 to 0.64)	0.73 (0.68 to 0.77)	0.69 (0.60 to 0.78)
Race/ethnicity				
Non-Hispanic white	261	0.58 (0.51 to 0.65)	0.73 (0.67 to 0.78)	0.69 (0.59 to 0.79)
Hispanic	85	0.49 (0.35 to 0.62)	0.63 (0.52 to 0.75)	0.73 (0.47 to 0.98)
Non-Hispanic black	34	0.55 (0.34 to 0.76)	0.74 (0.61 to 0.87)	0.64 (0.18 to 1.00)
Asian	44	0.53 (0.35 to 0.71)	0.71 (0.59 to 0.83)	0.55 (0.31 to 0.79)
Other	12	0.68 (0.37 to 0.99)	0.77 (0.53 to 1.00)	1.00 (1.00 to 1.00)

by family history, race/ethnicity, and study site (Table 6). Girls were more likely to report a higher pubic Tanner stage compared with their mother (Table 5, and also Supplemental Table 9 for details).

DISCUSSION

Our study demonstrates that the validity of assessments of pubertal

development milestones differs by source of information. Compared with clinical reports, both mothers and daughters were more likely to underreport breast Tanner stage. Compared with the gold standard of clinical assessment, maternal assessment had a higher sensitivity and higher specificity in girls aged 10 years. At age ≥ 11 years, self-assessment had a higher sensitivity and specificity for breast Tanner

staging compared with clinical report. Our results suggest that maternal assessment of breast onset before age 11 years is more accurate compared with self-assessment. For girls aged ≥ 11 years, self-assessment is more accurate. We did not have a clinical assessment for pubic hair development. Maternal and self-assessment had moderate agreement and had a similar range for both breast and pubic Tanner

assessment. Girls were much more likely to report higher breast and pubic hair stage. Therefore, studies using only maternal assessment in older girls will result in a higher average age at these pubertal milestones.

Accuracy of Breast Development Versus Clinical Assessment as Gold Standard

Agreement between clinical raters was almost perfect in our study, with κ ranging from 0.85 to 1.00 for T1 to T5 and 0.94 to 1.00 for T2+ compared with T1. These κ values were slightly higher than those reported in other studies, where estimates have ranged from 0.67 to 0.90, indicating substantial agreement,^{20,21,32,33,39} with some exceptions.⁴⁰ We found that almost three-quarters of mothers accurately assessed breast Tanner stage compared with the clinical assessment. Mothers are generally found to be more reliable reporters of breast development than daughters, compared with physician ratings.^{32,41}

In contrast, the majority of the girls in our study did not correctly stage their own breast development, which is consistent with other small studies of girls with similar age ranges.^{40,42,43} In our study, girls tended to underreport their own breast development, perhaps as a result of embarrassment about bodily changes and breast development during puberty.⁴⁴ The literature on bias in self-assessment has been inconsistent,^{32,33,43} but there is some evidence that suggests that age and stage of development influence the direction of bias, with younger, less developed girls more likely to overestimate breast development and older, more developed girls more likely to underestimate breast development.^{43,45,46} We observed that girls aged ≥ 11 years were more accurate compared with the clinical gold standard than their mothers.

Maternal Versus Self-Staging of Pubic Hair Development

We assessed agreement between maternal and self-assessment for pubic hair measurements but did not have clinical measurements for validity. Previous studies comparing clinical and self-assessment reported agreement ranging from 0.37 to 0.91.^{40,43,45-47} Previous studies have shown a wide range of accuracy for self-assessed pubic hair staging,^{40,45,47,48} although 2 studies that also examined mother report suggest that self-staging may be more reliable than maternal staging for pubic hair development.^{33,41} We found that girls were more likely to report a higher stage of pubic hair development than their mothers.

Age-Related Differences Between Maternal and Self-Assessments

Our study can help reconcile opposing conclusions between 2 recent reports on the reliability and validity of Tanner staging in contemporary cohorts. In a Danish study, the authors argued that although clinical measures are preferred, self-assessments could be used in large epidemiologic studies if the main purpose was to determine whether the onset of puberty occurred (breast Tanner 2+ vs T1).³³ The Chilean study concluded that maternal reports could be used for cohorts without clinical measures and that these maternal measures did not differ by the daughters' BMI.³² Our findings help explain the different conclusions from these studies because the Danish study was conducted in older girls (median age 10.9, range 6.2 to 14.7),³³ compared with ours (median age 9.5, range 6.0 to 15.1). In our older girls, we also found that self-assessments are preferred for greater accuracy. We disagree with the conclusion of the Danish study³³ that epidemiologic studies can use self-assessment for distinguishing

between prepuberty (T1) and puberty (T2+). The higher sensitivity and specificity for pubertal onset in the Danish cohort, which concluded that self-assessment is accurate, is based on an older age distribution and a much smaller percentage of their cohort still in prepuberty. Comparing the 3 studies in terms of percentage of girls still in prepuberty (T1) determined by clinical assessment, the Chilean study had 83.9% of girls still in T1, compared with 56.4% in our study and only 19.8% in the Danish study. Thus, self-assessment may be useful for older girls in terms of the feasibility of data collection and more accurate than maternal assessment for girls aged ≥ 11 years based on our results, but it may be less useful for determining the onset of puberty, which, for many girls, takes place at younger ages.

Other Factors Affecting Accuracy and Agreement

After considering the age differences discussed earlier, only BMI was related to the discordance in our study between maternal and clinical assessments. A previous study reported poor reliability between clinical and self-staging in overweight girls,⁴² but others did not.^{32,33} In overweight girls with more fat tissue, it may be especially difficult to distinguish glandular breast tissue from fat tissue using visual assessment only.⁴² We overcame this limitation through our clinical ratings, which used visual assessment with palpation when necessary.⁴⁹ However, our maternal and self-assessments differed from the clinical assessments, particularly in overweight girls. Thus, we disagree with the conclusion by the Chilean study³² that mothers can be used when clinical assessments are not available without adjusting the maternal assessments for the level of sensitivity and specificity. Maternal and self-staging of pubic

hair development did not differ by BMI, likely a result of body size not influencing the appearance of pubic hair.

We also assessed whether accuracy differed by family history of breast cancer, given the higher breast cancer worry compared with those without a family history.^{50,51} The sensitivity of maternal assessment for breast Tanner stage was modestly higher in families with a breast cancer family history than in families without (80% vs 74%), but this difference was not statistically significant. Similar to an earlier study,⁵² we did not observe statistically significant differences in reliability and validity by race/ethnicity. Because most girls at the New York and Utah study sites were non-Hispanic white or Hispanic, we lacked sufficient statistical power to detect differences in maternal report of breast development for other racial/ethnic groups. Other studies in more diverse populations have found that black or Hispanic adolescents were less accurate in staging their breast and pubic hair development than were non-Hispanic white adolescents.^{40,48}

Our results suggest that findings from studies that rely on maternal or self-staging of pubertal development may be biased. Validity studies such as ours can be used to adjust the estimates from epidemiologic studies because they can be used to determine

the direction and the magnitude of the bias.^{53,54} We illustrate this by using the data reported from the Chilean study that stratified reliability measures by child's BMI and observed a similar κ between maternal assessment and clinical assessment (by trained personnel) for overweight girls as for average weight girls ($\kappa = 0.74$ compared with $\kappa = 0.71$, respectively).³² Even though they reported similar reliability measures, the validity measures using the results from the trained personnel were different (sensitivity = 0.87 and 0.92 for average weight and overweight girls, specificity = 0.94 and 0.90, respectively). Thus, using maternal reports in this case would result in a higher estimate of the association between being overweight and breast onset (OR = 1.39, 95% CI 0.82 to 2.3) compared with the association using the results from the clinical assessment (OR = 1.18, 95% CI 0.66 to 2.12). Thus, validity studies conducted within a subcohort provide essential data to understand the impact of measurement error when clinical assessments are not available for the entire cohort. Our study did not have a clinical assessment for pubic hair development, and thus our validity findings were limited to breast only, whereas our reliability findings evaluated both breast and pubic hair development.

CONCLUSIONS

Our findings have implications for the interpretation of pubertal development data across pubertal cohorts because many collect information on pubertal development only from a single source^{20,21,28-30} and/or change sources over time.³¹ Specifically, our results support that for breast development, maternal report is more accurate for girls younger than 11 years and that self-assessment alone should not be used in epidemiologic studies of pubertal onset. For girls aged ≥ 11 years, self-assessment is more accurate for breast development. In studies lacking clinical breast Tanner for the whole cohort, sensitivity analyses adjusting for the validity of maternal and self-assessments should be used to understand the impact measurement error may have on the overall study conclusions.

ACKNOWLEDGMENTS

The authors thank the LEGACY girls and their family members for their continuing contributions to the study and our colleagues at the participating family genetics and oncology clinics

ABBREVIATIONS

CI: confidence interval
OR: odds ratio

participated in the collection and assembly of data, analysis, and interpretation and manuscript writing; Drs Schwartz, Keegan, Houghton, and Knight participated in the collection and assembly of data and manuscript writing; Ms Goldberg and Ms Schechter participated in the analysis and interpretation of the data and manuscript writing; Dr Chung, Ms White, and Ms O'Toole conducted clinical Tanner assessments and participated in the collection and assembly of data, analysis and interpretation of data, and reviewed and revised the manuscript; and all authors approved the final manuscript as submitted.

DOI: 10.1542/peds.2015-4571

Accepted for publication Mar 28, 2016

Address correspondence to Mary Beth Terry, PhD, Columbia University Mailman School of Public Health, Department of Epidemiology, 722 West 168th St, New York, NY 10032. E-mail: mt146@columbia.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2016 by the American Academy of Pediatrics

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

FUNDING: This work was supported by grants from the National Cancer Institute at the National Institutes of Health (R01 CA138638 to Dr John, R01 CA138819 to Dr Daly, R01 CA138822 to Dr Terry, and R01 CA138844 to Dr Andrulis) and the Canadian Breast Cancer Foundation (Dr Andrulis). Dr Andrulis holds the Anne and Max Tanenbaum Chair in Molecular Medicine at Mount Sinai Hospital and the University of Toronto. Funded by the National Institutes of Health (NIH).

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

- Garland M, Hunter DJ, Colditz GA, et al. Menstrual cycle characteristics and history of ovulatory infertility in relation to breast cancer risk in a large cohort of US women. *Am J Epidemiol.* 1998;147(7):636–643
- Okasha M, McCarron P, Gunnell D, Smith GD. Exposures in childhood, adolescence and early adulthood and breast cancer risk: a systematic review of the literature. *Breast Cancer Res Treat.* 2003;78(2):223–276
- Petridou E, Syrigou E, Toupadaki N, Zavitsanos X, Willett W, Trichopoulos D. Determinants of age at menarche as early life predictors of breast cancer risk. *Int J Cancer.* 1996;68(2):193–198
- Rockhill B, Moorman PG, Newman B. Age at menarche, time to regular cycling, and breast cancer (North Carolina, United States). *Cancer Causes Control.* 1998;9(4):447–453
- Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol.* 2012;13(11):1141–1151
- Gong TT, Wu QJ, Vogtmann E, Lin B, Wang YL. Age at menarche and risk of ovarian cancer: a meta-analysis of epidemiological studies. *Int J Cancer.* 2013;132(12):2894–2900
- Jordan SJ, Webb PM, Green AC. Height, age at menarche, and risk of epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev.* 2005;14(8):2045–2048
- Currie C, Ahluwalia N, Godeau E, Nic Gabhainn S, Due P, Currie DB. Is obesity at individual and national level associated with lower age at menarche? Evidence from 34 countries in the Health Behaviour in School-aged Children Study. *J Adolesc Health.* 2012;50(6):621–626
- Dreyfus JG, Lutsey PL, Huxley R, et al. Age at menarche and risk of type 2 diabetes among African-American and white women in the Atherosclerosis Risk in Communities (ARIC) study. *Diabetologia.* 2012;55(9):2371–2380
- Mendle J, Turkheimer E, Emery RE. Detrimental psychological outcomes associated with early pubertal timing in adolescent girls. *Dev Rev.* 2007;27(2):151–171
- van Jaarsveld CH, Fidler JA, Simon AE, Wardle J. Persistent impact of pubertal timing on trends in smoking, food choice, activity, and stress in adolescence. *Psychosom Med.* 2007;69(8):798–806
- Joinson C, Heron J, Lewis G, Croudace T, Araya R. Timing of menarche and depressive symptoms in adolescent girls from a UK cohort. *Br J Psychiatry.* 2011;198(1):17–23
- Kaltiala-Heino R, Rimpelä M, Rissanen A, Rantanen P. Early puberty and early sexual activity are associated with bulimic-type eating pathology in middle adolescence. *J Adolesc Health.* 2001;28(4):346–352
- Striegel-Moore RH, McMahon RP, Biro FM, Schreiber G, Crawford PB, Voorhees C. Exploring the relationship between timing of menarche and eating disorder symptoms in Black and White adolescent girls. *Int J Eat Disord.* 2001;30(4):421–433
- Stice E, Presnell K, Bearman SK. Relation of early menarche to depression, eating disorders, substance abuse, and comorbid psychopathology among adolescent girls. *Dev Psychol.* 2001;37(5):608–619
- Jacobsen BK, Heuch I, Kvåle G. Association of low age at menarche with increased all-cause mortality: a 37-year follow-up of 61,319 Norwegian women. *Am J Epidemiol.* 2007;166(12):1431–1437
- Tamakoshi K, Yatsuya H, Tamakoshi A; JACC Study Group. Early age at menarche associated with increased all-cause mortality. *Eur J Epidemiol.* 2011;26(10):771–778
- DiVall SA, Radovick S. Pubertal development and menarche. *Ann N Y Acad Sci.* 2008;1135:19–28
- Wyshak G, Frisch RE. Evidence for a secular trend in age of menarche. *N Engl J Med.* 1982;306(17):1033–1035
- Biro FM, Galvez MP, Greenspan LC, et al. Pubertal assessment method and baseline characteristics in a mixed longitudinal study of girls. *Pediatrics.* 2010;126(3). Available at: www.pediatrics.org/cgi/content/full/126/3/e583
- 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
- Bodicoat DH, Schoemaker MJ, Jones ME, et al. Timing of pubertal stages and breast cancer risk: the Breakthrough Generations Study. *Breast Cancer Res.* 2014;16(1):R18
- de Muinich Keizer SM, Mul D. Trends in pubertal development in Europe. *Hum Reprod Update.* 2001;7(3):287–291
- Kaplowitz P. Pubertal development in girls: secular trends. *Curr Opin Obstet Gynecol.* 2006;18(5):487–491
- Biro FM, Greenspan LC, Galvez MP, et al. Onset of breast development in a longitudinal cohort. *Pediatrics.* 2013;132(6):1019–1027
- Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child.* 1969;44(235):291–303
- Morris NM, Udry JR. Validation of a self-administered instrument to assess stage of adolescent development. *J Youth Adolesc.* 1980;9(3):271–280
- Hui LL, Leung GM, Wong MY, Lam TH, Schooling CM. Small for gestational age and age at puberty: evidence from Hong Kong's "Children of

- 1997" birth cohort. *Am J Epidemiol*. 2012;176(9):785–793
29. Aksglaede L, Sørensen K, Petersen JH, Skakkebaek NE, Juul A. Recent decline in age at breast development: the Copenhagen Puberty Study. *Pediatrics*. 2009;123(5). Available at: www.pediatrics.org/cgi/content/full/123/5/e932
 30. Gillman MW, Rifas-Shiman S, Berkey CS, Field AE, Colditz GA. Maternal gestational diabetes, birth weight, and adolescent obesity. *Pediatrics*. 2003;111(3). Available at: www.pediatrics.org/cgi/content/full/111/3/e221
 31. Joinson C, Heron J, Araya R, et al. Association between pubertal development and depressive symptoms in girls from a UK cohort. *Psychol Med*. 2012;42(12):2579–2589
 32. Pereira A, Garmendia ML, González D, et al. Breast bud detection: a validation study in the Chilean growth obesity cohort study. *BMC Womens Health*. 2014;14:96
 33. Rasmussen AR, Wohlfahrt-Veje C, Tefre de Renzy-Martin K, et al. Validity of self-assessment of pubertal maturation. *Pediatrics*. 2015;135(1):86–93
 34. Bandera EVWM, Marcella S, Donaldson A, et al. Assessing breast development in The Jersey Girl Study: agreement between physician and mom assessment [SER Abstract 047]. *Am J Epidemiol*. 2010;171(Suppl 11):S12
 35. John EM, Terry MB, Keegan TH, et al. The LEGACY Girls Study: Growth and development in the context of breast cancer family history. *Epidemiology*. 2016
 36. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174
 37. Division of Nutrition PA, and Obesity, National Center for Chronic Disease Prevention and Health Promotion. A SAS Program for the 2000 CDC Growth Charts (ages 0 to <20 y). Available at: www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm. 2014. Accessed June 30, 2014
 38. Barlow SE; Expert Committee. Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. *Pediatrics*. 2007;120(suppl 4):S164–S192
 39. Britton JA, Wolff MS, Lapinski R, et al. Characteristics of pubertal development in a multi-ethnic population of nine-year-old girls. *Ann Epidemiol*. 2004;14(3):179–187
 40. Hergenroeder AC, Hill RB, Wong WW, Sangi-Haghpeykar H, Taylor W. Validity of self-assessment of pubertal maturation in African American and European American adolescents. *J Adolesc Health*. 1999;24(3):201–205
 41. Brooks-Gunn J, Warren MP, Rosso J, Gargiulo J. Validity of self-report measures of girls' pubertal status. *Child Dev*. 1987;58(3):829–841
 42. Bonat S, Pathomvanich A, Keil MF, Field AE, Yanovski JA. Self-assessment of pubertal stage in overweight children. *Pediatrics*. 2002;110(4):743–747
 43. Desmangles JC, Lappe JM, Lipaczewski G, Haynatzki G. Accuracy of pubertal Tanner staging self-reporting. *J Pediatr Endocrinol Metab*. 2006;19(3):213–221
 44. Brooks-Gunn J, Newman DL, Holderness C, Warren MP. The experience of breast development and girls' stories about the purchase of a bra. *J Youth Adolesc*. 1994;23(5):539–565
 45. 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
 46. Jaruratanasirikul S, Kreetapirom P, Tassanakijpanich N, Sriplung H. Reliability of pubertal maturation self-assessment in a school-based survey. *J Pediatr Endocrinol Metab*. 2015;28(3–4):367–374
 47. Duke PM, Litt IF, Gross RT. Adolescents' self-assessment of sexual maturation. *Pediatrics*. 1980;66(6):918–920
 48. 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
 49. Euling SY, Herman-Giddens ME, Lee PA, et al. Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. *Pediatrics*. 2008;121(suppl 3):S172–S191
 50. Gibbons A, Groarke A. Can risk and illness perceptions predict breast cancer worry in healthy women [published online ahead of print February 23, 2015]? *J Health Psychol*. pii 1359105315570984
 51. Cohen M. Breast cancer early detection, health beliefs, and cancer worries in randomly selected women with and without a family history of breast cancer. *Psychooncology*. 2006;15(10):873–883
 52. Neinstein LS. Adolescent self-assessment of sexual maturation: reassessment and evaluation in a mixed ethnic urban population. *Clin Pediatr (Phila)*. 1982;21(8):482–484
 53. Thompson WD. Kappa and attenuation of the odds ratio. *Epidemiology*. 1990;1(5):357–369
 54. Thompson WD, Walter SD. A reappraisal of the kappa coefficient. *J Clin Epidemiol*. 1988;41(10):949–958

Comparison of Clinical, Maternal, and Self Pubertal Assessments: Implications for Health Studies

Mary Beth Terry, Mandy Goldberg, Sarah Schechter, Lauren C. Houghton, Melissa L. White, Karen O'Toole, Wendy K. Chung, Mary B. Daly, Theresa H.M. Keegan, Irene L. Andrulis, Angela R. Bradbury, Lisa Schwartz, Julia A. Knight, Esther M. John and Sandra S. Buys

Pediatrics 2016;138;

DOI: 10.1542/peds.2015-4571 originally published online June 8, 2016;

Updated Information & Services	including high resolution figures, can be found at: http://pediatrics.aappublications.org/content/138/1/e20154571
References	This article cites 52 articles, 13 of which you can access for free at: http://pediatrics.aappublications.org/content/138/1/e20154571#BIBL
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): Developmental/Behavioral Pediatrics http://www.aappublications.org/cgi/collection/development:behavioral_issues_sub Growth/Development Milestones http://www.aappublications.org/cgi/collection/growth:development_milestones_sub Endocrinology http://www.aappublications.org/cgi/collection/endocrinology_sub Puberty http://www.aappublications.org/cgi/collection/puberty_sub
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://www.aappublications.org/site/misc/Permissions.xhtml
Reprints	Information about ordering reprints can be found online: http://www.aappublications.org/site/misc/reprints.xhtml

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN®



PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

Comparison of Clinical, Maternal, and Self Pubertal Assessments: Implications for Health Studies

Mary Beth Terry, Mandy Goldberg, Sarah Schechter, Lauren C. Houghton, Melissa L. White, Karen O'Toole, Wendy K. Chung, Mary B. Daly, Theresa H.M. Keegan, Irene L. Andrulis, Angela R. Bradbury, Lisa Schwartz, Julia A. Knight, Esther M. John and Sandra S. Buys

Pediatrics 2016;138;

DOI: 10.1542/peds.2015-4571 originally published online June 8, 2016;

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/138/1/e20154571>

Data Supplement at:

<http://pediatrics.aappublications.org/content/suppl/2016/06/06/peds.2015-4571.DCSupplemental>

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, 345 Park Avenue, Itasca, Illinois, 60143. Copyright © 2016 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 1073-0397.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN®

