

Review of methods for determining pubertal status and age of onset of puberty in cohort and longitudinal studies

Janis Baird, Inna Walker, Clare Smith, Hazel Inskip

MRC Lifecourse Epidemiology Unit, University of Southampton

April 2017



To cite this report, please use the following reference:

Baird J, Walker I, Smith C, Inskip H. Review of methods for determining pubertal status and age of onset of puberty in cohort and longitudinal studies. London, UK: CLOSER; 2017.

CLOSER

UCL Institute of Education
20 Bedford Way
London
WC1H 0AL
United Kingdom

Tel: +44 (0)20 7331 5102

Email: closer@ucl.ac.uk

Web: www.closer.ac.uk

Twitter: [@CLOSER_UK](https://twitter.com/CLOSER_UK)

YouTube: [CLOSER](https://www.youtube.com/CLOSER)

Contents

Introduction	2
Aims	3
Methods	3
Results	4
Clinical (Gold standard) assessment	4
Self-assessment	4
Growth	6
Hormonal assessment	7
Gonadotrophins and sex hormones	7
Leptin	8
Voice.....	8
Workshop to discuss pubertal assessment methods (see appendix 2) – summary of findings.....	9
Lay panel to obtain the views of young people (see appendix 3) – summary of findings.....	9
Discussion.....	10
Summary of findings.....	10
Strengths and limitations	11
Need for further research.....	11
Table 1: Description of studies using self-assessment methods to assess pubertal status.....	12
Table 2: Description of studies using growth methods to assess pubertal status.....	21
Table 3: Description of studies using gonadotrophin or gonadal hormone methods to assess pubertal status.....	28
Table 4: Description of studies using leptin methods to assess pubertal status.....	33
Table 5: Description of studies using voice methods to assess pubertal status	36
Appendix 1: Sexual Maturation Scale, based on Tanner and Marshall (1969).....	38
Boys - development of external genitalia	38
Girls - breast development.....	39
Boys and girls - pubic hair	40
Appendix 2: Report of workshop	41
Appendix 3: Pubertal assessment – the views of the University Hospital Southampton Children’s Panel.....	48
References.....	51

Introduction

In recent years, there has been increasing interest in adolescent health (1), and, notably, it was identified as a priority by the Chief Medical Officer for England in her 2012 annual report (2). Its importance in lifecourse epidemiology has recently been highlighted by authors who are involved in two CLOSER studies (the National Survey for Health and Development (NSHD) and the National Child Development Study (NCDS)) (3). Many cohort studies are interested in body composition assessment through the lifecourse. However, puberty is a dynamic period of development marked by rapid changes in body size, shape, and composition, all of which are sexually dimorphic (4). Cohort studies tend to assess participants at particular ages; at any given age in adolescence, children will be at varying stages of puberty and so account needs to be taken of pubertal stage in analyses of adolescent body composition. Relationships between timing of puberty and body composition through adolescence and beyond have been examined in the two CLOSER studies mentioned above (5, 6), and there have been various analyses of the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort in relation to puberty, not least examining the influences on timing of puberty itself in boys and girls (7, 8).

Stage of puberty is also valuable in the assessment of various health outcomes in adolescents. The hormonal changes associated with puberty can impact both on physical and mental well-being. As assessing pubertal stage is difficult, various studies have focused on a particular pubertal stage in girls (9, 10), namely age at menarche, which can often be determined retrospectively, but there is no easily obtained comparable measure for boys. This limits analyses to girls only, and to one measure of puberty.

There are a number of factors that complicate the measurement of pubertal status, including the complex nature of the pubertal process, individual differences in the pattern of pubertal change, lack of precise measurement techniques, and problems with securing permission to use the most precise measure – clinical assessment. Children and their parents are less likely to consent to clinical assessment than to other forms of assessment, such as questionnaire-based approaches. Schools share that reluctance and, in addition, have concerns about the use, within schools, of pictures or photographs depicting the development of secondary sexual characteristics (11). Chronological age is not a reliable parameter for determining the biological characterisation of individuals. To assess levels of sexual maturity in growing children, many clinicians rely on physical examination. However, clinical assessment within a cohort study is challenging. As a result, many cohort studies have relied on self-assessment methods or parental report. Within the Southampton Women's Survey, for example, a clinical assessment of puberty is being conducted at the 11-13 year follow-up, with a self-assessment questionnaire given to those who refuse the clinical assessment.

Aims

In this review, our aims were:

- 1. To identify the methods that have been used to assess pubertal stage and onset of puberty in cohort and other studies around the world**
- 2. To assess the validity of these methods and identify similarities among methods that will contribute to harmonisation of datasets**
- 3. To identify barriers to these assessments and acceptability of the various approaches, through consultation with a children's Patient and Public Involvement (PPI) committee**
- 4. To synthesise the evidence summarising approaches to pubertal assessment and their quality in a review article to inform those conducting cohort and other studies of adolescents.**

Methods

We sought studies that had assessed the validity of methods of assessing pubertal stage used in cohort and cross-sectional studies. We searched online databases of abstracts, with assistance from an information specialist. These included Medline, PsycInfo, Scopus, Sociological Abstracts, CINAHL and ERIC. In addition, we consulted experts in the field, including a paediatric endocrinologist, and hand-searched key journals.

While this synthesis of evidence was not a systematic review, we used systematic review methods to extract data related to the validation of each method of pubertal assessment and to give an overview of study quality.

The preliminary findings of the review were presented at a workshop, in September 2016, at which researchers from CLOSER cohort studies were the predominant attendees. The workshop provided an opportunity to discuss the approaches to pubertal assessment taken within cohorts and to consult researchers on the methods they felt would work most effectively within cohort studies.

In order to assess the acceptability to young people of the approaches we identified, and to find out what they considered might be barriers to assessment, we consulted a panel of young people aged 10 to 16 years. Using some of the main findings of our review, we showed them material relating to different approaches (e.g. self-assessment questionnaires) and asked what they thought of them and how they thought other young people might view them.

Results

We screened 11,935 abstracts and assessed 157 papers in detail. Of these, 30 reported data comparing two or more methods of pubertal assessment. Not all studies were validation studies in that some compared two methods but did not conduct a comparison with the gold standard of clinical assessment.

Data were extracted from the 30 studies using a standardised proforma and summarised within tables. The majority of studies had methodological weaknesses. Many were based on small samples and selection bias was frequent – for example, basing studies on highly selected children with a particular background such as attendance at grammar schools.

Review findings are grouped according to the method of pubertal assessment, with summary tables presented for each section.

Clinical (Gold standard) assessment

Tanner staging from a physical examination of the adolescent is considered to be the gold standard measurement of pubertal stage (12). There are five stages, separately defined for boys and girls, determined by pubic hair growth for both sexes and breast and testicular development for girls and boy respectively. The five stages are summarised in the sexual maturation scale (SMS) – a set of photographs that those performing the assessment use to judge the stage that a young person has reached in relation to each of their secondary sexual characteristics (see appendix 1). However, interrater reliability between examiners can be as low as 76% (13), and therefore it is important to consider who the examiners are and their consistency.

Self-assessment

Twelve relevant studies were identified – the largest group of studies on approaches to pubertal assessment. Most compared either self-assessment with the sexual maturation scale (SMS) or the pubertal development scale (PDS) against the gold standard of Tanner staging derived from clinical assessment. The SMS was developed by Tanner in 1962. It assesses secondary sexual characteristics using a series of photographs (used originally to guide clinicians in the gold-standard measurement) and it has been shown that the identification of these characteristics, using the scale, is related to endocrine changes, growth and other pubertal changes (11). While the photographic version of the SMS has been used for self-assessment, a line drawing version was produced to make it more acceptable to young people (14). The PDS was developed in response to reluctance on the part of schools to show depictions of secondary sexual characteristics to their pupils (11). The PDS is an interview-based continuous measure of pubertal development. It includes a series of questions about growth, body hair and skin changes. There are additional gender-specific questions, including facial hair and voice change for boys and breast development and menarche for girls (11). There are four response options: 'not yet started, barely started, definitely started, seems complete'. The questionnaire does not contain any pictures or diagrams.

Other methods for self-assessment reported in the literature included a global question, about stage of puberty in relation to their peers, a visual analogue scale and questions about age at menarche. Many of the studies identified in this review were conducted in the United States.

Comparing self-assessment questionnaires with clinical assessment, precise agreement in Tanner stage has been reported in a range from 43% to 81%. For both the SMS and the PDS there were higher levels of agreement with clinical assessment to within one pubertal stage: Schmitz demonstrated 85% agreement to within one pubertal stage for the SMS. For the PDS, agreement to within one stage ranged from 85 to 100% (13, 15). In general, level of agreement for both scales was better at younger and older ages, perhaps corresponding to earlier and later stages of puberty when secondary sexual characteristics might be easier for young people to assess and report. Girls tended to be more accurate in their reporting of pubertal characteristics than boys. Certain aspects of development were more accurately reported by boys and girls than others: pubic hair development tended to correlate well with clinical assessment for both boys and girls, whereas breast and testicular development were more weakly correlated with clinical staging.

Comparison of the SMS self-assessment approach with the PDS questionnaire showed perfect agreement between the two measures in 39% of males and 56% of females (16). However there was agreement to within one Tanner stage for 97% of girls and 89% of boys. Agreement was higher for girls than boys and for 12-13 year olds than 10-11 year olds or 14-15 year olds. These findings differ from those where each measure is compared with the gold standard of clinical assessment. Boys tended to rate themselves as more mature according to SMS drawings compared to their PDS score and girls tended to rate themselves as less mature. This suggests that viewing the images of pubertal stage might encourage a socially desirable response among young people (16).

Berg-Kelly et al used a global question to assess pubertal stage. Adolescents were asked "Considering your bodily development, how do you rate yourself compared to your classmates: very late, somewhat late, similar, somewhat early or very early?" (17). The authors reported a high concordance with physician clinical assessment using Tanner staging: 95% for boys and 93.5% for girls.

Schmitz et al developed a visual analogue scale to assess pubertal stage (12). This allowed a continuum of assessment across the pubertal stages. To assess the perception of degree of progress through puberty, participants drew a line on a piece of paper between the two extreme Tanner stages for each secondary sexual characteristic. The measure was scored as a continuous measure in 1/16-inch increments with a standard ruler. When compared with self-assessment using the PDS or the SMS, it was found to be less accurate.

Some studies combined adolescent self-assessment with parental assessment of their child (15, 18). There was some evidence to support triangulation of assessments from young people with those of parents. Mothers were more likely to correctly predict pubertal stage for their daughters than for their sons. Brooks-Gunn et al demonstrated that mothers of girls rated their daughter's Tanner stage as higher than clinicians (15). This study was quite extensive, comparing self-report

using the SMS and the PDS, mother-report using the SMS, and clinician assessment. The authors stated that parental assessment of their sons may be less accurate; a finding supported by expert opinion from the ALSPAC cohort study team. In the more recent study by Lum et al, the self-assessments of parents were compared with those of young people performed a year earlier (18). The assessments of parents were more closely aligned with the findings of clinical assessment than those of young people. However, the different timing of the assessments made this study difficult to interpret.

Growth

We identified nine relevant studies. Studies of growth generally focused on two parameters – age at take-off, which indicates the start of pubertal growth spurt, and age at peak height velocity, which indicates the intensity of the pubertal growth spurt. Serial measurement of height is the most frequently reported approach to assessing pubertal stage and can be used to identify both age at take-off and peak height velocity. There is also some evidence to suggest that height velocity correlates with secondary sexual characteristics – a study by Bundak demonstrated high correlation of height velocity with testicular volume (19). Regular height measurement can also be used to identify the pre-pubertal growth spurt (20).

One approach to growth curve analysis is the SITAR method (21). It has the advantage, over conventional approaches, of taking account of characteristics that differ from one individual to the next - namely mean height, timing and rate of puberty. As well as being used for height, the SITAR method can be applied to growth in other parameters, such as foot length, and to the development of secondary sexual characteristics – breast, testicular and pubic hair development. The SITAR method produces three measurements representing differences in mean size and growth tempo and a measure of growth velocity.

Recent studies of change in foot length suggest that serial measurements of the foot might offer an effective method of detecting pubertal stage, and one that is likely to be acceptable to young people, given that it is less intrusive than a clinical examination. Busscher et al studied shoe shop foot size data and demonstrated that mean age at shoe size peak velocity occurred between 10 and 11 years of age in girls and between 11 and 12 years of age in boys (22). A study of girls by Ford et al demonstrated high correlation between age at increase in foot velocity and age at take off in height (23). Similarly, Mitra et al studied young people in India and showed that a rapid increase in foot length corresponded to stage 2 in the SMS based on clinical assessment (24).

Radiological approaches to assessing development have been proposed since skeletal maturation can be a proxy for stage of pubertal assessment. Studies of the cervical vertebra, olecranon and digits have been carried out. There is some evidence to suggest that a cervical-vertebral index indicates level of skeletal maturation, and that radiological images of the olecranon correlate with those of the digits, there is a lack of any comparison with a reference standard in these studies (25, 26). The absence of data on validity, combined with the acceptability issues, mean that these methods are unlikely to be appropriate for use in cohort, or other research, studies as a means of assessing pubertal stage.

Age at menarche has been used in clinical studies as a proxy for pubertal stage in girls. We did not identify any studies that validated this a measure of pubertal status, although a study based on the National Study of Health and Development used self or maternal-reported age at menarche as a predictor of bone development (27).

Hormonal assessment

Gonadotrophins and sex hormones

A number of studies discuss the potential for measurement of gonadotrophins as a means of assessing pubertal stage. At the onset of puberty, there is an increase in the overnight pulsatile release of LH suggesting that early morning urinary LH, adjusted for creatinine, might be a way of discriminating Tanner stage in girls (28, 29). Serum testosterone is aromatised to oestrogen in fat, and it is oestrogen that triggers growth hormone (GH) secretion in both sexes. Pulsatile GH secretion leads to increase in IGF-1 and insulin, which persist until around age 25 years when they begin to fall (30). Low dose oestrogen primes GH secretion, but high doses close epiphyses (31). Oestrogen levels in girls are less discriminatory in the early stages of puberty, even with good assays. Inhibin B is released from the ovary in pubertal girls, rising in early puberty. Inhibin A is slower to rise (32). Testosterone is low in girls, rising with puberty, but always remaining lower than levels observed in boys. For boys, testosterone rises throughout puberty with the steepest rise seen between stages 3 and 4.

There has been little validation of hormonal assays against clinical staging and we identified only four studies that reported a comparison between hormonal measurement and alternative methods of pubertal assessment. In a study examining the correlation of 3-monthly urinary oestradiol, testosterone and LH, with self-reported SMS Tanner staging, the levels of all three hormones were correlated with Tanner staging at baseline and 12 months later (33). The authors noted, however, that single sample testing is of limited value and that frequent longitudinal sampling might be necessary to determine the hormonal changes associated with pubertal stage. Added to this, they stated that urinary hormone changes may not be progressive and will likely be subject to considerable within-in person variation thus rendering longitudinal assessment difficult to interpret if, for example, only two samples were used.

For boys, a comparison of 24-hour testosterone levels (both total and free serum levels) with testicular volume suggested that testosterone levels in later puberty correlate reasonably well with testicular volume. Serum concentrations of testosterone increased progressively throughout puberty with a marked increase occurring between early and mid-puberty. The onset of puberty was marked by accentuation of the diurnal rhythm of testosterone release due to increased release of testosterone at night (34). However, assays of urinary sex steroids in boys are difficult to interpret due to within-person variability and repeated measures would be required to detect particular phases of pubertal development (35). For girls, assays of gonadotrophins and oestrogen levels give a broad indication of Tanner stage (36). Inhibin A and B have been seen as potential markers of pubertal status for boys and girls (33, 36). However, the studies by Chada et al and Crofton et al (35, 36) demonstrated the complexity of relationships between gonadotrophins, sex

hormones and inhibins in boys and girls, respectively, demonstrating the challenges of interpreting such data.

Leptin

Leptin is an adipocyte hormone that is important in regulating energy homeostasis. Leptin interacts with the reproductive axis at multiple sites with stimulatory effects at the hypothalamus and pituitary, and inhibitory action on the gonads (37). Evidence is accumulating that leptin potentially affects the regulation of GnRH and LH secretion during puberty, pregnancy and lactation (38).

Mantzoros et al examined serum leptin to assess whether levels could represent the hormonal signals responsible for triggering the onset of puberty in humans, in a cohort of eight pre-pubertal boys (39). They reported a 50% increase in serum leptin levels just before the onset of puberty, and a decrease to approximately baseline after the initiation of puberty. However, these findings were based on a small sample and, given that another study of school children showed that leptin levels vary widely from one individual to the next (40), it would be hard to design an appropriate approach to pubertal assessment using leptin assay. Added to this, the study by Carlsson et al showed no correlation between serum leptin levels and age in boys while there was a significant correlation in girls (41).

Other studies have examined the potential to assay urinary leptin, demonstrating that leptin excretion over six months (measured by three consecutive first morning urine samples) was higher in girls than boys (42). The study suggested that leptin was higher in children advancing into puberty compared to children remaining pre-pubertal, but the measure of pubertal status with which leptin levels were compared was very basic and poorly described. This study demonstrated that urine could be used as an alternative method for measuring pubertal status and is less invasive than blood. In another study of urinary leptin, Zaman et al related urinary leptin to pubertal stage (43). Early morning urine was used, and children were assessed using clinician-rated pubertal assessment using the Sexual Maturation Scale. The study showed that urinary leptin correlated with serum leptin, and circadian variability was observed. Urinary leptin was similar in the two sexes: in boys it increased significantly from stage 1 to stage 2, peaked in stage 3 and then declined for stage 4 and 5 while in girls there was a linear relationship between leptin levels and pubertal development.

Overall, the evidence relating sex hormones and leptin indicates that there is considerable individual variability in leptin, inhibin and other hormonal levels suggesting that serial measures might be required to make an assessment of pubertal stage. Furthermore, the evidence relating to patterns of leptin and inhibin in boys and girls during puberty is inconsistent.

Voice

The maturation of the human voice, as a function of age, is characterised by changes in pitch, loudness and a variety of tone qualities (44). In both sexes, the voice drops throughout childhood as the larynx grows. Voice breaking in boys usually occurs as a distinct event during late puberty due to the increased length of the vocal cords that follows the growth spurt of the larynx and represents a further non-invasive measure of pubertal timing (45). A rapid drop in voice occurs during

Tanner stages 3 and 4, usually around 12-15 years (46). The maximum change in the male voice takes place at puberty (44).

Cooksey et al defined a six-stage pattern of pubertal voice development based on the singing range in boys, which has been previously validated (47). Harries et al studied the relationship between the Cooksey classification of voice and the Tanner staging of pubertal development among Cambridge public school boys, demonstrating a clear correlation (44). The study assessed boys' voices at 3 monthly intervals and examined testosterone levels in saliva. Change in fundamental voice frequency was correlated with testicular volume but not with serum testosterone levels.

Workshop to discuss pubertal assessment methods (see appendix 2) – summary of findings

A workshop, held in September 2016, was attended by 21 people, most of whom were researchers with involvement, or an interest, in UK cohort studies. Presentations from experts in growth, biological approaches to, and clinical assessment of, pubertal status and a summary of the preliminary findings of this review, were followed by discussions of the best approaches to pubertal assessment within cohort and other research studies. There was general agreement that clinical assessment still represents the gold standard for assessment, but recognition that compliance among study participants was often low. Assessment of growth in height or foot size was agreed to be a promising approach, albeit that there was relatively little evidence relating to foot growth, and one that is feasible within cohort studies provided biannual measurements can take place. Such an approach would not be feasible within a cross-sectional study. Self-assessment, while much easier to administer within cohort studies and possibly the only realistic method other than clinical assessment in cross-sectional studies, was seen as a rather crude method of assessment given the likelihood that it will only accurately categorise pubertal status to within one Tanner stage. Assessment of voice was seen as a potentially promising approach, particularly given the availability of a free app to do this. However, the need for research to assess this, particularly given the likely differences in accuracy between boys and girls, was acknowledged. Hormonal approaches to assessment were seen as holding promise, but the need for repeated measures in each young person and the potential cost of assays, were seen as current barriers to their use in cohort studies. Since this is a rapidly changing field, it is possible that hormonal approaches might be more realistic methods of pubertal assessment in the future.

Lay panel to obtain the views of young people (see appendix 3) – summary of findings

We talked to a panel of 10 young people aged from 10 to 16 years – six boys and four girls, presenting them with information about the different approaches to

pubertal assessment and asking for their view on each. The young people had a preference for questionnaire approaches rather than clinical assessment - this applied to the boys and girls who were interviewed separately. They also preferred paper versions of questionnaires rather than digital approaches. The boys thought that measurements of height and foot size were acceptable and likely to be so for other boys that they knew. Likewise, having tested the voice app, they thought this was an acceptable approach and did not believe that boys would exaggerate the depth of their voice, provided the assessment was done individually. The views of girls were similar to those of the boys in relation to self-assessment questionnaires and measurements of height and foot size. They were generally happy to be asked about age at menarche, provided the person asking was a professional, but they thought their mothers might be more accurate in their reporting of this. Both boys and girls made suggestions about how to increase compliance with clinical assessment. They agreed that young people would be much more likely to agree to an assessment if it were made by a professional of the same gender as them. They felt that good communication that this was a quick assessment and did not require complete removal of their clothes, or for them to be touched by the person doing the assessment, would also be advantageous.

Discussion

Summary of findings

Both the review of evidence, and expert opinion, concluded that clinical assessment was still the most accurate approach to pubertal assessment at a given point in a young person's development. Evidence of inter-rater variability in those carrying out the assessment could be counteracted by thorough training and rigorous protocols within a research study. Evidence from young people suggested that good communication, coupled with an assessor of the same gender as the young person, could optimise clinical assessment.

Assessments of growth, particularly if utilising the SITAR method to take account of tempo, were seen as accurate approaches to pubertal assessment within longitudinal studies, provided biannual assessments of the parameter were feasible within the cohort study. While most of the current evidence relates to growth in height, the SITAR methods can be used for other parameters and, based on evidence from recent studies, repeated measurement of foot size might offer a promising method of growth assessment within cohort studies particularly if reported shoe size were shown to be a good proxy measure of size.

Utilisation of hormonal assays would be challenging at present given the need for repeated longitudinal assessment and the likelihood of within-person variability. Advances in this field might yield acceptable and reliable methods in the future.

Assessment of the voice is a method that is acceptable to young people taking part in research studies, particularly with the advent of app-based assessment. Further evidence is required to demonstrate the validity of this approach.

Strengths and limitations

In this review, we combined a synthesis of published evidence, with expert opinion and the views of young people. While this was not a systematic review, we used a number of the approaches that are conventionally employed in systematic reviews. These included extensive searches of the published literature, with input from an information specialist, careful screening of potential relevant abstracts and papers, and detailed data extraction and quality assessment of included studies. Our review does not cover all research published about approaches to pubertal assessment: our approach to identifying studies did not include a search of unpublished literature and so publication bias is likely and, given that abstract screening was conducted by a single reviewer, it is likely that not all relevant published studies were identified. Nevertheless, triangulation of the main findings of the literature review with the views of experts, and of young people, strengthens our interpretation of findings. There was considerable consistency between the views of the expert and the evidence from the literature.

Need for further research

In relation to clinical assessment, evaluation of the effect on compliance of the changes suggested by young people (assessor of the same gender as the young person, improved communication) would be of value and relatively easy to assess within existing CLOSER cohort studies. Further research on self-assessment could focus on the effect of including assessments both by the young person and their parents, perhaps with inclusion of paternal assessments for boys. Studies of growth, employing the SITAR method, for parameters other than height would be invaluable. This would be particularly beneficial when it comes to foot size since this is a measure that will be acceptable to young people, based on feedback from the lay panel. In addition, a comparison of measured foot size with self-reported shoe size would determine whether there is scope for measuring growth using repeated self-administered questionnaires. There is a need for further research to assess the validity of hormonal assays to assess pubertal status, with particular attention to the required frequency of measurement and to the potential for using less invasive methods, such as hair or urine sampling.

Table 1: Description of studies using self-assessment methods to assess pubertal status

First author and year	Article title	Study setting and sample	Study aims and methods of assessment	Validation and/or correlation	Quality assessment of the study	Key findings and conclusions
Lum, 2015 (18)	Assessing pubertal status in multi-ethnic primary school children	London, UK, multi-ethnic school children aged 8-11 years, and their parents. 246 parent-child pairs, as part of the Size and Lung function In Children (SLIC) study	To investigate the feasibility of assessing the attainment of secondary sex characteristics as a proxy for pubertal status, and to explore ethnic differences in rates of pubertal attainment. Cross-sectional survey of children using an illustrated Tanner questionnaire, with a yes/no question on the children's secondary sexual characteristics, answered by parents 12 months later.	None of the children underwent a clinical assessment. Child self-report is compared with the parental report.	Comparing two different methodologies at two different time points (one year apart). Therefore impossible to conduct a meaningful comparison between the methods. Paired data available for 246 children/parents.	Agreement between parental report and self-reporting of pubertal status in at least 68% of children. Overestimation by self-report compared to parents in 17% of children and underestimation in 15% of children. The later could be because of a change in development over a year. 25% girls and 62% boys were unsure of some aspects of their pubertal development. Children of Black African origin were more likely to have attained puberty at any of the ages studied than other ethnic groups. <u>Conclusion:</u> This paper concludes that parental assessment may be more reliable than self-assessment.
Rabbani, 2013 (48)	Reliability of pubertal self-assessment method: an Iranian study	Three contrasting regions within Iran, 190 boys	To compare a Persian Tanner Stages self-assessment questionnaire with	Comparison with clinical assessment	Of note the boys were allowed to examine themselves and complete the questionnaire, and then	Substantial agreement between self-assessment of pubertal status and clinical assessment. Complete agreement occurred in 72% of cases. This varied depending on

CLOSER Resource Report: Review of methods for determining pubertal status

		aged 8-16 years	physical examination.		they were examined by a clinician. Reasonable quality.	pubertal stage. 46% agreement with those in stage 3, 89% in stage 4 and 85% in stage 5. <u>Conclusion:</u> Self-assessment can be used in Iranian populations for children in later stage of pubertal development.
Norris, 2008 (49)	Are there shortcuts to pubertal assessment? Self-reported and assessed group differences in pubertal development in African adolescents.	South Africa, black youth, urban setting, 182 participants (49% female), aged 10-18 years	To compare self-assessment according to pubertal development scale (PDS) with clinical assessment using the sexual maturity scale (SMS).	Clinical assessment was used to validate the self-assessment methodologies	Self-assessment is not as reliable as a method for pubertal assessment as physical assessment. However, the self-assessment was done via a member of staff because of concerns regarding literacy and understanding. This could have biased the results. Little detail given regarding the study methodologies.	Females: 56% agreement between the PDS and the SMS. 28% females underestimated and 16% overestimated. Males: 26% agreement between PDS and the SMS, 70% males underestimated and 4% overestimated. <u>Conclusion:</u> Self-assessment is less reliable in multi-ethnic community-based research. This differs from outcomes in studies of Caucasian populations. There is no shortcut to reliable pubertal assessment.
Bond, 2006 (50)	A comparison of self-reported puberty using the Pubertal Development Scale and the Sexual Maturation Scale in a school-based epidemiologic survey	Australia, students aged 9-16 years from primary and secondary schools, multi-centre, 1486 girls and 1378 boys	Comparison of two self-assessment measures: SMS and PDS	Neither measure was compared against a physician rating.	Pupils completed the SMS and the PDS. The stage of pubertal development was then compared between the two. To examine the acceptability of these methods the authors considered the proportion of missing data.	Perfect agreement between the SMS and the PDS in 39% of males and 56% of females. 97% of females and 89% of males had concordance to within one category. Post-hoc categorisation into early, middle and late puberty improved the agreement. <u>Conclusion:</u> Due to better response rate and acceptability in schools, the use of PDS is to be preferred for self-assessment.

CLOSER Resource Report: Review of methods for determining pubertal status

Desmangles, 2006 (51)	Accuracy of pubertal Tanner staging self-reporting	USA, multicentre, urban setting, 240 children aged 6-16	SMS self-assessed and then compared with physician assessment. In the boys, pubic hair was used as surrogate marker of overall pubertal development in boys, this does not completely follow the SMS assessment	Comparison of self-assessment and physician made using levels of agreement and kappa statistic	No justification for only using pubic hair assessment for the boys. Children were allowed to assess their own development before completing the questionnaire. Large differences in the number of girls in each group e.g. there were 5 girls in Tanner stage 2 and 60 girls in Tanner stage 1.	For breast development 60% agreement between self-assessment and clinical rating. Of the 40% not in agreement, 48% overestimated and 52% underestimated. There was a higher correlation at 77% for pubic hair development. 61% of boys assessed pubic hair development correctly. Of the remainder, 73% overestimated and 27% underestimated. Higher degree of concordance amongst boys than other studies. <u>Conclusion:</u> Self-assessment did not appear reliable as a method of assessing pubertal development
Schmitz, 2004 (12)	A validation study of early adolescents' pubertal self-assessment	USA, children from various ethnic groups. Study 1: convenience sample of 178 children aged 8 to 18 years attending a diabetes and endocrine clinic, Study 2: 125 children aged 10 to 13 years	Four pubertal development self-assessment measures: PDS, SMS, attenuated three-point Tanner and a Tanner visual analogue scale. Different versions of the self-assessment methods were used for the two study groups. Study 1 children had a physician assessment and Study 2 children	Clinical assessment was used as the gold standard for Study 1. Correlations coefficients reported as well as kappa statistic.	Group 1 compared with physician rating but Group 2 compared with DXA bone mineral density. Multiple comparisons between all measurements are made. Correlations were calculated between pubertal measures and bone mineral density (all between 0.45 and 0.65). Very complex design that is hard to follow and every scale compared with every other one.	More than 85% agreement within one stage was obtained for all measures with the exception of breast assessment on the visual analogue scale. None of the self-assessment measures were statistically different from one another in the proportion of agreement with physician rating. While correlations were quite high with physician rating (apart from the PDS) the kappas were low (all below 0.5). <u>Conclusion:</u> The SMS self-assessment is possibly a more reliable marker of pubertal development than the PDS scale.

		recruited from an osteoporosis prevention study.	had a dual-energy X-ray absorptiometry (DXA) scan. 29 Study 1 participants repeated all four self-report measures three months later.			
Bonat, 2002 (52)	Self-assessment of pubertal stage in overweight children	USA, 244 non-obese and obese healthy children, aged 6-12 years	Determine the reliability of Tanner stage self-assessments in both non-obese and obese children. A standardised series of drawing with explanatory text for self-assessment. Breast and pubic hair for girls, pubic hair only for boys. Physical examination by a trained and experienced nurse or paediatrician.	Self-assessment against the references standard of physical examination	Selection and observer bias – for example, due to subject recruitment through mailing and no possibility of completely blinding assessors to the BMI of the child. Potential for overestimation of self-assessment accuracy due to a higher proportion of obese children than in the target population and younger age range in the sample. Interrater reliability 100% for breast and 98% for pubic hair due to substantial training.	Obese girls were more likely to overestimate breast stage than non-obese. Non-obese girls were not any more likely to over- or underestimate breast stage, and both obese and non-obese girls were not any more likely to over- or underestimate pubic hair stage. Both obese and non-obese boys were more likely to overestimate pubic hair stage. Majority of overestimations were in Tanner stage 1 or 2. <u>Conclusion:</u> Authors conclude that self-assessment of breast Tanner stage in young non-obese girls, and of pubic hair in all girls, is a reasonable substitute for a physical examination. The study addressed an under-researched area, but is limited by selection bias and a young age range in the sample.
Taylor, 2001 (53)	Performance of a new pubertal self-assessment	London, UK, 62 boys and 41 girls attending a	Self-administered SMS questionnaire compared with	Sequentially recruited children with 81% response	A high quality study allowing a direct comparison between clinician rating and self-	For pubic hair distribution there was agreement to within one Tanner stage in 88% of children. For female breast and male

CLOSER Resource Report: Review of methods for determining pubertal status

	questionnaire: a preliminary study	paediatric endocrinology clinic	clinician assessment.	rate. Agreement and kappa statistics used for comparisons.	rating, though the numbers are quite small.	genitalia stage there was agreement to within one Tanner stage for 76% of children. The children tended to underestimate their level of development. <u>Conclusion:</u> Overall good agreement between self-assessment SMS and clinical assessment for pubic hair, and moderate agreement for breast/genitalia stage. This questionnaire may prove useful in cohort and longitudinal studies.
Hergenroeder, 1999 (13)	Validity of self-assessment of pubertal maturation in African American and European American adolescents	USA, 107 healthy girls aged 8-17 years	To evaluate interobserver reliability of physician assessments of pubertal maturation in girls in a multi-ethnic sample. Next, to evaluate the validity of self-assessment compared to physician assessments of pubertal maturation. Assessment by one of two physicians in adolescent medicine, self-assessment using drawings and	Validation of self-assessment against reference standard of physical Tanner staging	Selection and observer bias, potential for measurement error – self-assessment appeared to be from memory, rather than after self-examination. Small numbers of ethnic minority girls, which were excluded from some of the analyses.	Interobserver reliability: 84% for pubic hair and 76% for breast stage. Self-assessment: tendency to overestimate pubic hair development and no consistent trend for breast stage. Validity of self-assessment versus physical examination: kappa 0.34 for breast and 0.37 for pubic hair, i.e. marginal. <u>Conclusion:</u> Authors conclude that self-assessment is not valid and interobserver reliability was low. Significant bias and confounding, small numbers.

			descriptions of five stages of development according to Tanner.			
Koo, 1997 (54)	Accuracy of short-term recall of age at menarche	Canada, 101 girls who had reported menarche in previous two annual follow-ups	To investigate the accuracy of recall by parents and daughter of age at menarche within interval of recall of 1-2 years. Parents and daughters were asked to recall age at menarche in the third follow-up.	Comparison of delayed recall with recall within 12 months of menarche	Selection bias and lack of basic information about recruitment and study participants. Small numbers, with only 88 usable responses.	Mean interval of recall 430 days. 59.1% able to recall month and year and 77.3% within one month of the originally reported date. More subjects able to recall correctly with a shorter interval of recall. No tendency to recall menarche occurring earlier or later than originally reported. <u>Conclusion:</u> Short report of a small study. Indicates an association of the time since menarche with reduced accuracy of recall by parents and daughters, even within the short period of three years.
Berg-Kelly, 1997 (17)	Self-assessment of sexual maturity by mid-adolescents based on a global question	Goteborg, Sweden, 4516 young people, multi-centre	Cross-sectional study of three different school year groups (average age 13.5, 15.5 and 17.5 years). A global question: 'Considering your bodily development, how do you rate yourself compared to your classmates: very late, somewhat late,	Comparison of the categories of the global question (excluding 'similar') was made with self-assessment and various markers of puberty (menarche, voice changes etc.), presented as percentages.	Large study of 4516 young people, although limited by its cross-sectional design. Small sample validated in terms of 'gold standard.' Those who were assessed volunteered (bias). Little evidence about the validation of the Q90 questionnaire. Self-assessment was done using descriptions of the Tanner staging rather than the usual	Two-thirds rated themselves as average in development, similar across age ranges. Boys under-rated their pubic hair growth in comparison with clinical assessment. Agreement between global question and clinician rating was 94%. The question is simple to apply, has universal application and does not require clinical assessment. <u>Conclusion:</u> The study suggests that the global question has potential as a measure of pubertal development, although it does not

			similar or very early?' Global question was compared with the Q90 questionnaire, which is a set of questions about pubertal development, based on Tanner stages and other aspects of growth. 100 volunteers also underwent clinical assessment.	Sub-group self-assessment was compared with professional assessment of pubertal stage, with percentage agreement reported.	pictorial questions so the validation of the global question was not against a recognised gold standard. In the validation sub-sample the self-assessment questions were compared with the professional clinical assessments and percentage agreement was reported.	provide an indication of time point within puberty. You would need to correlate that with age and predicted pubertal stage for that individual. Need to ensure that adolescents have adequate numbers of peers around them in-order to make comparisons.
Carskadon, 1993 (55)	A self-administered rating scale for pubertal development	USA, school children, 323 boys, 375 girls, their parents and teachers	A self-assessment scale, suitable for completion in the classroom, adapted from the PDS, without use of pictures or the need for an interview A parallel teacher and parent form of the scale was also developed	Comparison between assessments made by the teenagers themselves, their parents and their teachers. No comparison with a clinical assessment.	A pilot study was done with 38 students prior to this study. The correlation with physician assessment ($r = -0.85$) was good and therefore the self-assessment alone was used in this study. Response bias of the sampling group: children who returned the questionnaire were more socially mature and more able academically. 43% boys and 46% of girls had "missing data" and were not included in the study.	Significant correlations were found between parents and students for all the measures for 6 th graders and 5 th grade girls and several measures for 6 th grade boys. Correlations between parent and children were better for girls than for boys. Teachers' physical ratings demonstrated the lowest correspondence to student and parent ratings. <u>Conclusion:</u> The scales worked better for older children, particularly for boys. Authors concluded that the scales were useful measures of maturational status in settings where direct examinations, interviews or pictorial representations were not possible.

CLOSER Resource Report: Review of methods for determining pubertal status

<p>Petersen, 1988 (11)</p>	<p>A self-report measure of pubertal status: reliability, validity and initial norms</p>	<p>USA, two successive birth cohort samples, Caucasian, middle and upper class, data on 252 children</p>	<p>To assess a PDS questionnaire with height measurements. Longitudinal study with five measurements</p>	<p>Not validated, but a subset had a comparison with growth measurements, and correlations of peak height velocity with scores from the PDS were reported. Internal consistency assessed using alpha coefficients (assumed to be Cronbach's alpha).</p>	<p>335 boys and girls, data on 252 subjects. Repeated measurement of two cohorts of adolescents. Twice annual assessments, over three years. They completed the same self-assessment according to the PDS. Interview-based assessment. Correlations between age at peak height velocity and PDS scores at each age are all negative and hard to interpret.</p>	<p>The longitudinal data indicated that assessment with the PDS reflected the sequence of pubertal Tanner staging. 10% of girls and 6% of boys decreased their level of development. The highest amount of regression was observed with skin changes and growth. Of note acne is a temporary phase and therefore regression may be expected. Alpha coefficients ranged between 0.68 and 0.83 for the different ages of measurement indicating good internal consistency. <u>Conclusion:</u> A self-report measure may be useful if the researcher is interested in an approximate assessment of pubertal status.</p>
<p>Brooks-Gunn, 1987 (15)</p>	<p>Validity of self-report measures of girls pubertal status</p>	<p>USA, 151 girls aged 11-13 years recruited from a large cross-sectional study of female development</p>	<p>Self-completed PDS and SMS, and a maternal assessment using the SMS. Each measurement was compared with physician assessment.</p>	<p>Compared with physician assessment using correlations and alpha-coefficients</p>	<p>This study was done in 151 girls. It was not performed in boys because of opposite sex modesty, which emerges in adolescence and may render mothers an inaccurate source. Considers the potential influence of social desirability and the existence of a set of norms.</p>	<p>The correlation with physician ratings was 0.82 for the SMS self-ratings, 0.85 for ratings with the mother, and 0.64 for self-reports on the PDS. <u>Conclusion:</u> The PDS was less valid than the SMS for self-assessment of pubertal development. Mothers could rate their daughters' maturation well. The PDS may be useful in situations where the use of pictures is unacceptable.</p>

CLOSER Resource Report: Review of methods for determining pubertal status

<p>Duke, 1980 (56)</p>	<p>Adolescents self-assessment of sexual maturation</p>	<p>USA, 43 females aged 9 to 17, 23 males aged 11 to 18</p>	<p>To demonstrate that adolescents can accurately assess their own developmental stage according to Tanners standard photographs. Comparison of SMS by self-assessment with that by physician examination.</p>	<p>Validated against physician use of the SMS</p>	<p>Small numbers. Draws on the value of understanding the wide range of normal pubertal maturation and to follow their own sexual maturation over time. This outlines the benefit of self-assessment.</p>	<p>Agreement with the physician rating for breast development was 86% (kappa 0.81), female pubic hair 93% (kappa 0.91) and male genital stage 91% (kappa 0.88). <u>Conclusion:</u> Authors concluded that teenagers could accurately designate their own level of sexual maturation.</p>
<p>Morris, 1980 (14)</p>	<p>Validation of a self-administered instrument to assess stage of adolescent development</p>	<p>USA, 47 females and 48 males aged 12-16 years</p>	<p>SMS completed by the participants, which were then assessed by a physician who was blind to the self-rated levels</p>	<p>Validated against physician use of the SMS</p>	<p>Pearson's correlation coefficients used inappropriately for the analysis</p>	<p>Adolescent male estimates of genital development, and genital and underarm hair, correlated 0.6 with physician observation. Male facial hair correlated less well, at 0.3. Estimates of testicular volume did not correlate. No indication of agreement given but from a table of self and physician ratings, agreement of Tanner stage of girls' breast development can be calculated as 51%. <u>Conclusion:</u> Overall variable correlation between self and physician ratings, but hard to interpret these findings.</p>

Table 2: Description of studies using growth methods to assess pubertal status

First author and year	Article title	Study setting and sample	Study aims and methods of assessment	Validation and/or correlation	Quality assessment of the study	Key findings and conclusions
Kuh, 2016 (27)	Pubertal timing and bone phenotype in early old age: findings from a British cohort study	UK, cohort study of 866 women and 762 men, assessment aged 14-15 years	To investigate the effect of pubertal timing, assessed in adolescence, on bone size, strength and density in early old age. Pubertal assessment: examination and an interview by a school doctor aged 14-15 years old. Girls: age at menarche from mothers' reports if postmenarchal or from a postal questionnaire aged 48. Boys: pubertal staging based on voice breaking and the development of genitalia and pubic and axillary hair. SITAR (Superimposition by Translation and Rotation) model for	Correlation was with bone phenotype in early old age, rather than between methods of pubertal or growth assessment	In relation to pubertal and growth assessment only: no pubertal staging for girls, other than by age of menarche. Selection bias, including that resulting from limiting the analysis to clinic attendees. Observer bias in relation to clinical assessment by a school doctor. Recall bias for age of menarche and self-reported height at age 20 and 26 years. Somewhat limited generalisability due to the cohort being born in the post-war period.	No pubertal staging was done for girls, but mean age at menarche was 13 years. 26% boys were fully mature, 30% advanced, 34% at an early stage and 10% pre-adolescent by 14.5 years of age. No separate SITAR results given. Since validation or correlation of methods of pubertal assessment was not the aim of this study, few of the relevant results are reported separately. <u>Conclusion:</u> Pubertal assessment methods, although useful in predicting bone phenotype in later life, were not the focus of this study, and it is difficult to draw any conclusions on their quality, given the limited information provided in the article.

			growth curve analysis.			
Canavese, 2014 (26)	A comparison of the simplified olecranon and digital methods of assessment of skeletal maturity during the pubertal growth spurt	France, 44 boys and 78 girls in an orthopaedic clinic	To compare the accuracy of the simplified olecranon and digital methods during the pubertal growth spurt and investigate this in relation to the experience of the assessor. X-ray radiographs of the hand and elbow to determine skeletal age and correlate this with the stage of pubertal growth. These were taken during the follow-up of children whose standing height increases by more than 4cm in the last six months.	No pubertal assessment of secondary sexual characteristics. Correlation between the two methods, as per study aims.	Selection and observer bias – recruitment from an orthopaedic clinic; more girls than boys; decision to take radiographs made by the surgeon; method based on personal judgement of images; and measurement error. No comparison to a reference standard test in a study attempting to compare accuracy. Numbers appear small, but the quoted sample size calculations suggested a minimum of 20 measurements. Strong correlation between nine observers doing two reviews.	Correlation between the two methods was strong with $r=0.84$ for boys and 0.83 for girls. The pubertal growth spurt occurred at 11-13 years of skeletal age in girls and 13-15 years in boys. <u>Conclusion:</u> The olecranon method offered detailed information during the pubertal growth spurt. The digital method was as accurate, but less detailed, making it more useful after the pubertal growth spurt, once the olecranon has ossified. Lack of a reference standard test, a considerable potential for bias, and lack of information on confounders are of note. The study could have been expanded by including pubertal development staging or sitting and standing heights.
Busscher, 2011 (22)	The value of shoe size for prediction of the timing of the pubertal growth spurt	Netherlands, shoe shop data, 242 girls and 104 boys in early puberty: mean length of follow-up after	To describe the increase in shoe size during adolescence and determine whether the timing of the peak increase could be an early	Correlation of the increase in shoe size from this study with the increase in sitting height from an unrelated data	Longitudinal data, which are rarely otherwise available. Selection bias due to the use of data from shoe shops. Not clear whether retrospective or prospective study. Lack	Mean peak increase in shoe size for girls was 10.4 years and for boys was 11.5 years. Mean peak velocity of shoe size increase was 2.4 shoe sizes per year for girls and 2.6 shoe sizes for boys. Average age for girls (n=138) to reach a plateau

		8 years of age – 5.4 years for girls and 6.2 for boys	indicator of the peak growth velocity of sitting height. Use of longitudinal data of shoe sizes from two shoe shops.	source in the Netherlands	of demographic information and adjustment for confounding. Shoe size may be brand-dependent. Fewer boys than girls, and boys could have a second foot growth spurt not captured in this study.	phase in shoe size was 12 years. Average for boys (n=36) was 13.7 years. The mean peak increase in shoe size generally occurred 1.3 years and 2.5 years before the mean peak growth of sitting height, in girls and boys respectively, when correlating the results with sitting height from a different data source (Gerver and de Bruin, 2001) <u>Conclusion:</u> Authors conclude that shoe size increase can be a useful indicator for the timing of the sitting height growth spurt, however, this is based on cross-study correlation, therefore accuracy and reliability of this method remains to be proven.
Mitra, 2011 (24)	Foot length as a marker of pubertal onset	Urban India, 973 middle income healthy children, age 8-16 years	Correlation of foot length and full sexual maturity rating (SMR): breast and pubic hair development staging for girls and genital and pubic hair development staging for boys.	Sexual development based on Tanner staging of puberty	Cross-sectional study with only one measurement per subject. Potential for selection and measurement bias, with insufficient information provided. Sample size calculations included, but the study nonetheless underpowered. Full Tanner staging carried out by two types of secondary sexual characteristics, unlike	Foot length increased rapidly in stage 2 of SMR, coinciding with the onset of puberty, with no statistically significant variations in foot size in subsequent stages. <u>Conclusion:</u> Foot length measurement may be an acceptable and easy method, which can be used as an indicator of pubertal development. However, this study is limited by its cross-sectional design, and further research is needed.

					many other studies, which rely on just one. High observer agreement, although the size of the subset, in which this was studied, is not specified.	
Cole, 2010 (21)	SITAR – a useful instrument for growth curve analysis	UK, 3245 boys aged 9-19 years from Christ's Hospital School (1936–1964) and 105 girls with Turner syndrome aged 9-18 years from a randomised controlled trial of oxandrolone to increase final height	To fit the model described by Beath and show how effectively it summarises height increase around the time of puberty. To show how the estimated subject-specific parameters can be related to earlier exposures and later outcomes. Assessment of growth using the SITAR model. A shape invariant model combining three parameters of growth: size, tempo and velocity. Boys' height was measured twice a term between 9 and 19 years of age.	No correlation to stage of puberty. Using an already established method.	Selection, attrition and measurement bias, due to the study design, recruitment methods and unavoidable lack of blinding, in particular in the case of Turner syndrome. Substantial sample size for boys. External validity in girls is limited by the use of girls with Turner syndrome on various combinations of growth-enhancing treatment.	The SITAR model explained 99% of the variance in both datasets, matching the fit of individually-fitted Preece–Baines curves. SITAR is a shape invariant growth curve that summarises individual growth curves with a single summary curve and subject specific random effects. The random effects reflect each subject's size, growth, tempo and growth velocity. <u>Conclusion:</u> A valid approach to pubertal assessment, using a non-invasive method, but requiring frequent growth measurements, which is impractical in a cohort setting.

<p>Ford, 2009 (23)</p>	<p>Early markers of pubertal onset: height and foot size</p>	<p>USA, 86 girls aged 6-7 years</p>	<p>To determine whether change in foot size may be used as a marker of onset of puberty. Height, weight, foot size and maturation ratings obtained every six months. Breast and pubic hair staging by inspection and palpation (the latter presumably only for breast) by a 'trained healthcare professional'.</p>	<p>Correlation of velocity of foot size increase with height velocity increase and the age of onset of secondary characteristics.</p>	<p>Potential for selection and observer bias due to lack of information on recruitment and whether the same observers assessed height and pubertal development. Weight was measured, but not used in the analysis, in the form of body mass index (BMI) or otherwise. Adjusted for ethnicity, but not other confounders. Six-monthly assessment may not be sufficiently frequent when estimating onset of puberty with a precision of months.</p>	<p>Onset of secondary sexual characteristics was at a mean age of 8.79 years; foot velocity increase – 8.40 years; height velocity increase – 8.27 years. Statistically significant difference between age of onset of secondary sexual characteristics and age of foot velocity increase, as well as age of height velocity increase. No significant difference in age of foot velocity increase and height velocity increase. No significant race interaction, but African-American girls reached maximum height and foot velocities, and started developing secondary sexual characteristics on average of 0.42 years before their white American peers. <u>Conclusion:</u> Authors conclude that change in foot size could be an early marker to document onset of puberty, but there was no difference between the age of increased foot velocity and height velocity. Hence on the basis of this study there appears to be little benefit of using foot size over height, in particular as height-measuring equipment may be more readily available than foot-measuring devices.</p>
------------------------	--	-------------------------------------	--	---	---	---

<p>Bundak, 2007 (19)</p>	<p>Analysis of puberty and pubertal growth in healthy boys</p>	<p>Urban Turkey, 1112 healthy boys aged 8-18 years, high socioeconomic class</p>	<p>To provide normative data for the onset and tempo of puberty in Turkish boys, evaluate height velocity at each testicular volume stage and analyse the growth parameters in puberty. Biannual growth measurements (weight and height) and calculation of growth velocity and final height (FH) – velocity of less than 0.5 cm/year. Measurement of pubertal development (testis volume, pubic and axillary hair).</p>	<p>Compared growth to Tanner pubertal development stage</p>	<p>Benefits from longitudinal design, but individual boys were followed up for different time periods. Selection bias: for example, not stated how the schools were chosen; the sample consisted of the highest socio-economic class only; not stated how the rather small subsample of 30 boys for follow-up to FH was selected, or whether these were the only boys with a complete follow-up from a sample of 1112. Axillary hair assessment is referred to, but this is not normally part of Tanner staging. Good recognition of the role of inter-observer variability and the need for training and consistency. One observer performed the testicular volume measurements. Random subsamples of boys and girls are referred to, although the sample consisted of boys only.</p>	<p>Height at onset of puberty was positively correlated with FH and with duration of puberty. BMI at onset of puberty correlated negatively with age at onset of puberty and with the duration of puberty, but not with FH. Maximal height velocity was obtained after the testes attained a volume of 10ml. The testes reached this volume at a mean age of 12.9 years and a mean height of 155cm. <u>Conclusion:</u> Height at onset of puberty was the most important determinant of FH. Height and height velocity were correlated with each testicular volume stage in this study. This allows measurement of peak height velocity and determination of pubertal stage through the correlated testis volume, without directly measuring testis volume.</p>
--------------------------	--	--	--	---	--	---

<p>Ozer, 2006 (25)</p>	<p>A practical method for determining the pubertal growth spurt</p>	<p>150 patients, 9 to 19 years old, selected at random from 300 male patients referred for orthodontic treatment</p>	<p>To determine correlation of the cervical vertebrae maturation (CVM) index with the modified medial phalange index (MP3). Use of lateral cephalometric and hand phalange X-ray radiographs</p>	<p>No pubertal assessment of secondary sexual characteristics. Correlation of CVM with MP3</p>	<p>Selection bias due to recruitment from an orthodontic clinic. Small numbers in some of the maturation stages. Simple percentage agreement is given for both inter-rater reliability and correlation of the indices. Much of the article concentrated on describing cervical and phalangeal maturation stages, but analysis of the study findings is limited.</p>	<p>CVM and MP3 maturation stages had 89.3% simple percentage agreement. <u>Conclusion:</u> The authors conclude that the modified MP3 index correlates to the CVM index and can be a useful method to assess skeletal maturation. Whilst this may be so, methodological and analysis limitations restrict the usefulness of the findings of this study.</p>
<p>Karlberg, 2003 (20)</p>	<p>Pubertal growth assessment</p>	<p>Sweden, 81 boys and 64 girls aged 3-18 years, with annual examinations, from an international longitudinal growth study of healthy children</p>	<p>To introduce a pre-pubertal standard for the assessment of pre-pubertal height for children with late onset of puberty, to apply a final height prediction method, and to devise a method for assessing total pubertal height gain. By plotting height values for each child in a chart containing pre-pubertal reference values one can</p>	<p>No correlation to pubertal staging</p>	<p>There is no correlation of growth to stage of pubertal development. The study uses data from another study and concentrates on a small proportion of the results, raising questions about the methodology and data dredging. Therefore it is difficult to fully assess the quality of the paper.</p>	<p>Infancy-childhood-puberty (ICP)-constructed reference values for pre-pubertal growth agreed with those derived directly from the growth record data used in this study. Pre-pubertal growth and pubertal onset can be assessed in a more appropriate way using the ICP model than ordinary cross-sectional standards. <u>Conclusion:</u> This paper delineates a method by which clinicians can ascertain whether a child with late onset puberty is entering very early puberty, as identified by the prepubertal growth spurt. It also identifies a model of prediction of final height and a method of</p>

			identify the onset of the pubertal growth spurt, and thereafter apply a final height prediction method.			assessing total pubertal height gain. Since secondary sexual characteristics develop chronologically after the pre-pubertal growth spurt, it provides an earlier time point for assessment, making an argument for the value of regular height measurements and calculations of height velocity.
--	--	--	---	--	--	--

Table 3: Description of studies using gonadotrophin or gonadal hormone methods to assess pubertal status

First author and year	Article title	Study setting and sample	Study aims and methods of assessment	Validation and/or correlation	Quality assessment of the study	Key findings and conclusions
Singh, 2015 (33)	Urinary sex steroids and anthropometric markers of puberty – a novel approach to characterising within-person changes of puberty hormones	Australia, 104 healthy children (57 girls), age 10-12 years	To examine the feasibility of three-monthly urine collection in young adolescents, as well as the utility of liquid-chromatography-tandem mass spectrometry assays for urine and serum sex hormones. To examine the association	Hormone levels were correlated with Tanner staging by self-assessment	Potential for selection and measurement bias and/or error: very little information on recruitment and participants; no objective clinical examination; pubic hair assessment not included for reasons not specified. Seemingly small numbers, but the authors use the term 'large cohort', and	The levels of all three hormones positively correlated with Tanner staging at baseline and 12-month follow-up. Change in height was associated with changes in serum testosterone, and serum and urine luteinising hormone (LH) in females and both serum and urine testosterone in males. Change in weight was associated with changes in urine oestradiol, serum testosterone and serum LH in females. Serum testosterone and LH was associated with self-rated Tanner stage in males. Hormone concentrations increased through each Tanner stage and each

			<p>between the changes in anthropometry and self-reported Tanner stage with changes in hormones in both urine and serum over a 12-month period.</p> <p>Self-reported breast staging for girls and genital staging for boys using line drawings.</p>		<p>no details of sample size calculations are provided.</p>	<p>year of age. High compliance, low attrition rates.</p> <p><u>Conclusion:</u> Morning urine sampling can be advantageous over single blood sampling due to pulsatile hormone release, in particular in early puberty. However, single sample testing is of limited value and frequent longitudinal sampling may be necessary. In addition, urinary hormone changes may not be progressive, indicating within-subject variability. Tanner stage and anthropometrical changes lag behind hormonal changes. Yet urine sampling over 2-3 years may be useful in explaining biological basis for patterns of pubertal changes. This particular study is limited by the self-reported nature of the data used to derive Tanner staging, based on one parameter only. There were seemingly small numbers, and insufficient detail on participant characteristics to exclude selection bias.</p>
Ankarberg-Lindgren, 2004 (34)	<p>Changes of diurnal rhythm and levels of total and free testosterone secretion from pre to late puberty in boys: testis size of 3ml is a transition stage to puberty</p>	<p>Sweden, 55 boys, age 5.0-18.6 years, healthy (n=25) or investigated for short (n=26) or tall (n=4) stature</p>	<p>To establish levels for comparison for 24-h total and free serum testosterone before and throughout puberty, and to relate these values to testicular volume and</p>	<p>All subjects underwent hormone studies, orchidometry and pubic hair assessment according to Tanner</p>	<p>Small sample size, potential for selection and observer bias: little detail is given on recruitment; more than half of the sample had growth abnormalities; not clear whether there was adequate blinding of observer at various stages,</p>	<p>Based on the changes in testosterone and calculated free testosterone (calc-FT) in relation to testicular volume, the authors classified puberty into six stages. Serum concentrations of testosterone increased progressively to a constant high value in late puberty, with a marked increase between early and mid-puberty. Serum concentrations of calc-FT, however, increased continuously throughout puberty, with a striking increase between early puberty and the penultimate stage of puberty.</p>

			<p>pubic hair development. Tanner staging done by a clinician as part of investigation for short stature every three months from the time of enrolment. In healthy controls and tall boys staging done within one year before and after blood samples. 4 of 55 boys had samples taken repeatedly through puberty, 6 participated 2-3 times, and 45 only once.</p>		<p>including correlating blood test results, testicular size and Tanner staging; inconsistent frequency of blood sampling and Tanner staging. Caucasian sample, half of which investigated for abnormal stature; lack of information on recruitment and baseline characteristics.</p>	<p>The onset of puberty was characterised by an accentuation of the diurnal rhythm due to an early nocturnal increase in testosterone as compared with the prepubertal diurnal rhythm. Late puberty is characterized by a diminished diurnal rhythm compared with early and mid-puberty. <u>Conclusion:</u> This is a small study consisting of detailed investigations of diurnal variation and levels of testosterone before and during puberty, which correlated these to testicular volume and pubic hair stage,</p>
<p>Chada, 2003 (35)</p>	<p>Inhibin B, follicle stimulating hormone, luteinizing hormone and testosterone during childhood and puberty in males: changes in</p>	<p>The Czech Republic, 78 healthy boys throughout childhood and adolescence</p>	<p>To determine changes in serum inhibin B, follicle stimulating hormone (FSH), luteinizing hormone (LH) and testosterone concentrations during childhood</p>	<p>Hormone levels correlated with Tanner staging – genital only, no pubic hair assessment. No further details are given on how and by whom</p>	<p>Potential for selection and measurement bias and error. Insufficient information is given to enable judgement of the internal and external validity of the study. Critique of the laboratory methods is outside</p>	<p>Inhibin B increased with pubertal progression and reached a peak in stage G3 (third out of five Tanner stages of puberty), as well as observing a rise in serum gonadotropins and testosterone during pubertal development. The correlation of inhibin B to FSH, LH and testosterone changed during puberty in agreement with previous studies. In early puberty, serum inhibin B concentrations had a positive association with testosterone and LH, but little</p>

	serum concentrations in relation to age and stage of puberty		and puberty in males and investigate the relationship between these with regards to age and stage of puberty in boys.	Tanner staging was carried out.	of the scope of this review. Small numbers in each age category. Many values did not reach a statistically significant level for the difference of means of age categories.	relationship to FSH. In contrast, around mid-puberty (stage G3), inhibin B lost its positive correlation with LH and testosterone, but developed a strong negative correlation with FSH, which persisted into adulthood. <u>Conclusions:</u> The study may add to the evidence available on the concentrations and interrelationships of the hormones studied, which would be useful in the investigation of gonadal disorders in children, but there is less scope of using the study's findings for the assessment of pubertal staging in epidemiological studies, partly due to complicated methodology and individual variations in hormone levels.
Crofton, 2002 (36)	Dimeric inhibins in girls from birth to adulthood: relationship with age, pubertal stage, FSH and oestradiol	Scotland and Ireland, 345 healthy girls 0-18 years for age-related reference data and 80 premenarchal girls with full pubertal staging: 51 of these investigated for familial short stature, 40 out of which in addition	To investigate how dimeric inhibins change from birth to late adolescence in girls, to derive reference data and to explore their relation with pubertal stage, FSH, oestradiol and each other.	A single observer carried out pubertal assessment (seemingly breast staging only) in girls investigated for familial short stature. No details of the assessment in relation to the 29 girls who presented for	Selection bias. Only breast staging is used in the analysis, but misleadingly 'full pubertal assessment' is referred to in places. Most girls with pubertal staging were early on in pubertal development (only 14 girls at stages B3-5 of breast development). Half the girls with pubertal staging were being treated with GH as part of	Both inhibins varied markedly with age, and inhibins A, B, FSH and oestradiol all showed a marked relationship with pubertal stage. Detailed results of the hormone concentration fluctuations by age and puberty stage were also provided. <u>Conclusions:</u> Several conclusions are offered by the authors in relation to correlations and changes in the levels of the hormones studied. 'To summarise, in girls in late puberty, there may be complex interactions among FSH, oestradiol and both inhibins that would require more detailed longitudinal and interventional studies to elucidate'. The authors state that these age- and puberty-related data will be valuable in future studies on inhibins A and B as

CLOSER Resource Report: Review of methods for determining pubertal status

		treated with growth hormone (GH)		minor illnesses, other than 'their pubertal stage was appropriate for age'.	their usual care, and all the girls taking GH in stage B2 were removed from the analysis due to higher levels of inhibin B and FSH in this group. No information as to whether menarche had been reached and the stage of menstrual cycle. No information on confounders other than age. Not clear whether any blinding took place at any stage.	markers of follicular development. However, the variability of, and lack of information on, the recruitment methods and subjects' baseline characteristics limit internal and external validity of the study.
--	--	----------------------------------	--	---	--	---

Table 4: Description of studies using leptin methods to assess pubertal status

First author and year	Article title	Study setting and sample	Study aims and methods of assessment	Validation and/or correlation	Quality assessment of the study	Key findings and conclusions
Maqsood, 2007 (42)	The relationship between nocturnal urinary leptin and gonadotrophins as children progress towards puberty	Manchester, UK, 20 children (13 boys) judged to be close to the initiation of puberty	Three consecutive first morning urine samples were collected from each subject each month over 6 months. At the end of the study, the children were classified into those who remained physically prepubertal (n = 7) and those that had advanced in puberty (n = 13). Leptin and gonadotrophins were measured by immunoradiometric and immunofluorometric assay, respectively.	Arbitrary measure of puberty, looking at whether the child has remained physically pre-pubertal or advanced in puberty	Early morning urine samples collected for leptin. Oversimplified the assessment of pubertal development. Focused more on correlation between follicle stimulating hormone (FSH), luteinising hormone (LH) and leptin, than pubertal stage or development and leptin. Small numbers.	In children approaching and progressing into puberty, leptin was associated with changes in LH and FSH over the same timeframe. <u>Conclusion:</u> The data imply that leptin is an important facilitator in the early stages of pubertal development.
Wang, 2004 (40)	Serum leptin levels in healthy adolescents: effects of gender and growth	Japan, city population, 906 schoolchildren, aged 9 to 17 years	Cross-sectional study, incorporating height and other measurements, as well as a questionnaire. Fasting blood samples collected	Pubertal assessment – age of menarche in girls and maximum increment in height (MIA) used in boys. They then	The method of pubertal assessment does not appear to have been validated and the only references are to papers from their team published in local reports	The paper establishes standard age variation curves of serum leptin levels in healthy adolescents by calculating 25 th , 50 th and 75 th centiles for each age in both boys and girls. Leptin rises

			and serum leptin measured.	created a pubertal stage based upon growth.	No formal pubertal assessment using standard methods. Large numbers 906 children.	in boys before MIA then falls, but steadily rises in girls. <u>Conclusion:</u> These curves are a potentially useful method for evaluating serum leptin levels in adolescents, considering the impact of gender and growth, but the ranges of measurement at each age are wide, particularly for boys, and these are presented by chronological age, not pubertal stage.
Zaman, 2003 (43)	Leptin measurement in the urine in children and its relationship to other growth peptides in the serum and the urine.	Manchester, UK, 188 healthy school children aged 5-19 years, recruited as part of a larger study.	Cross-sectional study of urine leptin, measured in the first morning void urine and expressed as nanograms excreted overnight, and serum concentrations of leptin, IGF-1, IGF-11, IGFBP-3 and IGFBP-1 were determined.	Pubertal status assessed using the sexual maturation scale.	Potential for selection and measurement bias – no information on recruitment into this or the larger study, from which the data were used; not clear if pubertal staging was by self-assessment or an observer. Numbers are higher than is often the case in studies on hormones, but small numbers in some pubertal stages.	Urinary leptin showed similar changes through puberty to those of serum leptin, with levels rising in females throughout puberty, whereas levels in males peaked in stages 2/3 and then decreased. <u>Conclusion:</u> Urinary leptin is a valid marker of serum leptin concentration.

CLOSER Resource Report: Review of methods for determining pubertal status

<p>Mantzoros, 1997 (39)</p>	<p>A longitudinal assessment of hormonal and physical alterations during normal puberty on boys.</p>	<p>USA, 8 boys aged 9.75–11.9 at entry, recruitment method uncertain</p>	<p>A longitudinal assessment of leptin and testosterone levels measured every four months. Comparison of leptin was with the onset of puberty as determined by rise in testosterone levels.</p>	<p>Pubertal status assessed by a physical examination using the sexual maturation scale, rise in testosterone was taken as the onset of puberty</p>	<p>Small numbers. Examined longitudinally until Tanner stage 5 reached. Frequent sampling. Testosterone and dehydroepiandrosterone levels also obtained.</p>	<p>Leptin levels rose about 100% before or at the time of the onset of puberty, and decreased to baseline after the initiation of puberty. <u>Conclusion:</u> This method would be difficult to use as a one-off measure because levels before and after puberty are similar.</p>
<p>Carlsson, 1997 (41)</p>	<p>Serum leptin concentrations in relation to pubertal development</p>	<p>Sweden, 252 healthy children aged 1.7 to 18.6 years, source of participants unclear</p>	<p>Cross-sectional study, with a longitudinal element for 15 girls with known dates of menarche. Blood samples obtain between 10am and 2pm from which serum leptin concentrations were obtained.</p>	<p>Tanner staging through the Sexual Maturation Scale.</p>	<p>252 children – single sample of leptin. Subgroup of 15 girls with known dates for menarche was followed up longitudinally with two to seven repeated observations. Different findings to other studies. Many samples for boys were taken in pre-pubertal stages, with low numbers for the pubertal stage (124 vs 38).</p>	<p>30-fold variation in leptin concentration between participants. <u>Conclusion:</u> In girls leptin increased through puberty whereas there was no change in leptin with pubertal development in boys. The paper hypothesises this may be due to amount of adipose tissue.</p>

Table 5: Description of studies using voice methods to assess pubertal status

First author and year	Article title	Study setting and sample	Study aims and methods of assessment	Validation and/or correlation	Quality assessment of the study	Key findings and conclusions
Ong, 2012 (45)	Timing of voice breaking in males associated with growth and weight gain across the life course	London, UK, using data from the 1946 birth cohort, 2008 boys aged 14 years	In 1961, the boys underwent a pubertal assessment at school, which included secondary sexual characteristics and subjective assessment of voice. The child's school doctor rated the development of genitals, pubic hair and voice on a three-option rating scale, and appearance of axillary hair on a dichotomous scale.	Voice change timing, used as a marker of pubertal development and cross-referenced to later outcomes in adulthood	Large sample size, but subjects of white European descent only (which was the norm for those born in 1946 in the UK). Potential observer bias as children's school doctors carried out pubertal staging – only one observer per assessment. Crude pubertal development rating scale consisting of only two or three options. Subjective methods of voice assessment, despite availability of other methods.	Similar to females with earlier menarche, the trajectory to earlier sexual maturation in males was preceded by faster early postnatal growth and weight gain, and lead to higher adult BMI. Timing of pubertal maturation was seen to have potential relevance to adult disease risks in males. <u>Conclusion:</u> This study demonstrated how voice could be used as a marker of pubertal development in males.
Harries, 1997 (44)	Changes in the male voice at puberty	Cambridge, UK, 26 school boys aged 13-14 years.	To investigate the characteristics of the male speaking and signing voice in relation to other biological changes in puberty. Measurement of standing height, weight, pubertal stage by Tanner, and testicular volume;	Correlation between the speaking and signing voice and testosterone levels, testicular volume and Tanner stages of puberty.	Potential for selection bias – no details of recruitment are given, and observer bias – not clear who carried out the measurements. Small numbers. Perhaps a wider range of ages and a longer time-period would have been beneficial, although the group appear to have collated	Changes in voice fundamental frequencies correlated with testis volume, but not with testosterone levels. There was a clear correlation between Tanner stages and a Cooksey musical classification during male puberty. Voice-breaking is a late event in male puberty. <u>Conclusion:</u> This study shows a good correlation between Tanner stages and Cooksey stages of

			salivary testosterone profiling; acoustic and musical recordings. Three-monthly follow-up over 12 months.		information throughout puberty.	pubertal voice development. It provides an alternative method to assessing puberty in cohort and longitudinal studies.
--	--	--	---	--	---------------------------------	--

Appendix 1: Sexual Maturation Scale, based on Tanner and Marshall (1969)

Boys - development of external genitalia

Stage 1: Prepubertal

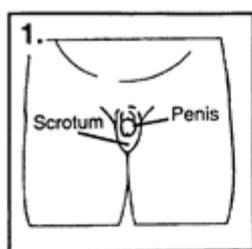
Stage 2: Enlargement of scrotum and testes; scrotum skin reddens and changes in texture

Stage 3: Enlargement of penis (length at first); further growth of testes

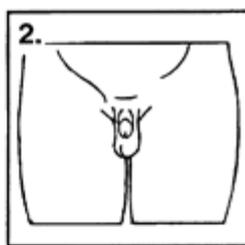
Stage 4: Increased size of penis with growth in breadth and development of glans; testes and scrotum larger, scrotum skin darker

Stage 5: Adult genitalia

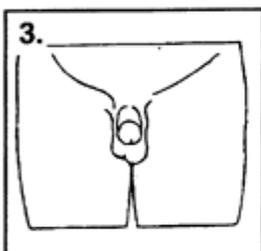
Corresponding line drawings for self-assessment, based on Taylor (2001)



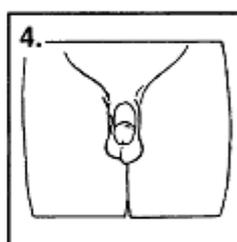
Scrotum and Penis same size as when you were younger



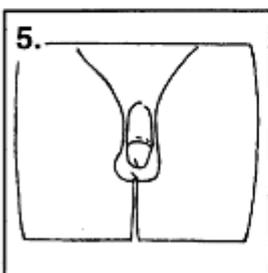
The Scrotum has lowered a bit and the Penis is a little larger



The Penis is longer and the Scrotum is larger



The Penis is longer and wider, the Scrotum is darker and bigger than before



The Penis and Scrotum are the size and shape of an adult

Girls - breast development

Stage 1: Prepubertal

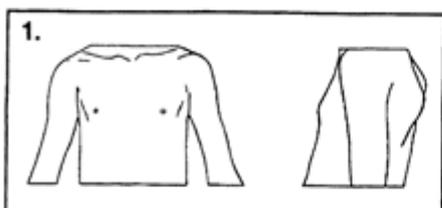
Stage 2: Breast bud stage with elevation of breast and papilla; enlargement of areola

Stage 3: Further enlargement of breast and areola; no separation of their contour

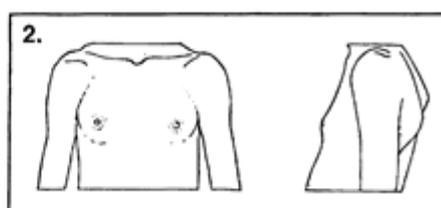
Stage 4: Areola and papilla form a secondary mound above level of breast

Stage 5: Mature stage: projection of papilla only, related to recession of areola

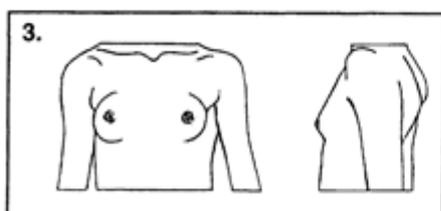
Corresponding line drawings for self-assessment, based on Taylor (2001)



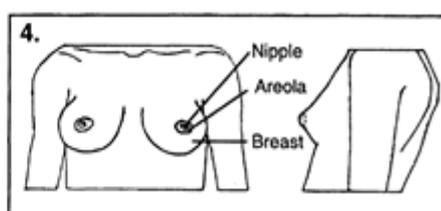
The breasts are flat.



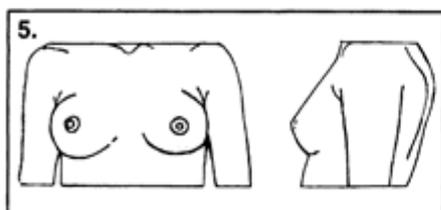
The breasts form small mounds.



The breasts form larger mounds than in 2.



The nipple and the surrounding part (the Areola) make up a mound that sticks up above the breast.



Only the nipple sticks out beyond the breast.

Boys and girls - pubic hair

Stage 1: Prepubertal (can see vellus hair similar to abdominal wall)

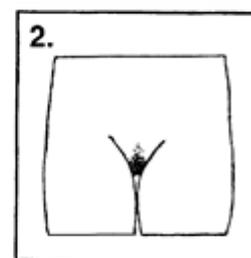
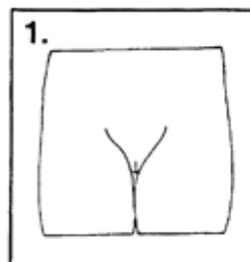
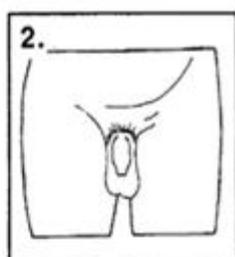
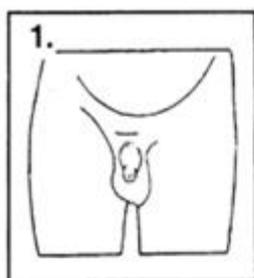
Stage 2: Sparse growth of long, slightly pigmented hair, straight or curled, at base of penis or along labia

Stage 3: Darker, coarser and more curled hair, spreading sparsely over junction of pubes

Stage 4: Hair adult in type, but covering smaller area than in adult; no spread to medial surface of thighs

Stage 5: Adult in type and quantity, with horizontal distribution

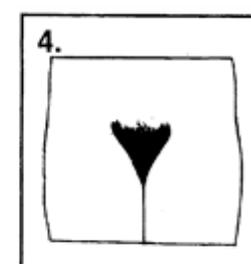
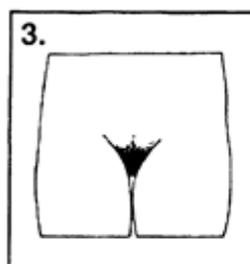
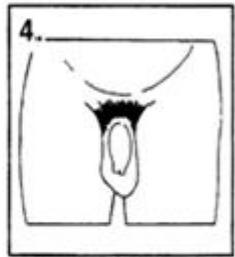
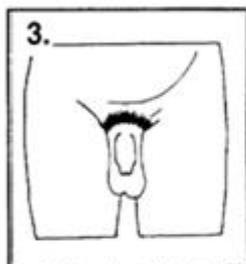
Corresponding line drawings for self-assessment, based on Taylor (2001)



Very little hair

No hairs

Very little hair

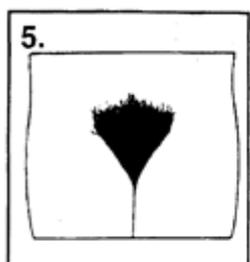


Quite a lot of hair

The hair has not spread over the thighs

Quite a lot of hair

The hair has not spread over the thighs



The hair has spread over the thighs

The hair has spread over the thighs

Appendix 2: Report of workshop

Workshop to discuss methods for determining pubertal status in cohort and longitudinal studies

The workshop was organised by Janis Baird, Clare Smith and Hazel Inskip. It was held on 28th September 2016 at the MRC Lifecourse Epidemiology Unit, University of Southampton and funded by grants from CLOSER and the Society for Social Medicine.

The workshop began with four oral presentations. Janis Baird, University of Southampton, gave an overview of preliminary findings from the CLOSER-funded evidence review of methods for pubertal assessment. Tim Cole, University College London, outlined how longitudinal data to assess growth can be used to indicate the timing and intensity of puberty. David Dunger, University of Cambridge, talked about the potential use of hormones and other biomarkers to indicate pubertal stage. Justin Davies, University Hospital Southampton, gave a clinical perspective on pubertal assessment.

Synopsis of presentations

Janis Baird - Preliminary findings of the review of pubertal assessment methods

The challenges for research are assessing pubertal stage at a given time and assessing rate of progress through puberty in longitudinal studies. The review funded by CLOSER focuses on identifying methods for assessing pubertal stage used in cohort, and other studies, and is looking for evidence of validation of these approaches.

To date 55 relevant studies have been identified from searches of bibliographic databases and contact with experts. Screening of article reference lists is ongoing and so more studies are likely to be identified. The largest group of studies identified (n=14) used self-assessment to assess pubertal stage. Exact agreement between self-assessment, in which adolescents are asked to rate themselves against photographs or drawings to answer specific questions about their development, and clinical Tanner stage ranged between 43% and 81% but was lower than 60% in the majority of studies. There was, however, good agreement with clinical assessment to within one Tanner stage (85 to 100%). Agreement was better at younger and older ages and for girls. Pubic hair self-assessment was generally more accurate for both sexes than breast and male genitalia. Studies that combined adolescent and parental reports combined provided some evidence to support triangulation, although parental assessments of boys' pubertal stage were less accurate than for girls. Self-reported age at menarche has been used retrospectively as an indicator of late puberty. Age at which the voice breaks is might provide a measure of late puberty in boys although the rate with which boys' voices change is variable.

Review of studies of growth is ongoing. Evidence reviewed to date indicates that peak height velocity is the most useful measure of rate of progress through puberty.

Serial measurements of height are required to determine peak height velocity. The SITAR method offers an approach that allows for individual variation in the timing and intensity of puberty. Other methods under review include assays of sex hormones, leptin and inhibin where there has been little validation against clinical staging, but one comparison of testosterone and testicular volume suggested that testosterone levels in later puberty correlate reasonably well with testicular volume. The evidence indicates that there is considerable individual variability in leptin and inhibin levels suggesting that serial measures might be required. There is some evidence that voice breaking correlates with testicular volume. Reported age at voice breaking has been determined retrospectively. More recently an app called Speechtest has been developed for phones and tablets.

Tim Cole - Assessing growth to determine the timing of puberty

Puberty is defined as the stage of adolescence in which an individual becomes physiologically capable of sexual reproduction. This is distinct from adolescence which is defined as the period of physical and psychological development from the onset of puberty to adulthood. The timing of puberty varies enormously from one individual to the next, as does rate of passage through puberty, known as intensity. An earlier age at onset of puberty is typically associated with faster progression through puberty. One of the most striking features of puberty is the growth spurt in height, the intensity of which is measured by peak height velocity and its timing by age at peak height velocity.

Professor Cole explained the concept of developmental age which differs from chronological age in that it takes account of pubertal stage. Conventional growth curves for a set of individuals differ in mean height, timing and rate of puberty. The SITAR method of growth curve analysis, which adjusts for these factors, provides estimates of the timing and intensity of pubertal height growth in individuals. Height measurements every two years might be sufficient to rank people by age at puberty, but work is needed to assess whether this is frequent enough or not. As well as being used for height, the SITAR method can also be applied to breast development in girls, to genital development in boys and to pubic hair in both sexes.

Growth markers at earlier stages of puberty include hand and foot size since they mature relatively early. A study by Busscher et al 2011 showed mean age at shoe size peak velocity between 10 and 11 years in girls and between 11 and 12 years in boys. This occurs earlier than peak height velocity. Regular measurements of foot size, or questions about shoe size, could be included within cohort studies with relative ease.

David Dunger - Metabolic approaches to pubertal assessment

In general, androgens stimulate the growth of pubic hair in both sexes, oestrogens are responsible for breast development and follicular stimulating hormone (FSH) and luteinizing hormone (LH) stimulate ovarian development. Puberty is associated with a doubling of lean body mass. Peak bone mass is achieved when people are in their early 20s. Even after pubertal stage 5, hormone levels remain high until age 25-30.

Onset of puberty has a genetic component: familial factors are responsible for 60-80% of the variation in timing and intensity. Genes: LIN28B is involved in timing of reproduction; other genes determine hormone levels and SNPs associated with weight gain are also important. Markers of pubertal onset might emerge from ongoing epigenetic studies.

Professor Dunger outlined hormonal changes during puberty and suggested how these might be potential biomarkers for pubertal stage if good assays were available. At puberty onset, there is an increase in the overnight pulsatile release of LH. Early morning urinary LH, adjusted for creatinine, might be good at discriminating Tanner stage (2016 data) in girls. Serum testosterone is aromatised to oestrogen in fat, and it is oestrogen that triggers growth hormone (GH) secretion in both sexes. Pulsatile GH secretion leads to increase in IGF-1 and insulin, which persist until age 25 when they begin to fall. Low dose oestrogen primes GH secretion, but high doses closes epiphyses. Oestrogen levels in girls are less discriminatory in the early stages of puberty, even with good assays. Inhibin B is released from the ovary in pubertal girls, rising in early puberty. Inhibin A is slower to rise). Testosterone is low in girls but rises with puberty and always remains lower than levels observed in boys.

For girls, assays of adrenal androgens give a broad indication of Tanner stage, and Inhibin-B is a useful marker of early puberty. For boys, serum testosterone levels correlate with testicular size. The potential to assay hormone levels in blood from early morning urine and saliva or from hair specimens means that such approaches to pubertal assessment could be used in research. However, suitable methods are not yet available.

Justin Davies - Clinical approaches to pubertal assessment

Dr Davies outlined some of the key hormonal changes during puberty. LH is released in pulses at night but this only seems to occur during phases of deep sleep (phases 3 and 4). The nature of gonadotrophin-releasing hormone (GnRH) pulses determines whether LH or FSH is released during sleep.

During the last trimester of pregnancy, GnRH release by the fetus is suppressed by placental oestrogen. Then in early infancy, children go through a mini-puberty at around 6 months of age in boys, and between 12 and 18 months in girls.

Premature thelarche (breast development) is of no concern unless before 8 years of age.

Recent research has shown that MKRN3 is an imprinted gene that controls the brake on puberty – deletions in the gene leading to precocious puberty. For girls, breast development is driven by oestrogens and pubic hair by androgens. For boys, pubic hair development is driven by androgens and testicular growth is driven by gonadotrophins.

There is some interest in ultrasound of the uterus in girls since measurement of the endometrial strip provides information about oestrogen levels. Ultrasound scanning of the uterus gives information on prior oestradiol exposure. It is not, however, used to stage puberty although some radiologists will comment whether the uterine

appearance is pre-pubertal, peri-pubertal or post-pubertal based on the dimensions and shape of the uterus and the size of the endometrial strip.

Dr Davies outlined the steps in clinical assessment of pubertal stage and showed a series of photographs demonstrating the features of the five Tanner stages in each sex, contrasting these with line drawings now widely used in self-assessment, as photographs are deemed to be too explicit.

Ideas arising from the workshop discussions

Discussions focused on methods for pubertal assessment that could be used in research. Participants split into three groups. The points listed below represent the main ideas and recommendations that arose from the group discussions:

Longitudinal measurement of height – it was felt that further research is needed to determine the frequency of measurements required to derive peak height velocity
Use of a voice app in boys should be validated against clinical measures. Research should determine how frequent measurements would need to be in order to monitor fluctuations in voice and identify passage through puberty.

Involving psychologists in explaining the process and framing questions for self-assessment could improve participation. Recent evidence from a study of colorectal screening achieved high response rate as a result of involving a research psychologist.

Biological samples – there is interest in early morning urine or saliva to assess hormone levels in the blood (LH and creatinine in girls and Inhibin and Testosterone in boys) although there was acknowledgement that serial measures would be required and further research would be needed to determine progress through puberty. One off blood samples would also be useful as an adjunct to these samples to validate assays based on saliva, urine and hair.

There was interest in the idea of a Global Question where adolescents are asked to compare themselves with their peers, although there was recognition that this might lead to overestimation of pubertal stage.

In Sweden, boys carry out ultrasound of their own testicular volume following advice from radiologists, thus maintaining their privacy, although there was recognition that this might be difficult to implement within many cohort studies.

Clinical assessment could be made more acceptable by looking at pubic hair for both sexes and genitalia for boys while carrying out other measurements such as skinfold thickness. Better explanation that the assessment could be done quickly and did not require complete removal of clothes was also seen as important in increasing uptake.

Researchers in the Millennium Cohort Study and in India had noted ethnic differences in puberty that make pubertal assessment more challenging; for example Asian boys tend to begin puberty early than other ethnic groups.

General conclusion

No clear consensus emerged about the best method for use in cohort and other research studies. The gold standard of clinical assessment is difficult to implement. Various methods show promise but need further development. Specifically, appropriate biomarkers may well be identified over the coming years, and

measurements of height and/or foot length, done with appropriate frequency, could provide less invasive approaches to determining pubertal status and age at puberty. Novel methods such as the Speechtest app might also have utility and require further evaluation.

Review of methods for determining pubertal status in cohort and longitudinal studies Programme

10:30 – 11.00	Registration, tea and coffee
11.00 – 11.10	Welcome and introductions – Hazel Inskip, Professor of Statistical Epidemiology
11.10 – 11.30	Janis Baird, Associate Professor of Public Health Preliminary findings of the review of pubertal assessment methods
11.30 – 12.00	Tim Cole, Professor of Medical Statistics, University College London Assessing growth to determine the timing of puberty
12.00 – 12.30	David Dunger, Professor of Paediatrics, University of Cambridge Metabolic approaches to pubertal assessment
12.30 – 13.15	Lunch (provided)
13.15 – 13.45	Justin Davies, Consultant Paediatric Endocrinologist, University Hospital Southampton Clinical approaches to pubertal assessment
13.45 – 14.45	Group activity: What cross sectional approach is most suitable to determine current pubertal stage? What pubertal features should we assess longitudinally (e.g. peak height velocity) and how should we measure them?
14.45 – 15.00	Tea and coffee
15.00 – 15.15	Feedback from groups
15.15 – 16.00	Defining some of the detail – breakout groups to allow people to discuss areas in more depth
16.00 – 16.30	Recommendations arising from breakout group discussions
16.30	Close

Participants

Pubertal assessment workshop 28th September 2016

Janis Baird, University of Southampton
Tim Cole, University College London
Sue Collins, University of Southampton
Sian Crossweller, University of Bristol
Beth Curtis, University of Southampton
Justin Davies, University Hospital Southampton
David Dunger, University of Cambridge
Caroline Fall, University of Southampton
Nicola Foster, University College London
Julia Hammond, University of Southampton
Nicholas Harvey, University of Southampton
Sue Higginbottom, University of Southampton
Hazel Inskip, University of Southampton
Yi Lu, University College London
Sue Macey, University of Southampton
Dian Rogers, University of Southampton
Keith Godfrey, University of Southampton
Kate Smith, Institute of Education
Kaitlin Wade, University of Bristol
Peter Whincup, St George's, University of London
Afshin Zilanawala, University College London

Janis Baird
21/10/16

Appendix 3: Pubertal assessment – the views of the University Hospital Southampton Children’s Panel

10 young people aged 10 to 16 attended the panel meeting. They chose to talk to researchers in two groups – boys and girls. A general issue in both groups was the need to protect children and young people’s anonymity when doing any kind of assessment.

Boys

There were six boys, aged 12 to 16.

Self-assessment questionnaire

The boys thought the questionnaire was acceptable and thought other boys would find it easy to complete. They thought the instruction about ticking the box should say ‘tick on top of the picture inside the box’. They thought boys would prefer this paper copy of a questionnaire to any kind of sliding scale on a computer or app because the paper copy is quick and easy and doesn’t require them to set anything up. However, they would be willing to use the sliding scale approach if set up on a computer in a clinic.

Question about age at first wet dream

They thought that most boys would be embarrassed by this question and that most would not remember the timing of their first wet dream.

Clinical assessment

They thought that clearer information might increase the numbers of boys who would agree to this. If a leaflet explained that the examination was very quick and that they didn’t have to undress – merely allow the examiner to look down their trousers, then this would be more acceptable. They also thought boys would be more likely to agree to the assessment if the assessors were male.

Orchidometer

They thought that boys would find it difficult to judge their own testicular size against the orchidometer without undressing. They would want somewhere private to do this if they were attending a clinic.

Shoe size and foot size

They thought boys would be happy to report their shoe size but they thought it best to ask about school shoes as the size of sports and casual shoes varies. They thought boys would be happy to have their feet measured regularly using ipads (as is now the norm in shoe shops) or other techniques and would also be willing to mark the size of their feet at home – perhaps using a sheet of paper and marking the end

of their big toe and their heel. One of the boys wondered whether measuring hand size would be an alternative to measuring feet.

Speech

The boys tried Speectest, an app that reports pubertal stage based on a boy's voice when he counts backwards from 20 to 1. They thought boys would be happy to do this and that it was best to use the counting backward approach as reading out text would be difficult if children had literacy problems.

Girls

There were 4 girls aged 10 to 13 years

Self-assessment questionnaire

They thought girls would be happy completing this although they thought that some girls might need help from their mothers.

Age at menarche

They thought girls would generally be happy answering questions about age at menarche and that they would be happy for their mums to be there. They thought mums might be more likely to answer correctly. They would be happy to discuss such an issue with life-long friends and family and people like doctors who were complete strangers. They would not want to talk to people they knew in other contexts.

Terminology

They wanted the language used to be direct and that the correct terminology to be used.

Clinical assessment

They thought this should not be done opportunistically as part of anthropometry. They thought having an information sheet to explain its importance on the day of a clinic would be helpful. The leaflet should say that the assessment is quick, that they don't need to undress and that, usually, they don't need to be touched. They thought it best for men to assess boys and women to assess girls.

Foot and shoe size

They thought girls would be happy to report this annually or to measure their feet annually and that school shoe size should be used because other types of shoe vary in size.

Speech

If the app was used for boys at a school, then the girls thought it would be very important that anonymity was protected.

Janis Baird, Tina Horsfall, Julia Hammond
13/10/16

References

1. Lancet. Adolescent Health 2012. Available from: <http://www.thelancet.com/series/adolescent-health-2012>.
2. Viner RM. Chapter 8: Life stage: adolescence. Chief Medical Officer's annual report 2012: Our Children Deserve Better: Prevention Pays. London: Department of Health; 2013.
3. Viner RM, Ross D, Hardy R, Kuh D, Power C, Johnson A, et al. Life course epidemiology: recognising the importance of adolescence. *J Epidemiol Community Health*. 2015.
4. Rogol AD, Roemmich JN, Clark PA. Growth at puberty. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*. 2002;31(6 Suppl):192-200.
5. Hardy R, Kuh D, Whincup PH, Wadsworth ME. Age at puberty and adult blood pressure and body size in a British birth cohort study. *Journal of hypertension*. 2006;24(1):59-66.
6. Power C, Lake JK, Cole TJ. Body mass index and height from childhood to adulthood in the 1958 British born cohort. *Am J Clin Nutr*. 1997;66(5):1094-101.
7. Monteilh C, Kieszak S, Flanders WD, Maisonet M, Rubin C, Holmes AK, et al. Timing of maturation and predictors of Tanner stage transitions in boys enrolled in a contemporary British cohort. *Paediatric and perinatal epidemiology*. 2011;25(1):75-87.
8. Rubin C, Maisonet M, Kieszak S, Monteilh C, Holmes A, Flanders D, et al. Timing of maturation and predictors of menarche in girls enrolled in a contemporary British cohort. *Paediatric and perinatal epidemiology*. 2009;23(5):492-504.
9. Gnatiuc L, Kato B, Matheson MC, Newson RB, Jarvis DL. The association of asthma with BMI and menarche in the 1958 British Birth Cohort. *The Journal of asthma : official journal of the Association for the Care of Asthma*. 2013;50(7):751-8.
10. Canoy D, Beral V, Balkwill A, Wright FL, Kroll ME, Reeves GK, et al. Age at menarche and risks of coronary heart and other vascular diseases in a large UK cohort. *Circulation*. 2015;131(3):237-44.
11. Petersen AC, Crockett L, Richards M, Boxer A. A self-report measure of pubertal status: Reliability, validity, and initial norms. *J Youth Adolesc*. 1988;17(2):117-33.
12. Schmitz KE, Hovell MF, Nichols JF, Irvin VL, Keating K, Simon GM, et al. A Validation Study of Early Adolescents' Pubertal Self-Assessments. *The Journal of Early Adolescence*. 2004;24(4):357-84.
13. 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. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*. 1999;24(3):201-5.
14. Morris NM, Udry JR. Validation of a self-administered instrument to assess stage of adolescent development. *J Youth Adolesc*. 1980;9(3):271-80.
15. Brooks-Gunn J, Warren MP, Rosso J, Gargiulo J. Validity of self-report measures of girls' pubertal status. *Child Dev*. 1987;58(3):829-41.
16. Bond L, Clements J, Bertalli N, Evans-Whipp T, McMorris BJ, Patton GC, et al. A comparison of self-reported puberty using the Pubertal Development Scale and

- the Sexual Maturation Scale in a school-based epidemiologic survey. *Journal of adolescence*. 2006;29(5):709-20.
17. Berg-Kelly K, Erdes L. Self-assessment of sexual maturity by mid-adolescents based on a global question. *Acta paediatrica (Oslo, Norway : 1992)*. 1997;86(1):10-7.
 18. Lum S, Bountziouka V, Harding S, Wade A, Lee S, Stocks J. Assessing pubertal status in multi-ethnic primary schoolchildren. *Acta Paediatr*. 2015;104(1):e45-8.
 19. Bundak R, Darendeliler F, Gunoz H, Bas F, Saka N, Neyzi O. Analysis of puberty and pubertal growth in healthy boys. *European journal of pediatrics*. 2007;166(6):595-600.
 20. Karlberg J, Kwan C-W, Gelander L, Albertsson-Wikland K. Pubertal growth assessment. *Hormone research*. 2003;60(Suppl 1):27-35.
 21. Cole TJ, Donaldson MDC, Ben-Shlomo Y. SITAR--a useful instrument for growth curve analysis. *International journal of epidemiology*. 2010;39(6):1558-66.
 22. Busscher I, Kingma I, Wapstra FH, Bulstra SK, Verkerke GJ, Veldhuizen AG. The value of shoe size for prediction of the timing of the pubertal growth spurt. *Scoliosis*. 2011;6:1-.
 23. Ford KR, Houry JC, Biro FM. Early markers of pubertal onset: height and foot size. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*. 2009;44(5):500-1.
 24. Mitra S, Samanta M, Sarkar M, Chatterjee S. Foot length as a marker of pubertal onset. *Indian pediatrics*. 2011;48(7):549-51.
 25. Ozer T, Kama JD, Ozer SY. A practical method for determining pubertal growth spurt. *American journal of orthodontics and dentofacial orthopedics : official publication of the American Association of Orthodontists, its constituent societies, and the American Board of Orthodontics*. 2006;130(2):131.e1-6.
 26. Canavese F, Charles YP, Dimeglio A, Schuller S, Rousset M, Samba A, et al. A comparison of the simplified olecranon and digital methods of assessment of skeletal maturity during the pubertal growth spurt. *The bone & joint journal*. 2014;96-B(11):1556-60.
 27. Kuh D, Muthuri SG, Moore A, Cole TJ, Adams JE, Cooper C, et al. Pubertal timing and bone phenotype in early old age: findings from a British birth cohort study. *Int J Epidemiol*. 2016;45(4):1113-24.
 28. Grumbach MM, Roth JC, Kaplan SL, Kelch RP. Hypothalamic-pituitary regulation of puberty in man: Evidence and concepts derived from clinical research. *Control of the onset of puberty*. New York: John Wiley & Sons; 1974. p. 115-66.
 29. Apter D, Bützow T, Laughlin GA, Yen SS. Accelerated 24-hour luteinizing hormone pulsatile activity in adolescent girls with ovarian hyperandrogenism: relevance to the developmental phase of polycystic ovarian syndrome. *J Clin Endocrinol Metab*. 1994;79(1):119-25.
 30. Søbørg T, Frederiksen H, Mouritsen A, Johannsen TH, Main KM, Jørgensen N, et al. Sex, age, pubertal development and use of oral contraceptives in relation to serum concentrations of DHEA, DHEAS, 17 α -hydroxyprogesterone, Δ 4-androstenedione, testosterone and their ratios in children, adolescents and young adults. *Clin Chim Acta*. 2014;437:6-13.
 31. Ahmed M. Longitudinal analysis of the pubertal growth, body composition, and endocrine development in young people with and without diabetes 2009.
 32. Sehested A, Andersson AM, Müller J, Skakkebaek NE. Serum inhibin A and inhibin B in central precocious puberty before and during treatment with GnRH agonists. *Horm Res*. 2000;54(2):84-91.

33. Singh GKS, Balzer BWR, Kelly PJ, Paxton K, Hawke CI, Handelsman DJ, et al. Urinary Sex Steroids and Anthropometric Markers of Puberty - A Novel Approach to Characterising Within-Person Changes of Puberty Hormones. *PLOS ONE*. 2015;10(11):e0143555.
34. Ankarberg-Lindgren C, Norjavaara E. Changes of diurnal rhythm and levels of total and free testosterone secretion from pre to late puberty in boys: testis size of 3 ml is a transition stage to puberty. *European journal of endocrinology / European Federation of Endocrine Societies*. 2004;151(6):747-57.
35. Chada M, Prusa R, Bronsky J, Kotaska K, Sidlova K, Pechova M, et al. Inhibin B, follicle stimulating hormone, luteinizing hormone and testosterone during childhood and puberty in males: changes in serum concentrations in relation to age and stage of puberty. *Physiological research / Academia Scientiarum Bohemoslovaca*. 2003;52(1):45-51.
36. Crofton PM, Evans AEM, Groome NP, Taylor MRH, Holland CV, Kelnar CJH. Dimeric inhibins in girls from birth to adulthood: relationship with age, pubertal stage, FSH and oestradiol. *Clinical endocrinology*. 2002;56(2):223-30.
37. Rockett JC, Lynch CD, Buck GM. Biomarkers for assessing reproductive development and health: Part 1--Pubertal development. *Environmental health perspectives*. 2004;112(1):105-12.
38. Brann DW, Wade MF, Dhandapani KM, Mahesh VB, Buchanan CD. Leptin and reproduction. *Steroids*. 2002;67(2):95-104.
39. Mantzoros CS, Flier JS, Rogol AD. A longitudinal assessment of hormonal and physical alterations during normal puberty in boys. V. Rising leptin levels may signal the onset of puberty. *The Journal of clinical endocrinology and metabolism*. 1997;82(4):1066-70.
40. Wang T, Morioka I, Gowa Y, Igarashi Y, Miyai N, Yamamoto H, et al. Serum leptin levels in healthy adolescents: Effects of gender and growth. *Environ Health Prev Med*. 2004;9(2):41-6.
41. Carlsson B, Ankarberg C, Rosberg S, Norjavaara E, Albertsson-Wikland K, Carlsson LM. Serum leptin concentrations in relation to pubertal development. *Archives of disease in childhood*. 1997;77(5):396-400.
42. Maqsood AR, Trueman JA, Whatmore AJ, Westwood M, Price DA, Hall CM, et al. The relationship between nocturnal urinary leptin and gonadotrophins as children progress towards puberty. *Hormone research*. 2007;68(5):225-30.
43. Zaman N, Hall CM, Gill MS, Jones J, Tillmann V, Westwood M, et al. Leptin measurement in urine in children and its relationship to other growth peptides in serum and urine. *Clinical endocrinology*. 2003;58(1):78-85.
44. Harries ML, Walker JM, Williams DM, Hawkins S, Hughes IA. Changes in the male voice at puberty. *Archives of disease in childhood*. 1997;77(5):445-7.
45. Ong KK, Bann D, Wills AK, Ward K, Adams JE, Hardy R, et al. Timing of voice breaking in males associated with growth and weight gain across the life course. *The Journal of clinical endocrinology and metabolism*. 2012;97(8):2844-52.
46. Hodges-Simeon CR, Gurven M, Cardenas RA, Gaulin SJC. Voice change as a new measure of male pubertal timing: a study among Bolivian adolescents. *Annals of human biology*. 2013;40(3):209-19.
47. Cooksey JM. The male adolescent changing voice: Some new perspectives. In: Runfola M, editor. *Research symposium on the male adolescent voice*: Buffalo: State University of New York Press; 1984. p. 4-59.

48. Rabbani A, Noorian S, Fallah JS, Setoudeh A, Sayarifard F, Abbasi F. Reliability of pubertal self assessment method: an Iranian study. *Iran J Pediatr.* 2013;23(3):327-32.
49. Norris SA, Richter LM. Are there short cuts to pubertal assessments? Self-reported and assessed group differences in pubertal development in African adolescents. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine.* 2008;42(3):259-65.
50. Bond L, Clements J, Bertalli N, Evans-Whipp T, McMorris BJ, Patton GC, et al. A comparison of self-reported puberty using the Pubertal Development Scale and the Sexual Maturation Scale in a school-based epidemiologic survey. *J Adolesc.* 2006;29(5):709-20.
51. Desmangles J-C, Lappe JM, Lipaczewski G, Haynatzki G. Accuracy of pubertal Tanner staging self-reporting. *Journal of pediatric endocrinology & metabolism : JPEM.* 2006;19(3):213-21.
52. Bonat S, Pathomvanich A, Keil MF, Field AE, Yanovski JA. Self-assessment of pubertal stage in overweight children. *Pediatrics.* 2002;110(4):743-7.
53. Taylor SJ, Whincup PH, Hindmarsh PC, Lampe F, Odoki K, Cook DG. Performance of a new pubertal self-assessment questionnaire: a preliminary study. *Paediatric and perinatal epidemiology.* 2001;15(1):88-94.
54. Koo MM, Rohan TE. Accuracy of short-term recall of age at menarche. *Annals of human biology.* 1997;24(1):61-4.
55. Carskadon MA, Acebo C. A self-administered rating scale for pubertal development. *J Adolesc Health.* 1993;14(3):190-5.
56. Duke PM, Litt IF, Gross RT. Adolescents' self-assessment of sexual maturation. *Pediatrics.* 1980;66(6):918-20.