Self-assessment of pubertal Tanner stage by realistic colour images in representative Chinese obese and non-obese children and adolescents

Ying Sun1,2, Fang-Biao Tao(taofangbiao@126.com)1,2, Pu-Yu Su1,2, China Puberty Research Collaboration*

1. Department of Maternal, Child and Adolescent Health, School of Public Health, Anhui Medical University, Hefei, Anhui Province, China
2. Anhui Provincial Key Laboratory of Population Health and Aristogensics, Anhui Medical University, Hefei, Anhui Province, China

ABSTRACT

Aim: This investigation aims to evaluate the validity of self-assessment of pubertal Tanner stage by realistic colour images in representative Chinese children and adolescents.

Methods: The study is a nationally representative cross-sectional survey in eight research sites in the large project entitled ‘China Puberty Research Collaboration’. Weight, height, self-assessed pubertal Tanner stage and physical examination of pubic hair in each gender, breast development in girls and genital development in boys were analysed.

Results: A large proportion of subjects aged 7.9–18.9 years old were capable of identifying their own pubertal Tanner stage accurately or close to the rater’s assessments. Obese group tends to overestimate their pubertal development compared to non-obese peers, except for genital assessment in boys. The k values for non-obese and obese girls were 0.619 (p < 0.0001) and 0.527 (p < 0.0001), respectively, while the k values for non-obese and obese boys were 0.503 (p < 0.0001) and 0.352 (p < 0.0001).

Conclusions: Self-assessment of the pubertal stage by using realistic colour images could be a better alternative assessment tool for large epidemiological puberty research compared with Tanner’s original black and white pictures.

INTRODUCTION

Accurate assessment of sexual maturity is an important consideration for many paediatric and adolescent investigations. Despite the criticisms against the clinical assessment of pubertal stage by this method, its use has been widely spread in outpatient routine. Some researchers regarded self-assessment of Tanner stage as a reasonable substitute for Tanner stage determined by physical examination, for the purpose of large-scale, epidemiologic research studies when physical examination may not be feasible (1,2).

Since early 1980s, many previous studies regarding self-assessment of pubertal stage in normal weight children and adolescents have shown inconsistent accuracy of self-assessment comparatively to the objective assessment of sexual maturation within each area and across the areas. The percentage agreement between physician assessment and self-assessment ranged from 49% to 86% for the Tanner breast stage (2–6), from 58% to 86% for girls’ Tanner pubic hair stage (2–6) and from 48% to 78% for boys’ Tanner pubic hair stage (2,4,6). Factors such as the method chosen, cultural characteristics and notions of self-image (influenced by the presence or not of overweight, nutritional condition that is knowingly associated with problems related to self-image perception) are certainly accountable for discrepant results in different populations.
Although high body mass index (BMI) seems to be associated with early pubertal development (7), little is known about the effects of fatness on self-assessments of Tanner stage. Only two studies examined the effect of adiposity on the sexual maturation assessment in adolescents recruited from communities (2,8).

To the best of our knowledge, no previous studies have examined self-reports of pubertal maturation in large sample and wide age range of normal weight and obese children and adolescents. We, therefore, compared the accuracy of Tanner stage self-assessments in a representative sample of Chinese obese and normal weight children and adolescents by using colour images of pubertal rating according to Tanner stages, using data from China Puberty Research Collaboration Survey which were conducted in eight areas across China.

**METHODS**

**Procedure and sample**

This research is framed within the larger project entitled ‘China Puberty Research Collaboration’. The multi-centred collaboration is initiated by School of Public Health, Anhui Medical University (Hefei) in Sep 2010, cooperated with other five medical universities and two school healthcare centres.

Those eight research sites are representative of eastern (Shenyang, Shanghai), central (Hefei, Wuhan, Zhengzhou), western (Chongqing, Kunming), southern (Guangzhou) of China. The minimum sample size required for each collaborating centres 50 girls and 50 boys of Han ethnicity in each grade (from grade 1 to grade 12) in rural and urban area, respectively. Because of the cooperation of local Bureau of Education, Kunming research centre only obtains 490 and 559 girls in urban and rural areas, respectively.

Figure S1 illustrated the selection procedure of subjects in this study. Written consent from each participant’s guardian and written assent from each participant were obtained after providing an explanation of the study.

**MEASUREMENT**

**Assessment of sexual maturity status**

Before subjects underwent physical examination, they were given a standardized series of realistic colour images with explanatory text to assess their own pubertal development individually. Those images of pubertal rating according to Tanner stages (Fig. S2) were proposed by Carel JC and Leger J published in New England Journal of Medicine in 2008 (9).

In girls, breast development is rated from 1 (preadolescent) to 5 (mature), and stage 2 (appearance of the breast bud) marks the onset of pubertal development. Pubic hair stages are rated from 1 (preadolescent, no pubic hair) to 5 (adult), and stage 2 marks the onset of pubic hair development.

Subjects from grade 4 to grade 12 were asked to read the brief descriptions of each stage after which they would check the box over the picture that best represented them for each component of development in a private room. Students were also given an explanation of some terms to describe secondary sex characteristics if they requested for this information.

The physician assessment was conducted on the same day in a room individually and privately, as part of the physical examination, by two trained examiners, both experienced in Tanner stage evaluation. The two examiners assessed only same-sex children and adolescents. Visual inspection and palpation for breast and genital examination were requested. The examiner was blinded to the subjects’ self-assessments. Substantial training was conducted to achieve high level of interrater reliability.

**Anthropometric measures**

Subjects were measured for height and weight by one observer using standard techniques with the children standing without shoes and lightly clothed.

BMI classification reference proposed by Working Group on Obesity in China (WGOC) was used as screening reference to calculate the prevalence and trends of overweight/obesity in these groups (10). BMI was categorized into three groups: BMI < 15 percentile (P15), normal weight group (BMI between P15 and P85), overweight and obese group (BMI ≥ P85). The latter two groups were included into data analysis.

**Statistical analysis**

The data were analysed by using SPSS 10.01, Chicago, IL, USA. The correlations between actual and self-reported Tanner stage ratings were examined using Kendall rank correlations, and the differences between actual and self-reported Tanner stage ratings were examined with the 1-sample sign test. Tanner stage data were analysed separately for girls and boys. Additional analyses examined whether Tanner stage was estimated with equal accuracy in obese and non-obese subjects. An alpha level of p < 0.05 was set a priori.

The validity of self-assessment versus physician assessment of pubertal Tanner stage was evaluated by per cent agreement of the two assessments and by kappa coefficient (k). Accuracy of self-assessment was defined as the proportion of subjects who gave the same assessment as the physician. The k coefficient measures the level of agreement between two independent sets of responses, in this case the physician versus subject assessments of the five Tanner stages.

Kappa ‘Benchmarks’ of the strength of agreement are as follows: k < 0.01 = poor; 0.01–0.20 = slight; 0.21–0.40 = fair; 0.41–0.60 = moderate; 0.61–0.80 = substantial; k > 0.8 = almost perfect (11).

**Ethical considerations**

The study is approved by the ethical committee of Anhui Medical University. The study is presented as a study on
growth and puberty timing. All participants and their parents gave informed consent.

RESULTS
Characteristic of study population
A total of 9,132 girls and 6,924 boys were included in our analysis. The age range of the girls and the boys was 8.0–18.9 and 7.9–18.9, respectively. The BMI ranged from 18.2 to 54.5 kg/m² (mean 25.1±3.7) in obese boys and 17.2–53.7 kg/m² (mean 23.1±3.4) in obese girls. The median ages at stage 2 for breast development of Chinese girls were 9.18 (95% CI 8.91 to 9.44) years and at stage 2 for pubic hair growth were 11.65 (95% CI 11.41–11.90) years. The median ages at stage 2 for genital development of Chinese boys were 11.24 (95% CI 11.08–11.41) years and at stage 2 for pubic hair growth were 12.67 (95% CI 12.46–12.88) years.

Comparison of self-assessment by subjects and physician assessment in girls
Table S1 showed the comparison of the physician and subject assessment for breast and pubic hair development in non-obese and obese girls.

Kendall rank correlation coefficient between breast Tanner stage self-reports and actual ratings for non-obese girls was 0.766 (p < 0.0001) and for obese girls was 0.735 (p < 0.0001; Fig. S3A). The k value for non-obese girls was 0.619 (p < 0.0001) and for obese girls was 0.527 (p < 0.0001), which indicated substantial reproducibility.

Kendall rank correlation coefficient between pubic hair Tanner stage self-reports and actual ratings for non-obese girls was 0.781 (p < 0.0001) and for obese girls was 0.794 (p < 0.0001; Fig. S3B). The k value for non-obese girls was 0.595 (p < 0.0001) and for obese girls was 0.573 (p < 0.0001), which indicated moderate reproducibility.

Figure S3A showed that 353 (17.9%) obese girls overestimated their breast Tanner stage, which is significantly higher than that of non-obese girls (9.3%) (x² = 116.25, p < 0.01). However, there were less obese girls than non-obese girls that underestimated their breast Tanner stage, with 18.6% vs. 20.1% (x² = 2.16, p = 0.05).

Figure S3B showed that 358 (18.2%) obese girls overestimated their pubic hair Tanner stage, which is significantly higher than that of non-obese girls (14.3%) (x² = 17.58, p < 0.01). However, there were more non-obese girls than obese girls that underestimated their pubic hair Tanner stage, with 17.7% vs. 14.8% (x² = 8.96, p < 0.01). There were no significant differences between accurate estimated breast and pubic hair Tanner stage between two groups.

Comparison of self-assessment by subjects and physician assessment in boys
Table S2 showed the comparison of the physician and subject assessment for genital and pubic hair development in non-obese and obese boys.

Kendall rank correlation coefficient between genital Tanner stages self-reports and actual ratings for non-obese boys was 0.778 (p < 0.0001) and for obese boys was 0.639 (p < 0.0001; Fig. S3C). Although self-assessed and measured Tanner stage ratings were highly correlated, the k value for non-obese boys was 0.505 (p < 0.0001) and for obese boys was 0.352 (p < 0.0001), which indicated fair to moderate reproducibility.

Kendall rank correlation coefficient between pubic hair Tanner stage self-reports and actual ratings for non-obese boys was 0.829 (p < 0.0001) and for obese boys was 0.738 (p < 0.0001; Fig. S3D). Although self-assessed and measured Tanner stage ratings were highly correlated, the k value for non-obese boys was 0.642 (p < 0.0001) and for obese boys was 0.446 (p < 0.0001), which indicated moderate to substantial reproducibility.

Figure S3C,D showed that obese boys tend to misclassify their genital and pubic hair Tanner stage compared with non-obese boys (p value < 0.01, except for underestimate for pubic hair stage). Non-obese boys accurately estimated their genital (60.5%) and pubic hair (72.2%) Tanner stage, which were significantly higher than that of obese boys (49.6% and 57.1%, respectively) (p < 0.01).

DISCUSSION
This study evaluated the validity of the self-assessment sexual maturation status (SMS) in 9,132 girls and 6,924 boys who are representative of mainland Chinese children and adolescents. Our results suggested that most children were capable of identifying their own SMS accurately or close to the rater’s assessments by using the realistic, coloured images of pubertal development, which resulting in a moderate to almost perfect agreement.

Regarding study participants and study methods, there were a number of differences between the current study and previous studies. First, the large and representative sample size allowed us to fully compare the differences in agreement of self-assessment and rater’s assessment in SMS in both non-obese and obese boys and girls. Most previous related studies only included <500 children and adolescents (1,8). For example, Chan et al. (1) studied 172 boys and 182 girls aged 8–18 years. So far, there were few studies examined the effect of adiposity on the sexual maturation assessment in adolescents. Lee et al. (8) only investigated 77 morbidly overweight children and adolescents in their study. Thus, low power affected by small sample size of their study must be considered.

Differences in the age groups or sexual maturation stages might be an important reason for these inconsistent findings. Several subjects in previous studies were younger than 15 years old and at the early stages of sexual maturation. Then, the accuracy of the self-assessment of sexual maturation in overweight adolescents older than 15 years and at later stages of sexual maturation could not be predicted. In our study, wide age range (7.9–18.9 years old) made it feasible to analyse both subjects in early Tanner stage and late Tanner stage.

While self-assessed SMS in most previous studies used black and white line drawings by Morris and Udry (2,3,12–14) or Tanner’s standard photographs (15–18), our study
used coloured images of pubertal rating according to Tanner stages, proposed by Carel JC and Leger J published in New England Journal of Medicine in 2008 (9). Those computer-generated, hyper-realistic pictures make it clear and acceptable for children and adolescents to self-assess their sexual maturation; thus, it has higher percentage agreement compared with those of studies using the drawings and Tanner’s standard photographs.

Our study adopted the weighted Kendall rank correlation statistic to analyse the data, which is a more appropriate method to evaluate the validity of the pubertal self-assessment, as it allows for differences in the importance of disagreements and gives different weighting to different responses. Kendall rank correlations between raters’ assessment and children’s self-reported pubertal Tanner stage were found to be from moderate to substantial (0.503–0.642), with genital assessment in obese boys the lowest (k = 0.352). The results are in consistent with several previous researches (14,17).

Our results supported previous findings that children of normal weight can accurately assess their own stage of sexual maturation using drawings and pictures and that both obese and non-obese girls assessed pubic hair Tanner stage without significant bias. For example, Chan NP et al. found a substantial agreement for breast Tanner stages, moderate agreement for genital Tanner stages and substantial to almost perfect agreement for pubic hair Tanner stages in both genders between self-assessments and rater’s assessments.

Similar to other studies (2,3,17), the agreement in obese group was found to be lower than that in non-obese group. Obese children and adolescents tended to assess their sexual maturity less accurately than did normal weight children and adolescents. The results might be explained by the difficulty to distinguish lipomastia from true breast tissue in obese girls.

The present study shows that, in the studied population, the use of realistic colour images yields high associations with objective assessment. For the purpose of large-scale, epidemiologic research studies when physical examination may not be feasible, more instructions and explanations should be given to obese children and adolescents, especially boys.

References


SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1 RStudy Procedure in flow diagram.

Figure S2 Pubertal rating according to Tanner stages.

Figure S3 Comparison of underestimate, correctly estimate and overestimate in breast (A), girls’ pubic hair (B), genital (C) and boys’ pubic hair (D) Tanner stage percentage in boys between normal weight and overweight and obese groups.

Table S1 Tanner stage for breast and pubic hair development in girls: Comparison of self-assessment by subjects and physician assessment.

Table S2 Tanner stage for genital and pubic hair development in boys: Comparison of self-assessment by subjects and physician assessment.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.