Pubertal development in Norwegian boys

Ultrasound assessment of testicular volume, hormonal references, and association with anthropometric measurements

Ninnie Bakken Oehme

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2021

UNIVERSITY OF BERGEN

Pubertal development in Norwegian boys

Ultrasound assessment of testicular volume, hormonal references, and association with anthropometric measurements

Ninnie Bakken Oehme

Thesis for the degree of Philosophiae Doctor (PhD)

at the University of Bergen

Date of defense: 19.03.2021

Year: 2021

Title: Pubertal development in Norwegian boys

Name: Ninnie Bakken Oehme

Print: Skipnes Kommunikasjon / University of Bergen

Scientific environment

The research work presented in this thesis was carried out from 2017 to 2020 at the Department of Clinical Science, Faculty of Medicine, University of Bergen, as part of affiliation with the WestPaed Research Group from the Department of Paediatrics and Youth Medicine, Haukeland University Hospital.

The study was funded by the University of Bergen and the Western Norway Regional Health Authority (Helse Vest) provided financial support for data collection as part of the study.

Professor Pétur B. Júlíusson is the principal investigator of Bergen Growth Study 2 and affiliated with the Department of Clinical Science, University of Bergen; Department of Paediatrics and Youth Medicine, Haukeland University Hospital; and Department of Health Registry Research and Development, Norwegian Institute of Public Health.

This PhD study was co-supervised by: PhD Mathieu Roelants, affiliated with Environment and Health, Department of Public Health and Primary Care, KU Leuven-University of Leuven, Belgium; Professor Robert Bjerknes, affiliated with the Department of Clinical Science, University of Bergen, and the Department of Paediatrics and Youth Medicine, Haukeland University Hospital, Bergen; and Professor Karen Rosendahl, affiliated with the Department of Radiology, University Hospital of North Norway, and the Department of Clinical Medicine, University of Tromsø.

Acknowledgements

I took an early interest in the field of endocrinology, and I was therefore very excited when I was invited in 2015 to join the research group behind Bergen Growth Study 2. During my first year as a member of the group, while still on maternity leave, I travelled around, actively recruiting schools to participate in the study, while carrying my youngest child with me.

Thanks to PhD funding from the University in Bergen, I was able to conduct research on a full-time basis starting in 2017. In addition, my research studies would not have been possible without the financial support for data collection from the Western Norway Regional Health Authority (Helse Vest).

There are so many who deserve appreciation and gratitude for their contribution to this study—first and foremost, all the children who volunteered to take part in the study and to their parents for giving their consent to their children's participation.

I was lucky to have had four great supervisors during this period. I am eternally grateful to my main supervisor Professor Pétur B. Júlíusson who, with his great patience and commitment, immense knowledge, and endless enthusiasm, has guided and supported me through these years. You have been an inspiration and a fantastic role model on how to be a great researcher, paediatrician, and endocrinologist. My co-supervisor Mathieu Roelants also deserves a big thank you. Despite living in Belgium, you have been very much present and contributed with your vast knowledge and statistical input, all of which has raised the quality of this work. Further, your accuracy and attention to detail have been valuable and your language editing skills are impressive. I also give my thanks to my co-supervisors Professor Karen Rosendahl and Professor Robert Bjerknes for their contribution to this project and thesis. Karen, thank you for sharing your knowledge on ultrasound and your experience in conducting studies. You always teach with calmness and patience, which I am appreciative of. I would like to thank Robert for always giving constructive feedback with the 'big picture' in mind and always giving new perspectives. I also would like to thank Professor Geir Egil Eide for his contributions to the statistical work. I am also grateful and greatly indebted to Magnus R. Sveen,

Abstract

Background

Puberty marks the transition from childhood to adulthood and is characterized by physiological and psychological changes leading to sexual maturity and reproductive function. Assessment of pubertal development in boys is challenging, due to the intimate and subjective nature of the examinations. Over the last decades, several studies have suggested a trend towards earlier puberty in boys, but data from Norway have been lacking. Up-to-date descriptive data allow to investigate secular trends and are required to define early or late puberty, both of which may impact on later health outcomes. The underlying mechanisms that influence the timing and progression of puberty are, particularly in boys, not fully elucidated. Overweight and obesity, as well as exposure to endocrine-disrupting chemicals, have been proposed as possible drivers for the trend towards earlier puberty.

Aims

The main aims of this study were to explore ultrasound as a reliable method for assessment of testicular volume and to establish references for the timing of pubertal development in Norwegian boys based on ultrasound-measured testicular volumes and the development of pubic hair. In addition, the study also aimed to establish references for serum levels of testosterone and other reproductive hormones in relation to ultrasound-derived testicular volumes and to examine the association between pubertal status and anthropometric measures in boys.

Materials and methods

This study is based on data from Bergen Growth Study 2 collected in 2016–2017. A total of 514 healthy boys aged 6–16 years were examined with ultrasound to measure the testicular volume, as well as clinically to assess for development of pubic hair according to the Tanner scale. In addition, anthropometric measurements, including height, weight, waist circumference, and subscapular skinfolds, as well as body composition, including body fat percentage, were recorded, and blood samples were collected for most of the participants.

Results

Results showed that ultrasound can be used to quantitate testicular volume in boys, without interference from surrounding scrotal tissue. The intra- and interobserver error was acceptable for clinical use. Prader orchidometry, compared to ultrasound, tended to overestimate smaller testicular volumes. Norwegian boys reached pubertal testicular volume at a mean (SD) age of 11.7 (1.1) years, and the onset of pubic hair development occurred, on average, at 11.8 (1.2) years. The study also found that testicular volume accounted for more variance in serum testosterone levels than chronological age, and that male pubertal hormone reference intervals benefited from stratification by testicular volume. Further, low body mass index (BMI) and small waist circumference for age, rather than high BMI and large waist circumference for age, influenced the timing of pubertal development. Boys with low BMI for age entered puberty approximately 8 months later than normal-weight or overweight boys.

Conclusion

This study demonstrates the usefulness and potential advantages of ultrasound as a method for evaluation of testicular volume in boys. Implementation of an ultrasound protocol has the added advantage of enabling more objective measurements on a continuous scale. In this study, the first references for clinical assessment of puberty in Norwegian boys were developed, which showed that Norwegian boys exhibited pubertal timing that is comparable with current Northern European references, and no apparent secular trend towards earlier puberty was observed over the last decades. Stratification of pubertal hormone references based on objective ultrasound assessments of testicular volume was shown to narrow the reference ranges and thus has the potential to increase the diagnostic value of traditional references based on chronological age. Finally, the study showed that low, but not high, BMI for age was associated with pubertal status, indicating that all weight classes should be taken into consideration when assessing sexual maturation in children and adolescents.

Table of Contents

List of publications

Paper I

Oehme NHB, Roelants M, Bruserud IS, Eide G.E, Bjerknes R, Rosendahl K, Juliusson P.B. Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys—intra- and interobserver agreement and comparison with Prader orchidometry. *Pediatr Radiol*. 2018;**48**:1771–8.

Paper II

Oehme NHB, Roelants M, Bruserud IS, Madsen A, Eide G.E, Bjerknes R, Rosendahl K, Juliusson P.B. Reference data for testicular volume measured with ultrasound and pubic hair in Norwegian boys are comparable with Northern European populations. *Acta Paediatr*. 2020;**109**:1612–19.

Paper III

Madsen A, Oehme NB, Roelants M, Bruserud I.S, Eide G.E, Viste K, Bjerknes R, Almås B, Rosendahl K, Sagen J.V, Mellgren G, Juliusson P.B. Testicular ultrasound to stratify hormone references in a cross-sectional Norwegian study of male puberty. *J Clin Endocrinol Metab*. 2020;**105**:dgz094.

Paper IV

Oehme NHB, Roelants M, Bruserud IS, Madsen A, Bjerknes R, Rosendahl K, Juliusson P.B. Low BMI, but not high BMI, influences the timing of puberty in boys. Submitted, manuscript under review.

Related papers

1.   Oehme NHB, Bruserud IS, Madsen A, Juliusson PB. Is Puberty starting earlier than before? *Tidsskr Nor Legeforen*. 2020;**140**doi: 10.4045/tidsskr.20.0043

2.   Madsen A, Bruserud IS, Bertelsen BE, Roelants, M., Oehme, NHB, *et al*. Hormone references for ultrasound breast staging and endocrine profiling to detect female onset of puberty. *J Clin Endocrinol Metab*. 2020;**105**:dgaa679.

3.   Bruserud IS, Roelants M, Oehme NHB, *et al*. References for ultrasound staging of breast maturation, Tanner breast staging, pubic hair and menarche in Norwegian girls. *J Clin Endocrinol Metab*. 2020;**105**:dgaa107.

4.   Bruserud IS, Roelants M, Oehme NHB, *et al*. Ultrasound assessment of pubertal breast development in girls; intra- and interobserver agreement. *Pediatr Radiol*. 2018;**48**:1576–83.

Abbreviations

%BF	body fat percentage
AOR	age-adjusted odds ratio
BGS1	Bergen Growth Study 1
BGS2	Bergen Growth Study 2
BIA	bioelectrical impedance analysis
BMI	body mass index
CALIPER	Canadian Laboratory Initiative for Pediatric Reference Intervals
CDC	Centers for Disease Control and Prevention
CDGP	constitutional delay of growth and puberty
CHH	congenital hypogonadotropic hypogonadism
CI	confidence interval
CLSI	Clinical Laboratory Standards Institute
CNS	central nervous system
CV	coefficient of variation
DEXA	dual-energy X-ray absorptiometry
DHEAS	dehydroepiandrosterone
EDC	endocrine-disrupting chemical
edf	equivalent degrees of freedom
EDTA	ethylenediaminetetraacetic acid
EWAS	epigenome-wide association studies
FSH	follicle-stimulating hormone
GAM	generalized additive model
GLM	generalized linear model
GnIH	gonadotropin-inhibiting hormone
GnRH	gonadotropin-releasing hormone
GOOD	Gothenburg Osteoporosis and Obesity Determinants (study)
GWAS	genome-wide association studies
HPA	hypothalamic–pituitary–adrenal
HPG	hypothalamic–pituitary–gonadal
IGF-1	insulin-like growth factor 1

IOTF	International Obesity Task Force
LC-MS/MS	liquid chromatography with tandem mass spectrometry
LH	luteinizing hormone
LL	lower limit
LOA	limits of agreement
MoBa	Norwegian Mother and Child Cohort Study
NHANES III	Third National Health and Nutrition Examination Survey
OM	orchidometer
OR	odds ratio
PH	pubic hair (Tanner staging)
PHV	peak height velocity
PROS	Pediatric Research in Office Settings (study)
ROC	receiver operating characteristic
SD	standard deviation
SE	standard error
SHBG	sex hormone-binding globulin
SPSS	Statistical Package for the Social Sciences
SSF	subscapular skinfolds
TEM	technical error of measurement
TV	testicular volume
UL	upper limit
US	ultrasound
USTV	ultrasound-determined testicular volume
USTVz	testicular volume-for-age z-score
WC	waist circumference

1.   Introduction

The Bergen Growth Study 1 (BGS1) conducted in 2003–2006, provided valuable information about contemporary growth in Norwegian children. However, it did not include information about puberty. The Bergen Growth Study 2 (BGS2) filled this gap by collecting data on pubertal development in Norwegian children. This thesis presents findings on boys who were examined as part of BGS2.

      Prior to the BGS2, only limited data on pubertal development in Norwegian boys were available, and the pubertal age references used on current Norwegian growth charts are based on Danish data collected between 1991 and 1993 (1). Contemporary pubertal references have both epidemiological and clinical relevance. From an epidemiological point of view, data on pubertal onset are needed to assess the timing of puberty initiation and possible secular trends in puberty development in a population over time. In the clinical setting, a pubertal reference allows to define early and late puberty, both of which are related to health risks for an individual child. In addition, pubertal studies are important for a better understanding of the underlying mechanisms that influence puberty, including its timing and progression, which, particularly in boys, are not fully elucidated.

1.1 Normal pubertal development in boys

Puberty marks the transition from childhood to adulthood and is characterized by the appearance of secondary sex characteristics, growth spurt, sexual maturation and subsequent fertility, and profound psychological changes. The psychological development during puberty is outside the scope of this study which here mainly focuses on the physical changes in puberty and the achievement of pubertal milestones.

1.1.1 The hypothalamic–pituitary–gonadal axis

The hypothalamic–pituitary–gonadal (HPG) axis comprises the hypothalamus, the pituitary gland, and the gonads (testes in boys and ovaries in girls) and is the control centre of the central pubertal development (Figure 1). Gonadotropin-releasing

hormone (GnRH) is secreted from the hypothalamus in a pulsatile fashion and stimulates the synthesis and release of the gonadotropins luteinizing hormone (LH) and follicle-stimulating hormone (FSH) from the anterior pituitary gland. LH and FSH act on the testes, with LH inducing the production of testosterone from Leydig cells and FSH promoting the secretion of inhibin B from Sertoli cells (2,3). Testosterone stimulates spermatogenesis (the development of sperm cells) and is also important for muscle development, voice deepening, and enlargement of the penis. Inhibin B nurtures and supports spermatogenesis (4).
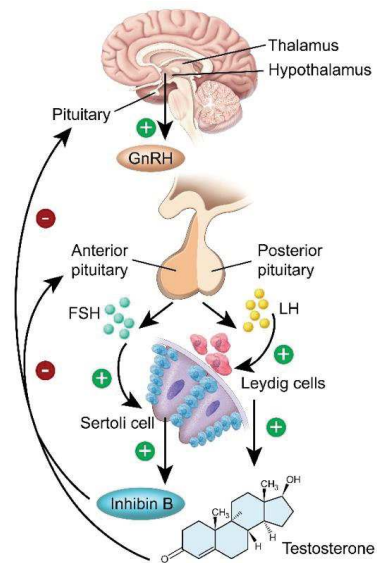
Figure 1 The HPG axis with positive and negative feedback signals. Illustration by Matthew Holt.

The HPG axis is tightly controlled and maintained by several feedback loops whereby the gonadotropins and sex steroids inhibit further GnRH and gonadotropin release. Kisspeptin and gonadotropin-inhibiting hormone (GnIH) are two hypothalamic neuropeptides that seem to play a critical role in the regulation of the reproductive axis. Kisspeptins act as stimulators of the reproductive axis (5) and are thought to have an essential role in the control of puberty (6), while GnIH is the inhibitory counterpart (7). Other hormones, e.g., the gastrointestinal regulatory

hormone leptin and insulin from the pancreas, also have stimulatory effects on the reproductive axis, whereas ghrelin, another gastrointestinal hormone, exerts an inhibitory effect on GnRH secretion (8), thus suggesting a tight correlation between reproductive axis and nutritional status. The HPG axis undergoes two activation phases throughout the lifespan. The first activation occurs as a transient surge in the first months of life, so-called mini-puberty (9), followed by a quiescent period during childhood, before the second activation (i.e. reactivation) upon puberty initiation (10). The mechanisms underlying the suppression and triggering of HPG axis reactivation are not fully known. In essence, increased pulsatile GnRH secretion at puberty represents the cumulative effects of highly complex hypothalamic interactions that are influenced by both genetic factors and environmental signals (11).

Independently of the onset of sex steroid secretion from the gonads (gonadarche), the adrenal glands are activated through the hypothalamic–pituitary–adrenal (HPA) axis, marked by elevated levels of androgens, e.g. dehydroepiandrosterone (DHEAS) (12). This process is called adrenarche and can begin years before gonadarche, at a mean age of 7–9 years in boys (13). Adrenarche, together with testosterone, is responsible for the growth of pubic and axillary hair (pubarche) and the development of adult body odour and acne and may cause a transient acceleration of linear growth and bone maturation.

1.1.2 Physical changes during puberty

Pubertal physical changes occur following gonadarche. Boys undergo progressive masculinization, which includes scrotal maturation, increasing penile length and width, voice deepening, development of male hair pattern, accelerated growth, and changes to the musculoskeletal system. These changes are known as *pubertal milestones* and result from an increase in sex hormone synthesis by the gonads under the control of the HPG axis. Puberty progression usually occurs in an ordered sequence, with testicular enlargement as the first pubertal milestone, followed by pubic hair growth and penile growth. Longitudinal studies have shown that it takes approximately 4–5 years for boys to reach full development of adult male genitalia from the first signs of genital growth (14).

1.1.2.1 Testicular and scrotal development

The testes are two small oval-shaped organs that are contained within the scrotum and are responsible for secreting testosterone and producing sperm (Figure 2). Testosterone is necessary for normal masculinization and spermatogenesis. Prenatal and postnatal activation of the HPG axis is associated with testicular growth and testicular descent into the scrotum. Testicular size increases from early childhood to prepuberty from around 1 mL to 3 mL, as measured with the Prader orchidometer (15), with peak testicular growth achieved during puberty (16-18), before a testicular size of ≥15 mL is reached in adulthood. There is a large interindividual variation in the timing of testicular growth, as well as in adult testicular volume (TV) (19). Starting in the first trimester of fetal life, the testis contains two compartments that gradually differentiate through to adulthood to comprise the seminiferous tubules containing Sertoli and germ cells, and the interstitial tissue containing Leydig cells (20). Pubertal testicular size increase to a final volume of 15–25 mL and is largely dependent on the action of FSH inducing germ cell proliferation and growth of the seminiferous tubules, whereas LH and testosterone are essential players in completion of spermatogenesis (21). FSH, LH, and testosterone work in synergy, with all three hormones needed for normal spermatogenesis. In a recent study, adult TV has been shown to correlate with sperm output and concentration (22).
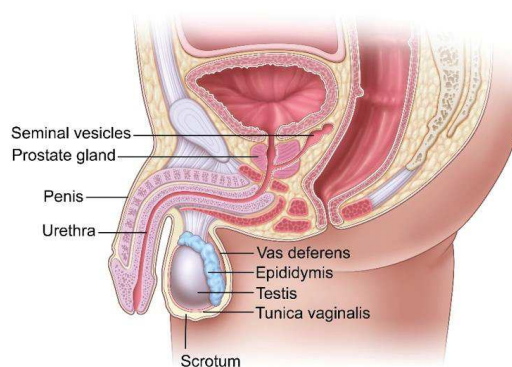
Figure 2 Anatomy of the male reproductive system. Illustration by Matthew Holt.

The testes are closely associated with several structures within the scrotum, namely the tunica vaginalis, the epididymis, and the vas deferens. The tunica vaginalis is a membrane that covers the testes. The epididymis is a long tube which moves sperm from the testicle, to the vas deferens where the sperm is stored before it is carried out of the scrotal sac. The vas deferens connects the epididymis and the urethra. Structures outside the scrotum that are also part of the male reproductive system include the seminal vesicles and the prostate. The seminal vesicles lie behind the bladder and produce and release seminal fluid rich in fructose and proteins. Seminal fluid is a constituent of semen, contributing about 50–80% of the ejaculatory volume (23,24). The prostate gland surrounds the neck of the bladder and urethra and secretes an alkaline fluid, also a constituent of semen.

1.1.2.2 Penile growth

The growth curve of the penis differs from that of the rest of the body (25). The penis starts to grow from birth for 3–4 years and thereafter changes little until pubertal onset. Penile growth is growth hormone (GH)- and testosterone-dependent; it occurs first in length and then in diameter, and it is a relatively early pubertal sign, beginning gradually from about age 10 years (26). In the average male, adult penile length is reached by age 16 or 17 years, although with considerable variation.

1.1.2.3 Pubarche/adrenarche

Pubarche refers to the first appearance of pubic hair and is considered a manifestation of adrenarche. Pubic hair is dark, long, and eventually curly. In a longitudinal study from Denmark, 90 healthy boys were examined every 6 months for 5 years. They found that only 25% experienced pubic hair development before they reached a pubertal TV of >3 mL (pubarche pathway) and that 60% achieved pubertal TV before the development of pubic hair (testicular pathway) (27). Pubarche does not necessarily represent evidence of gonadotropin-dependent puberty (through the HPG axis). However, the role of adrenal androgens in central initiation of normal puberty in boys remains unknown. In addition to pubic hair growth, boys will also undergo the development of axillary and facial hair and adult body odour, with some also

experiencing the development of acne, all of which are clinical consequences of adrenarche.

1.1.2.4 Voice break

Voice break is a result of lengthening of the vocal cords that follows the growth spurt of the larynx, thus causing an abrupt decrease in the fundamental voice frequency (28). Voice changes in boys occurs about 2 years after pubertal onset (defined by TV ≥4 mL) and become obvious when the TV is around 12 mL (29). Voice break can be used as a marker for late puberty, as over 30% of boys complete voice break by age 14 years (30).

1.1.2.5 Gynaecomastia

Gynaecomastia is enlargement or swelling of the breast tissue in males. It can be unilateral or bilateral and occurs to some degree during puberty in 39–75% of boys (31,32). It is thought to be due to a relative imbalance between free oestrogen and free androgen actions in the breast tissue, and lower serum free testosterone levels have been observed (32). Physiological gynaecomastia usually resolves as puberty progresses and testosterone levels increase.

1.1.2.6 Skeletal growth and body composition

GH and insulin-like growth factor 1 (IGF-1) are markedly increased during puberty. Along with sex steroids (especially oestradiol, which is aromatized from testosterone in the growth plate), both GH and IGF-1 contribute to pubertal growth spurt. This spurt is the most rapid growth phase since the neonatal period, following a reduced growth rate in late childhood. Up to 20% of adult height is achieved during puberty, and the total height gained from the take-off point to cessation of growth averages 28 cm in boys (33,34), with an incremental rate of 9.5 cm/year (35). Early-maturing boys often have a large pubertal growth spurt, but a shorter period of childhood growth, while late-maturing boys experience a less pronounced pubertal growth spurt, but a longer period of childhood growth. Therefore, the timing of pubertal onset does not greatly influence adult height (30).

As boys go through puberty, total body bone mass and fat-free mass continue to increase, resulting in an increased body mass index (BMI) in puberty. The increase in lean body mass starts at around age 10 years in boys and is the earliest change in body composition in puberty (36). Puberty in males is characterized by greater gain in fat-free mass, compared to fat mass (37), as well as greater gain in central fat relative to total body fat (38).

1.2 Assessment of puberty in boys

It can be challenging to determine when the first signs of puberty appear or even to know what signals pubertal onset. Unlike in girls in whom menarche is a clear marker of puberty, there is no similar convenient marker of puberty in boys that can be assessed for over an interview consultation. Studies on boys must therefore rely on physical examination, preferably with accurate staging of TV, pubic hair, and genital development (39). Such studies are difficult to implement on a large scale, which could explain why there are only few studies on pubertal development in boys. Data on pubertal development gathered from a physical examination are often unavailable from epidemiological studies, and are commonly based either on surrogate markers, such as peak height velocity, voice break, and timing of the first conscious ejaculation (40-42), or on self-assessment of pubertal status (43,44). The validity of such data has been debated (45).

1.2.1 Tanner stages for genital and pubic hair development

The commonest way to assess puberty in clinical practice is based on a classification system of secondary sex characteristics developed in the late 1960s by the British paediatricians William Marshall and James Tanner (46). Based on longitudinal photographic observations of genital development in a rather small sample of 228 boys living in a children's home, Marshall and Tanner developed a five-grade scale for the development of external genitalia (Tanner G) and pubic hair development (Tanner PH). Tanner stages for genital and pubic hair development are determined by visual inspection of individual boys, with comparison to pictures or sketches (Figure 3). Tanner stages G1 and PH1 are considered prepubertal, while Tanner stage G2

(enlargement of the scrotum and testes, i.e. TV ≥4 mL; scrotum skin reddening; and changes in texture) marks the onset of puberty. Visual inspection of scrotal and penile changes may be inconsistent and is considered a subjective assessment, as subtle changes in penile size and scrotal skin texture at the onset of puberty can be difficult to detect (47). Therefore, Tanner stage G2 alone is considered a poor index of pubertal onset (48). Tanner stage PH2 (sparse growth of long, slightly pigmented hair at the base of the penis) is considered pubertal and is often the easiest physical change to observe. However, the appearance of pubic hair alone may not indicate the onset of gonadal activity but instead reflect adrenal androgen secretion. Tanner and Marshall reported a mean age of 11.64 years for Tanner stage G2 and 13.44 years for Tanner stage PH2 (as assessed from photographs in cases where the first appearance of pubic hair is difficult to see) (46). Tanner stages G5 and PH5 mark the adult phenotype.
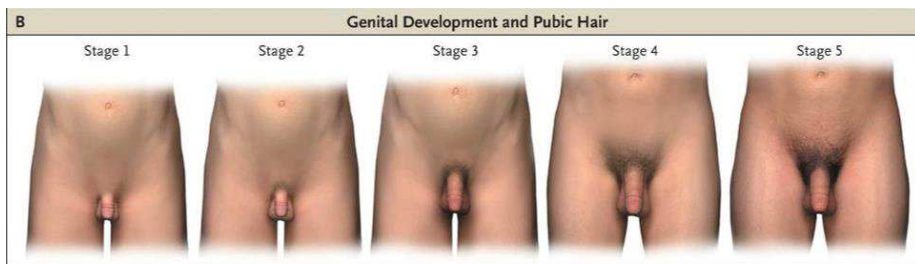
Figure 3 In boys, genital development is rated from Tanner stage 1 (prepubertal) to Tanner stage 5 (adult); stage 2 marks the onset of pubertal development and is characterized by scrotal and testicular enlargement, as well as by a change in the texture and reddening of the scrotal skin. Pubic hair development is rated from Tanner stage 1 (prepubertal, no pubic hair) to Tanner stage 5 (adult); stage 2 marks the onset of pubic hair development. Reproduced with permission from Carel and Léger, *N Engl J Med* 2008, Copyright Massachusetts Medical Society.

1.2.2 Measurement of testicular volume

Measurements of the TV are considered more objective and quantifiable, compared to Tanner staging of genital development, with less interobserver variation (47). Reliable and accurate measurements of the TV are of great importance for examining pubertal development, and for diagnosing and monitoring treatment for cryptorchidism, hypogonadotropic hypogonadism, and varicocele, as well as

testicular damage from cytotoxic agents, and for estimating potential male infertility (21). A simple visual inspection is not sufficient, and both palpation and actual measurements are required for an accurate determination of the TV. The TV is usually measured with a Prader orchidometer, simple calliper, or ruler, or with ultrasound.

1.2.2.1 Prader orchidometer

The Prader orchidometer, introduced by Andrea Prader in 1966 (49), is the most widely used clinical tool to assess the TV. It consists of a chain of 12 solid ellipsoid beads of different sizes (1–6, 8, 10, 12, 15, 20, and 25 mL) (Figure 4). TV measurement using the Prader orchidometer is performed by holding the orchidometer in one hand, tightening the scrotal skin around a testicle with the other hand, and identifying the best size-matched bead on comparing with the testicle. The Prader orchidometer tends to overestimate small TVs, when compared to the methods of water displacement and ultrasound, due to potential interference from surrounding structures such as the scrotal skin, epididymis, and tunica vaginalis (50). Studies have shown that the accuracy of measurements of testicular size is also highly dependent on the operator's experience (51). Moreover, highly significant interobserver variation has been found among users of the Prader orchidometer (52).
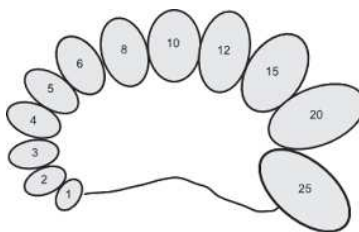
Figure 4 Prader orchidometer. Bead size 1–3 mL = prepubertal; 4–6 mL = early puberty; 8–10 mL = mid puberty; 12–15 mL= advanced puberty; >15 mL = adult. With permission from Wikimedia.org.

The first reliable marker for central pubertal onset in boys is a TV of >3 mL, often expressed as ≥4 mL when measured with an orchidometer (46). This is usually consistent with Tanner stage G2. In addition to the Prader orchidometer, punched-out

orchidometers, callipers, and ordinary rulers also have been used to estimate the TV (53,54) but have not gained the same clinical standing as the Prader orchidometer.

1.2.2.2 Ultrasound

Ultrasound uses high-frequency sound waves to produce images of tissues and organs. It can differentiate between tissue types, thus allowing direct observation of the testicle and their size measurements, while excluding the scrotal skin and epididymis, as well as of structural features of potentially pathologic conditions such as hydrocele or varicocele. Several authors have suggested that ultrasound is the gold standard for TV measurements and that it should be the method of choice when accuracy of TV measurements is of particular importance (55,56). The length (L), width (W), and depth (D) of the testicle are measured from the mid-sagittal and mid-transverse planes and the TV is calculated using these three dimensions. The formula for a prolate ellipsoid ($L \times W^2 \times 0.52$) has been widely used, but several studies have shown that the Lambert formula ($L \times D \times W \times 0.71$) (57) is more accurate and its calculated TV corresponds better to the true TV (58,59). While ultrasound has been used for decades to detect scrotal pathology, its use to assess the pubertal stage or to establish reference ranges has been a more recent development. Thus, a Dutch study in 2011 (60) was the first to present reference values for the TV in healthy children and adolescents using ultrasound. Ultrasound imaging is regarded as safe, with no associated ionizing radiation exposure or undesirable side effects (61), and can thus be widely implemented for evaluation of pubertal development.

1.2.3 **Other pubertal markers**

1.2.3.1 Peak height velocity

Age at peak height velocity (PHV) is the age at which a child experiences the greatest increase in stature during the adolescent growth spurt. It requires the collection of longitudinal measurements until near cessation (e.g. <2 cm/year) of linear growth to estimate the velocity and acceleration of height and the age at PHV (62). PHV is one of the most commonly used indicators of puberty in population studies of pubertal development and adolescent maturation (41,63), as it is non-invasive and objective,

particularly compared to Tanner staging or TV assessment. Onset of pubertal growth spurt ('take-off') usually occurs when the TV measured with a Prader orchidometer is about 8 mL, while PHV coincides with a TV of about 12 mL, at a mean age of 13.5 years (33). Take-off and PHV are therefore relatively late signs of puberty. A recent study from Denmark reported the age at PHV to be 13.7 years (29). Because there is substantial variability in the timing of PHV across Tanner stages, PHV might not be a good marker for the degree of pubertal development, but rather for the tempo of growth and rate of maturation (48,64).

1.2.3.2 Voice break

The age at voice break has also been used to determine timing of puberty in population studies (30,42,65). Assessment of age at voice break can be conducted by direct observation of an examiner as a 'yes' or 'no' outcome, or by using Cooksey classification of voice analysis (66), or by self-reporting of either unintentional falsetto notes or voice deepening by individual boys themselves. As with PHV, voice break is also a late pubertal milestone, with mean age of 13.6 [95% confidence interval (CI) 13.5–13.8] years (29).

1.2.3.3 Spermarche and ejacularche

Spermarche in boys, the counterpart of menarche in girls, is the onset of release of spermatozoa (sperm cells). Spermarche is usually identified by detecting the presence of spermatozoa in the urine. Spermarche is seen as an early pubertal event (67) that occurs between the ages of 11 and 15 years. In a longitudinal study of 40 healthy Scottish boys over a period of 7 years, sperm was detected in early-morning urine samples at a median age of 13.4 (range 11.7–15.3) years (67). The first conscious ejaculation (which is discharge of semen from the male reproductive tract as a result of an orgasm), called ejacularche, was self-reported to occur at a mean age ± standard deviation (SD) of 13.3 ± 1.1 years in a total of 1582 Bulgarian boys (40). While adult sperm morphology, motility, and concentration are only observed when the bone age advances to around 17 years (68), it is possible for boys with an immature physical appearance to be fertile.

1.2.4 **Paediatric endocrine references**

Measurement of serum testosterone levels has proven to be an accurate predictor of pubertal development in boys (69). The availability of appropriate population-based endocrine reference intervals (i.e. normative values) is crucial for clinicians to aid in disease diagnosis and treatment, as well as patient follow-up, and measurements of hormones involved in puberty are part of the general assessment in the clinical setting. Endocrine references can help in identifying endocrinopathies, with either excessive or impaired production of different hormones.

Establishing reference intervals is particularly challenging in the field of paediatrics, due to continuous physiological changes that occur throughout childhood and adolescence (70), in addition to 24-hour variation in hormone levels (71). Reference intervals must therefore be stratified or partitioned in terms of both sex and age, and preferably also in terms of the stage of pubertal development. Robust sample sizes and appropriate age ranges are needed to develop reliable estimates for the normal range and 90% CIs (72). In a series of publications on the challenges of, and proposed solutions for, establishing paediatric reference intervals, the Canadian Laboratory Initiative for Pediatric Reference Intervals (CALIPER) consortium provided a comprehensive set of guidelines and mathematical framework to address these challenges (73). In addition, the CALIPER consortium provided a comprehensive database of age- and sex-specific reference intervals for >100 biomarkers of paediatric diseases. Reference intervals stratified according to ultrasound-measured TVs have until now not been available.

To summarize, the TV is regarded as the most reliable marker in assessing pubertal onset in boys, supplemented with measurement of serum testosterone levels. Other pubertal markers, such as age at PHV and age at voice break are, however, often used in epidemiological studies.

1.3 Timing of puberty in boys

1.3.1 Secular trends

The timing of pubertal onset has a near-normal distribution in the general population, with too early or delayed puberty being statistically defined, using 2–3 SDs below and above the population mean age of onset of puberty, respectively. Puberty is usually said to be physiological when it begins between the ages of 9 and 14 years in boys (74). Evidence of secular trends (the changing distribution of a population parameter over time) in male pubertal development is limited by the small number of studies and the use of different pubertal markers making comparisons difficult. Studies from Europe before the twenty-first century did not show the same trend towards earlier puberty in boys (1,75) as that seen in girls (76,77). However, as reported in the Third National Health and Nutrition Examination Survey (NHANES III), age at onset of Tanner stage G2 seemed to occur much earlier in the United States, compared with Europe at around the same time period, with a mean age at Tanner stage G2 onset of 10.1 years in white American boys (78), compared to 11.8 years in Danish boys (1). In addition, it was also reported that age at achievement of Tanner stage G2 in the United States declined from around 11.6 to 10.1 years among non-Hispanic white boys from 1966–70 to 1988–94 (45). Studies from the last few decades, on the other hand, have suggested a possible trend towards earlier puberty also in European boys. The Copenhagen Puberty Study reported a decline in age at onset of puberty of 3 months between 1991 and 2006 (79), while a Greek study found no evidence of a secular trend between 1996 and 2009 (80). The only previous published data on pubertal development in Norwegian boys came from a small study including 109 boys aged 1.9–16.9 years from the 1970s, which demonstrated a mean age at pubertal onset of just below 12 years, defined by a TV of 4 mL measured using the Prader orchidometer (81). Table 1 summarizes selected literature on pubertal timing in boys (1,27,29,60,75,79,80,82-88).

Table 1 Selected published literature on pubertal timing in boys

Study	Year	Boys (N)	Country	Pubertal marker (mean or median age, years)			Comments
				Testicular volume	Tanner stage PH2	Tanner stage G2	
Lindgren, 1996 (82)	1980	116	Sweden		12.7	11.6	
Juul et al., 2006 (1)	1991–93	826	Denmark	11.9 (>3 mL)	11.9	11.8	
Herman-Giddens (NHANES III), 2006 (83)	1988–94	2481	United States		12.0 (w) 11.2 (b)	10.1 (w) 9.5 (b)	Visual inspection only
De Simone et al., 2004 (84)	1991–94	535	Italy		11.5	11.2	
Mul et al., 2001 (75)	1997	~2000	The Netherlands	11.5 (≥4 mL)	11.7	11.5	
Monteilh et al. (ALSPAC), 2011 (85)	1999	3940	United Kingdom		11.4		Longitudinal, self-reported data
Roelants et al., 2009 (86)	2002–04	4219	Belgium	11.4 (≥4 mL)	11.9	11.4	
Sorensen et al., 2010 (79)	2006–08	704	Denmark	11.7 (>3 mL)	12.4	11.6	
Papadimitriou et al., 2011 (80)	2007–09	932	Greece	11.3 (≥4 mL)	11.2		
Goede et al., 2011 (60)	2007–09	769	The Netherlands	11.6 (≥4mL)			TV examined both with PO and US
Herman-Giddens et al. (PROS), 2012 (87)	2005–10	4131	United States	11.5 (≥4 mL) (w) 11.8 (≥4 mL) (b)	11.5 (w) 10.3 (b)	10.1 (w) 9.1 (b)	
Mouritsen et al., 2013 (27)	2006–10	90	Denmark	11.5 (>3 mL)	11.6		Longitudinal
Wohlfahrt-Veje et al., 2016 (88)	2006–13	846	Denmark	11.6 (>3 mL)	11.9	11.5	Longitudinal
Busch et al., 2019 (29)	2006–14	714	Denmark	11.6 (≥4 mL)	12.2	11.6	Mixed cross-sectional/longitudinal

Abbreviations: ALSPAC, Avon Longitudinal Study of Parents and Children; (b), black boys; NHANES III, Third National Health and Nutrition Examination Survey; PO, Prader orchidometer; PROS, Pediatric Research in Office Settings; TV, testicular volume; US, ultrasound; (w), white boys.

1.3.2 **Disorders of pubertal timing**

Pubertal development usually follows a predictable pattern of onset, sequence, and tempo. Lack of, or premature, development of pubertal milestones outside the defined limits warrants further investigations.

1.3.2.1 Early puberty

Early puberty, or pubertas praecox, is defined as testicular enlargement with a TV of 4 mL or more before the age of 9 years in boys. Early puberty can be either central (involving activation of the HPG axis) or peripheral (most often caused by a gonadal or adrenal gland disorder). A population-based study from the Danish National Registry with data collected between 1993 and 2001 showed an incidence rate of precocious puberty of 1–2 in 10 000 boys (89). Idiopathic central pubertas praecox is very uncommon in boys (ten times less frequent, compared to girls) (90) and is more likely to have an underlying pathology in the central nervous system (CNS) (91) such as tumours, congenital malformations, or infections (10).

1.3.2.2 Delayed puberty

Delayed puberty, or pubertas tarda, is defined as absence of testicular enlargement beyond the age of 14 years. It is commoner in boys than in girls, with constitutional delay of growth and puberty (CDGP) as the commonest cause (in up to 83% of boys with pubertal delay) (92), and typically has a familial component. CDGP is characterized by sex hormones and gonadotropins levels correlating with bone age, rather than with chronological age. It is, however, seen as a normal variant if puberty later initiates spontaneously, after the upper age limit. Delayed puberty can also be caused by psychosocial stress, malnutrition, endocrine or gastrointestinal disorders, or renal failure (92-94).

1.3.2.3 Puberty failure

It may be difficult to distinguish CDGP from puberty failure, where in the latter puberty will not spontaneously begin. Congenital hypogonadotropic hypogonadism

(CHH) is a rare disorder with a prevalence estimated at 1 in 4000 to 1 in 10 000 males (95) and is characterized by inability to produce LH and FSH in the pituitary gland. When CHH is associated with hyposmia or anosmia (~60% of cases), it is known as Kallman syndrome (96). Hypogonadotropic hypogonadism can also be acquired, as it can be caused by intracranial tumours, nutritional deficiencies (anorexia nervosa), and autoimmune diseases such as sarcoidosis (97).

Further, pubertal failure can be a result of primary gonadal failure with decreased testosterone production, leading to a lack of negative feedback to the hypothalamus, in turn causing hypergonadotropic hypogonadism with high LH and FSH and low testosterone levels (10). This is seen in Klinefelter's syndrome (47,XXY), gonadal dysgenesis, cryptorchidism, and post-radiation therapy or chemotherapy (93). Puberty failure is usually treated with lifelong testosterone supplementation.

1.3.2.4 Premature adrenarche

Premature adrenarche in boys is when androgenic signs appear before the age of 9 years, together with circulating DHEAS concentrations above the low prepubertal level. The incidence rate in boys was found to be 1.8% in a Finnish study (98). The most revealing sign of premature adrenarche is the appearance of pubic or axillary hair, but the development of adult body odour, acne, and accelerated growth also might be observed (99).

1.4 Factors influencing the timing of puberty

The precise genetic pathways which regulate the age at onset of puberty are largely unknown, but in addition to genetic influence, it is likely that environmental factors, such as BMI, nutritional status, psychosocial factors, and endocrine-disrupting chemicals (EDCs), also impact pubertal development.

1.4.1 Genetics

The timing of pubertal onset is highly heritable and polygenic. Studies have shown that around 50-80% of the variation in pubertal onset might be genetically determined

(74,100,101). Support to these findings has come from twin studies (102), showing a higher correlation for age at onset of puberty and age at PHV in monozygotic than in dizygotic twin boys (r = 0.9 vs 0.4, respectively) (63). Further, studies on delayed or absent puberty have revealed the involvement of ~20 genes (95), and only a few genes implicated in precocious puberty (103,104), thus furthering our understanding of the genetic regulation of puberty timing in males. In recent years, genome-wide association studies (GWAS) have uncovered the potential involvement of an increasing number of genes in the normal variation in pubertal timing, although all seem to have small effect sizes (105). The largest GWAS on male puberty timing to date, including >200 000 men, identified 76 independent signals for puberty timing (106). This study also showed a genetic association between male puberty timing and adverse health outcomes and, by contrast, a longer lifespan in boys with later puberty corresponding to 9 months longer life per year of later puberty. An overlap of genes involved in puberty timing and adiposity has also been found (107,108), and epidemiological studies have proposed the existence of a pathway for early infancy growth and earlier puberty (30). However, in contrast to girls, in whom alleles associated with increased BMI correlated with earlier breast development, there was an association in boys between some BMI-increasing alleles and earlier sexual development and other alleles with delayed sexual development (109). These findings are in line with epidemiological studies showing conflicting correlations between prepubertal BMI and timing of puberty in boys. One of the most frequently reported associations of a genetic locus with puberty timing is *LIN28B* on chromosome 6q21. *LIN28B* has been reported to show an association with voice break status at age 15 years, more advanced pubic hair stage at ages 13 and 15 years, and faster height growth at age 10 years (110).

The secular trend towards earlier puberty is unlikely to be caused by rapid genetic alterations, but rather by changes in non-genetic factors (74), and some effects are thought to be epigenetically modulated (105).

1.4.2 **Body composition**

Energy homeostasis represents an important factor in the central neuroendocrine system influencing puberty timing. It is well known that chronic malnutrition and chronic illness delay the onset of puberty and slow its progression (111,112) and that adequate nutrition is a key factor for normal puberty timing and tempo. However, it remains unclear to what extent overnutrition, leading to overweight and obesity, influences puberty timing in boys. The secular increase in overweight and obesity in children and adolescents over the last decades has received special attention as a potential aetiological factor for the concurrent secular trend towards earlier puberty onset, especially seen in girls (76,113-117). The effect of obesity on early puberty in boys, however, is more ambiguous, with studies reporting conflicting results (44,79,118-123).

The timing of weight gain and increased BMI seems to influence puberty timing in different ways. Fast weight gain from 0 to 6 months and during childhood was found to be associated with advanced puberty in boys and girls in a Jamaican study, although there was no similar association with large birth size alone (124). The same study also found that elevated fat mass at 8 years of age was associated with advanced puberty; by contrast, at age 11, it was elevated lean mass, and not fat mass, that showed this association. Dunger *et al*. concluded in their review that infancy probably is the most important age period during which weight gain influences the tempo of growth and timing of puberty onset (125). It seems to be the *change* in BMI, rather than the *absolute* BMI, in an individual child that most influences puberty timing.

There has been little focus on the effects of low body weight on later onset of puberty in boys. However, a few studies have found evidence of delayed puberty in leaner boys, compared to normal-weight and overweight children (126,127).

The satiety hormone leptin has been suggested to be one of the links between weight status and puberty timing (128). Leptin is produced by adipocytes and its levels rise gradually with age and are increased in subjects with high body fat (129). Leptin is thought to have a permissive role, rather than being a trigger for the onset of puberty. It acts on the hypothalamus by modulating the Kiss1/Kiss1R system (130).

In boys, leptin levels seem to rise transiently and then decrease after Tanner stage 2 to prepubertal levels, corresponding to the reduction in body fat seen in boys during puberty (37). The importance of leptin in normal functioning of the HPG axis is shown in patients with either leptin or leptin receptor deficiency presenting with hypogonadotropic hypogonadism (131,132).

Insulin resistance, which is commonly observed in overweight and obese boys, has been proposed in some studies as a causative factor explaining why these boys enter puberty *earlier* (133-135). In the presence of insulin resistance, compensatory hyperinsulinaemia usually results in reduced levels of sex hormone-binding globulin (SHBG), consequently increasing the bioavailability of sex steroids, which, in turn, can change the onset and tempo of puberty (133-135).

1.4.2.1 Assessment of body composition

Defining overweight, obesity, or excess body fat in children is more difficult than in the case of adults, as normal body fat not only differs between the sexes, but also varies with age and the maturity of the child. BMI, calculated as kg/m^2, is the most used measure of weight status in population studies, as well as in clinical settings. It is easy to measure, with relatively low interobserver variation, and facilitates comparison across studies. The International Obesity Task Force (IOTF) established an international definition of paediatric overweight and obesity, based on the widely used cut-offs for overweight and obesity at age 18 years, i.e. BMI of 25 kg/m^2 and 30 kg/m^2, respectively, creating centiles for children aged between 2 and 18 years (136). However, national references are still used. Thus, in the United States and the United Kingdom, the 85th and 95th BMI-percentiles from the Centers for Disease Control and Prevention (CDC) (137) and the UK, 1990 (138) growth charts are often used to define overweight and obesity, respectively. In Norway, the national BMI references include the IOTF cut-offs (139).

Because BMI includes both fat and lean body mass, it is therefore not the most sensitive marker for detecting excess body fat. However, BMI has shown strong correlations with other measures of body fat mass, e.g. skinfold measurements and body fat assessment by dual-energy X-ray absorptiometry (DEXA) (140).

Waist circumference (WC) is regarded as a measure of abdominal fat, which is associated with increased metabolic risk in adulthood such as dyslipidaemia, hypertension, and hyperglycaemia (141). Correlation between WC and truncal fat in children has been confirmed with DEXA (142,143), and Brannsether *et al*. demonstrated that WC was the one measure that most strongly correlated with BMI at all ages (144). Norwegian references for WC have been established (145).

Skinfold thickness is considered as a direct measure of subcutaneous fat and is most commonly assessed on triceps skinfolds and subscapular skinfolds (SSF) in children. In contrast to BMI, skinfold thickness measurement is more prone to interobserver variability and is therefore not suitable in routine clinical practice (146). However, skinfold thickness has been shown to correlate with body fat percentage (%BF), with a higher sensitivity, compared to BMI and WC, in determining excess body fat (140,147). Norwegian references for triceps skinfolds and SSF have been established (148).

DEXA and underwater weighing both give information about body composition, including the proportion of fat tissue, and DEXA has been proposed as the gold standard for evaluation of body composition (149). However, due to limited accessibility of DEXA in many settings, its relative high running costs, and associated exposure to low-dose radiation, bioelectrical impedance analysis (BIA) may be an easier alternative to determine %BF in epidemiological studies (150). BIA has been proposed as a more precise tool than skinfold thickness measurement to determine fat mass in epidemiological studies (151). BIA provides data on %BF and body fat mass by sending an electric current that passes quickly through water normally stored in muscle tissue but meets with resistance when it hits fat tissue. This resistance, known as impedance, is measured, and used to calculate body composition. BIA is a cost-effective, rapid, and non-invasive method to estimate body composition in children and adolescents.

To summarize, there are conflicting data on the association between weight status and pubertal onset. Leptin and insulin are, among others, two hormones suggested to be the link between weight status and pubertal timing. There is a range of different weight-related anthropometric measures, with BMI as the most

commonly used in epidemiological studies. However, anthropometric measures that describe adiposity, such as WC and triceps skinfolds and SSF, in addition to %BF estimated by BIA, might add valuable information on the association between weight status and timing of pubertal development.

1.4.3 Endocrine-disrupting chemicals

EDCs are either naturally occurring or synthetic substances that can interfere with normal endocrine function (152). For example, some EDCs have anti-androgenic and anti-oestrogenic effect, whereas others are aromatase inhibitors (153). The effects of EDCs on puberty timing has been an ongoing concern (154); EDCs are thought to play a causative role in the recent decline in sperm counts and impaired fertility worldwide, in addition to the increase in reproductive cancers in some geographical areas (155). One study found that genital and pubic hair development in boys was inversely associated with the serum concentrations of some EDCs, namely polychlorinated biphenyls (156). The effects of EDCs can manifest right before puberty, as well as much earlier in life, including during neonatal, and even fetal life through pregnant women's exposure to EDCs (157). Consequently, different exposure timings, together with different types and levels of EDCs exposed to, make it difficult and complex to precisely investigate the effects of EDCs on puberty timing. Studies have shown that prenatal and early postnatal exposure, compared to prepubertal exposure, can result in different effects of EDCs on puberty timing in boys, i.e. either early or late pubertal onset and progression, depending on the specific 'culprit' EDCs (158).

1.4.4 Stress and socio-economic factors

Several studies reported that psychosocial stress during prepuberty or puberty in girls may cause pubertal delay or arrest (159), whereas advanced puberty has been described in girls who experience such stress in early postnatal life or infancy (160). Only a few studies have examined these same effects in boys, mostly due to the difficulties of assessing puberty timing in epidemiological studies. However, a Danish study found an increased risk of developing precocious puberty in foreign-

adopted boys, compared to boys with Danish background, with an incidence rate ratio of 13.4 (95% CI 5.8–31.1; p <0.001) (161). Further, foreign-adopted boys also had an increased risk of developing precocious puberty, compared to foreign boys who immigrated with their families, thus indicating the potential effect of stress on puberty timing.

1.5 Consequences of altered puberty timing

As for other aspects of puberty timing in boys, only few studies have been conducted on adult health outcomes of altered puberty timing. However, data from a few epidemiological studies showed that early pubertal onset in men is a risk indicator for adult disease, e.g. angina pectoris, type 2 diabetes, and hypertension (162,163). One study also reported an increased risk of testicular cancer in boys with early onset of puberty (164), whereas a meta-analysis including 12 studies found a protective effect of *later* puberty against testicular cancer (165). An association between puberty timing and other male reproductive cancers, such as prostate cancer, is more uncertain. However, one study found markers of delayed puberty, e.g. delayed growth spurt, to be associated with a decreased risk of prostate cancer (166), while other studies showed that early age at first sexual intercourse, which may reflect early pubertal development, was associated with an increased risk of prostate cancer (167). Late pubertal onset in males have also been associated with reduced semen quality (168).

Studies on mental health issues and the effect of altered pubertal timing have shown alarming results. A Norwegian study reported that boys with off-time puberty timing (both early and late) had an increased suicidal risk (169). Young boys with earlier maturation were more likely to have functional or depressive symptoms, as well as more frequent sexual encounters and substance use (170). Further, delayed puberty has been linked to being bullied, poor self-esteem, and psychosocial distress (171).

2. Aims and hypotheses

The overall aim of the work presented in this thesis was to provide the first comprehensive data on pubertal development in Norwegian boys. The research focused on estimating the degree of male pubertal development by measuring the TV with ultrasound and assessing pubic hair development. Further, the study aimed to establish hormone references in relation to TVs. Finally, the study also aimed to identify associations between anthropometric measurements and pubertal status.

The specific study aims and hypotheses, as described in the four research papers presented in this thesis, are as follows:

Paper I: To assess the intra- and interobserver agreement of ultrasound measurements of TV in prepubertal and pubertal boys and to compare this method of measurement with the use of the Prader orchidometer.

Hypothesis: Ultrasound is a reliable method for assessment of TV, with acceptable intra- and interobserver agreement, making it useful for constructing pubertal references.

Paper II: To estimate the first pubertal references of TV and Tanner stages of pubic hair development in Norwegian boys and to compare the pubertal development with data from similar populations.

Hypothesis: Pubertal timing in Norwegian boys does not differ from that in boys from comparable populations, which implies an absence of a secular trend.

Paper III: To establish references for serum testosterone and key hormones of the male pituitary–gonadal signalling pathway, stratified by the pubertal stage based on the TV.

Hypothesis: Pubertal stage and TV are more strongly associated with testosterone levels than with chronological age, and endocrine references stratified by TV represent a feasible alternative.

Paper IV: To investigate the associations between anthropometric measurements and pubertal development.

Hypothesis: Overweight and obese boys enter puberty at a younger age and have a larger TV adjusted for age than their normal-weight peers and compared to BMI, adiposity, measured using WC, skinfolds, or %BF, show a stronger association with pubertal status.

3. Materials and methods

3.1 Study design

The work presented in this thesis is based on data from BGS2, a cross-sectional study on pubertal development in Norwegian children. Data reported in Papers II, III, and IV (main study) were collected from January to May 2016 in a mix of six randomly selected combined primary and secondary public schools, which were mostly urban and stratified by town area in the municipality in Bergen, Norway. Data presented in Paper I, which is a test–retest study, were collected from boys recruited from an additional seventh school in Bergen in February 2017 and from boys recruited from a local sports club in June 2017. Participants in BGS2 included both girls and boys, but the work presented here is based on boys only. All children in the selected schools were invited to participate, with signed informed consent from either the child's parent or their legal guardian, as well as assent from the child themselves, as a prerequisite for participation.

Participation rates for boys in the main study varied across schools, ranging from 27% to 46%, and across grades, ranging from 27% among third graders to 51% among second graders.

All boys recruited from the schools (both main study and test–retest study) were examined during school hours in their respective school. Boys recruited from the sports club were examined at Haukeland University Hospital.

3.2 Childhood populations

For Paper I, a random sample of 130 boys aged 6–16 years were invited to participate. Of these, 58 agreed to take part (34 from the selected seventh school and 24 from the sports club). The mean age was 12.0 (range, 6.5–16.4) years. One boy with a history of cryptorchidism was excluded from the study, with the remaining 57 boys eligible for examination.

In the main study (Papers II, III, and IV), all 1329 boys aged ≥6 years from the six selected schools were invited to participate. Parental informed consent was obtained for 493 (37%) boys. On the day of examination, two boys refused to give

their assent, six did not attend, and eight were excluded as their medical history included a clinical condition that could affect their growth and development (coeliac disease, cancer, benign glioma, Down's syndrome, di George syndrome, ulcerative colitis, rheumatoid arthritis, and epilepsy with ongoing antiepileptic drug therapy). In addition, 20 boys were also excluded due to past or newly discovered scrotal pathology: 4 bilateral cryptorchidism; 11 unilateral cryptorchidism; 2 retractile testes (inguinal canal); 1 hydrocele; 1 operated retractile testis; and 1 microlithiasis. Therefore, taking into account the exclusions listed above, the main study population from the six selected schools comprised 457 boys. For the reference paper (Paper II), the 57 boys from the test–retest study were also added to the main study population ($n = 514$). For Paper III, only 414 (90.6%) of the 457 boys from the main study were included, due to lack of blood samples from the other 43 boys. For the association paper (Paper IV), only boys ≥9 years were included ($n = 324$).

Of a total of 328 (71.8%) boys with a known country of origin for both parents, 254 (77.4%) had both parents from Norway, 33 (10.1%) had one or both parents from another European country, and 41 (12.5%) had either one or two non-European parents, mostly from Asia ($n = 18$), Africa ($n = 8$), or South America ($n = 10$). Analyses described in all four papers included data on *all* boys, from now on referred to 'Norwegian boys', regardless of their parents' country of origin, as logistic regression analysis of reaching pubertal TV showed no statistically significant differences between boys of Norwegian origin and other boys of European ($p = 0.17$) or non-European origin ($p = 0.11$).

Of the 336 boys with information about parental education, the highest educational level achieved by either parent was classified as: no secondary education (2.7%); secondary education (high school: 15.8%); and higher education (college or university degree: 81.6%—28% <4 years and 53.6% ≥4 years).

According to the IOTF BMI cut-off points (172), 7.7% of participating boys were classified as underweight (IOTF-BMI ≤18.5 kg/m^2), 80.5% as normal weight (IOTF-BMI >18.5 to 24.9 kg/m^2), 11.8% as overweight (IOTF-BMI ≥25 kg/m^2), and 1.9% as obese (IOTF-BMI ≥30 kg/m^2).

3.3 Ultrasound measurements

All ultrasound examinations were performed by a male radiographer with more than 8 years' experience. Before study start, an introductory ultrasound course was given by an experienced paediatric radiologist (Professor Karen Rosendahl (K.R.), also acting as second observer) with over 25 years' experience in paediatric ultrasound. Further, the first 30 ultrasound examinations were performed under supervision by K.R. A SonoSite Edge Ultrasound machine (Fujifilm SonoSite, USA) was used for examinations performed in the schools, and a SonoSite M-Turbo® HFL50× machine (Fujifilm SonoSite, USA) for examinations carried out at the local sports club; both devices were equipped with the same 15-6 MHz linear probe.

With the boy supine, the length (L), width (W), and depth (D) of the right testicle were measured according to a standardized protocol prepared beforehand. The left testicle was also measured if deemed larger on visual inspection ($n = 3$), and the volume of the largest testicle was recorded. First, the ultrasound probe was placed in the mid-sagittal testicular plane, perpendicular to the skin surface. Second, the examiner gently moved the ultrasound probe slightly back and forth until the largest diameter was obtained, namely the length. Third, the probe was rotated 90° and the width and depth measured in the mid-transverse plane (Figure 5). The TV was then calculated at a later time using the empirical Lambert formula (TV = $L \times W \times D \times 0.71$) (57).

(a) (b)
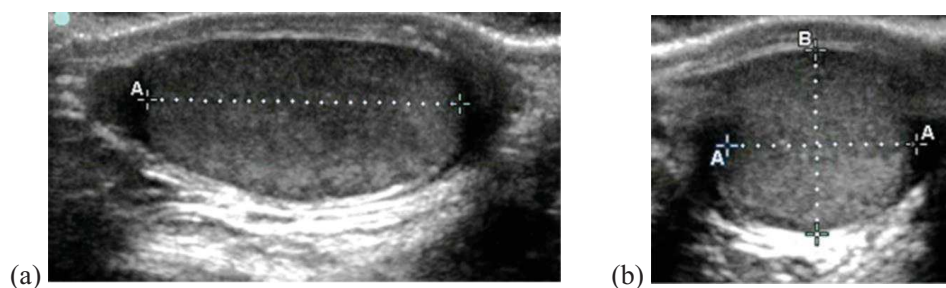
Figure 5 Ultrasound scan of the testis in a 12-year-old boy using a 15-6 MHz linear probe. (a) Length measured (dotted line A), mid-sagittal view. (b) Width and depth (dotted line B), mid-transverse view. Reprinted with permission. Oehme NHB, Roelants M, Bruserud IS, *et al*. Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys—intra- and interobserver agreement and comparison with Prader orchidometry. *Pediatr Radiol*. 2018;**48**:1771–8'

In the test–retest study (Paper I), the TV was measured twice by the main observer, with a time interval of at least 20 minutes between the two measurements during which examination of at least three other participants was performed. This was done to minimize the risk of recall of the first measurement. The participating boys were examined once by the second observer who was blinded to the results obtained by the first observer.

3.4 Pubertal assessments

Tanner staging of pubic hair (PH) development was determined by the main observer. Tanner PH stages were visually assessed with respect to the quantity, characteristics, and distribution of pubic hair. The boys were examined in the supine position, using pictures from the original work of Marshall and Tanner as reference (46). This assessment was done only as part of the main study, and data were recorded for 452 of the 457 boys.

For the test–retest study, TV measurements of the right testicle was performed using a Prader orchidometer by a paediatric endocrinologist (Professor Pétur Júlíusson) who has more than 18 years' experience. The boys were examined in a warm room in a standing position. The best matching volume was recorded by comparative palpation. If the testicular size was perceived to be in between two consecutive beads, the mean volume of these beads was recorded.

3.5 Blood samples

Venepuncture was carried out from an antecubital vein by an experienced biomedical laboratory scientist during school hours. About half of all blood samples were collected between 09:00 and 11:00, and 90% before 13:00. For each subject, non-fasting blood samples were collected as follows: 2×6.6-mL serum gel tubes; 1×6.6-mL ethylenediaminetetraacetic acid (EDTA) plasma tube; and 1×7-mL EDTA blood tube. All blood samples were transported on dry ice to Haukeland University Hospital in Bergen and stored at $-80°C$ in a biobank freezer for subsequent analyses. Blood

analyses from 414 (90.6%) boys (excluding those with a chronic disease or scrotal pathology) were included in Paper III.

Androgen levels were analysed using state-of-the-art liquid chromatography with tandem mass spectrometry (LC-MS/MS), as described previously (173), and peptide hormone levels (LH, FSH, SHBG) were quantified using the Siemens IMMULITE® 2000 XPi radioimmunoassay platform (Siemens Healthcare,.). The Hormone Laboratory at Haukeland University Hospital is accredited according to ISO 15189 standards. The hormones analysed and described in this thesis include testosterone, LH, FSH, and SHBG. Assay performance in terms of coefficients of variation (interassay CV%) and limits of quantification are as described in the Method section in Paper III.

3.6 Anthropometric measurements and bioelectrical impedance analysis

All anthropometric measurements were taken by the same observer. Height was measured using a Harpenden Portable Stadiometer (Holtain Ltd, Crosswell, UK). The boys were asked to stand straight, and their height recorded to the nearest 0.1 cm. Weight was measured using Tanita MC-780MA electronic scale (Tanita Corporation of America Inc., Arlington Heights, IL, USA), with subjects in their underwear only. BMI was calculated as weight in kilograms divided by height in metres squared.

SSF was measured on the left side using a Holtain Tanner/Whitehouse Skinfold Caliper® (Holtain Ltd). The skinfold was picked up with two fingers inferomedially, just below the inferior angle of the left scapula. The caliper was placed about 1 cm below the edge of the fingers, and the measurement recorded to the nearest 0.1 mm after 2–3 seconds of full pressure from the caliper.

WC was measured at the narrowest level between the 10th rib and the iliac crest using a non-elastic metal measuring tape. If no 'narrowest level' could be identified, the level halfway between the 10th rib and the iliac crest was used as reference. Measurements were recorded to the nearest 0.1 cm at the end of normal expiration. WC measurements were performed according to the same protocol used in BGS1 (174).

For BIA, the Tanita MC-780MA body composition analyser (Tanita Corporation of America Inc.) was used. The boys were instructed to stand with bare feet on the device platform and to hold the handgrips/electrodes, one in each hand, until the %BF is displayed.

3.7 Questionnaire

A parental questionnaire was distributed to all participating boys enrolled in the main study. Of the 457 boys included in the main study population, 340 (74.4%) completed and returned the questionnaire. The questionnaire included items such as parents' country of origin and history of chronic disease and previous genital pathology. It also included information about parental educational level, which was categorized into 'no secondary education' (only primary school), 'secondary education' (e.g. senior high school), or 'higher education' (college or university).

3.8 Quality control

All measuring equipment were checked and calibrated every morning over the data sampling period. The exact same equipment and methodology were used in BGS2 as in BGS1 for data sampling. BGS1 reported the technical error of measurement (TEM) to be 0.28 cm for height, 0.80 cm for WC, and 0.64 mm for SSF (175).

3.9 Statistical analysis

Statistical analyses used on data presented in all four papers were carried out using IBM Statistical Package for the Social Sciences (SPSS) versions 24 and 25 (IBM Corp., Armonk, NY, USA), R version 3.5 (R Foundation for Statistical Computing, Vienna, Austria), and GraphPad Prism v7 (GraphPad Software, San Diego, CA, USA).

This section gives a detailed overview of the main statistical methods used. Descriptive statistics are reported as means with 95% CIs and SDs for continuous data, and as frequencies and percentages for categorical data. Further details on the statistical methods used are described in the respective papers.

The z-scores for anthropometric measures (height, weight, BMI, WC, and SSF) were calculated according to the Norwegian growth references, which are based on data collected in BGS1 in 2003–2006 (139,145,148). The z-scores for %BF were calculated using the reference described by McCarthy *et al*. (176).

3.9.1 Observer agreement

3.9.1.1 Bland–Altman plots

Bland–Altman plots were used to assess agreement between two continuous measurements of the same quantity. This can be repeated measurements by the same observer (repeatability) or by different observers (reproducibility), or measurements using different equipment or methods (177,178). The plot shows the difference between two measurements on the *y*-axis and the mean of both measurements on the *x*-axis, and thus allows to visually assess if the measurement variance is homogeneous (equal distribution) and unbiased (no trend). In the absence of a trend, a one-sample *t*-test can be used to assess if the mean difference of the measurements is significantly different from zero, which would be indicative of systematic bias. Likewise, linear regression can be used to detect a bias that depends on the measurement. The measurement variation is quantified by the 95% limits of agreement (LOA), which is the mean difference ± 1.96 times the SD of the differences. These are usually marked as horizontal lines on the plot. The LOA indicate between which extremes 95% of the differences are located (178), and thus show by how much a repeated measurement by the same observer or by another observer or with a different method can differ. If the 95% LOA are sufficiently narrow, one can conclude that the observers or methods agree sufficiently to be used interchangeably (179). For TV, a clear dependency of the measurement variation on the mean volume was observed, and the differences (d) between measurements, either by the same or different observers or methods, were therefore expressed as percentages of the mean TV: $\%d = 100 \times (TV_1 - TV_2)/[(TV_1 + TV_2)/2]$. The mean of %d was used as a measure of systematic bias, and the SD of %d, denoted as $s_{\%d}$, as a measure of variability expressed as a percentage. Twice the $s_{\%d}$ indicates how far a measurement can be from the true value and is an index of reliability and is

considered acceptable in our study if <15%. In addition to the $s_{\%d}$, the TEM was also calculated. While the $s_{\%d}$ shows the variability between two measurements or methods, both of which are prone to measurement error, the TEM shows the variability of a single observation due to measurement error. Since both $s_{\%d}$ and TEM are calculated relative to the measurement, they can be interpreted as a CV.

3.9.1.2 Derivation of a formula to convert ultrasound measured testicular volume to Prader orchidometer volume and vice versa

The Bland–Altman plots of TVs measured by Prader orchidometry and ultrasound showed a tendency of Prader orchidometry to overestimate the volume of small prepubertal testicles. This overestimation decreased with increasing TV and levelled off at around a TV of 10 mL. This relationship was further explored, and a logarithmic transformation of both variables showed a clear linear trend with constant variance (Figure 6).
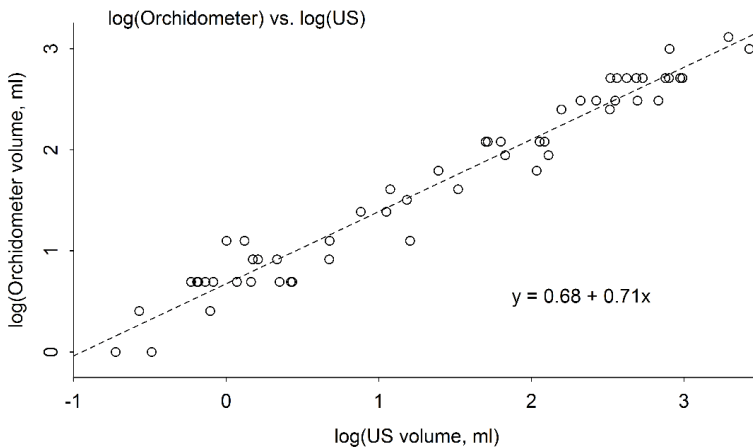
Figure 6 Linear regression of the logarithm of testicular volumes measured by Prader orchidometry versus the logarithm of testicular volumes measured by ultrasound ($L \times W \times D \times 0.71$). The variance is homogeneous across the entire range of fitted values. US, ultrasound.

A back-transformation of the log–log regression model coefficients generated the equation to predict the equivalent orchidometer volume based on ultrasound measurements on the measurement scale. The formula is log(Prader orchidometer volume) = 0.68 + 0.71 × log(US volume); residual SD = 0.18 on the log scale, which translated to (orchidometer volume = 1.96 × US volume$^{0.71}$) on the measurement scale.

3.9.2 **Pubertal references**

3.9.2.1 The LMS method

A growth reference is a statistical summary of anthropometry or another continuous variable in a reference sample of children, usually presented as a contiguous frequency distribution at consecutive ages (180). The reference sample is usually representative of a specific geographical region at a particular time. Data collection is usually cross-sectional. The distribution is usually summarized by the mean and SD, or alternatively by the median and selected percentiles, by age and sex. The LMS method is currently one of the most used methods for growth curve estimation (180-182). It was specifically designed for data that are not normally distributed, as in the case of the TV. The TV distribution was summarized by three smoothed curves: the L-curve representing the Box-Cox power transformation needed to convert the data to a normal (Gaussian) distribution at each age; the M-curve, or median curve, by age; and the S-curve which is the approximate CV (SD/mean). The information contained in these three curves allows to calculate the distribution of a measurement at a given age (i.e. the 'growth curve') and to convert a measurement into a z-score or percentile. The amount of smoothing of the LMS curves is expressed in terms of equivalent degrees of freedom (edf). For TV measurements described in Paper II, 8 edf were used for the M-curve and 4 for the S-curve. The optimal Box-Cox power L was determined to be constant at 0.5.

3.9.2.2 Probit regression

Age references like those for pubertal development (e.g. reaching TVs corresponding to selected discrete Prader orchidometer volumes and for each of the Tanner PH

stages) require a different approach because although it is known that the event has occurred, the event itself cannot be observed. For these references, cumulative incidence curves were estimated using probit regression within generalized linear models (GLMs) which assume a normal distribution of ages, and with generalized additive models (GAMs) which are non-parametric. Since using both models gave identical results, data were reported from the GLM probit models, allowing to summarize the age distribution by the mean and SD.

3.9.3 **Endocrine references**

The Clinical Laboratory Standards Institute (CLSI) has provided a protocol for establishing reference intervals that meet the minimum requirements for reliability and usefulness (183). Reference intervals should ideally include 120 or more observations in each partition, but a minimum of 40 observations are tolerated when robust statistical methods are used. To assist in deciding where and whether to partition reference intervals (e.g. between two adjacent age intervals), Harris and Boyd suggested calculating the statistical significance of the difference between subgroup means using a pairwise standard normal deviate test (184). This test was applied in the work presented in Paper III to justify partitioning of reference intervals. A minimum of 40 observations were included, and as many as 120 when possible. The Harris and Boyd approach considers sample size, mean analyte value, and variance in consecutive partitions, and the difference in distribution between two partitions is expressed as a z-score which is compared to a critical value. Adjacent partitions were considered justifiably separated if the Harris–Boyd z-score exceeded the corresponding 'critical' z* which penalizes the result for low n observations by the formula $z^* = 3(n_1 + n_2/120)^{1/2}$ (185). For these analyses, log-transformed data were used to achieve an approximate Gaussian distribution. When the sample size was ≥120, the non-parametric method was used to calculate the central 95% reference intervals and corresponding 90% CIs of the lower limits (LL, i.e. 2.5th percentile) and upper limits (UL, i.e. 97.5th percentile). The robust method was used when the sample size was between 40 and 120. Non-parametric estimations were based on the

binominal distribution of observation ranks, whereas the robust method was based on 500 bootstrapped samples.

3.9.4 **Association analyses**

In the work presented in Paper IV, we classified the boys based on their testicular volume-for-age z-score (USTVz) as early, average, or late maturing. Those boys in the upper tertile (>67th percentile) were considered early maturing, those in between percentiles 33 and 67 as average, and those with the smallest TV for age (<33rd percentile) as late maturing. Further, we stratified the boys into three groups, according to their BMI-, WC-, SSF-, and %BF-for-age z-scores to assess the effect of adiposity and body composition on the timing of puberty and degree of maturation. A z-score below −1 was considered as low, and a z-score above 1 as high.

3.9.4.1 Logistic regression

Logistic regression is a suitable method to analyse the association between a dichotomous outcome variable and one or more categorical of continuous predictor variables. The results are expressed as an odds ratio (OR), usually reported with 95% CI. An OR above 1 indicates a higher probability of occurrence, whereas an OR below 1 indicates a lower probability. We estimated the age-adjusted odds ratio (AOR) for having reached a pubertal level of either TV, serum testosterone, or pubic hair, using a high (>1) or low (<−1) z-score of the different anthropometric measurements (BMI, WC, SSF, and %BF) as a predictor. A z-score in the normal range (between −1 and 1) was used as the reference category, to which boys with a high (>1) or low (<−1) z-score were compared. If the CI excludes the value 1, the difference between groups is considered as statistically significant.

3.9.4.2 Proportional odds regression

The proportional odds regression model is an extension of logistic regression for ordinal dependent variables. The advantage is that a single OR is estimated, independent of which value of the ordinal variable is used as cut-off, but the validity depends on the assumption of a comparable OR for each cut-off (proportional odds

assumption). It was possible to use this model since the OR for late-maturing boys (ultrasound-determined testicular volume (USTV) <33rd percentile) versus average- or early-maturing boys (USTV >33rd percentile) was comparable to the OR for late- or average-maturing boys (USTV <67th percentile) versus early-maturing boys (USTV >67th percentile). Using this regression model, the association between the level of maturity (early, average, or late, based on the USTVz) and the grouped anthropometric measurements was studied, comparing boys with a 'low' (<−1 SD) or 'high' (>1 SD) value to those with an average value for each measure separately.

3.9.4.3 Cumulative incidence curves

To show the mean ages for reaching pubertal onset (USTV ≥2.7 mL, serum testosterone ≥0.5 nmol/L, and Tanner stage PH2) according to weight class (BMIz <−1, −1< BMIz <1, and BMIz >1), cumulative incidence curves were plotted. The curves were estimated using a GAM with a binary outcome and probit link function (see Section 3.9.2.2). The degree of smoothing was determined with generalized cross-validation using the 'mgcv' package in R. The mean age at reaching maturity (USTV 2.7 mL) was obtained by inverse prediction.

3.9.4.4 ROC curves

To determine the serum level of testosterone that marks the onset of puberty (presented in Paper IV), receiver operating characteristic (ROC) curve analysis was performed. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 − specificity) of different cut-off values. Values in the upper left corner of the curve have both a high sensitivity and a high specificity, which is usually the desired result. The optimal choice to discriminate between prepubertal and pubertal boys was determined using the Youden index as 0.5 nmol/L (Figure 7).
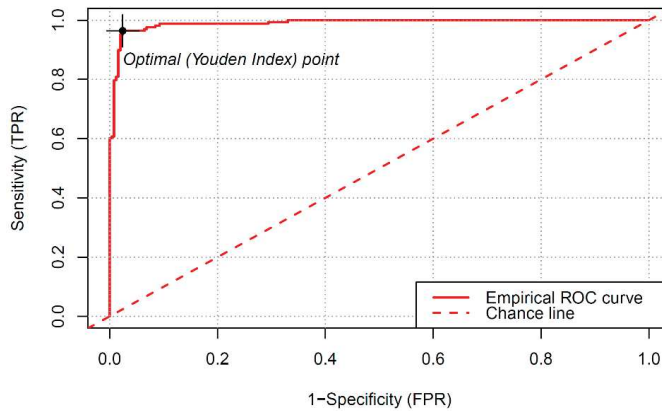
Figure 7 The pROC package in R was used to construct ROC curves, using total serum testosterone levels as a biomarker of pubertal onset. Boys were dichotomized as prepubertal (TV <2.7 mL) or pubertal (TV ≥2.7 mL), and the pROC algorithm was used to retrieve the optimal confusion matrix and corresponding level of testosterone to distinguish between the two groups. FPR, false positive rate; TPR, true positive rate.

3.9.5 Power calculations

The planned sample size of BGS2 was 1000 boys (about 100 per year of age) in order to estimate the mean (median) age at transition from one pubertal stage to the next (e.g. from PH1 to PH2, or reaching pubertal TV), with a standard error (SE) of ~1 month and the normal limits with an SE between 2 and 3 months, assuming an SD of about 1 year. Due to a response below expectation, logistic challenges, and time constraints, only half of the boys were measured by the end of the study period. This increased the SE of the mean to 1.5 months for the mean age, and 3 months for the normal limits, which was still considered as acceptable.

3.10 Ethics

This study was approved by the Norwegian Regional Committee for Medical and Health Research Ethics West (REC-WEST 2015/128). The study also was conducted in accordance with the Declaration of Helsinki (186). Written informed consent was obtained from a parent or the legal guardian of each study participant, as well as assent from the participants themselves. All children received age-appropriate

information in writing, as well as verbally, from the main observer prior to their participation. A cinema voucher was given as an incentive.

4. Summary of results

4.1 Paper I

Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys—intra- and interobserver agreement and comparison with Prader orchidometry

Paper I describes the methodological study aimed at estimating the intra- and interobserver agreement of ultrasound measurements of TV and comparing use of ultrasound with Prader orchidometry in TV measurement. The mean age of the 57 participating boys was 12.0 (range, 6.5–16.4) years. As the degree of measurement variation increased with mean TV, the differences between measurements, observers, and methods were reported as relative differences. Intra-observer agreement, which is the measure of repeatability, showed a mean difference (bias) of −2.2% ($p = 0.08$), which indicated no systematic bias. The corresponding 95% LOA ranged from −20.3% to 15.9%, with a variability of 9.2% and a TEM of 6.5%. The differences were <15% in 89% of measurement pairs.

Interobserver agreement, which is the measure of reproducibility, showed a small bias of 4.8% ($p = 0.052$). The 95% LOA were wider, compared to the intra-observer agreement, and ranged from −35.7% to 45.3%, with a variability of 20.7% and a TEM of 14.6%. The index of variability ($2* s_{\%d}$) thus exceeded the 15% limit that was set as acceptable a priori.

Comparison of TV measurements using ultrasound versus Prader orchidometry revealed that the overall mean and SD of TVs measured with a Prader orchidometer were highly comparable to the corresponding ultrasound measurements. However, when plotted against the mean volume, there was a clear tendency towards larger volumes with the Prader orchidometer in prepubertal boys, although less so in larger testicles (Figure 8a). The relationship between ultrasound-measured TV and TVs measured by Prader orchidometry was further explored on the log scale. A corresponding equation to predict Prader orchidometer volume from ultrasound volume on the measurement scale was derived as: $Vol_{OM} = 1.96 \times Vol_{US}^{0.71}$. Consequently, a TV of 2.7 mL measured by ultrasound corresponded to a Prader

orchidometer volume of 4 mL which marks pubertal onset. The Bland–Altman plot of TVs measured with the Prader orchidometer and the corresponding TVs predicted from the ultrasound measurements is shown in Figure 8b.

(a)

(b)
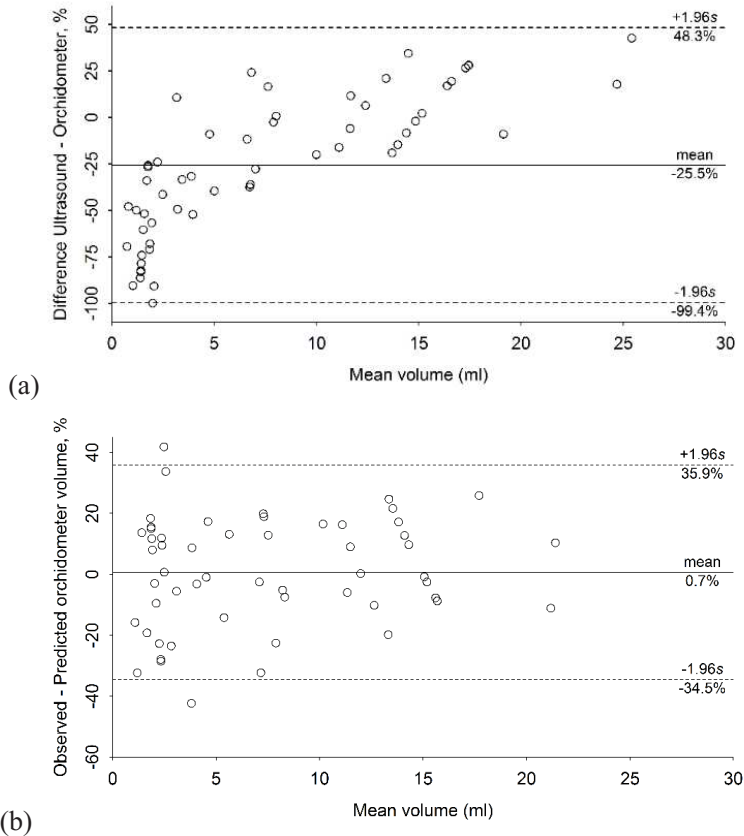
Figure 8 Bland–Altman mean difference plots of testicular volume in 56 Norwegian boys aged 6.5–16.4 years, as measured with (a) ultrasound (US) ($L \times W \times D \times 0.71$) and with the Prader orchidometer (OM) and (b) measured with the OM and the corresponding volume predicted from the US measurements using the equation $Vol_{OM} = 1.96 \times Vol_{US}^{0.71}$. Differences between methods are expressed as percentage of the mean [$100 \times (US - OM)$/mean volume] or [$100 \times (observed - predicted)$/mean volume], since the variance increases with volume. Plot (a) shows a clear upward trend and very wide limits of agreement, whereas plot (b) shows no bias and narrower limits of agreement and an overall SD of differences of 18% that decreases slightly with volume. Horizontal lines indicate the mean difference and 95% limits of agreement (mean $\pm 1.96 \times$ SD). With permission, from Paper I, Oehme NHB, Roelants M, Bruserud IS, *et al*. Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys—intra- and interobserver agreement and comparison with Prader orchidometry. *Pediatr Radiol*. 2018;**48**:1771–8'

4.2 Paper II

Reference data for testicular volume measured with ultrasound and pubic hair in Norwegian boys are comparable with Northern European populations

Paper II presents pubertal references for boys living in Norway. The continuous reference curve of TV for age was estimated using the LMS method. This curve and the corresponding age references were based on 514 boys with a mean age of 11.0 (range, 6.1–16.4) years. Pubertal onset was defined as the achievement of a USTV of ≥2.7 mL in at least one testicle, which corresponds to a TV of ≥4 mL when measured with the Prader orchidometer. Tabulated values of L, M, and S for age (see Section 3.9.2.1 for detailed description) are presented, giving all information needed to calculate any percentile or to convert the measurements into z-scores. The mean age (SD) for attainment of a USTV of 2.7 mL was 11.7 (1.1) years, and the 3rd and 97th percentiles were 9.7 and 13.7 years, respectively. In addition, cumulative incidence curves for reaching selected discrete Prader orchidometer volumes are also presented (Figure 9).
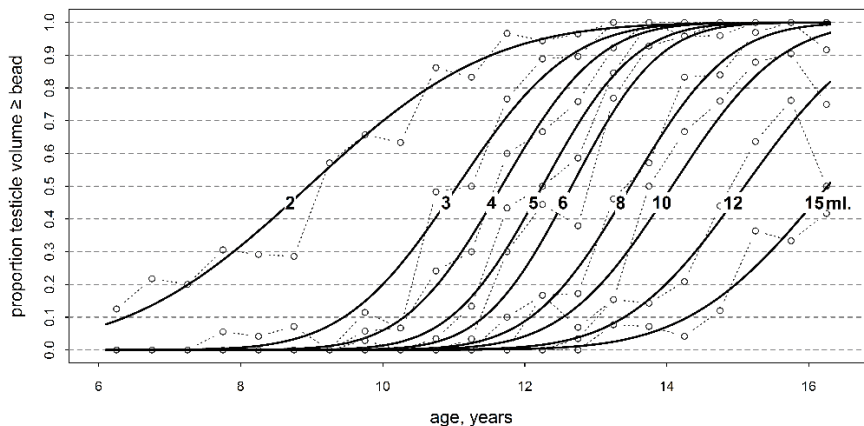
Figure 9 Cumulative incidence of reaching selected equivalent Prader orchidometer volumes estimated with probit regression in 514 healthy Norwegian boys aged 6–16 years. Connected markers on dotted lines show the empirical data, and bold lines the corresponding probit models. With permission, from Paper II, Oehme NHB, Roelants M, Bruserud IS, *et al*. Reference data for testicular volume measured with ultrasound and pubic hair in Norwegian boys are comparable with Northern European populations. *Acta Paediatr*. 2020;**109**:1612–19'

The pubertal reference for pubic hair development was based on 452 boys with a mean age of 10.9 (range, 6.1–16.3) years. The mean age (SD) of pubarche (Tanner stage PH2) was 11.8 (1.2) years, with the 3rd and 97th percentiles of 9.5 and 14.1 years, respectively. Further, more boys achieved pubertal TV (≥ 2.7 mL) before they developed pubic hair (Tanner stage PH2), compared to boys who developed pubic hair as the first sign of puberty (14.0 versus 8.1%, respectively).

Further, there was no indication that Norwegian boys entered puberty earlier than boys from comparable European countries.

4.3 Paper III

Testicular ultrasound to stratify hormone references in a cross-sectional Norwegian study of male puberty

In Paper III, references for circulating serum levels of total testosterone, LH, FSH, and SHBG are presented in relation to ultrasound-measured TV, Tanner PH stages, and chronological age. Serum levels of total testosterone were assayed by LC-MS/MS, while the levels of LH, FSH, and SHBG were determined using immunoassays. These endocrine endpoints were available for 414 of the 457 boys from the main study.

In pubertal boys (USTV ≥ 2.7 mL), the TV accounted for more variance in serum testosterone levels than age (Spearman $r = 0.75$, $p < 0.001$ versus $r = 0.69$, $p < 0.01$). The Harris–Boyd test recommended stratification of reference intervals by pubertal status based on the TV in the age window of puberty transition (10–13 years) despite overlapping ages. This generated two separate and statistically acceptable reference intervals.

The LH:testosterone ratio was included as a marker of Leydig cell function, and the transition from prepubertal (USTV < 2.7 mL) to pubertal TV (USTV ≥ 2.7 mL) was found to be characterized by a unidirectional ratio shift. The mean age of pubertal onset estimated to be 11.7 (SD 1.1) years in Paper II agreed very well with the age at which serum testosterone levels started to increase.

4.4 Paper IV

Low BMI, but not high BMI, influences the timing of puberty in boys

Paper IV presents findings on the association between pubertal onset (defined as pubertal TV (USTV ≥2.7 mL), or serum testosterone levels of ≥0.5 nmol/L, or Tanner stage PH2) and anthropometric measures (BMI, WC, SSF) and %BF in 324 boys aged ≥9 years (including 180 pubertal boys). The boys were also stratified according to the degree of pubertal development based on USTVz. Boys with a low BMI (<−1 SD) were found to have a lower probability of having pubertal TV (USTV ≥2.7 mL), compared to boys with average BMI (AOR 0.3; 95% CI 0.1, 0.9; $p = 0.038$). Boys with a high BMI (>1 SD), on the other hand, did not have a significantly higher probability of having pubertal TV (AOR 1.3; 95% CI 0.4, 3.9; $p = 0.691$). Similar associations were also found for low and high WCs, with an AOR of 0.2 (95% CI 0.0, 0.6; $p = 0.008$) and 0.9 (95% CI 0.3, 2.9; $p = 0.918$), respectively. A similar trend was observed for low SSF and low %BF, with a lower probability of having reached any of the markers for pubertal onset, although none of these associations were statistically significant. Cumulative incidence curves for pubertal TV showed that boys with a low BMI entered puberty (USTV ≥2.7 mL) at a mean age of 12.34 years versus a mean age of 11.66 years in normal-weight boys, i.e. a puberty delay of about 8 months. Further, no significant associations were found between Tanner stage PH2 and BMI, WC, SSF, or %BF z-scores.

Ordinal logistic regression also showed that boys with low BMI or low WC for age had a significant lower probability of being in a higher category of testicular volume-for-age (USTVz >33rd percentile, average, or early maturing) compared with those with average BMIz (OR 0.3; 95% CI 0.2, 0.5; $p < 0.001$) or WCz (OR 0.2; 95% CI 0.1,0.4; $p < 0.001$). However, boys with high BMI or high WC for age did not have an increased probability of being in a higher category of testicular volume for age, as a sign of being more mature.

5. Discussion

Results presented in this thesis show that ultrasound is a reliable method for TV measurement, allowing for the construction of reference intervals by age. Novel pubertal references for boys in Norway were estimated, with age references for TV and pubic hair development. Puberty timing in Norwegian boys did not differ from that in boys from other Northern European countries. In addition, USTV and age were used to stratify serum testosterone levels and other key hormones of the male pituitary–gonadal pathway, yielding novel reference intervals for pubertal hormones in relation to TV. Finally, the studies showed that low BMI and WC for age were associated with a delay in pubertal onset, whereas no significant change in puberty timing was observed in boys with higher BMI or WC for age.

5.1 Methodological considerations

Bias is any systematic deviation from the truth, whether in data collection, analysis, interpretation, or publication. Bias in research can occur either intentionally or unintentionally. When estimating references, the study sample should be representative of the target population. A population-based sample is preferred where no subgroup is more or less likely to participate. However, this may be challenging, especially when investigating children and particularly when studying pubertal development.

5.1.1 Childhood populations

5.1.1.1 Participation rates

The overall participation rate in the main study was 37.1%. A somewhat higher participation rate was observed among primary school boys aged 6–12 years (39.6%) than among boys attending middle school (30.8%). A total participation rate of about 37% might seem rather low, but this is comparable to, and even somewhat higher than, other studies on pubertal assessments, with a 35% participation rate in the 1991–1993 Copenhagen Puberty Study cohort and 24.7% in the 2006 cohort (79). The participation rate in BGS2 was lower, compared to BGS1 (70% among primary

school children and 55% in middle school), but this may be attributable to the different examinations involved in these two studies. While BGS1 focused on normal growth, BGS2 focused on pubertal development, including assessment of the external genitals, which, obviously, could be an embarrassing experience to potential participants. Further, it might be that boys maturing early or late, compared to their peers, were less inclined to participate.

The prevalence of boys with overweight and obesity in BGS2 was 11.8% and 1.9%, respectively, as defined by the IOTF. This matches the 12.5% and 2.1% prevalence rates, respectively, reported for Norwegian boys in the same age range in BGS1 from 2003 to 2006 (174). In addition, a Norwegian population-based study of 8- to 9-year-old boys found a similar prevalence of overweight and obesity of 13% (including a rate of 2% for obesity) in 2015 (187). Based on these prevalence rates, the study sample described in this thesis is considered to be representative of the Norwegian population of boys.

5.1.1.2 Age range

The age range of the study participants was 6–16.4 years. In other words, the study participants included boys attending either primary or middle school. Inclusion of boys aged from 6 years onwards ensured proper coverage of the prepubertal period. Unfortunately, not including older boys meant the final growth period was missed out and not all boys in the oldest age group had yet reached the final pubertal stages. A study from Belgium showed a median height of 175.8 cm at 16 years, while the median height at 21 years was 181.0 cm (86), demonstrating continued growth and development beyond age 16 years. Without doubt, more precise estimates, and a better fit of the curve near the upper age range would have been obtained if more older boys were included in the study presented here. Likewise, it would have been preferable to increase the accuracy of age estimates for larger TVs and Tanner stages PH4 and 5. However, including older boys required data sampling in high schools; due to limited financial resources, priority was focused on covering the ages around pubertal onset, as it was felt this was more important than examining the final adult stages.

5.1.1.3 Socio-economic status

A survey on living conditions in 2016 found that living conditions in the municipality of Bergen was overall very similar to the rest of the Norwegian population in terms of its socio-economic structure (188). The seven schools from which the participants were recruited were randomly selected after stratification based on town area, and covered various socio-economic environments, although they were all located in mostly urban areas of the municipality of Bergen. Parental educational level was used as a measure of socio-economic status, as this has been shown to be a valid variable for this purpose (189). In our study, among boys who provided information about parental education level (336 of the 340 who completed the study questionnaire), as many as 81.5% reported to have at least one parent with higher education (college or university degree). This is an overrepresentation, compared to the general population in Bergen with a prevalence of higher education of ~40% in 2019 (190). Such overrepresentation in health-related surveys is well documented (191,192). However, logistic regression analysis of pubertal onset (yes/no) according to age and parental education level (college/university, yes/no) did not show any statistically significant differences ($p = 0.159$).

Further, there were no significant difference in the prevalence of overweight or obesity based on parental education level (9.3% with overweight/obesity among boys whose parents did *not* have higher education versus 8.4% in the group whose parents *did have* higher education). However, among those children with no information provided on parental education level, 20.7% were classified as overweight or obese, based on IOTF classification. Further, there were no statistically significant differences in BMI z-scores between native Norwegian boys and their non-Norwegian counterparts ($p = 0.53$).

5.1.1.4 Country of origin

Information about country of origin was also gathered in the questionnaire survey. Among those boys with a known country of origin, 77.4% had both parents from Norway, 10% had one or both parents from another European country, and 12.5%

had one or both parents from a non-European country. This is a fair reflection of the demographic composition of the general Norwegian population in 2016 (6% European and 10% non-European) (193). Previous studies have shown that adult TV differ between ethnicities with larger testes in Caucasian, compared to Asian, populations (194,195). However, testicular size in young boys aged 0–6 years has not been shown to differ between ethnic groups (56). Further, studies from the USA have demonstrated earlier achievement of pubertal TV (≥4 mL) in white boys, compared to African American boys (11.46 versus 11.75 years, respectively), whereas achievement of pubarche and Tanner stage PH2 seems to occur earlier in African American boys (83,87). The population sample in the study presented here was recruited with the designated aim to assess pubertal development in healthy boys living in Norway and to construct pubertal references for clinical use for all boys. Logistic regression did not indicate any significant differences when comparing age at attainment of pubertal TV or pubarche in native Norwegian versus non-European immigrant boys ($p = 0.11$ and $p = 0.59$, respectively), and all boys regardless of country of origin were therefore included.

5.1.1.5 Chronic illness, scrotal pathology, and prematurity

Information about chronic illness and previous scrotal pathology were collected through the questionnaire survey. Thus, a total of eight boys (1.6%) with an illness or condition that could affect growth and development were excluded from the study. However, it is possible that boys with a chronic illness were included in the study if they did not answer the questionnaire.

A total of 20 boys were excluded due to past or current history of scrotal pathology. On examination, four boys were found to have bilateral, and 11 with unilateral, cryptorchidism. In addition, the testicle was found in the inguinal canal in two boys. In a cross-sectional study with >1500 school-aged boys (aged 7–12 years), the prevalence of cryptorchidism and retractile testis was 0.73% and 3.9%, respectively (196). In the study presented here, taking together these two conditions resulted in a prevalence of 3.6%. It is likely that many of the study participants recorded as having either bilateral or unilateral cryptorchidism probably had retractile

testes, but due to time constraints and the limited clinical experience of our main observer, these testes were not located.

From the questionnaire survey, 22 of the boys in the main cohort were born before gestational week 37 or had a birthweight of <2500 g. A recent review of the association between preterm birth and puberty timing concluded that premature birth did not lead to a significant acceleration in the onset of puberty (197). Logistic regression analysis did not indicate any significant differences in age at attaining pubertal TV when comparing boys born preterm with those born full term ($p = 0.391$). Boys with known premature birth were therefore included in the study.

To summarize, participating boys in this study are considered to be representative of the current population of children and adolescents in Norway, with respect to weight status and origin. Overrepresentation of boys with parents with higher educational level did not seem to influence the study results. Taken together, these are strong indications that findings from this study may be generalizable to the rest of the Norwegian male population.

5.1.2 Pubertal assessment

5.1.2.1 Ultrasound examination

For the test–retest study presented in Paper I, all boys were examined with ultrasound twice by the main observer and once by a second observer. While an experienced examiner should acquire a high degree of precision and accuracy over time, less experienced examiners will probably produce greater variability in their ultrasound measurements. Since there were only two observers in the study presented here, it is possible that the expected discrepancy in volume estimates between any pair of observers is considerably higher or lower than described here. Analysis based on replicated measurements from more than two observers, each with different experience and training, would, of course, provide a better estimate of both intra- and interobserver agreement (198). However, such a study with several observers would be costly and time-consuming, and not least be an additional burden for the participating boys.

As described in Paper I, the time interval of at least 20 minutes between the first and second ultrasound examination performed by the main observer included examination of at least three boys, but in some cases as many as five or six other boys. It may be argued that this was not sufficient to mitigate the risk of recall bias for the intra-observer agreement. However, the TV was calculated at a later stage, and recall after a time interval of >20 minutes would have involved recalling three separate dimensions, rather than a single volume (i.e. the TV). Therefore, the risk of recall bias was considered to be small.

Standardization of methods is always necessary, so that a method is independent of a single observer. Interobserver variations may be, in part, due to different techniques used, e.g. the amount of pressure applied onto the testis with the ultrasound probe. In this study, the main observer, on average, measured greater widths and depths and shorter lengths, compared to the second observer, indicating that the main observer did not compress the testicle as much as the second observer. This could be overcome by better method standardization. In addition, use of a gel-pad or application of generous amounts of coupling gel might help eliminate some errors arising from individual differences in pressure applied during examinations.

In the work presented in Paper II, measurements were taken only for the right testicle, except in three cases where both testicles were measured since the left testicle appeared visually larger. On measurement, the left testicle was found to be larger in only one of the three cases. In the literature, the consensus is that there is no significant difference between the left and right TVs in healthy males (199,200). However, one study found that one out of five healthy boys aged 11–16 years with no scrotal pathology had smaller left TVs (201). Therefore, the choice of examining only the *right* testicle seems justifiable. Further, although interesting, measuring both the left and right TVs in >500 boys would add considerably to the workload for the observer and represent an additional burden for the boys without necessarily providing any reasonable advantage for data analysis.

The decision to use the Lambert formula (57) for TV calculations was based on several studies showing that this formula is more accurate than other commonly

used formulae, when calculated TVs are compared to true TVs measured by water displacement (50,59).

5.1.2.2 Tanner staging

The assessment of pubic hair development described in Papers II, III, and IV was based on Tanner staging (46). The fact that the original photographs by Marshall and Tanner, together with the description of the different stages, were always made available to the observer likely enhanced the validity of the assessments. Assigning a Tanner stage for genital (G) development has been reported by others to be difficult and more subjective (47,202), with significant interobserver variation (52). In the study presented here, assigning the participating boys with a Tanner G stage, albeit very interesting and desirable, would have added to the burden of exposure for the boys during examination, which would have likely led to recruitment difficulties. In addition, the main observer only had minimal experience in Tanner genital staging. An alternative to clinical assessment could be collecting self-reported pubertal data. However, there is strong evidence that direct assessments by health professionals provide the most accurate accounts from subjects (202,203).

As described in Paper I, TV measurements with the Prader orchidometer were performed by a highly experienced paediatric endocrinologist. This process thus ensured blinding to the ultrasound measurements carried out by the main observer. Using orchidometry to assess male pubertal development may be prone to inaccuracy due to subjectivity of this method of assigning the assessor's impression based on visual inspection and palpation to a set of numbered beads, each representing different testicular sizes. Studies have shown that the Prader orchidometer systematically overestimates small TVs, probably due to the inability of the instrument to differentiate the central testicle from surrounding tissues, e.g. the epididymis, scrotal skin, and the tunica vaginalis (55,58,204,205). However, studies have reported a high correlation ($r = 0.85$) between TVs obtained by orchidometry and ultrasound volumes (205). Nevertheless, the measurement variation was found to increase with testicular size (16), thereby showing poor agreement between the two methods of TV measurement.

In this study, a mean age of pubertal onset (defined as a Prader orchidometer volume ≥4 mL, using the conversion formula presented in Paper I, $Vol_{OM} = 1.96 \times Vol_{US}^{0.71}$) of 11.7 years was obtained, which was comparable to Joustra *et al.*'s result of 11.6 years in Dutch boys (15). The authors assessed the TV both with ultrasound and by Prader orchidometry and found a mean USTV of 1.4 mL (calculated using the formula for an ellipsoid $L \times W \times D \times 0.52$) at 11.6 years when the boys reached an orchidometer volume of 4 mL. In the study presented here, the results were compared with these Dutch references (60) by multiplying their TVs by a factor of 0.71/0.52, which showed highly comparable mean TVs for each age group between both studies (not shown). Therefore, this supports the study findings here that ultrasound is a useful method for construction of growth references of TV.

5.1.3 Blood sampling and analyses

Blood sampling was carried out at participants' individual schools during school hours, as shown in Table 2. The children were not requested to fast in advance. It is known that diurnal variation in testosterone secretion is more pronounced in early and mid-puberty (206). However, we believe that the diurnal variation's influence on the results and conclusions is minimal because intra-individual diurnal fluctuations from early morning to afternoon are negligible, compared to interindividual variations (71). Another similar study also collected blood samples between 08:00 and 13:00 and provided references for FSH, LH, and testosterone from their data (207). Further, the reference intervals published from the highly regarded CALIPER study did not take diurnal fluctuations of gonadotropins or testosterone into consideration and fasting was not required (73). Therefore, for both clinical and academic purposes, the study presented here strongly support that intraindividual fluctuations can be disregarded when data are collected in a sufficiently large cohort for the purpose of constructing references.

Table 2 Blood draw times for indicated amounts of samples

Sample time (hour:min)	Number of prepubertal samples (USTV <2.7 mL)	Number of pubertal samples (USTV ≥2.7 mL)	Total number of combined samples (prepubertal and pubertal) (%)
08:20 to 09:00	30	9	39 (9.5%)
09:01 to 10:00	71	29	100 (24.5%)
10:01 to 11:00	57	38	95 (23.3%)
11:01 to 12:00	35	19	54 (13.2%)
12:01 to 13:00	41	38	79 (19.4%)
13:01 to 14:00	16	22	38 (9.3%)
14:01 to 15:00	0	3	3 (0.7%)
Average time of blood draw (hour:min)	10:41	11:19	10:57

An important consideration from the CLSI guidelines is the formalized requirement of $n \geq 120$ observations to construct each valid reference interval for clinical use. However, although partitions should ideally be defined by ≥ 120 observations, a minimum sample of 40 is sufficient to estimate a reference interval with robust resampling (185). In Paper III, continuous centiles were calculated from no fewer than 40 observations in the data set tail end and the 90% CIs associated with the 2.5th and 97.5th percentiles were calculated by resampling using *referenceIntervals* package in *R*.

In this study, LC-MS/MS was used, which quantifies the molecular analyte directly from its fragmented mass/charge ratio signature and offers greater analytical sensitivity and specificity required to assess gonadal function and therapeutic drug monitoring in children and adolescents (173,208-210). The traditional immunoassay techniques, however, quantify the hormones indirectly, with particular challenges in terms of analytical specificity and sensitivity, especially in the lowest range, such as in prepubertal children. In this study, the detection limit for testosterone using LC-MS/MS was 0.01 nmol/L. For comparison, the Copenhagen Puberty Study reported a detection limit of 0.23 nmol/L measured by radioimmunoassay (29).

5.1.4 **Anthropometric measurements**

All anthropometric measurements in BGS2 were carried out by the main observer, which limits variability due to interobserver variability. In addition, height, WC, and SSF were measured using identical equipment and methods as in BGS1, from which anthropometric z-scores were estimated.

Good correlation has been shown between the BMI and body fat, but the accuracy of the BMI in predicting overweight and obesity varies with the degree of fatness, with higher accuracy in obese children and lower accuracy in leaner children (137). This may potentially overlook the so-called 'normal-weight obese' children (136). Furthermore, the correlation between the BMI and body fat is much lower in boys than in girls (211). This may be because during male puberty, increasing muscle mass associated with the anabolic effect of rising testosterone levels causes an increase in weight and BMI, independent of any increase in body fat. In this study, WC measurements were also included as a measure of abdominal fat, and SSF as a direct measure of truncal fat. Importantly, as with the BMI, the WC is also regarded only as a proxy for increased adiposity. The WC does not distinguish between central adiposity and muscle mass, meaning that boys with higher muscle mass may have increased WC measurement, partially due to muscular core development, as seen with increasing maturity (212).

As described in Paper IV, BIA was used for evaluation of %BF. BIA is generally considered to be more accurate than anthropometric-based indices such as BMI (213). However, BIA may overestimate %BF in leaner subjects and underestimate %BF in fatter subjects (214). For the calculation of %BF z-scores, references previously published by others were used (176). When calculating z-scores based on measurements in this study, the %BF z-scores for boys aged ≥9 years (not shown) matched very well with the z-scores obtained by McCarthy *et al*. (176).

Based on the study presented here, measuring not only BMI, but also WC, SSF, and %BF, would enable the study of variables of both body shape (BMI and WC) and body composition (SSF and %BF), thus adding a new dimension to this association study.

5.1.5 **Statistical considerations**

5.1.5.1 Bland–Altman plots

Agreement within and between observers and methods was assessed using Bland–Altman plots, as described in Paper I. Previous studies have mainly examined the correlation between the two methods of TV measurements, namely orchidometry and ultrasound (205,215-217). However, using Pearson and intra-class correlation only measures the strength of an association but gives no information about agreement. Correlation does not take into account, for instance, systematic bias between two observers or methods, and consequently, even a high correlation between two methods does not guarantee a clinically acceptable agreement (178,218). Therefore, for the purpose of this study, the Bland–Altmann plot was considered as the appropriate statistical method. Moreover, it provides statistics (e.g. LOA) that are useful in the interpretation of measurement variability in a clinical or research context.

5.1.5.2 GAMs and GLMs

In this study, probit regression was used with a GLM when estimating the cumulative incidence curves for reaching selected TVs and PH stages. To compare the timing of puberty, some studies express their results in terms of the mean ± SD under the assumption that the data are normally distributed. Others report in medians and percentiles. This issue is important, because the mean and median are only comparable when the distribution of data is symmetric. In this study, the GLM probit models were compared to the corresponding non-parametric GAMs as an informal test of normality, which confirmed the assumption of a normal age distribution at different pubertal milestones.

Since this is a cross-sectional study, the individual progression between different pubertal stages will be faster than the intervals between the median ages shown on the probit curves. Likewise, the testicular growth curve of an individual boy will be steeper than the reference curves in this study. Progression of pubertal development is therefore outside the scope of the current work.

5.1.6 **Ethical considerations**

When including children and adolescents in research, ethical concerns inevitably arise. The Declaration of Helsinki emphasizes the importance of voluntary participation in research (186). The intimate nature of this study demanded particular consideration of the ethical aspect of the research. In light of this, all boys were informed that they could withdraw from the study at any time, without any consequences. In addition, when conducting research on such a vulnerable subject group, making sure to include only the necessary number of children (as determined by power calculations) will limit the number of participating children undergoing examination, thereby minimizing the risk of causing unnecessary harm.

5.2 Discussion of results

5.2.1 **Observer agreement**

To estimate the reference values for USTV, it was important to report on observer error. This was done by evaluating the intra- and interobserver agreement for USTV in 57 boys, as described in Paper I. As mentioned previously, reporting agreement using Bland–Altman plots enables detection of any systematic bias, which is not possible with correlation analysis (218). It is well known that accuracy of TV measurement using the Prader orchidometer is dependent on observer experience (49,52). The same issue applies to ultrasound examinations (51). In this study, results showed that intra-observer agreement (reliability) was unbiased, with a variability of 9.2% (SD of the relative differences). This was highly comparable to earlier studies reporting an intra-observer variability ranging between 7.0% and 9% (16,51,56). The intra-observer 95% LOA were −20.3% to 15.9%, or ± 18.4%, which were close to the previously set value of ±15% considered to be acceptable for clinical use. This cut-off value was a clinical decision, rather than based on statistics. Analysis of interobserver agreement (reproducibility) showed that the measurement of TV volume with ultrasound was slightly biased, with a mean difference of 4.8%. In contrast to the acceptable intra-observer agreement, the agreement between the two observers in this study showed a variability of 20.7%, which was higher than the CVs of 11.7% and 15% reported in earlier studies (16,56). This may be explained partly by the fact that

Kuijper *et al.* tested the observer variation in boys aged 0–6 years (56), which limited the variation in testicular size, and that Sadov *et al.* used the formula of an ellipsoid ($L \times W^2 \times 0.52$), which does not include depth measurement, thus disregarding the potential measurement error of this variable (16). Results from both the study presented in this thesis and Sadov *et al.*'s study found that the between-measurement variation clearly increased with testicular size (16). The TV in the 57 boys described in Paper I ranged from 0.5 mL to 30.8 mL. This high variability resulted in wider LOA ($\pm 41.4\%$). This poorer interobserver agreement, compared to the intra-observer agreement, may be, in part, due to the different clinical experience between the two observers (the radiographer with eight years' experience versus the radiologist with over 25 years' experience) and minor differences in examination techniques. The better intra-observer agreement, however, shows the potential of improving and standardizing the method, which, in turn, will lead to improved interobserver agreement for better longitudinal follow-up of the individual boys. Since statistically significant bias was not found, the data suggest that ultrasound is a reliable method for construction of growth references of TV.

5.2.2 Method comparison

To determine the between-methods agreement, TVs were measured once with ultrasound and once with the Prader orchidometer. A comparable mean and SD of the TVs measured with the two methods was found. However, there was a clear tendency towards overestimating the smaller TVs with the Prader orchidometer, with a mean difference of 59.1% (SD 27.5%) for prepubertal TVs, but only 5.2% (SD 24.4%) for pubertal TVs. It is well known that ultrasound agrees better with the true TV measured by water displacement or by direct weighing (51,55,58,219) and that the Prader orchidometer tends to overestimate the volumes, compared to USTVs and true TVs (204,217,219). To overcome this non-linear and volume-dependent relation, regression analysis of logarithmic-transformed volumes was performed, resulting in the conversion formula $Vol_{OM} = 1.96 \times Vol_{US}^{0.71}$. The variation between the predicted and observed volumes was of the same magnitude as the estimated interobserver

agreement for ultrasound measurements and did not exceed the reported variation for orchidometer measurements, with a CV of 20.4% (55).

TV measurement is an important tool in diagnosing a variety of medical and surgical conditions. The Prader orchidometer is regarded as having too low sensitivity for detection of minor volume differences, and ultrasound is therefore recommended for routine use when investigating testicular growth impairment (55,205,220). For example, a difference in volume of 20–25% between the right and left testicles has been used as an indicator for surgical correction in boys with varicocele (221). Precise determination of the TV is therefore of utmost importance in such cases and only ultrasound can detect such small differences. In addition to detecting volume variations, use of ultrasound also has the advantage of detecting testicular pathology, such as microlithiasis, varicocele, and hydrocele, and can locate a retractile testis during examination.

5.2.3 **Pubertal timing**

Up-to-date references defining pubertal onset is important for detection of any secular trend of earlier development within a population (1) and for defining altered pubertal timing, since this might have public health implications for the individual boy (171). The definition of normal physiological range for pubertal development is usually set as 2.5 SD below and above the mean (74). In this study, the range limits were 9.0 and 14.3 years, in line with the current definition of normal pubertal onset between 9 and 14 years.

The onset of puberty is defined as reaching a USTV of ≥2.7 mL. Only one earlier study has defined pubertal timing using USTV (15). In this study, pubertal onset occurred at a mean age of 11.7 (1.1) years, as reported in Paper II. This is very similar to the mean ages in comparable countries like Denmark (11.7 years) (79) and The Netherlands (11.6 years) (15,60) where data were collected almost a decade before this study. Several population-based studies have also been conducted in the USA. The most recent Pediatric Research in Office Settings (PROS) study with data collected from 2005 to 2010 showed a mean age of pubertal onset of 11.5 years in non-Hispanic white boys (87), which is also comparable to the results of the already

mentioned European countries. Based on these findings, there was no evidence of a secular trend over the past decade. However, comparing results from this present study to a very small study conducted by Waaler *et al*. in the 1970s that included 109 Norwegian boys aged 1.9–16.9 years (81) showed that contemporary boys reached pubertal TV ~2–3 months earlier than four decades ago. However, this study included only 72 boys at pubertal age (>9 years). The mean age at Tanner stage PH2 in the present study cohort was 11.8 (1.2) years, which also corresponded well to the reported mean ages from Denmark (27,88), Italy (84), Belgium (86), and The Netherlands (222), varying between 11.5 and 11.9 years. Deviating from this trend are two studies from Denmark, both part of the Copenhagen Puberty Study, reporting a mean age at Tanner stage PH2 of 12.4 and 12.2 years, meaning a *delay* of 5 months in the 2006 cohort, compared to the 1991 cohort (29,79). However, the authors found that the age at entry into Tanner stages PH4 and PH5 was reached significantly earlier in 2006, indicating a shorter interval between the stages. When comparing mean ages at more mature pubertal stages in the present study with the 2006 Copenhagen Puberty Study cohort, results showed that the mean ages (± 2 SD) at reaching Tanner stages PH4 and PH5 were 13.46 (11.7–15.2) years and 14.42 (12.6–16.2) years in BGS2, compared to 13.67 (11.56–15.78) years and 14.45 (12.51–16.39) years in the Copenhagen Puberty Study (79). Moreover, when comparing the mean ages for larger TVs from this study with data from Belgium, it was found that age at the 50th percentile (3rd to 97th percentiles) for a TV of 10 mL (measured with a Prader orchidometer) was 14.1 (11.8–16.3) years in this study, whereas the Belgian study reported younger ages of 13.2 (11.2–15.6) years (86). The same trend was seen for TVs of 12 and 15 mL.

To summarize, results from this present study showed that the mean ages at pubertal onset in Norwegian boys, defined as reaching pubertal TV (USTV ≥2.7 mL) and Tanner stage PH2, were similar to comparable data from other countries in Northern Europe obtained over the last few decades, with no indication of an ongoing secular trend. However, based on a small study from Norway in the 1970s, pubertal onset is now observed about 2–3 months earlier than four to five decades ago.

5.2.3.1 Comparison of studies

It must be noted that several authors and expert panels have previously concluded that the lack of specifying methods for pubertal assessment in boys presents significant challenges to the interpretability and replicability of many studies (223), thus complicating the evaluation of any secular trends in male pubertal development (224). Due to the lack of an easily measured and reliable marker in boys, like the menarche in girls, several different proxy markers of pubertal onset and progression have been used. The commonest, besides Tanner staging and Prader orchidometry, being PHV (41) and age at voice break (30,42). Indisputably, the best and most objective clinical marker for male puberty is TV measurement (225). This is seldom evaluated in population-based studies because of the intimate nature of the examination, which means that visual grading of genital development (Tanner G) has been used instead (83). The results from the NHANES III study in the USA (1988–1994) (78), which assessed pubertal development by visual inspection only, have been debated due to the quite large discrepancies in mean ages at pubertal onset compared to the majority of studies that were published at around the same time (79), as well as when comparing mean ages at Tanner stage G2 and Tanner stage PH2 (10.1 versus 12.0 years, respectively). These discrepancies are thought to be because pubertal development was classified solely by visual grading and use of Tanner G staging by multiple examiners, without TV assessment, possibly leading to an overclassification of prepubertal boys as being at Tanner stage G2. Due to the intimate nature of assessing the male genitals either visually or by palpation, other studies used self-reported data (85,226). However, self-assessment of pubertal development has only shown fair to moderate agreement with clinical examination (227-229).

5.2.4 Stratification of endocrine references

Paper III is the first study to present CALIPER and CLSI-compliant reference intervals in relation to the more objective measure of TV using ultrasound. Results showed that the studied hormones varied both with age and with puberty progression, as assessed with gonadal development. The finding that TV accounted for more

variation in testosterone levels than age in pubertal boys illustrates the biological relevance of TV during puberty and prompted the establishment of an additional set of references for hormone levels in relation to TV.

Reference intervals stratified by sex and age are essential when interpreting results from paediatric blood tests. For sex hormones especially, levels vary not only with age, but also according to gender and the degree of pubertal maturation. Current paediatric endocrine references are often based on a small sample size and inpatient tests and are usually not representative of a healthy paediatric population. They are mostly presented in relation to chronological age (230). However, studies have shown complex changes in fertility hormones both during the first year of life and especially throughout adolescence, highlighting the importance of stratifying reference intervals by pubertal stages (70). The CALIPER reference intervals for the HPG axis were partitioned based on self-reported Tanner stages (73).

According to the CLSI guidelines outlined in the document *EP28-A3c*, the CALIPER studies set a new standard for presenting sex- and age-specific reference intervals, with their white paper article covering over 100 biomarkers for paediatric diseases (73). The CALIPER initiative further quantified the effects of pre-analytical factors such as interindividual variation (72) and within-day biological variation (231).

The partitioning of data in the present study was chosen to best support clinical decision-making. The age range of 10–13 years is crucial for testicular maturation, since most boys attain pubertal TV during these years. In the study data set, there were only two boys who reached pubertal TV prior to the age of 10 years and only one boy with a prepubertal TV after the age of 13 years. In light of this, it made clinical sense to dichotomize the participants in this age interval as prepubertal/pubertal according to their TV and to partition the sample accordingly. When stratifying the boys aged between 10 and 13 years by attainment of pubertal TV, all tested hormones were satisfactorily separated using the Harris–Boyd and z* criteria. This approach was not possible in other age partitions. In an attempt to keep the analysis simple and clinically reasonable, the introduction of many combinations of age and TV was minimized in this study.

5.2.5 **Association between pubertal maturation and anthropometric measures**

The work presented in paper IV examined the associations between the timing of pubertal maturation and either a high or a low weight status, using the measurements of BMI, WC, SSF, and %BF for age. Results showed that low BMIz and low WCz were associated with a delay in pubertal onset and smaller TVs for age. However, no significant association was found between higher weight status and the level of pubertal development.

The association between overweight and early pubertal maturation in boys remains debated, with fewer available data in boys compared to girls, presumably as data on pubertal development in boys are more difficult to determine on a large scale. In addition, comparison between studies might also be hampered by the lack of a uniform definition of overweight and obesity (232). This may contribute to the conflicting results of association studies. Some studies report an association between increased BMI and earlier age at pubertal maturation (29,79,114,118,127,233), whereas others found a reverse association, i.e. higher BMI with *later* puberty (44,121-123,234). Further, in line with the study results here, some studies found no association between obesity and early pubertal onset (113,235,236). Lastly, some groups have even reported diverging results between overweight and obese boys, with overweight boys maturing earlier and obese boys maturing later, both compared to normal-weight boys (133,237). These studies all underline the uncertainty of an association between weight status and pubertal development in boys. Conflicting results have also been observed regarding the association with pubic hair development. In this present study, no significant association was found between pubic hair development and any of the anthropometric measurements (BMI, WC, and SSF) or %BF. Busch *et al*. did not observe a significant difference in the timing of pubarche between a control group and the obese study cohort (118). However, Sørensen *et al*. observed earlier pubarche (79), and Kleber *et al*. later pubarche, with increasing BMI (44).

Most studies on pubertal timing and the influence of adiposity are based on cross-sectional data, which makes it difficult to identify the direction of the association between pubertal timing and weight status. This association might be bidirectional, in that obesity might influence puberty, or vice versa. However, contradictory results have been reported in the few published longitudinal studies. Ong *et al*. showed that boys with more advanced voice break at age 14 (as a sign of earlier pubertal development) had a higher mean BMI at ages 2 and 14, compared with other boys (30). He and Karlberg found that an increase in BMI between ages 2 and 8 years, called adiposity rebound, was associated with earlier growth spurt in boys (41), which was confirmed in the Swedish Gothenburg Osteoporosis and Obesity Determinants (GOOD) study a few decades later (238). Aksglaede *et al*. also found that the heavier boys were at age 7, the earlier the age at onset of growth spurt and PHV (119). Additional studies have also confirmed this trend (239-241). In contrast, Buyken *et al*. did not find any independent associations between BMI z-scores 1 or 2 years before age of onset of pubertal growth in a cohort of 108 German boys (242). One must bear in mind that age at PHV and voice break, which are the pubertal markers assessed in most of these studies, do not describe the pubertal onset, but pubertal milestones occurring years later, complicating the cause/effect analysis. The early occurrence of PHV and voice break might represent a more rapid progression of puberty, rather than an earlier initiation of puberty, or simply might represent accelerated growth, independent of puberty, as obese children have demonstrated accelerated growth throughout childhood, independent of the timing of puberty (243).

The secular trend towards earlier puberty, especially in girls over the last decades, has been associated with the increasing number of obese children (244). Therefore, there has been a special focus on the association between overweight and pubertal timing. However, fewer studies have focused on the effect of underweight and low fat on pubertal timing. Results in Paper IV demonstrated an association between low BMI (<−1 SD) and low WC for age and pubertal timing, with a delay in achieving pubertal TV of about 8 months, compared to normal weight boys. At the same time, no significant association was found between high BMI (> 1 SD) or WC,

in contrast with the study hypothesis. In line with the study findings, a study including cohorts from both Germany and Switzerland found that PHV occurred significantly later in lean, compared to normal-weight, children, with no significant difference between obese and normal-weight children (126). Tomova *et al*. also found in a study including >4000 boys that underweight boys (BMI <12th percentile) reached every stage of puberty (TV, Tanner PH, and penis length and circumference) at older ages, when compared to normal-weight boys (127), and at the same time reported that overweight and obese boys reached every pubertal stage at younger ages, compared to normal-weight peers. In the study presented here, the fact that no significant association was found between anthropometric measures and pubertal onset for boys with high BMI and WC for age does not exclude that such associations do exist, but that no such associations could be demonstrated here. This may be linked to the limited number of participating boys, as well as the cross-sectional study design.

It is well known that body composition itself changes during puberty and that BMI increases with maturation, even without changes in adiposity (35). Thus, early puberty from any cause can produce a physiological increase in parameters used to define obesity, such as BMI for age, which could potentially lead to overestimation of adiposity in early-maturing boys (245). This lack of specificity of BMI as a marker for adiposity contributes to the difficulties of proving a direct and causal relationship between excessive adiposity and early puberty (133). Body fat distribution, rather than body weight, might play an important role in the timing of onset of puberty. It was interesting therefore to investigate in this study the association of more direct measures of adiposity with pubertal timing. However, no such associations were found, whether with high or low SSF or %BF or with high WC for age.

Recent large-scale genetic studies have identified genetic correlations between puberty timing and BMI, in addition to later adverse health outcomes like diabetes, cardiovascular disease, and a shorter lifespan (106,162). Findings from these genetic studies, together with results from epidemiological studies, including the study by Ong *et al*. (30), support the theory of an overlapping of genes involved in pubertal timing and adiposity, and the existence of a pathway of early infancy weight gain and

a faster tempo of growth throughout infancy and childhood, leading to earlier pubertal development. Higher childhood body weight and BMI are thought to stimulate earlier pubertal onset and progression through the actions of leptin, insulin resistance, or other hormonal mechanisms such as effects of EDCs (125,134,158). To further complicate these theories, studies have also described later puberty in obese boys. This association is thought to be mediated through increased aromatization of androgens to oestrogens in adipose tissue (246). Towards whichever direction the association is trending, nutritional factors, epigenetics, and EDCs are all potential mediators linking the onset of puberty to weight status (133).

In summary, research to date has shown conflicting results regarding how weight status affects puberty timing in boys. This may be due to varying definitions of puberty and even due to different definitions of obesity, together with differences between populations. These inconsistencies emphasize the need for future longitudinal research in this area.

6. Conclusions

Based on the study aims and hypotheses, the conclusions of the study can be summarized as follows:

- Ultrasound is a reliable method for assessing TV, with acceptable intra-observer agreement and no significant bias, and is suitable for construction of references. However, wider interobserver agreement warrants better standardization of the measurements and better calibration of the observers.

- There was a tendency for Prader orchidometry to overestimate the smaller TVs, compared to use of ultrasound. A nonlinear conversion formula ($Vol_{OM} = 1.96 \times Vol_{US}^{0.71}$) was used to calculate the predicted orchidometer volumes from their corresponding ultrasound measurements, to overcome the volume-dependent relationship.

- The mean age for achieving pubertal TV was 11.7 (1.1) years, and 11.8 (1.2) years for reaching Tanner stage PH2. The age distribution for reaching pubertal milestones was comparable with data from other Northern European countries. Results showed no indication of a secular trend towards earlier puberty over the past few decades.

- Ultrasound assessment of TV was useful for stratification of hormone references, to aid in differentiating between prepubertal and pubertal boys. TV accounted for more variance in serum testosterone levels than chronological age in pubertal boys, whereas the opposite was found in prepubertal boys.

- Low BMI for age (<-1 SD) was associated with delayed onset of puberty, with boys with low BMI for age reaching pubertal TV about 8 months later than normal-weight boys. No association between high BMI and pubertal onset was found.

- Boys with low WC for age (<-1 SD) had a lower probability of being pubertal, compared to boys with normal BMI. The other direct measurements of body fat (SSF and %BF) did not show significant associations with the timing of pubertal onset or the level of maturity.

- Boys with low BMI and low WC for age were delayed over the whole pubertal age range as demonstrated by smaller testicular volume by age. Boys with high BMI and high WC for age did not demonstrate larger testicular volume for age.

- Associations with pubertal onset found for BMI and WC, but not for SSF and %BF, suggest that BMI and WC could be more suitable markers for maturity, rather than adiposity.

7. Future perspectives

This is the first pubertal reference study conducted in Norway, with the aim to assess pubertal development and construct pubertal references for Norwegian boys. The knowledge and definition of normal puberty timing for today's boys are crucial, with important clinical implications for the individual child. Up-to-date references will lead to better guidance for when to investigate boys with early, or delayed, or absent pubertal development. Alterations in the timing of growth and sexual maturation at a population level are also important, particularly in relation to potential later adverse health outcomes, such as diabetes and cardiovascular disease, with implications in terms of public health and behavioural interventions.

This study has demonstrated that ultrasound is a suitable method for evaluating testicular size. Ultrasound may be perceived as technically more challenging than assessment with the Prader orchidometer. However, since ultrasound is becoming increasingly available, has the advantage of detecting testicular pathology, and now also has allowed for references to be established on a continuous scale, it is likely that more clinicians will make use of this method.

The relationship and cause and effect between body fat and timing of puberty onset are still unclear. Further longitudinal studies are required to better understand the physiological link between body fat during infancy, childhood, and peripuberty and the timing of pubertal onset. For this purpose, as a continuation of the present study, further work is already under way. Written consent has been obtained from most of the participating boys in this study for retrieval of their early growth data, from birth till age 6 years, from well-baby clinics. This work will potentially help to better understand the potential link between childhood growth and puberty. In addition, further genetic analyses may aid in finding the etiology of altered puberty timing and thereby lead to better genetic counselling. Many of the participating boys in this study also took part in the Norwegian Mother and Child Cohort Study (MoBa) where maternal, paternal, as well as cord, blood samples have been collected. Together with the data on pubertal development, there are plans to perform a trio-analysis based on the genetic material collected in BGS2 and MoBa. This will be

conducted by means of GWAS, epigenome-wide association studies (EWAS), and methylation analyses.

Finally, identification of modifiable influencing factors on puberty timing is important. The effects of EDCs on puberty timing have remained an ongoing concern. With that in mind, additional blood samples from each study participant were collected, with the aim to analyse several EDCs known to potentially influence pubertal development. This work is expected to commence soon.

References

1. Juul A, Teilmann G, Scheike T, *et al*. Pubertal development in Danish children: comparison of recent European and US data. *Int J Androl*. 2006;**29**:247–55; discussion 286–90.

2. Clavijo RI, Hsiao W. Update on male reproductive endocrinology. *Transl Androl Urol*. 2018;**7**(Suppl 3):S367–72.

3. Grumbach MM. The neuroendocrinology of human puberty revisited. *Horm Res*. 2002;**57**(Suppl 2):2–14.

4. Kelsey TW, Miles A, Mitchell RT, Anderson RA, Wallace WHB. A normative model of serum inhibin B in young males. *PLoS One*. 2016;**11**:e0153843.

5. Skorupskaite K, George JT, Anderson RA. The kisspeptin-GnRH pathway in human reproductive health and disease. *Hum Reprod Update*. 2014;**20**:485–500.

6. Pinilla L, Aguilar E, Dieguez C, Millar RP, Tena-Sempere M. Kisspeptins and reproduction: physiological roles and regulatory mechanisms. *Physiol Rev*. 2012;**92**:1235–316.

7. Ubuka T, Son YL, Tobari Y, *et al*. Central and direct regulation of testicular activity by gonadotropin-inhibitory hormone and its receptor. *Front Endocrinol (Lausanne)*. 2014;**5**:8.

8. Comninos AN, Jayasena CN, Dhillo WS. The relationship between gut and adipose hormones, and reproduction. *Hum Reprod Update*. 2014;**20**:153–74.

9. Kuiri-Hänninen T, Sankilampi U, Dunkel L. Activation of the hypothalamic–pituitary–gonadal axis in infancy: minipuberty. *Horm Res Paediatr*. 2014;**82**:73–80.

10. Abreu AP, Kaiser UB. Pubertal development and regulation. *Lancet Diabetes Endocrinol*. 2016;**4**:254–64.

11. Burt Solorzano CM, McCartney CR. Obesity and the pubertal transition in girls and boys. *Reproduction*. 2010;**140**:399–410.

12. Sklar CA, Kaplan SL, Grumbach MM. Evidence for dissociation between adrenarche and gonadarche: studies in patients with idiopathic precocious puberty, gonadal dysgenesis, isolated gonadotropin deficiency, and

constitutionally delayed growth and adolescence. *J Clin Endocrinol Metab*. 1980;**51**:548–56.

13. Xing Y, Lerario AM, Rainey W, Hammer GD. Development of adrenal cortex zonation. *Endocrinol Metab Clin North Am*. 2015;**44**:243–74.

14. Susman EJ, Houts RM, Steinberg L, *et al*. Longitudinal development of secondary sexual characteristics in girls and boys between ages 91/2 and 151/2 years. *Arch Pediatr Adolesc Med*. 2010;**164**:166–73.

15. Joustra SD, van der Plas EM, Goede J, *et al*. New reference charts for testicular volume in Dutch children and adolescents allow the calculation of standard deviation scores. *Acta Paediatr*. 2015;**104**:e271–8.

16. Sadov S, Koskenniemi JJ, Virtanen HE, *et al*. Testicular growth during puberty in boys with and without a history of congenital cryptorchidism. *J Clin Endocrinol Metab*. 2016;**101**:2570–7.

17. Zachmann M, Prader A, Kind HP, Häfliger H, Budliger H. Testicular volume during adolescence. Cross-sectional and longitudinal studies. *Helv Paediatr Acta*. 1974;**29**:61–72.

18. Largo RH, Prader A. Pubertal development in Swiss boys. *Helv Paediatr Acta*. 1983;**38**:211–28.

19. Koskenniemi JJ, Virtanen HE, Toppari J. Testicular growth and development in puberty. *Curr Opin Endocrinol Diabetes Obes*. 2017;**24**:215–24.

20. Mattioli G, Lazzeroni P, Paraboschi I, Di Iorgi N, Napoli F, Maghnie M. Testis development and descent. In: Simoni M, Huhtaniemi IT, eds. *Endocrinology of the Testis and Male Reproduction*. Cham: Springer International Publishing; 2017:273–311.

21. De Sanctis V, Soliman AT, Di Maio S, Millimaggi G, Kattamis C. For debate: testicular volume development along ages: evaluation by different methods. *Pediatr Endocrinol Rev*. 2019;**16**:421–30.

22. Hart RJ, Doherty DA, McLachlan RI, *et al*. Testicular function in a birth cohort of young men. *Hum Reprod*. 2015;**30**:2713–24.

23. Clement P, Giuliano F. Anatomy and physiology of genital organs—men. *Handb Clin Neurol*. 2015;**130**:19–37.

24. King BF, Hattery RR, Lieber MM, Berquist TH, Williamson B Jr, Hartman GW. Congenital cystic disease of the seminal vesicle. *Radiology.* 1991;**178**:207–11.

25. Schonfeld WA, Beebe GW. Normal growth and variation in the male genitalia from birth to maturity. *J Urol*. 1942;**48**:759.

26. Tomova A, Deepinder F, Robeva R, Lalabonova H, Kumanov P, Agarwal A. Growth and development of male external genitalia: a cross-sectional study of 6200 males aged 0 to 19 years. *Arch Pediatr Adolesc Med*. 2010;**164**:1152–7.

27. Mouritsen A, Aksglaede L, Soerensen K, *et al*. The pubertal transition in 179 healthy Danish children: associations between pubarche, adrenarche, gonadarche, and body composition. *Eur J Endocrinol*. 2013;**168**:129–36.

28. Kahane JC. Growth of the human prepubertal and pubertal larynx. *J Speech Hear Res*. 1982;**25**:446–55.

29. Busch AS, Hollis B, Day FR, *et al*. Voice break in boys-temporal relations with other pubertal milestones and likely causal effects of BMI. *Hum Reprod*. 2019;**34**:1514–22.

30. Ong KK, Bann D, Wills AK, *et al*. Timing of voice breaking in males associated with growth and weight gain across the life course. *J Clin Endocrinol Metab*. 2012;**97**:2844–52.

31. Braunstein GD. Gynecomastia. *N Engl J Med*. 1993;**328**:490–5.

32. Biro FM, Lucky AW, Huster GA, Morrison JA. Hormonal studies and physical maturation in adolescent gynecomastia. *J Pediatr*. 1990;**116**:450–5.

33. Tanner JM, Whitehouse RH, Marubini E, Resele LF. The adolescent growth spurt of boys and girls of the Harpenden growth study. *Ann Hum Biol*. 1976;**3**:109–26.

34. Abbassi V. Growth and normal puberty. *Pediatrics*. 1998;**102**(2 Pt 3):507–11.

35. Veldhuis JD, Roemmich JN, Richmond EJ, *et al*. Endocrine control of body composition in infancy, childhood, and puberty. *Endocr Rev*. 2005;**26**:114–46.

36. Loomba-Albrecht LA, Styne DM. Effect of puberty on body composition. *Curr Opin Endocrinol Diabetes Obes*. 2009;**16**:10–15.

37. Ahmed ML, Ong KK, Morrell DJ, *et al*. Longitudinal study of leptin concentrations during puberty: sex differences and relationship to changes in body composition. *J Clin Endocrinol Metab*. 1999;**84**:899–905.

38. Taylor RW, Grant AM, Williams SM, Goulding A. Sex differences in regional body fat distribution from pre- to postpuberty. *Obesity (Silver Spring)*. 2010;**18**:1410–16.

39. Tinggaard J, Mieritz MG, Sorensen K, *et al*. The physiology and timing of male puberty. *Curr Opin Endocrinol Diabetes Obes*. 2012;**19**:197–203.

40. Tomova A, Lalabonova C, Robeva RN, Kumanov PT. Timing of pubertal maturation according to the age at first conscious ejaculation. *Andrologia*. 2011;**43**:163–6.

41. He Q, Karlberg J. BMI in childhood and its association with height gain, timing of puberty, and final height. *Pediatr Res*. 2001;**49**:244–51.

42. Juul A, Magnusdottir S, Scheike T, Prytz S, Skakkebaek NE. Age at voice break in Danish boys: effects of pre-pubertal body mass index and secular trend. *Int J Androl*. 2007;**30**:537–42.

43. Brix N, Ernst A, Lauridsen LLB, *et al*. Childhood overweight and obesity and timing of puberty in boys and girls: cohort and sibling-matched analyses. *Int J Epidemiol*. 2020;**49**:834–44.

44. Kleber M, Schwarz A, Reinehr T. Obesity in children and adolescents: relationship to growth, pubarche, menarche, and voice break. *J Pediatr Endocrinol Metab*. 2011;**24**(3–4):125–30.

45. Sun SS, Schubert CM, Liang R, *et al*. Is sexual maturity occurring earlier among U.S. children? *J Adolesc Health*. 2005;**37**:345–55.

46. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child*. 1970;**45**:13–23.

47. Slora EJ, Bocian AB, Herman-Giddens ME, *et al*. Assessing inter-rater reliability (IRR) of Tanner staging and orchidometer use with boys: a study from PROS. *J Pediatr Endocrinol Metab*. 2009;**22**:291–9.

48. Reiter EO, Lee PA. Have the onset and tempo of puberty changed? *Arch Pediatr Adolesc Med*. 2001;**155**:988–9.

49. Prader A. Testicular size: assessment and clinical importance. *Triangle*. 1966;**7**:240–3.

50. Sakamoto H, Saito K, Oohta M, Inoue K, Ogawa Y, Yoshida H. Testicular volume measurement: comparison of ultrasonography, orchidometry, and water displacement. *Urology*. 2007;**69**:152–7.

51. Behre HM, Nashan D, Nieschlag E. Objective measurement of testicular volume by ultrasonography: evaluation of the technique and comparison with orchidometer estimates. *Int J Androl*. 1989;**12**:395–403.

52. Carlsen E, Andersen AG, Buchreitz L, *et al*. Inter-observer variation in the results of the clinical andrological examination including estimation of testicular size. *Int J Androl*. 2000;**23**:248–53.

53. Sotos JF, Tokar NJ. Testicular volumes revisited: a proposal for a simple clinical method that can closely match the volumes obtained by ultrasound and its clinical application. *Int J Pediatr Endocrinol*. 2012;**2012**:17.

54. Taskinen S, Taavitsainen M, Wikstrom S. Measurement of testicular volume: comparison of 3 different methods. *J Urol*. 1996;**155**:930–3.

55. Rivkees SA, Hall DA, Boepple PA, Crawford JD. Accuracy and reproducibility of clinical measures of testicular volume. *J Pediatr*. 1987;110:914–17.

56. Kuijper EA, van Kooten J, Verbeke JI, van Rooijen M, Lambalk CB. Ultrasonographically measured testicular volumes in 0- to 6-year-old boys. *Hum Reprod*. 2008;**23**:792–6.

57. Lambert B. The frequency of mumps and of mumps orchitis and the consequences for sexuality and fertility. *Acta Genet Stat Med*. 1951;**2**(Suppl 1):1–166.

58. Paltiel HJ, Diamond DA, Di Canzio J, Zurakowski D, Borer JG, Atala A. Testicular volume: comparison of orchidometer and US measurements in dogs. *Radiology*. 2002;**222**:114–19.

59. Hsieh ML, Huang ST, Huang HC, Chen Y, Hsu YC. The reliability of ultrasonographic measurements for testicular volume assessment: comparison of three common formulas with true testicular volume. *Asian J Androl*. 2009;**11**:261–5.

60. Goede J, Hack WW, Sijstermans K, *et al*. Normative values for testicular volume measured by ultrasonography in a normal population from infancy to adolescence. *Horm Res Paediatr*. 2011;**76**:56–64.

61. Arthurs OJ, Bjørkum AA. Safety in pediatric imaging: an update. *Acta Radiol*. 2013;**54**:983–90.

62. Simpkin AJ, Sayers A, Gilthorpe MS, Heron J, Tilling K. Modelling height in adolescence: a comparison of methods for estimating the age at peak height velocity. *Ann Hum Biol*. 2017;**44**:715–22.

63. Silventoinen K, Haukka J, Dunkel L, Tynelius P, Rasmussen F. Genetics of pubertal timing and its associations with relative weight in childhood and adult height: the Swedish Young Male Twins Study. *Pediatrics*. 2008;**121**:e885–91.

64. Granados A, Gebremariam A, Lee JM. Relationship between timing of peak height velocity and pubertal staging in boys and girls. *J Clin Res Pediatr Endocrinol*. 2015;**7**:235–7.

65. Andersen E. Skeletal maturation of Danish school children in relation to height, sexual development, and social conditions. *Acta Paediatr Scand*. 1968:**Suppl 185**:11+.

66. Harries ML, Walker JM, Williams DM, Hawkins S, Hughes IA. Changes in the male voice at puberty. *Arch Dis Child*. 1997;**77**:445–47.

67. Nielsen CT, Skakkebaek NE, Richardson DW, *et al*. Onset of the release of spermatozoa (spermarche) in boys in relation to age, testicular growth, pubic hair, and height. *J Clin Endocrinol Metab*. 1986;**62**:532–5.

68. Janczewski Z, Bablok L. Semen characteristics in pubertal boys. II. Semen quality in relation to bone age. *Arch Androl*. 1985;**15**(2–3):207–11.

69. Wu FC, Brown DC, Butler GE, Stirling HF, Kelnar CJ. Early morning plasma testosterone is an accurate predictor of imminent pubertal development in prepubertal boys. *J Clin Endocrinol Metab*. 1993;**76**:26–31.

70. Konforte D, Shea JL, Kyriakopoulou L, *et al*. Complex biological pattern of fertility hormones in children and adolescents: a study of healthy children from the CALIPER cohort and establishment of pediatric reference intervals. *Clin Chem*. 2013;**59**:1215–27.

71. Ahokoski O, Virtanen A, Huupponen R, *et al*. Biological day-to-day variation and daytime changes of testosterone, follitropin, lutropin and oestradiol-17beta in healthy men. *Clin Chem Lab Med*. 1998;**36**:485–91.

72. Bailey D, Colantonio D, Kyriakopoulou L, *et al*. Marked biological variance in endocrine and biochemical markers in childhood: establishment of pediatric reference intervals using healthy community children from the CALIPER cohort. *Clin Chem*. 2013;**59**:1393–405.

73. Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N. The Canadian Laboratory Initiative on pediatric reference intervals: a CALIPER white paper. *Crit Rev Clin Lab Sci*. 2017;**54**:358–413.

74. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr Rev*. 2003;**24**:668–93.

75. Mul D, Fredriks AM, van Buuren S, Oostdijk W, Verloove-Vanhorick SP, Wit JM. Pubertal development in The Netherlands 1965–1997. *Pediatr Res*. 2001;**50**:479–86.

76. Aksglaede L, Sorensen K, Petersen JH, Skakkebaek NE, Juul A. Recent decline in age at breast development: the Copenhagen Puberty Study. *Pediatrics*. 2009;**123**:e932–9.

77. Herman-Giddens ME, Slora EJ, Wasserman RC, *et al*. Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings network. *Pediatrics*. 1997;**99**:505–12.

78. Herman-Giddens ME, Wang L, Koch G. Secondary sexual characteristics in boys: estimates from the national health and nutrition examination survey III, 1988–1994. *Arch Pediatr Adolesc Med*. 2001;**155**:1022–8.

79. Sørensen K, Aksglaede L, Petersen JH, Juul A. Recent changes in pubertal timing in healthy Danish boys: associations with body mass index. *J Clin Endocrinol etab*. 2010;**95**:263–70.

80. Papadimitriou A, Douros K, Kleanthous K, Papadimitriou DT, Attilakos A, Fretzayas A. Pubertal maturation of contemporary Greek boys: no evidence of a secular trend. *J Adolesc Health*. 2011;**49**:434–6.

81. Waaler PE, Thorsen T, Stoa KF, Aarskog D. Studies in normal male puberty. *Acta Paediatr Scand Suppl*. 1974;**249**:1–36.

82. Lindgren G. Pubertal stages 1980 of Stockholm schoolchildren. *Acta Paediatr*. 1996;**85**:1365–7.

83. Herman-Giddens ME. Recent data on pubertal milestones in United States children: the secular trend toward earlier development. *Int J Androl*. 2006;**29**:241–6; discussion 286–90.

84. De Simone M, Danubio ME, Amicone E, Verrotti A, Gruppioni G, Vecchi F. Age of onset of pubertal characteristics in boys aged 6–14 years of the Province of L'Aquila (Abruzzo, Italy). *Ann Hum Biol*. 2004;**31**:488–93.

85. Monteilh C, Kieszak S, Flanders WD, *et al*. Timing of maturation and predictors of Tanner stage transitions in boys enrolled in a contemporary British cohort. *Paediatr Perinat Epidemiol*. 2011;**25**:75–87.

86. Roelants M, Hauspie R, Hoppenbrouwers K. References for growth and pubertal development from birth to 21 years in Flanders, Belgium. *Ann Hum Biol*. 2009;**36**:680–94.

87. Herman-Giddens ME, Steffes J, Harris D, *et al*. Secondary sexual characteristics in boys: data from the Pediatric Research in Office Settings Network. *Pediatrics*. 2012;**130**:e1058–68.

88. Wohlfahrt-Veje C, Mouritsen A, Hagen CP, *et al*. Pubertal onset in boys and girls is influenced by pubertal timing of both parents. *J Clin Endocrinol Metab*. 2016;**101**:2667–74.

89. Teilmann G, Pedersen CB, Jensen TK, Skakkebaek NE, Juul A. Prevalence and incidence of precocious pubertal development in Denmark: an epidemiologic study based on national registries. *Pediatrics*. 2005;**116**:1323–8.

90. Cesario SK, Hughes LA. Precocious puberty: a comprehensive review of literature. *J Obstet Gynecol Neonat Nurs*. 2007;**36**:263–74.

91. Choi KH, Chung SJ, Kang MJ, *et al*. Boys with precocious or early puberty: incidence of pathological brain magnetic resonance imaging findings and factors related to newly developed brain lesions. *Ann Pediatr Endocrinol Metab*. 2013;**18**:183–90.

92. Howard SR, Dunkel L. Delayed puberty-phenotypic diversity, molecular genetic mechanisms, and recent discoveries. *Endocr Rev*. 2019;**40**:1285–317.

93. Alotaibi MF. Physiology of puberty in boys and girls and pathological disorders affecting its onset. *J Adolesc*. 2019;**71**:63–71.

94. Bordini B, Rosenfield RL. Normal pubertal development: part II: clinical aspects of puberty. *Pediatr Rev*. 2011;**32**:281–92.

95. Silveira LF, Latronico AC. Approach to the patient with hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*. 2013;**98**:1781–8.

96. Boehm U, Bouloux PM, Dattani MT, *et al*. Expert consensus document: European Consensus Statement on congenital hypogonadotropic hypogonadism—pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol*. 2015;**11**:547–64.

97. Fraietta R, Zylberstejn DS, Esteves SC. Hypogonadotropic hypogonadism revisited. *Clinics (Sao Paulo)*. 2013;68 Suppl 1(Suppl 1):81-88.

98. Mäntyselkä A, Jääskeläinen J, Lindi V, *et al*. The presentation of adrenarche is sexually dimorphic and modified by body adiposity. *J Clin Endocrinol Metab*. 2014;**99**:3889–94.

99. Utriainen P, Laakso S, Liimatta J, Jääskeläinen J, Voutilainen R. Premature adrenarche—a common condition with variable presentation. *Horm Res Paediatr*. 2015;**83**:221–31.

100. Palmert MR, Hirschhorn JN. Genetic approaches to stature, pubertal timing, and other complex traits. *Mol Genet Metab*. 2003;**80**(1–2):1–10.

101. Wehkalampi K, Widen E, Laine T, Palotie A, Dunkel L. Patterns of inheritance of constitutional delay of growth and puberty in families of adolescent girls and boys referred to specialist pediatric care. *J Clin Endocrinol Metab*. 2008;**93**:723–8.

102. Hauspie RC, Bergman P, Bielicki T, Susanne C. Genetic variance in the pattern of the growth curve for height: a longitudinal analysis of male twins. *Ann Hum Biol*. 1994;**21**:347–62.

103. Abreu AP, Dauber A, Macedo DB, *et al*. Central precocious puberty caused by mutations in the imprinted gene *MKRN3*. *N Engl J Med*. 2013;**368**:2467–75.

104. Silveira LG, Noel SD, Silveira-Neto AP, *et al*. Mutations of the *KISS1* gene in disorders of puberty. *J Clin Endocrinol Metab*. 2010;**95**:2276–80.

105. Cousminer DL, Widén E, Palmert MR. The genetics of pubertal timing in the general population: recent advances and evidence for sex-specificity. *Curr Opin Endocrinol Diabetes Obes*. 2016;**23**:57–65.

106. Hollis B, Day FR, Busch AS, *et al*. Genomic analysis of male puberty timing highlights shared genetic basis with hair colour and lifespan. *Nat Commun*. 2020;**11**:1536.

107. Howard SR, Guasti L, Poliandri A, *et al*. Contributions of function-altering variants in genes implicated in pubertal timing and body mass for self-limited delayed puberty. *J Clin Endocrinol Metab*. 2018;**103**:649–59.

108. Day FR, Perry JR, Ong KK. Genetic regulation of puberty timing in humans. *Neuroendocrinology*. 2015;**102**:247–55.

109. Cousminer DL, Stergiakouli E, Berry DJ, *et al*. Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. *Hum Mol Genet*. 2014;**23**:4452–64.

110. Ong KK, Elks CE, Li S, *et al*. Genetic variation in *LIN28B* is associated with the timing of puberty. *Nat Genet*. 2009;**41**:729–33.

111. Pozo J, Argente J. Delayed puberty in chronic illness. *Best Pract Res Clin Endocrinol Metab*. 2002;**16**:73–90.

112. Muñoz-Calvo MT, Argente J. Nutritional and pubertal disorders. *Endocr Dev*. 2016;**29**:153–73.

113. Denzer C, Weibel A, Muche R, Karges B, Sorgo W, Wabitsch M. Pubertal development in obese children and adolescents. *Int J Obes (Lond)*. 2007;**31**:1509–19.

114. Ribeiro J, Santos P, Duarte J, Mota J. Association between overweight and early sexual maturation in Portuguese boys and girls. *Ann Hum Biol*. 2006;**33**:55–63.

115. Rosenfield RL, Lipton RB, Drum ML. Thelarche, pubarche, and menarche attainment in children with normal and elevated body mass index. *Pediatrics*. 2009;**123**:84–8.

116. Himes JH, Obarzanek E, Baranowski T, Wilson DM, Rochon J, McClanahan BS. Early sexual maturation, body composition, and obesity in African-American girls. *Obes Res*. 2004;**12** Suppl:64s–72s.

117. Currie C, Ahluwalia N, Godeau E, Nic Gabhainn S, Due P, Currie DB. Is obesity at individual and national level associated with lower age at menarche? Evidence from 34 countries in the Health Behaviour in School-aged Children Study. *J Adolesc Health*. 2012;**50**:621–6.

118. Busch AS, Hojgaard B, Hagen CP, Teilmann G. Obesity is associated with earlier pubertal onset in boys. *J Clin Endocrinol Metab*. 2020;**105**:dgz222.

119. Aksglaede L, Juul A, Olsen LW, Sorensen TI. Age at puberty and the emerging obesity epidemic. *PLoS One*. 2009;**4**:e8450.

120. Sandhu J, Ben-Shlomo Y, Cole TJ, Holly J, Davey Smith G. The impact of childhood body mass index on timing of puberty, adult stature and obesity: a follow-up study based on adolescent anthropometry recorded at Christ's Hospital (1936–1964). *Int J Obes (Lond)*. 2006;**30**:14–22.

121. Biro FM, Khoury P, Morrison JA. Influence of obesity on timing of puberty. *Int J Androl*. 2006;**29**:272–7; discussion 286–90.

122. Crocker MK, Stern EA, Sedaka NM, *et al*. Sexual dimorphisms in the associations of BMI and body fat with indices of pubertal development in girls and boys. *J Clin Endocrinol Metab*. 2014;**99**:E1519–29.

123. Wang Y. Is obesity associated with early sexual maturation? A comparison of the association in American boys versus girls. *Pediatrics*. 2002;**110**:903–10.

124. Boyne MS, Osmond C, Fraser RA, *et al*. Developmental origins of cardiovascular risk in Jamaican children: the Vulnerable Windows Cohort study. *Br J Nutr*. 2010;**104**:1026–33.

125. Dunger DB, Ahmed ML, Ong KK. Early and late weight gain and the timing of puberty. *Mol Cell Endocrinol*. 2006;**254–5**:140–5.

126. Heger S, Korner A, Meigen C, *et al*. Impact of weight status on the onset and parameters of puberty: analysis of three representative cohorts from central Europe. *J Pediatr Endocrinol Metab*. 2008;**21**:865–77.

127. Tomova A, Robeva R, Kumanov P. Influence of the body weight on the onset and progression of puberty in boys. *J Pediatr Endocrinol Metab*. 2015;**28**(7–8):859–65.

128. Kiess W, Reich A, Meyer K, *et al*. A role for leptin in sexual maturation and puberty? *Horm Res*. 1999;**51**(Suppl 3):55–63.

129. Argente J, Barrios V, Chowen JA, Sinha MK, Considine RV. Leptin plasma levels in healthy Spanish children and adolescents, children with obesity, and adolescents with anorexia nervosa and bulimia nervosa. *J Pediatr*. 1997;**131**:833–8.

130. Fernandez-Fernandez R, Martini AC, Navarro VM, *et al*. Novel signals for the integration of energy balance and reproduction. *Mol Cell Endocrinol*. 2006;**254–5**:127–32.

131. Farooqi IS, Wangensteen T, Collins S, *et al*. Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. *N Engl J Med*. 2007;**356**:237–47.

132. Strobel A, Issad T, Camoin L, Ozata M, Strosberg AD. A leptin missense mutation associated with hypogonadism and morbid obesity. *Nat Genet*. 1998;**18**:213–15.

133. Reinehr T, Roth CL. Is there a causal relationship between obesity and puberty? *Lancet Child Adolesc Health*. 2019;**3**:44–54.

134. Ahmed ML, Ong KK, Dunger DB. Childhood obesity and the timing of puberty. *Trends Endocrinol Metab*. 2009;**20**:237–42.

135. Sørensen K, Aksglaede L, Munch-Andersen T, *et al*. Sex hormone-binding globulin levels predict insulin sensitivity, disposition index, and cardiovascular risk during puberty. *Diabetes Care*. 2009;**32**:909–14.

136. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ*. 2000;**320**:1240–3.

137. Freedman DS, Sherry B. The validity of BMI as an indicator of body fatness and risk among children. *Pediatrics*. 2009;**124**(Suppl 1):S23–34.

138. Cole TJ, Freeman JV, Preece MA. Body mass index reference curves for the UK, 1990. *Arch Dis Child*. 1995;**73**:25–9.

139. Juliusson PB, Roelants M, Nordal E, *et al*. Growth references for 0–19 year-old Norwegian children for length/height, weight, body mass index and head circumference. *Ann Hum Biol*. 2013;**40**:220–7.

140. Wohlfahrt-Veje C, Tinggaard J, Winther K, *et al*. Body fat throughout childhood in 2647 healthy Danish children: agreement of BMI, waist circumference, skinfolds with dual X-ray absorptiometry. *Eur J Clin Nutr*. 2014;**68**:664–70.

141. Kelishadi R, Mirmoghtadaee P, Najafi H, Keikha M. Systematic review on the association of abdominal obesity in children and adolescents with cardio-metabolic risk factors. *J Res Med Sci*. 2015;**20**:294–307.

142. Taylor RW, Jones IE, Williams SM, Goulding A. Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk fat mass, as measured by dual-energy X-ray absorptiometry, in children aged 3–19 y. *Am J Clin Nutr*. 2000;**72**:490–5.

143. Wang J, Thornton JC, Bari S, Williamson B, Gallagher D, Heymsfield SB, Horlick M, Kotler D, Laferrère B, Mayer L, Pi-Sunyer FX, Pierson RN, Jr. Comparisons of waist circumferences measured at 4 sites. *Am J Clin Nutr*. 2003;**77**:379–84.

144. Brannsether B, Eide GE, Roelants M, Bjerknes R, Juliusson PB. Interrelationships between anthropometric variables and overweight in childhood and adolescence. *Am J Hum Biol*. 2014;**26**:502–10.

145. Brannsether B, Roelants M, Bjerknes R, Juliusson PB. Waist circumference and waist-to-height ratio in Norwegian children 4–18 years of age: reference values and cut-off levels. *Acta Paediatr*. 2011;**100**:1576–82.

146. Ulijaszek SJ, Kerr DA. Anthropometric measurement error and the assessment of nutritional status. *Br J Nutr*. 1999;**82**:165–77.

147. Freedman DS, Ogden CL, Blanck HM, Borrud LG, Dietz WH. The abilities of body mass index and skinfold thicknesses to identify children with low or elevated levels of dual-energy X-ray absorptiometry-determined body fatness. *J Pediatr*. 2013;**163**:160–6.e161.

148. Brannsether B, Roelants M, Bjerknes R, Juliusson PB. References and cutoffs for triceps and subscapular skinfolds in Norwegian children 4–16 years of age. *Eur J Clin Nutr*. 2013;**67**:928–33.

149. Sopher AB, Thornton JC, Wang J, Pierson RN Jr, Heymsfield SB, Horlick M. Measurement of percentage of body fat in 411 children and adolescents: a comparison of dual-energy X-ray absorptiometry with a four-compartment model. *Pediatrics*. 2004;**113**:1285–90.

150. Jensky-Squires NE, Dieli-Conwright CM, Rossuello A, Erceg DN, McCauley S, Schroeder ET. Validity and reliability of body composition analysers in children and adults. *Br J Nutr*. 2008;**100**:859–65.

151. Pecoraro P, Guida B, Caroli M, *et al*. Body mass index and skinfold thickness versus bioimpedance analysis: fat mass prediction in children. *Acta Diabetol*. 2003;**40**(Suppl 1):S278–81.

152. Zawatski W, Lee MM. Male pubertal development: are endocrine-disrupting compounds shifting the norms? *J Endocrinol*. 2013;**218**:R1–12.

153. Almstrup K, Fernández MF, Petersen JH, Olea N, Skakkebaek NE, Leffers H. Dual effects of phytoestrogens result in u-shaped dose–response curves. *Environ Health Perspect*. 2002;**110**:743–8.

154. Mouritsen A, Aksglaede L, Sørensen K, *et al*. Hypothesis: exposure to endocrine-disrupting chemicals may interfere with timing of puberty. *Int J Androl*. 2010;**33**:346–59.

155. Sharpe RM, Skakkebaek NE. Are oestrogens involved in falling sperm counts and disorders of the male reproductive tract? *Lancet*. 1993;**341**:1392–5.

156. Den Hond E, Roels HA, Hoppenbrouwers K, *et al*. Sexual maturation in relation to polychlorinated aromatic hydrocarbons: Sharpe and Skakkebaek's hypothesis revisited. *Environ Health Perspect*. 2002;**110**:771–6.

157. Bourguignon JP, Juul A, Franssen D, Fudvoye J, Pinson A, Parent AS. Contribution of the endocrine perspective in the evaluation of endocrine disrupting chemical effects: the case study of pubertal timing. *Horm Res Paediatr*. 2016;**86**:221–32.

158. Parent AS, Franssen D, Fudvoye J, Gerard A, Bourguignon JP. Developmental variations in environmental influences including endocrine disruptors on pubertal timing and neuroendocrine control: revision of human observations and mechanistic insight from rodents. *Front Neuroendocrinol*. 2015;**38**:12–36.

159. Tahirovic HF. Menarchal age and the stress of war: an example from Bosnia. *Eur J Pediatr*. 1998;**157**:978–80.

160. Wierson M, Long PJ, Forehand RL. Toward a new understanding of early menarche: the role of environmental stress in pubertal timing. *Adolescence*. 1993;**28**:913–24.

161. Teilmann G, Pedersen CB, Skakkebaek NE, Jensen TK. Increased risk of precocious puberty in internationally adopted children in Denmark. *Pediatrics*. 2006;**118**:e391–9.

162. Day FR, Elks CE, Murray A, Ong KK, Perry JR. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. *Sci Rep*. 2015;**5**:11208.

163. Ohlsson C, Bygdell M, Nethander M, *et al*. Early puberty and risk for type 2 diabetes in men. *Diabetologia*. 2020;63(6):1141-1150.

164. [No authors listed]. Aetiology of testicular cancer: association with congenital abnormalities, age at puberty, infertility, and exercise. United Kingdom Testicular Cancer Study Group. *BMJ*. 1994;**308**:1393–9.

165. Maule M, Malavassi JL, Richiardi L. Age at puberty and risk of testicular cancer: a meta-analysis. *Int J Androl*. 2012;**35**:828–34.

166. Giles GG, Severi G, English DR, *et al*. Early growth, adult body size and prostate cancer risk. *Int J Cancer*. 2003;**103**:241–5.

167. Honda GD, Bernstein L, Ross RK, Greenland S, Gerkins V, Henderson BE. Vasectomy, cigarette smoking, and age at first sexual intercourse as risk factors for prostate cancer in middle-aged men. *Br J Cancer*. 1988;**57**:326–31.

168. Jensen TK, Finne KF, Skakkebaek NE, *et al*. Self-reported onset of puberty and subsequent semen quality and reproductive hormones in healthy young men. *Hum Reprod*. 2016;**31**:1886–94.

169. Wichstrøm L. Predictors of adolescent suicide attempts: a nationally representative longitudinal study of Norwegian adolescents. *J Am Acad Child Adolesc Psychiatry*. 2000;**39**:603–10.

170. Michaud PA, Suris JC, Deppen A. Gender-related psychological and behavioural correlates of pubertal timing in a national sample of Swiss adolescents. *Mol Cell Endocrinol*. 2006;**254–5**:172–8.

171. Golub MS, Collman GW, Foster PM, *et al*. Public health implications of altered puberty timing. *Pediatrics*. 2008;**121**(Suppl 3):S218–30.

172. Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatr Obes*. 2012;**7**:284–94.

173. Methlie P, Hustad SS, Kellmann R, *et al*. Multisteroid LC-MS/MS assay for glucocorticoids and androgens, and its application in Addison's disease. *Endocr Connect*. 2013;**2**:125–36.

174. Juliusson PB, Roelants M, Eide GE, Hauspie R, Waaler PE, Bjerknes R. Overweight and obesity in Norwegian children: secular trends in weight-for-height and skinfolds. *Acta Paediatr*. 2007;**96**:1333–7.

175. Juliusson PB. *Overweight and obesity in Norwegian children. Trends, current prevalence, effect of socio-demographic factors and parental perception*. Doctoral thesis. Bergen: University of Bergen; 2010.

176. McCarthy HD, Cole TJ, Fry T, Jebb SA, Prentice AM. Body fat reference curves for children. *Int J Obes (Lond)*. 2006;**30**:598–602.

177. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol*. 2003;**22**:85–93.

178. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;**1**:307–10.

179. Bland JM, Altman DG. Agreed statistics: measurement method comparison. *Anesthesiology*. 2012;**116**:182–5.

180. Cole TJ. The development of growth references and growth charts. *Ann Hum Biol*. 2012;**39**:382–94.

181. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med*. 1992;**11**:1305–19.

182. Cole TJ. The LMS method for constructing normalized growth standards. *Eur J Clin Nutr*. 1990;**44**:45–60.

183. Horowitz GL, Altaie S, Boyd JC, *et al*. *EP28-A3C. Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline—Third edition*. Wayne, PA: Clinical and Laboratory Standards Institute; 2010.

184. Harris EK, Boyd JC. On dividing reference data into subgroups to produce separate reference ranges. *Clin Chem*. 1990;**36**:265–70.

185. Horn PS, Pesce AJ. Reference intervals: an update. *Clin Chim Acta*. 2003;**334**(1–2):5–23.

186. World Medical Association. Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;**310**:2191–4.

187. Norwegian Institute of Public Health. Barn og overvekt 2017. Available from https://www.fhi.no/nettpub/hin/levevaner/overvekt-og-fedme/.

188. Bergen Kommune. Levekår og helse i Bergen 2016. Available from https://www.bergen.kommune.no/hvaskjer/tema/folkehelse/statistikk-og-rapporter/folkehelseoversikt-for-bergen-rapporter.

189. Lien N, Kumar BN, Holmboe-Ottesen G, Klepp KI, Wandel M. Assessing social differences in overweight among 15- to 16-year-old ethnic Norwegians from Oslo by register data and adolescent self-reported measures of socio-economic status. *Int J Obes (Lond)*. 2007;**31**:30–8.

190. SSB (Statistisk Sentralbyrå). Personer 16 år og over, etter region, kjønn, alder, utdanningsnivå, statistikkvariabel og år: Statistics Norway 2019. Available from https://www.ssb.no/statbank/table/08921.

191. Reinikainen J, Tolonen H, Borodulin K, *et al*. Participation rates by educational levels have diverged during 25 years in Finnish health examination surveys. *Eur J Public Health*. 2018;**28**:237–43.

192. Langhammer A, Krokstad S, Romundstad P, Heggland J, Holmen J. The HUNT study: participation is associated with survival and depends on socioeconomic status, diseases and symptoms. *BMC Med Res Methodol*. 2012;**12**:143.

193. SSB (Statistisk Sentralbyrå) Innvandrere og norskfødte med innvandrerforeldre. Available from https://www.ssb.no/statbank/table/05196/.

194. Mittwoch U. Ethnic differences in testis size: a possible link with the cytogenetics of true hermaphroditism. *Hum Reprod*. 1988;**3**:445–9.

195. Diamond JM. Ethnic differences. Variation in human testis size. *Nature*. 1986;**320**:488–9.

196. Inan M, Aydiner CY, Tokuc B, *et al*. Prevalence of cryptorchidism, retractile testis and orchiopexy in school children. *Urol Int*. 2008;**80**:166–71.

197. James E, Wood CL, Nair H, Williams TC. Preterm birth and the timing of puberty: a systematic review. *BMC Pediatr*. 2018;**18**:3.

198. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol*. 2008;**31**:466–75.

199. Bahk JY, Jung JH, Jin LM, Min SK. Cut-off value of testes volume in young adults and correlation among testes volume, body mass index, hormonal level, and seminal profiles. *Urology*. 2010;**75**:1318–23.

200. Tatsunami S, Matsumiya K, Tsujimura A, *et al*. Inter/intra investigator variation in orchidometric measurements of testicular volume by ten investigators from five institutions. *Asian J Androl*. 2006;**8**:373–8.

201. Vaganee D, Daems F, Aerts W, *et al*. Testicular asymmetry in healthy adolescent boys. *BJU Int*. 2018;**122**:654–66.

202. Dorn L, Dahl RE, Woodward HR, Biro F. Defining the boundaries of early adolescence: a user's guide to assessing pubertal status and pubertal timing in research with adolescents. *Appl Dev Sci*. 2010;**10**:30–56.

203. Coleman L, Coleman J. The measurement of puberty: a review. *J Adolesc*. 2002;**25**:535–50.

204. al Salim A, Murchison PJ, Rana A, Elton RA, Hargreave TB. Evaluation of testicular volume by three orchidometers compared with ultrasonographic measurements. *Br J Urol*. 1995;**76**:632–5.

205. Diamond DA, Paltiel HJ, DiCanzio J, *et al*. Comparative assessment of pediatric testicular volume: orchidometer versus ultrasound. *J Urol*. 2000;**164**(3 Pt 2):1111–14.

206. 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. *Eur J Endocrinol*. 2004;**151**:747–57.

207. Andersson AM, Juul A, Petersen JH, Müller J, Groome NP, Skakkebaek NE. Serum inhibin B in healthy pubertal and adolescent boys: relation to age, stage of puberty, and follicle-stimulating hormone, luteinizing hormone, testosterone, and estradiol levels. *J Clin Endocrinol Metab*. 1997;**82**:3976–81.

208. Ankarberg-Lindgren C, Dahlgren J, Andersson MX. High-sensitivity quantification of serum androstenedione, testosterone, dihydrotestosterone, estrone and estradiol by gas chromatography-tandem mass spectrometry with sex- and puberty-specific reference intervals. *J Steroid Biochem Mol Biol*. 2018;**183**:116–24.

209. Honour JW. Steroid assays in paediatric endocrinology. *J Clin Res Pediatr Endocrinol*. 2010;**2**:1–16.

210. Jannetto PJ, Fitzgerald RL. Effective use of mass spectrometry in the clinical laboratory. *Clin Chem*. 2016;**62**:92–8.

211. Daniels SR, Khoury PR, Morrison JA. The utility of body mass index as a measure of body fatness in children and adolescents: differences by race and gender. *Pediatrics*. 1997;**99**:804–7.

212. Burns R, Hannon JC, Brusseau TA, Shultz B, Eisenman P. Indices of abdominal adiposity and cardiorespiratory fitness test performance in middle-school students. *J Obes*. 2013;**2013**:912460.

213. Houtkooper LB, Lohman TG, Going SB, Howell WH. Why bioelectrical impedance analysis should be used for estimating adiposity. *Am J Clin Nutr*. 1996;**64**(3 Suppl):436s–48s.

214. Eisenmann JC, Heelan KA, Welk GJ. Assessing body composition among 3- to 8-year-old children: anthropometry, BIA, and DXA. *Obes Res*. 2004;**12**:1633–40.

215. Schiff JD, Li PS, Goldstein M. Correlation of ultrasonographic and orchidometer measurements of testis volume in adults. *BJU Int*. 2004;**93**:1015–17.

216. Karaman MI, Kaya C, Caskurlu T, Guney S, Ergenekon E. Measurement of pediatric testicular volume with Prader orchidometer: comparison of different hands. *Pediatr Surg Int*. 2005;**21**:517–20.

217. Sakamoto H, Saito K, Ogawa Y, Yoshida H. Testicular volume measurements using Prader orchidometer versus ultrasonography in patients with infertility. *Urology*. 2007;**69**:158–62.

218. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med*. 1990;**20**:337–40.

219. Fuse H, Takahara M, Ishii H, Sumiya H, Shimazaki J. Measurement of testicular volume by ultrasonography. *Int J Androl*. 1990;**13**:267–72.

220. Atabek ME. Prader orchidometer and ultrasound can be used for monitoring testicular growth: which is a more valid method? *Horm Res Paediatr*. 2011;**76**:144.

221. Diamond DA, Gargollo PC, Caldamone AA. Current management principles for adolescent varicocele. *Fertil Steril*. 2011;**96**:1294–8.

222. Fredriks AM, van Buuren S, Burgmeijer RJ, *et al*. Continuing positive secular growth change in The Netherlands 1955–1997. *Pediatr Res*. 2000;**47**:316–23.

223. Dorn L, Dahl RE, Woodward HR, Biro F. Defining the boundaries of early adolescence: a user's guide to assessing pubertal status and pubertal timing in research with adolescents. *Appl Dev Sci*. 2010;**10**:30–56.

224. Euling SY, Herman-Giddens ME, Lee PA, *et al*. Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. *Pediatrics*. 2008;**121**(Suppl 3):S172–91.

225. Biro FM, Lucky AW, Huster GA, Morrison JA. Pubertal staging in boys. *J Pediatr*. 1995;**127**:100–2.

226. Brix N, Ernst A, Lauridsen LLB, *et al*. Timing of puberty in boys and girls: a population-based study. *Paediatr Perinat Epidemiol*. 2019;**33**:70–8.

227. Chan NP, Sung RY, Kong AP, Goggins WB, So HK, Nelson EA. Reliability of pubertal self-assessment in Hong Kong Chinese children. *J Paediatr Child Health*. 2008;**44**:353–8.

228. Ernst A, Lauridsen LLB, Brix N, *et al*. Self-assessment of pubertal development in a puberty cohort. *J Pediatr Endocrinol Metab*. 2018;**31**:763–72.

229. Campisi SC, Marchand JD, Siddiqui FJ, Islam M, Bhutta ZA, Palmert MR. Can we rely on adolescents to self-assess puberty stage? A systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2020;**105**:2846–56.

230. Elmlinger MW, Kühnel W, Wormstall H, Döller PC. Reference intervals for testosterone, androstenedione and SHBG levels in healthy females and males from birth until old age. *Clin Lab*. 2005;**51**(11–12):625–32.

231. Bailey D, Bevilacqua V, Colantonio DA, *et al*. Pediatric within-day biological variation and quality specifications for 38 biochemical markers in the CALIPER cohort. *Clin Chem*. 2014;**60**:518–29.

232. Sweeting HN. Measurement and definitions of obesity in childhood and adolescence: a field guide for the uninitiated. *Nutr J*. 2007;**6**:32.

233. De Leonibus C, Marcovecchio ML, Chiavaroli V, de Giorgis T, Chiarelli F, Mohn A. Timing of puberty and physical growth in obese children: a longitudinal study in boys and girls. *Pediatr Obes*. 2014;**9**:292–9.

234. Lee JM, Kaciroti N, Appugliese D, Corwyn RF, Bradley RH, Lumeng JC. Body mass index and timing of pubertal initiation in boys. *Arch Pediatr Adolesc Med*. 2010;**164**:139–44.

235. Laron Z. Is obesity associated with early sexual maturation? *Pediatrics*. 2004;**113**(1 Pt 1):171–2; author reply 171–2.

236. Karpati AM, Rubin CH, Kieszak SM, Marcus M, Troiano RP. Stature and pubertal stage assessment in American boys: the 1988–1994 Third National Health and Nutrition Examination Survey. *J Adolesc Health*. 2002;**30**:205–12.

237. Lee JM, Wasserman R, Kaciroti N, *et al*. Timing of puberty in overweight versus obese boys. *Pediatrics*. 2016;**137**:e20150164.

238. Ohlsson C, Lorentzon M, Norjavaara E, Kindblom JM. Age at adiposity rebound is associated with fat mass in young adult males—the GOOD study. *PLoS One*. 2012;**7**:e49404.

239. Mills JL, Shiono PH, Shapiro LR, Crawford PB, Rhoads GG. Early growth predicts timing of puberty in boys: results of a 14-year nutrition and growth study. *J Pediatr*. 1986;**109**:543–7.

240. Hui LL, Wong MY, Lam TH, Leung GM, Schooling CM. Infant growth and onset of puberty: prospective observations from Hong Kong's 'Children of 1997' birth cohort. *Ann Epidemiol*. 2012;**22**:43–50.

241. Chen LK, Wang G, Bennett WL, *et al.* Trajectory of BMI from Ages 2 to 7 Years and Age at Peak Height Velocity in Boys and Girls. *J Pediatr*. 2020.

242. Buyken AE, Karaolis-Danckert N, Remer T. Association of prepubertal body composition in healthy girls and boys with the timing of early and late pubertal markers. *Am J Clin Nutr*. 2009;**89**:221–30.

243. Vignolo M, Naselli A, Di Battista E, Mostert M, Aicardi G. Growth and development in simple obesity. *Eur J Pediatr*. 1988;**147**:242–4.

244. Wagner IV, Sabin MA, Pfaffle RW, *et al*. Effects of obesity on human sexual development. *Nat Rev Endocrinol*. 2012;**8**:246–54.

245. Sorensen K, Juul A. BMI percentile-for-age overestimates adiposity in early compared with late maturing pubertal children. *Eur J Endocrinol*. 2015;**173**:227–35.

246. Kley HK, Deselaers T, Peerenboom H, Kruskemper HL. Enhanced conversion of androstenedione to estrogens in obese males. *J Clin Endocrinol Metab*. 1980;**51**:1128–32.

Errata

Page 38 Wrong data: "The mean age was 11.0 (range, 6.1-16.4) years" – corrected to "The mean age was 12.0 (range 6.5-16.4) years".

Paper I, Table 1 Wrong unit of measurement: "mm" corrected to "cm".

Paper IV, Figure 1 Y2 label: "Orchidometer volume, ml" corrected to "Equivalent orchidometer volume, ml".

Appendix

II

REGULAR ARTICLE

ACTA PÆDIATRICA
NURTURING THE CHILD
WILEY

Reference data for testicular volume measured with ultrasound and pubic hair in Norwegian boys are comparable with Northern European populations

Ninnie Helen Bakken Oehme[1,2] | Mathieu Roelants[3] | Ingvild Særvold Bruserud[1,2] | Andre Madsen[1,4] | Geir Egil Eide[5,6] | Robert Bjerknes[1,2] | Karen Rosendahl[7,8] | Petur B. Juliusson[1,2,9]

[1]Department of Clinical Science, University of Bergen, Bergen, Norway

[2]Department of Pediatrics, Haukeland University Hospital, Bergen, Norway

[3]Environment and Health, Department of Public Health and Primary Care, KU Leuven–University of Leuven, Leuven, Belgium

[4]The Hormone Laboratory, Haukeland University Hospital, Bergen, Norway

[5]Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway

[6]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway

[7]Department of Radiology, Haukeland University Hospital, Bergen, Norway

[8]Department of Clinical Medicine, University of Bergen, Bergen, Norway

[9]Department of Health Registries, Norwegian Institute of Public Health, Bergen, Norway

Correspondence
Ninnie Helen Bakken Oehme, Department of Clinical Science, University of Bergen, N-5020 Bergen, Norway.
Email: ninnie.oehme@uib.no

Funding information
The Western Norway Regional Health Authority, Grant number 911975, supported this study.

Abstract

Aim: To estimate references for testicular volume measured with ultrasound and Tanner stages of pubic hair in Norwegian boys, and to compare the timing of puberty with data from similar populations.

Methods: Testicular volume was derived from ultrasound measurements of testicular volume in a cross-sectional study of 514 healthy boys. A continuous testicular volume for age reference curve was estimated with the LMS method. Tanner stages for pubic hair were clinically assessed in 452 boys. Age references for pubertal milestones were estimated with probit regression.

Results: Puberty onset, defined by an ultrasound testicular volume of 2.7 mL, equivalent to an orchidometer volume of 4 mL, occurred at a mean (SD) age of 11.7 (1.1) years. The reference range was 9.7 (3rd) to 13.7 years (97th percentile). Pubic hair (Tanner stage 2) appeared on average at 11.8 (1.2) years with a corresponding reference range of 9.5-14.1 years.

Conclusion: The references for testicular volume measured with ultrasound are continuous in age and allow for the quantification of pubertal development. The age distribution of reaching pubertal milestones was comparable with data from other Northern European countries.

KEYWORDS
puberty, secular trend, testis, ultrasonography

1 | INTRODUCTION

During the past two decades, several authors have demonstrated renewed trends towards earlier puberty in girls, after a relatively stable period of almost 60 years.[1,2] Although results in boys are more equivocal,[3,4] some studies have suggested similar trends.[5,6] Overweight and obesity have been proposed as possible drivers for this renewed trend,[6-8] as well as exposure to endocrine-disrupting chemicals.[9]

Population-based studies of puberty are more challenging in boys than in girls, due to the lack of an easily measured, yet reliable pubertal marker like menarche.[4] Testicular examination with a Prader orchidometer is useful and widely used in clinical practice, but it is regarded as impractical for population studies.[10] In boys, attainment of a testicular volume (TV) of ≥4 mL when measured with a Prader orchidometer is considered the best indicator for the onset of male puberty.[11] It is therefore desirable that population studies also include assessments of TV. However, measuring TV with a Prader orchidometer may be perceived as intrusive outside of clinical context. In such situations, the use of ultrasound (US) to measure TV could be more acceptable because of the more technical nature of the examination. US is also the preferred method when the accuracy of TV is important.[12] In addition, this method has the advantage of detecting testicular pathology, which may explain developmental differences of testicular growth. We have previously shown that the measurement of TV with US is methodologically feasible and appropriate for the generation of pubertal references for TV.[13]

The aim of the current study was to estimate references for TV based on US measurements of the testicle and Tanner pubic hair (PH) staging in a representative cohort of healthy Norwegian boys.

2 | MATERIALS AND METHODS

2.1 | Childhood population

This study is a part of the Bergen Growth Study 2 on pubertal growth and development in Norway. All boys attending one of six randomly selected schools that provide primary and secondary education in the city of Bergen, Norway, were invited to participate in the study from January through June 2016. Parental consent was obtained for 493 (37%) out of 1329 eligible boys, but two boys did not assent and six were absent on the day of examination. In addition, we included US measurements from 58 boys who participated in a reliability study in 2017.[13] The age at examination was calculated from date of birth and date of examination. Eight boys with a disease or condition that could affect growth and 21 boys with a history of or present scrotal pathology (including cryptorchidism, hydrocele and microlithiasis) were excluded, and the Tanner PH stage was not registered in 62 boys. A parental questionnaire was obtained for 340 of the 514 (66.1%) boys included in the analysis. The questionnaire contained items on origin, chronic disease and previous genital pathology. Origin was grouped as both parents from Norway, one or both

Key notes

- The testicular volume is usually determined with a Prader orchidometer; however, ultrasound examinations are possibly a better alternative.
- We estimated up-to-date references for testicular volume, measured with ultrasound, and pubic hair in healthy Norwegian boys.
- Implementation of ultrasound for assessing testicular volume implies a transition towards an objective measurement on a continuous scale, which allows to detect smaller changes in the testicular volume and to quantify pubertal development.

parents from the European region, and one or both parents from outside the European region. Height and weight were measured in 457 of the boys included in the analysis. Based on the International Obesity Task Force (IOTF) body mass index (BMI) reference values,[14] 11.8% of boys were classified as overweight (IOTF-BMI ≥25 kg/m^2) and 1.9% as obese (IOTF-BMI ≥30 kg/m^2). This closely matches the 12.8% and 2.1% reported for Norwegian boys in this age range.[15]

2.2 | Ultrasound

All US examinations of the testis were performed by a single technician using a Sonosite Edge US machine with a 15-6 MHz linear probe. The length (L), width (W) and depth (D) of the right testicle were measured with the boy in a supine position according to a standardised protocol.[13] If the left testicle appeared larger by visual inspection, this was also measured, and the volume of the largest testicle was registered. The TV was calculated from the length, width and depth using the Lambert equation as TV = L×W×D×0.71. The observer variability of this method is 9.2% with a technical error of measurement (TEM) of 6.5%.[13] An empirical equation to predict Prader orchidometer volume from US volume was previously derived as $Vol_{OM} = 1.96 \times Vol_{US}^{0.71}$. The Prader orchidometer volume of ≥4 mL that defines puberty onset is thus equivalent to an US measured volume ≥2.7 mL.[13] A preliminary logistic regression analysis of pubertal onset (yes/no) according to age and origin showed no statistically significant differences between boys of Norwegian origin and European (P = .17) or non-European boys (P = .11). All boys were therefore included in the analysis of TV during puberty.

2.3 | Tanner staging

Tanner PH stages were visually assessed in the supine position by the same observer performing the US examinations. Illustrated descriptions based on the work of Marshall and Tanner served as a reference.[16] A preliminary analysis showed no statistical significant

difference in the timing of pubarche (Tanner PH stage 2) between boys of Norwegian origin and European (P = .82) or non-European boys (P = .59). All boys were therefore included in the analysis of pubic hair stages.

2.4 | Statistical analyses

A reference curve of the continuous US testicular volume for age was estimated with the LMS method.[17] The LMS method normalises the distribution of a variable by applying a Box-Cox power transformation to remove skewness from the data. The reference is summarised by three curves, representing the Box-Cox power to remove skewness (L), the mean (M) and the approximate coefficient of variation (S) along the independent covariate age. The amount of smoothing is expressed in terms of smoothing parameters or equivalent degrees of freedom (edf). For the TV, the optimal Box-Cox power L was determined to be constant at 0.5 (ie a square root transformation), the M-curve was fitted with 8 edf, and the S-curve with 4 edf. The tabulated values of L, M and S by age contain all the information that is needed to calculate any percentile, or to convert measurements into z-scores. Because L is a constant of 0.5, centiles can be derived using the simplified formula $C_\alpha = M(1 + Sz_\alpha/2)^2$, and z-scores as $Z = 2 \times (\sqrt{X/M} - 1)/S$, where z_α is the normal equivalent deviate that corresponds to the desired percentile.

In addition to the LMS reference curves for the continuous TV, we used probit regression within a generalised linear model (GLM) to estimate cumulative incidence curves for reaching TVs that correspond to selected discrete Prader orchidometer volumes, and for each of the Tanner PH stages. Non-parametric generalised additive models (GAM) provided identical results, which confirmed our assumption of a normal age distribution at the different pubertal milestones (data not shown).

All statistical analyses were performed using R version 3.4 (R foundation for Statistical Computing) or IBM SPSS statistics version 24 (IBM Corp).

2.5 | Ethical considerations

Written informed consent was obtained from a parent or legal guardian of each participant in the study, as well as assent from the participants themselves. A cinema voucher was given as an incentive. The study was approved by the Regional Committee for Medical and Health Research Ethics West (REC-WEST 2015/128).

3 | RESULTS

3.1 | Childhood population

The number of boys varied from 28 to 66 per age year between six and 16 years, and 12 boys were 16 years of age. Based on

information from the questionnaire, 77.4% had two Norwegian parents, 10% had one or two parents from another European country, and 12.5% had one or two parents from outside the European region. US examination of the scrotum revealed microlithiasis in one boy and testis located in the inguinal canal in two boys. In addition, we observed twelve cases of unilateral and six cases of bilateral cryptorchidism.

3.2 | Testicular volume

A total of 514 boys with a mean age of 11.0 years (range: 6.1-16.4 years) were included for the references. Figure 1 shows the US testicular volumes by age and the fitted median and ±2 SD lines. The corresponding L, M and S values are listed in Table 1, and selected percentiles are provided in Table S1. Figure 2 shows the cumulative incidence curves of selected discrete Prader orchidometer volumes by age, which were derived from the US volumes using the formula given in the methods section. The corresponding age quantiles are listed in Table 2 as SD scores and in the Table S2 as age percentiles. The mean age (SD) for attainment of a US measured TV of 2.7 mL (equivalent to a Prader orchidometer volume of 4 mL) was 11.7 (1.1) years, and the 3rd and 97th percentiles were respectively 9.7 and 13.7 years.

3.3 | Pubic hair

Tanner PH stage was determined in 452 (88.0%) boys with a mean age of 10.9 years (range 6.1-16.3 years). Figure 3 shows the cumulative incidence curves when boys reach Tanner PH stages 2-5. The
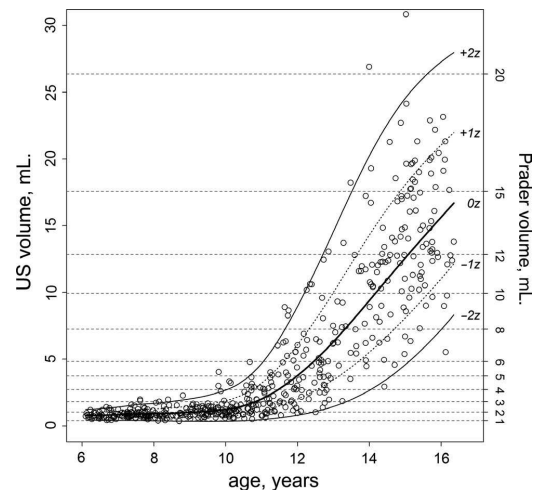

FIGURE 1 LMS-smoothed reference chart of ultrasound (US) measured testicular volume in 514 healthy Norwegian boys, aged 6-16 y. Corresponding equivalent Prader orchidometer volumes are shown on the right axis

Age	L	M	S	−2 SD	−1 SD	Mean	+1 SD	+2 SD
6	0.5	0.78	0.24622	0.4	0.6	0.8	1.0	1.2
7	0.5	0.83	0.30388	0.4	0.6	0.8	1.1	1.4
8	0.5	0.91	0.36146	0.4	0.6	0.9	1.3	1.7
9	0.5	1.00	0.41844	0.3	0.6	1.0	1.5	2.0
10	0.5	1.23	0.47844	0.3	0.7	1.2	1.9	2.7
11	0.5	2.02	0.52162	0.5	1.1	2.0	3.2	4.7
12	0.5	3.77	0.54007	0.8	2.0	3.8	6.1	8.9
13	0.5	6.30	0.51981	1.5	3.5	6.3	10.0	14.6
14	0.5	9.39	0.46762	2.7	5.5	9.4	14.3	20.2
15	0.5	12.58	0.39673	4.6	8.1	12.6	18.1	24.5
16	0.5	15.64	0.32065	7.2	11.0	15.6	21.1	27.0

TABLE 1 Age-specific references for ultrasound testicular volume (mL) estimated from 514 healthy Norwegian boys aged 6-16 y in 2016-2017 using the LMS method[a]

Abbreviations: L, skewness parameter; M, mean; S, coefficient of variation; SD, standard deviation.
[a]The LMS model was fitted on the original age scale with squared root transformed volumes (L = 0.5) and 8 and 4 equivalent degrees of freedom for the M and S curves.

FIGURE 2 Cumulative incidence of reaching selected equivalent Prader orchidometer volumes estimated with probit regression in 514 healthy Norwegian boys aged 6-16 y. Connected markers show the empirical data and bold lines the corresponding probit models

Prader	USV	M	SE	SD	−2 SD	−1 SD	+1 SD	+2 SD
2	1.0	8.95	0.16	2.01	4.9	6.9	11.0	13.0
3	1.8	11.05	0.11	1.25	8.6	9.8	12.3	13.6
4	2.7	11.67	0.11	1.07	9.5	10.6	12.7	13.8
5	3.7	12.25	0.11	1.01	10.2	11.2	13.3	14.3
6	4.8	12.66	0.11	0.97	10.7	11.7	13.6	14.6
8	7.2	13.49	0.12	1.10	11.3	12.4	14.6	15.7
10	9.9	14.08	0.13	1.20	11.7	12.9	15.3	16.5
12	12.8	15.10	0.15	1.32	12.5	13.8	16.4	17.7
15	17.6	16.27	0.32	1.54	13.2	14.7	17.8	19.4

TABLE 2 Age distribution (y) for testicular volumes corresponding to equivalent Prader orchidometer volumes (mL) in 514 healthy Norwegian boys aged 6-16 y in 2016-2017

Abbreviations: M, mean; Prader, equivalent Prader orchidometer volumes; SD, standard deviation; SE, standard error; USV, Ultrasound volume.

mean age (SD) of pubarche (Tanner PH stage 2) was 11.8 (1.2) years, and the corresponding reference range defined by the 3rd and 97th percentiles was 9.5-14.1 years. The mean ages and reference quantiles of other PH stages are listed in Table 3 as SD scores and in Table S3 as age percentiles.

The distribution of the continuous US TVs in boys who were classified as Tanner PH stages 1-5 is shown in Figure 4. There is both a substantial spread within, and overlap between, groups in terms of TV. When boys were classified as pre-pubertal (1-3 mL), pubertal (4-14 mL) or adult (≥15 mL) based on the equivalent Prader orchidometer volumes, 14.0% of boys with a pubertal testicular volume were characterised as Tanner PH1 (ie no pubarche), while amongst boys with a pre-pubertal testicular volume, 8.1% were characterised as ≥Tanner PH2 (pubarche). All boys with Tanner PH

FIGURE 3 Cumulative incidence of Tanner stages for pubic hair in 452 healthy Norwegian boys aged 6-16 y. Connected markers show the empirical observations and bold lines the corresponding probit models
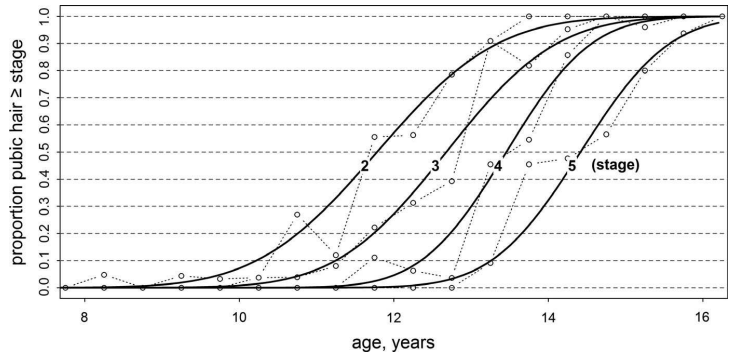

TABLE 3 Age distribution (y) by Tanner pubic hair stage (PH) in 452 healthy Norwegian boys aged 6-16 y in 2016-2017

PH	Mean	SE	SD	-2 SD	-1 SD	+1 SD	+2 SD
2	11.78	0.12	1.22	9.3	10.6	13.0	14.2
3	12.68	0.12	1.12	10.4	11.6	13.8	14.9
4	13.46	0.11	0.86	11.7	12.6	14.3	15.2
5	14.42	0.12	0.90	12.6	13.5	15.3	16.2

Abbreviations: PH, Tanner pubic hair stage; SD, standard deviation; SE, standard error.

stages 4 or 5 had attained either a pubertal or an adult volume of the testicles.

4 | DISCUSSION

In the current study, we present contemporary references for TV, obtained with US, and Tanner PH stages in 6- to 16-year-old Norwegian boys. By using LMS centile curves to summarise TV by age, the assessment of pubertal growth can now be quantified on a continuous scale since measurements can easily be converted to age-adjusted SD scores. In addition, we present age percentiles from probit analyses that document the cumulative incidence of reaching milestones of the development of TV and pubic hair.

In this paper, we report the very first references for pubertal development in Norwegian boys. Although it has been shown that, the timing of puberty in Northern European populations is very similar,[18] national reference data collected at regular intervals can help to detect secular trends earlier.[3] Up-to-date references are also important because early or late puberty may have consequences for the health of individual boys. Studies have found a protective effect of later puberty on testicular cancer,[19] but delayed puberty has also been linked to bullying, poor self-esteem and psychosocial distress.[20]

A testicular volume of 4 mL measured with a Prader orchidometer is commonly considered as a robust marker of the start of puberty.[11] However, we[13] and others [21] have previously shown that the orchidometer overestimates the true volume near this range and that the actual volume at the start of puberty is about 2.7 mL when measured

with ultrasound.[13] This also corresponded nicely with increased sex hormone levels.[22] While our references of TV are primarily based on US measurements, we previously devised a conversion formula that allows a seamless conversion from one method to the other.[13] Apart from being closer to the true TV, the US method has the additional advantage that the volume is measured on a continuous scale, contrary to the Prader orchidometer method which is limited to reaching a discrete set of volumes making it difficult to estimate volumes in between two consecutive beads or beyond all available beads.
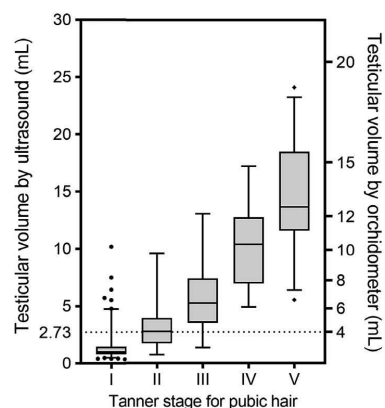

FIGURE 4 Box and whiskers plots of testicular volumes at different Tanner pubic hair stages in 452 Norwegian boys aged 6.1-16.4 y in 2016-2017. The boundaries of the box are the 1st and 3rd quartile. The median is identified by a line inside the box. The length of the box is the interquartile range (IQR)

The onset of puberty, defined by an US TV of 2.7 mL (4 mL with the Prader orchidometer), was reached at a mean age of 11.7 years. This is highly comparable with the 11.6 years observed in Dutch boys by Goede et al[23] and later remodelled by Joustra et al,[24] which is the only US reference for TV for adolescent boys published to date. In order to compare our data with these references, we multiplied the Dutch estimates with a factor 0.71/0.52 because these studies used the ellipsoid formula (L×W×D×0.52) to calculate volumes from testicular dimensions, while we used the Lambert formula (L×W×D×0.71) because it was found to give a better approximation.[25]

Our data are also in agreement with the age of attainment of an equivalent Prader orchidometer volume of 4 mL in other European countries, for example 11.4 years in Belgium,[26] 11.5 years in the Netherlands[27] and 11.7 years in Denmark.[6] While the Copenhagen Puberty Study reported a decline in age at onset of puberty of 3 months between 1991 and 2006, this trend was no longer significant after adjustment for BMI.[6] In the United States, the PROS study from 2005 to 2010 reported a mean age of 11.5 years in the non-Hispanic white population and 11.8 years in the African American population.[28] A comparison of our data with these references does not suggest a secular trend towards earlier puberty in boys over the last decade. However, compared with data from 109 Norwegian boys aged 1.9-16.9 years collected by Waaler in the 1970s,[29] contemporary boys reach a pubertal testicular volume approximately 2-3 months earlier, that is a rate of <1 month per decade.

Studies in European populations like Denmark,[30,31] Belgium,[26] Italy [32] and the Netherlands[27] reported relatively narrow range of mean age at pubarche (Tanner PH2) of 11.5-11.9 years. Our finding of 11.8 years corresponds with this range. One study from Denmark reported a mean age of 12.4 years, but this was a surprising finding because it implied a slow down with approximately five months between 1991 and 2006,[6] while other studies during the same period in Denmark reported stable average ages at pubarche of 11.6[31] and 11.9 years.[30] The PROS and NICHD studies from the United States both reported a mean age at Tanner PH2 in the non-Hispanic white population of 11.5 years,[28,33] which is about 3 months earlier compared with Norwegian boys.

Longitudinal studies have shown that around 46%-90% of boys enter puberty by the 'testicular' pathway, that is gonadal enlargement before the appearance of pubic hair (pubarche pathway).[11,31,33] The mean ages of attainment of a pubertal TV (11.7 years) and Tanner PH2 (11.8 years) in our study are consistent with this. Because our data are cross-sectional, we cannot estimate the duration of each stage nor the pace of progression throughout the various pubertal stages. However, a direct comparison of different pubertal markers showed that 14.0% of the boys reached a pubertal TV before pubic hair appeared, whereas only 8.1% showed pubic hair (Tanner PH2) prior to reaching a pubertal TV. Importantly, careful assessment of TV is the most reliable method to detect the earliest signs of puberty, whereas PH

staging alone may lead to misclassification of some boys in the earliest segment of pubertal maturation.[11]

Previous population studies have defined the normal physiological range for pubertal development in boys as 2.5 or 3 time the standard deviation below and above the mean.[18] In our study, the reference age range (mean ± 2.5 SD) of reaching a pubertal TV (US measured TV of 2.7 mL) is bounded by the ages of 9.0 and 14.3 years. We therefore recommend adhering to the current definition of normal pubertal onset in boys between 9 and 14 years.

A major strength of our study is the use of US for the measurement of TV in a population-based study. US provides the opportunity to obtain more accurate estimates of TV in comparison with a Prader orchidometer, without interference of surrounding tissues, such as the scrotal skin, the epididymis or the tunica vaginalis. Furthermore, US provides a continuous measure of volume, in contrast to the discrete ordinal Prader orchidometer beads, which allows for semiparametric data modelling and calculation of z-scores. US has the additional benefit of detecting testicular pathology, which may explain alterations in the timing of testicular growth, as in our cohort, we found one patient with testicular microlithiasis and two boys with testis located in the inguinal canal. Further, the majority (58%) of the participating boys from the test/retest study reported to prefer the examination of TV with US to direct palpation of the testicle. This may be explained by the less intrusive positioning of the examiner, facing the US machine rather than the scrotum directly, and that there is no direct contact between examiner's hand and the scrotum. Since US equipment and protocols are becoming more user-friendly and accessible, and because they are known to be safe and without a risk of ionising radiation, they might more readily be adopted for routine use by paediatricians and other clinicians.

Some limitations to the study needs to be addressed. Only 37% of the invited boys agreed to participate. This makes a selection bias possible, for instance, if boys maturing early or very late were less inclined to participate. In addition, only boys up to 16.4 years were included, potentially omitting the stabilisation of testicular growth at the adult range in our reference curve. Due to difficulties of recruitment and a potential reluctance regarding palpation of the testicles, examination with a Prader orchidometer was only performed in the reliability study. Based on these examinations, we were able to estimate a conversion equation to calculate Prader orchidometer volume from US volume. Another limitation was that only the right testicle was measured, except when the left testicle appeared larger by visual inspection. However, no statistical significant differences between left and right TV have been found in previous studies.[34]

5 | CONCLUSION

We have presented references for testicular growth based on US measurements of testicular dimensions, and for the clinical assessment of Tanner PH stages. Prader orchidometer has long been

considered a subjective clinical tool that is limited to an ordinal scale. Our implementation of an US protocol implies a transition towards an objective measurement on a continuous scale that allows to detect smaller changes in the TV and allows to quantify pubertal development. Further, US was the preferred examination method amongst the majority of the boys. The high degree of similarity of our data with previously published estimates of puberty onset in boys does not suggest an ongoing secular trend during the past decade.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

ORCID

Ninnie Helen Bakken Oehme https://orcid.org/0000-0002-0610-9571

REFERENCES

1. Aksglaede L, Sorensen K, Petersen JH, Skakkebaek NE, Juul A. Recent decline in age at breast development: the Copenhagen Puberty Study. *Pediatrics*. 2009;123(5):e932-e939.
2. Herman-Giddens ME, Slora EJ, Wasserman RC, et al. Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings network. *Pediatrics*. 1997;99(4):505-512.
3. Juul A, Teilmann G, Scheike T, et al. Pubertal development in Danish children: comparison of recent European and US data. *Int J Androl*. 2006;29(1):247-255.
4. Euling SY, Herman-Giddens ME, Lee PA, et al. Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. *Pediatrics*. 2008;121(suppl 3):S172-S191.
5. Herman-Giddens ME, Wang L, Koch G. Secondary sexual characteristics in boys: estimates from the national health and nutrition examination survey III, 1988–1994. *Arch Pediatr Adolesc Med*. 2001;155(9):1022-1028.
6. Sorensen K, Aksglaede L, Petersen JH, Juul A. Recent changes in pubertal timing in healthy Danish boys: associations with body mass index. *J Clin Endocrinol Metab*. 2010;95(1):263-270.
7. Reinehr T, Roth CL. Is there a causal relationship between obesity and puberty? *Lancet Child Adolesc Health*. 2019;3(1):44-54.
8. Lee JM, Wasserman R, Kaciroti N, et al. Timing of puberty in overweight versus obese boys. *Pediatrics*. 2016;137(2):e20150164.
9. Parent AS, Franssen D, Fudvoye J, Pinson A, Bourguignon JP. Current changes in pubertal timing: revised vision in relation with environmental factors including endocrine disruptors. *Endocr Dev*. 2016;29:174-184.
10. Abreu AP, Kaiser UB. Pubertal development and regulation. *Lancet Diabetes Endocrinol*. 2016;4(3):254-264.
11. Biro FM, Lucky AW, Huster GA, Morrison JA. Pubertal staging in boys. *J Pediatr*. 1995;127(1):100-102.
12. Kuijper EA, van Kooten J, Verbeke JI, van Rooijen M, Lambalk CB. Ultrasonographically measured testicular volumes in 0- to 6-year-old boys. *Hum Reprod*. 2008;23(4):792-796.
13. Oehme NHB, Roelants M, Bruserud IS, et al. Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys - intra- and interobserver agreement and comparison with Prader orchidometry. *Pediatr Radiol*. 2018;48(12):1771-1778.
14. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ*. 2000;320(7244):1240-1243.
15. Juliusson PB, Roelants M, Eide GE, Hauspie R, Waaler PE, Bjerknes R. Overweight and obesity in Norwegian children: Secular trends in weight-for-height and skinfolds. *Acta Paediatr*. 2007;96(9):1333-1337.
16. Rasmussen AR, Wohlfahrt-Veje C, Tefre de Renzy-Martin K, et al. Validity of self-assessment of pubertal maturation. *Pediatrics*. 2015;135(1):86-93.
17. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med*. 1992;11(10):1305-1319.
18. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr Rev*. 2003;24(5):668-693.
19. Weir HK, Kreiger N, Marrett LD. Age at puberty and risk of testicular germ cell cancer (Ontario, Canada). *Cancer Causes Control*. 1998;9(3):253-258.
20. Golub MS, Collman GW, Foster PM, et al. Public health implications of altered puberty timing. *Pediatrics*. 2008;121(suppl 3):S218-S230.
21. Diamond DA, Paltiel HJ, DiCanzio J, et al. Comparative assessment of pediatric testicular volume: orchidometer versus ultrasound. *J Urol*. 2000;164(3 Pt 2):1111-1114.
22. Madsen A, Oehme NB, Roelants M, et al. Testicular ultrasound to stratify hormone references in a cross-sectional Norwegian study of male puberty. *J Clin Endocrinol Metab*. 2019.
23. Goede J, Hack WW, Sijstermans K, et al. Normative values for testicular volume measured by ultrasonography in a normal population from infancy to adolescence. *Horm Res Paediatr*. 2011;76(1):56-64.
24. Joustra SD, van der Plas EM, Goede J, et al. New reference charts for testicular volume in Dutch children and adolescents allow the calculation of standard deviation scores. *Acta Paediatr*. 2015;104(6):e271-e278.
25. Hsieh ML, Huang ST, Huang HC, Chen Y, Hsu YC. The reliability of ultrasonographic measurements for testicular volume assessment: comparison of three common formulas with true testicular volume. *Asian J Androl*. 2009;11(2):261-265.
26. Roelants M, Hauspie R, Hoppenbrouwers K. References for growth and pubertal development from birth to 21 years in Flanders, Belgium. *Ann Hum Biol*. 2009;36(6):680-694.
27. Fredriks AM, van Buuren S, Burgmeijer RJ, et al. Continuing positive secular growth change in the Netherlands 1955–1997. *Pediatr Res*. 2000;47(3):316-323.
28. Herman-Giddens ME, Steffes J, Harris D, et al. Secondary sexual characteristics in boys: data from the pediatric research in office settings network. *Pediatrics*. 2012;130(5):e1058-e1068.
29. Waaler PE, Thorsen T, Stoa KF, Aarskog D. Studies in normal male puberty. *Acta Paediatr*. 1974;63(s249):1-36.
30. Wohlfahrt-Veje C, Mouritsen A, Hagen CP, et al. Pubertal onset in boys and girls is influenced by pubertal timing of both parents. *J Clin Endocrinol Metab*. 2016;101(7):2667-2674.
31. Mouritsen A, Aksglaede L, Soerensen K, et al. The pubertal transition in 179 healthy Danish children: associations between pubarche, adrenarche, gonadarche, and body composition. *Eur J Endocrinol*. 2013;168(2):129-136.
32. De Simone M, Danubio ME, Amicone E, Verrotti A, Gruppioni G, Vecchi F. Age of onset of pubertal characteristics in boys aged 6–14 years of the Province of L'Aquila (Abruzzo, Italy). *Ann Hum Biol*. 2004;31(4):488-493.

33. Susman EJ, Houts RM, Steinberg L, et al. Longitudinal development of secondary sexual characteristics in girls and boys between ages 91/2 and 151/2 years. *Arch Pediatr Adolesc Med*. 2010;164(2):166-173.

34. Tatsunami S, Matsumiya K, Tsujimura A, et al. Inter/intra investigator variation in orchidometric measurements of testicular volume by ten investigators from five institutions. *Asian J Androl*. 2006;8(3):373-378.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Reference data for testicular volume measured with ultrasound and pubic hair in Norwegian boys are comparable with Northern European populations

Ninnie Helen Bakken Oehme, Mathieu Roelants, Ingvild Særvold Bruserud, Andre Madsen, Geir Egil Eide, Robert Bjerknes, Karen Rosendahl, Petur B. Juliusson

Supplementary tables

Table S1 Percentiles of the ultrasound testicular volume (ml) reference by age, based on 514 healthy Norwegian boys aged 6-16 years

Age* (years)	P3	P10	P25	P50	P75	P90	P97
6	0.5	0.6	0.7	0.8	0.9	1.1	1.2
7	0.4	0.5	0.7	0.8	1.0	1.2	1.4
8	0.4	0.5	0.7	0.9	1.1	1.4	1.6
9	0.4	0.5	0.7	1.0	1.3	1.6	1.9
10	0.4	0.6	0.9	1.2	1.7	2.1	2.6
11	0.5	0.9	1.4	2.0	2.8	3.6	4.5
12	0.9	1.6	2.5	3.8	5.3	6.8	8.6
13	1.6	2.8	4.3	6.3	8.7	11.2	14.0
14	2.9	4.6	6.7	9.4	12.6	15.9	19.5
15	4.9	7.0	9.4	12.6	16.2	19.8	23.7
16	7.6	9.9	12.4	15.6	19.2	22.7	26.5

* Exact age; values derived from a continuous reference curve.

Table S2 Age percentiles (P) for attaining equivalent Prader orchidometer volumes based on a sample of 514 healthy Norwegian boys aged 6-16 years

Prader	USV	P3	P10	P25	P50	P75	P90	P97
2	1.0	5.2	6.4	7.6	8.9	10.3	11.5	12.7
3	1.8	8.7	9.4	10.2	11.1	11.9	12.7	13.4
4	2.7	9.7	10.3	10.9	11.7	12.4	13.0	13.7
5	3.7	10.4	11.0	11.6	12.2	12.9	13.5	14.1
6	4.8	10.8	11.4	12.0	12.7	13.3	13.9	14.5
8	7.2	11.4	12.1	12.7	13.5	14.2	14.9	15.6
10	9.9	11.8	12.5	13.3	14.1	14.9	15.6	16.3
12	12.8	12.6	13.4	14.2	15.1	16.0	16.8	17.6
15	17.6	13.4	14.3	15.2	16.3	17.3	18.2	19.2

Abbreviations and symbols: Prader = testicular volume measured with Prader orchidometer; USV = Ultrasound volume

Table S3 Age percentiles (P) for Tanner pubic hair stages (PH) 2 – 5 based on a sample of 452 healthy Norwegian boys aged 6-16 years

PH	P3	P10	P25	P50	P75	P90	P97
2	9.5	10.2	11.0	11.8	12.6	13.3	14.1
3	10.6	11.2	11.9	12.7	13.4	14.1	14.8
4	11.8	12.4	12.9	13.5	14.0	14.6	15.1
5	12.7	13.3	13.8	14.4	15.0	15.6	16.1

III

IV

Low BMI, but not high BMI, influences the timing of puberty in boys

Short title: Boys with a lower BMI enter puberty later

 *Ninnie Helen Bakken Oehme[1,2], MD, Mathieu Roelants[3], PhD, Ingvild Særvold Bruserud[1,2], MSN, PhD, Andre Madsen[4], PhD, Robert Bjerknes[1,2], MD, PhD, Karen Rosendahl[5,6], MD, PhD, Petur B. Juliusson[1,2,7], MD, PhD

1 Department of Clinical Science, University of Bergen, Bergen, Norway
2 Department of Pediatrics, Haukeland University Hospital, Bergen, Norway
3 Environment and Health, Department of Public Health and Primary Care, KU Leuven – University of Leuven, Belgium
4 Hormone Laboratory, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway
5 Department of Radiology, University Hospital of North Norway, Tromsø, Norway
6 Department of Clinical Medicine, University of Tromsø, The Arctic University of Norway, Tromsø, Norway
7 Department of Health Registry Research and Development, Norwegian Institute of Public Health

*Corresponding author:
Ninnie Oehme
Department of Clinical Science, University of Bergen
N-5020 Bergen, Norway
Telephone (mobile): +47 90540556
Email: ninnie.oehme@uib.no
Keywords: Associations, BMI, Puberty, Testicular volume, weight class

Abstract

Background: Previous studies investigating the association between weight status and onset of puberty in boys have been equivocal. It is currently unclear to what extent weight class influences puberty onset and progression.

Objectives: To explore the relationship between degree of sexual maturation and anthropometric measures in Norwegian boys.

Methods: The following endpoints were collected in a Norwegian cross-sectional study of 324 healthy boys aged 9-16 years: ultrasound-determined testicular volume (USTV), total serum testosterone, Tanner pubic hair stage, height, weight, waist circumference (WC), subscapular skin folds (SSF) and body fat percentage (%BF). Testicular volume-for-age z-scores for all boys were used to classify 'early', 'average', or 'late' maturing boys. Ordinal logistic regression analyses with a proportional odds model were applied to analyze the association between anthropometric variables and age-adjusted degree of pubertal development, with results expressed as age-adjusted odds ratio's (AOR). Cumulative incidence curves for reaching pubertal milestones were stratified by BMI.

Results: Boys with a low BMI-for-age (BMIz < -1) were less likely to have reached a pubertal testicular volume (USTV ≥ 2.7 mL) or a pubertal serum level of testosterone (≥ 0.5 nmol/L) compared to normal weight boys (AOR 0.3, p=0.038, AOR 0.3, p=0.026, respectively), and entered puberty on average with a delay of approximately eight months. Boys with high BMI-for-age (BMIz > 1) exhibited a comparable timing as normal weight boys. The same was found for WC. Pubertal markers were not associated with the SSF or %BF.

Conclusion: We found that a low BMI or a low WC for age were associated with a delayed timing of pubertal development in boys, whereas no significant association was observed for a high BMI or WC. Moreover, no significant effects of SSF or %BF were observed.

Introduction

Several studies have shown secular trends towards earlier puberty onset in girls during the past decades [1, 2]. Some studies suggest similar trends in boys [3, 4], but results are more equivocal [5]. The mechanism behind the onset of puberty and factors influencing this process are still not fully unraveled. Identification of modifiable causes of early puberty is however of great interest as early puberty is a known risk indicator for disease in adult men, such as type 2 diabetes, cardiovascular disease, and reproductive cancers [6, 7].

It has long been known that an adequate nutritional status is a requirement for a timely initiation of central pubertal development [8], and the secular increase in overweight and obesity has also received special attention as a potential driving factor for the concurrent secular trend towards earlier age at pubertal onset [9, 10]. Several studies have demonstrated earlier puberty in girls with a high BMI or obesity [11, 12, 13, 14], but findings in boys are more ambiguous. While some studies show that the BMI is negatively correlated with pubertal timing in overweight and obese boys [4, 15] others demonstrate *later* pubertal development in obese boys [16]. One study showed earlier puberty in overweight boys but delayed in obese [17].

The lack of consistent evidence regarding the effect of weight status on pubertal timing in boys might be due to difficulties obtaining reliable measures of pubertal timing or because these measures represent different benchmarks of puberty. A few studies report the testicular volume measured using a Prader orchidometer or a genital assessment using Tanner stages (Tanner G) [18], while others use proxy markers of pubertal onset and progression, such as peak height velocity [19] or age at voice breaking [20]. Attainment of a testicular volume ≥ 4 mL using the Prader orchidometer is the most widely used clinical marker for onset of puberty in boys, but the use of a Prader orchidometer is regarded as impractical for larger population studies [21]. At the same time, testicular ultrasound is considered to be a more precise method for volume assessment [22-25] and the implementation of an ultrasound protocol has the advantage of being a more objective measurement on a continuous scale [26], but may suffer from the same impracticality as the Prader assessment.

The aim of the current study was to investigate the relationship between anthropometric measures and age-adjusted degree of sexual maturation in Norwegian boys. In line with the literature, we hypothesized that boys with overweight or obesity would present with a more advanced pubertal development compared to boys with an average weight. Because of previous findings in the literature, boys with a low weight status were considered as a separate group in the analysis.

Materials and Methods

Childhood population: Participants were recruited as part of the Bergen Growth Study 2, a cross-sectional study of pubertal development and growth in Norwegian children. A total of 1329 boys between 6 and 16 years of age from six randomly selected public schools in Bergen, Norway, were invited to participate. Parental consent was obtained for the 493 (37%) boys included. The present analyses included 342 boys aged ≥ 9 years, to eliminate the strictly prepubertal population. One boy did not assent on the day of examination, and four boys were absent. In addition, four boys were excluded due to a condition or a disease likely to affect growth and development, and nine boys were excluded due to past or ad hoc evidence of scrotal pathology including cryptorchidism, hydrocele or microlithiasis, leaving 324 eligible boys for analysis. Evidence of scrotal pathology was coupled with personal referrals to our affiliated regional hospital for follow-up. The mean (range) age of the final sample was 12.3 (9.0-16.3) years. A parental questionnaire was obtained for 228 (70.4%) of the boys included in the analysis. The questionnaire contained items on country of origin, chronic disease, and previous genital pathology. Of the 217 (67%) with known country of origin of both parents, 165 (76.0%) had both parents from Norway, 22 (10.1%) had one or two European parents, and 30 (13.8%) had one or two non-European parents, mostly from Asia (n=11), Africa (n=8) or South America (n=7). The analyses include data from all boys, regardless of their country of origin.

Pubertal development and testicular volume: A trained pediatric radiographer performed all ultrasound examinations and anthropometric measurements. Length, depth, and width of the right testicle were measured with the boy in the supine position using a Sonosite Edge ultrasound machine with a 15-6 MHz linear probe according to a standardized protocol [27]. The testicular volume (TV) was calculated using the Lambert equation TV=length×width×depth×0.71 [28]. The intra-observer variability was 9.2% and the technical error of measurement 6.5% [27]. An empirical equation to predict the equivalent Prader orchidometer volume from ultrasound volume was previously derived as $Vol_{OM} = 1.96 \times Vol_{US}^{0.71}$, and the Prader orchidometer volume of ≥ 4 mL that defines puberty onset is thus equivalent to an ultrasound measured testicular volume ≥ 2.7 mL (USTV)[27]. The boys with a testicular volume below this cut-off (USTV < 2.7 mL, corresponding to Prader orchidometer volume of < 4 mL) were considered as prepubertal. Further, the boys were classified as early, average, or late maturing based on their testicular volume-for-age z-score (USTVz). The boys in the upper tertile (> 67th percentile) were considered as early maturing, those between percentiles 33-67 as average, and boys with the smallest testicular volume for age (< 33rd percentile, lower tertile) as late maturing (Fig. 1).

Tanner stages of pubic hair (PH) development were visually assessed in the supine position using descriptions based on the work of Marshall and Tanner as a reference [29] (n=321 boys). Tanner stage PH2 defined pubarche.
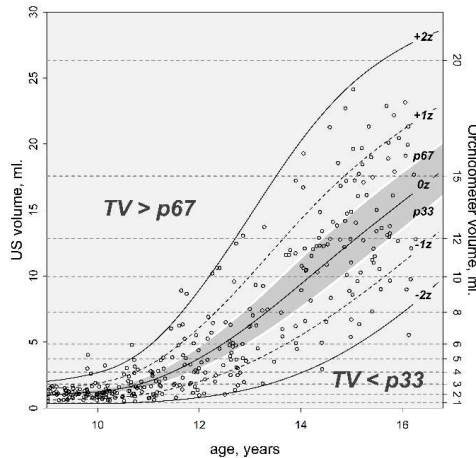

Figure 1 Grouping of boys as early (z-score > p67), average (p33 ≤ z-score ≤ p67) or late (z-score < p33) maturing based on testicular volume (TV) measured with ultrasound (US)[26]. The equivalent orchidometer volumes on the Y2 axis are calculated from the ultrasound measurements as $Vol_{OM} = 1.96 \times Vol_{US}^{0.71}$ (see text for details)[27]

Anthropometry: Height was measured in the standing position with a Harpenden Portable Stadiometer (Holtain Ltd Crosswell, UK) and recorded to the nearest 0.1 cm. Weight was measured in light clothing with an electronic scale (Tanita MC-780MA, Tanita Corp. of America, Inc. Illinois, USA) with a precision of 0.1 kg. Body mass index (BMI) was calculated as weight (kg) divided by the square of height (m^2). The waist circumference (WC) and subscapular skinfold (SSF) were measured according to the protocol used in the Bergen Growth Study 1 [30]. Further, the percentage of body fat (%BF) was assessed with bioelectrical impedance analysis (BIA), using a Tanita MC-780MA (Tanita corp. of America, Inc. Illinois, USA). The anthropometric measurements (BMI, WC and SSF) were converted to z-scores using the Norwegian growth reference charts from 2003-2006[31-33] while %BF z-scores were calculated using the references by McCarthy et al. [34]. Boys with a BMI z-score < -1 were classified as having a "low" BMIz, with a BMI z-score between -1 and 1 as "average", and those with a BMI z-score > 1 as having a "high" BMIz. The same cut-offs (z-scores -1 and 1) were also used for WC, SSF and %BF (WCz, SSFz and %BFz).

Blood test: Blood samples from 299 (92.3%) boys were collected between 0800 and 1400h and processed according to a protocol for blood sampling and analysis that was previously described [35]. Total testosterone was assayed by LC-MS/MS as described previously [36]. The analytical inter-assay coefficient of variation (CV%) was 4% in the range 1.5-37 nmol/L and limit of detection (LOD) was 0.01 nmol/L. A concentration of 0.5 nmol/L or more was used as an alternative marker for the start of puberty. This cutoff was determined with a ROC analysis of total testosterone to predict the onset of puberty defined as USTV ≥ 2.7 mL in 240 prepubertal and 180 pubertal boys in the BGS2. The area under the curve (AUC) was 0.9778 (95% CI; 0.96 to 0.99) and the positive and negative predictive values were 91.3% and 97.6%, respectively.

Statistical analysis: Continuous variables were compared between groups with a t-test and categorical variables with a chi-squared test. Multiple logistic regression with age as a covariate was used to estimate the odds ratio (OR) for having reached a pubertal level of either testicular volume (USTV≥ 2.7 mL), pubic hair (Tanner PH2), or serum testosterone (≥ 0.5 nmol/L) in boys with a high (> 1) or low (< -1) versus average (between -1 and 1) z-score for the different anthropometric measurements separately. Proportional odds logistic regression was used to study the association between the level of maturity (early, average or late based on the USTV z-scores) and the grouped anthropometric measurements, comparing boys with a "low" or "high" value to those with an average value for each measure separately. An OR larger than 1 means that boys in the tested group had a higher probability to be more advanced with respect to USTV for age. A non-significant score test indicated that the assumption of proportional odds was valid. Further we present the cumulative incidence curves for the three different pubertal markers in the three different weight groups BMIz < -1, - 1 ≤ BMIz ≥ 1 and BMIz > 1. The curves were estimated with a generalized additive model with a binary outcome and probit link function. The degree of smoothing was determined with generalized cross validation using the mgcv package in R. The mean age at reaching maturity (USTV 2.7 mL) was obtained by inverse prediction. All statistical analyses were performed using IBM SPSS statistics version 25 (IBM Corp) and R version 3.4 (R foundation for Statistical Computing).

Ethical considerations: This study was approved by the Norwegian Regional Committee for Medical and Health Research Ethics West (REC-WEST 2015/128). Written informed consent was obtained from a parent or legal guardian of each participant in the study, as well as assent from the participants themselves. A cinema voucher was given as an incentive.

Results

Of the 324 boys included in the analysis, 180 boys exhibited pubertal testicular volume USTV ≥ 2.7 mL (equivalent to ≥ 4 mL by orchidometer) and 144 had a volume USTV < 2.7 mL and were thus considered prepubertal. The youngest pubertal boy was 9.8 years, and the oldest prepubertal boy 13.1 years. Twenty-one boys presented with a prepubertal testicular volume, while pubic hair had already advanced to Tanner stage PH2. Only two of these had a pubertal serum testosterone level (≥ 0.5 nmol/L). The mean and SD of the z-scores for height, weight, BMI, WC and SSF for the whole group were not significantly different from the reference population in the Bergen growth study 1. Further, the z-scores for all anthropometric measures showed no significant difference between the prepubertal and pubertal boys. Based on the IOTF-criteria, 37 boys were defined as being overweight, and six as being obese. Further, 20 boys were defined as being underweight grade 1, and four as underweight grade 2. While BMI z-scores were not significantly different between the groups (p=0.310), the proportion of boys with a high BMI for age was larger in pubertal boys (16.7% vs 11.8%) but this difference was not statistically significant (p=0.267). Further, pubertal boys exhibited statistically significant lower %BF compared to the prepubertal boys (p=0.010).

Multiple logistic regression analysis with age as a covariate, showed that boys with a low BMIz had a lower probability of being pubertal (USTV ≥ 2.7 mL; AOR 0.3; 95% CI 0.1, 0.9; p=0.038) compared to boys with average BMIz (Table 1). Boys with a high BMIz did not have a significant higher probability of being pubertal (AOR 1.3; 95% CI 0.4,3.9; p=0.691). The same was observed for WC which showed a strong association with a low WCz, but not with high WCz. When these analyses were repeated for the other pubertal markers (serum testosterone ≥ 0.5 nmol/L and Tanner PH2), we could confirm the trend of an association with a low value for the BMIz and WCz but no clear association with a high BMIz or WCz, but it was only statistically significant for serum testosterone ≥ 0.5 nmol/L and not for Tanner PH2. No significant associations were found between SSF or %BF and any of the pubertal markers (Table 1).

Ordinal logistic regression showed that boys with low BMI or low WC for age had a significant lower probability of being in a higher category of testicular volume-for-age compared to those with average BMIz (OR 0.3; 95% CI 0.2,0.5; p<0.001) or WCz (OR 0.2; 95% CI 0.1,0.4; p<0.001) (Table 2). However, boys with high BMI or high WC for age did not have an increased probability of being in a higher category of testicular volume for age, as a sign of being more mature for age. We did not find any significant associations for SSF and %BF with the degree of maturation (Table 2).

Table 1 Age adjusted logistic regression analysis of having reached pubertal status according to different anthropometric measurements and markers of puberty

		USTV ≥ 2.7mL (N=324)				Serum testosterone ≥ 0.5 nmol/L (N=299)				Tanner PH2 (N=321)			
		N	A-OR	95%CI	p-value	N	A-OR	95%CI	p-value	N	A-OR	95%CI	p-value
BMI	Low	54	0.3	0.1,0.9	**0.038**	54	0.3	0.1,0.8	**0.026**	54	0.4	0.1,1.1	0.070
z-score	High	43	1.3	0.4,3.9	0.691	40	1.0	0.3,3.4	0.997	42	1.1	0.4,3.3	0.889
Waist	Low	36	0.2	0.0,0.6	**0.008**	35	0.2	0.1,0.9	**0.039**	36	0.3	0.1,1.1	0.079
z-score	High	45	0.9	0.3,2.9	0.918	42	1.1	0.3,3.7	0.850	44	1.2	0.4,3.5	0.761
SSF	Low	50	0.6	0.2,1.9	0.412	49	0.8	0.2,2.7	0.731	50	0.6	0.2,1.6	0.284
z-score	High	61	1.4	0.6,3.7	0.462	57	1.6	0.6,4.4	0.377	60	1.3	0.5,3.3	0.588
%BF	Low	32	0.5	0.1,2.1	0.363	28	0.8	0.2,3.6	0.724	32	1.5	0.4,5.8	0.555
z-score	High	51	1.6	0.6,4.7	0.387	47	1.6	0.5,5.0	0.456	51	1.1	0.4,3.2	0.811

AOR: Age adjusted odds ratio; USTV ≥ 2.7 mL: Pubertal testicular volume of 2.7 mL or more (ultrasound) or T4 mL (orchidometer); Tanner PH2: Pubarche; Low z-score: < -1; High z-score: > 1; BMI, body mass index; Waist, waist circumference; SSF, subscapular skinfold, %BF, body fat percentage. BMI, WC and SSF were converted to z-scores using the Norwegian growth reference from 2003-2006[31-33] while %BF z-scores were calculated using the references by McCarthy et al. [34]

The cumulative proportion of boys having attained a pubertal testicular volume in each of the three BMIz-groups separately is shown in Figure 2a. A comparison of the weight specific curves at the levels of the 50% attainment confirms that boys with low BMI for age (BMIz < -1) entered puberty with a delay of approximately eight months compared to normal weight boys, while the timing in boys with a high BMI for age (BMIz > 1) was comparable. The mean age of reaching a pubertal testicular volume was 12.34, 11.66, and 11.54 years in boys with a low, average, and high BMI for age respectively (Fig. 2a). Similar trends were observed for the attainment of a serum testosterone level above the threshold associated with puberty onset (serum testosterone ≥ 0.5 nmol/L; Figure 2b) and for the appearance of pubic hair (Tanner PH2; Figure 2c). For both pubertal markers, there is a clear delay in boys with a low BMIz, and a slight advancement in boys with a high BMIz. Also, the variability was smaller in these groups which resulted in steeper curves (Fig. 2b-c).

Table 2 Logistic regression and proportional odds logistic regression analysis of having a high (early maturing) or low (late maturing) testicular volume for age according to anthropometric measures

		N	USTV>p33			USTV>p67			Higher USTV tertile (proportional odds)		
			OR	95%CI	p-value	OR	95%CI	p-value	OR	95%CI	p-value
BMI z-score	Low	54	0.3	0.2,0.5	**<0.001**	0.2	0.1,0.5	**0.002**	0.3	0.2,0.5	**<0.001**
	High	43	1.0	0.5,2.1	0.981	1.2	0.6,2.3	0.627	1.1	0.6,2.1	0.731
Waist z-score	Low	36	0.2	0.1,0.4	**<0.001**	0.2	0.1,0.6	**0.008**	0.2	0.1,0.4	**<0.001**
	High	45	1.1	0.5,2.3	0.838	1.3	0.7,2.5	0.476	1.2	0.7,2.2	0.538
SSF z-score	Low	50	0.8	0.4,1.5	0.510	0.7	0.3,1.3	0.250	0.8	0.4,1.3	0.310
	High	61	1.0	0.6,1.9	0.913	1.1	0.6,2.0	0.720	1.1	0.6,1.8	0.774
%BF z-score	Low	32	0.7	0.3,1.5	0.308	1.0	0.4,2.1	0.903	0.8	0.4,1.6	0.478
	High	51	1.0	0.5,1.9	0.955	0.9	0.5,1.7	0.771	1.0	0.6,1.7	0.863

OR: Odds ratio; USTV>33p: this corresponds to the odds for being average or early vs. late maturing based on ultrasound measured testicular volume for age; USTV >67p: this corresponds to the odds for being early vs. average or late maturing; for the proportional odds model this corresponds to the odds for being in a higher category; Low z-score: < -1; High z-score: > 1; BMI, body mass index; Waist, waist circumference; SSF, subscapular skinfold, %BF, body fat percentage. BMI, WC and SSF were converted to z-scores using the Norwegian growth reference from 2003-2006[31-33] while %BF z-scores were calculated using the references by McCarthy et al. [34]
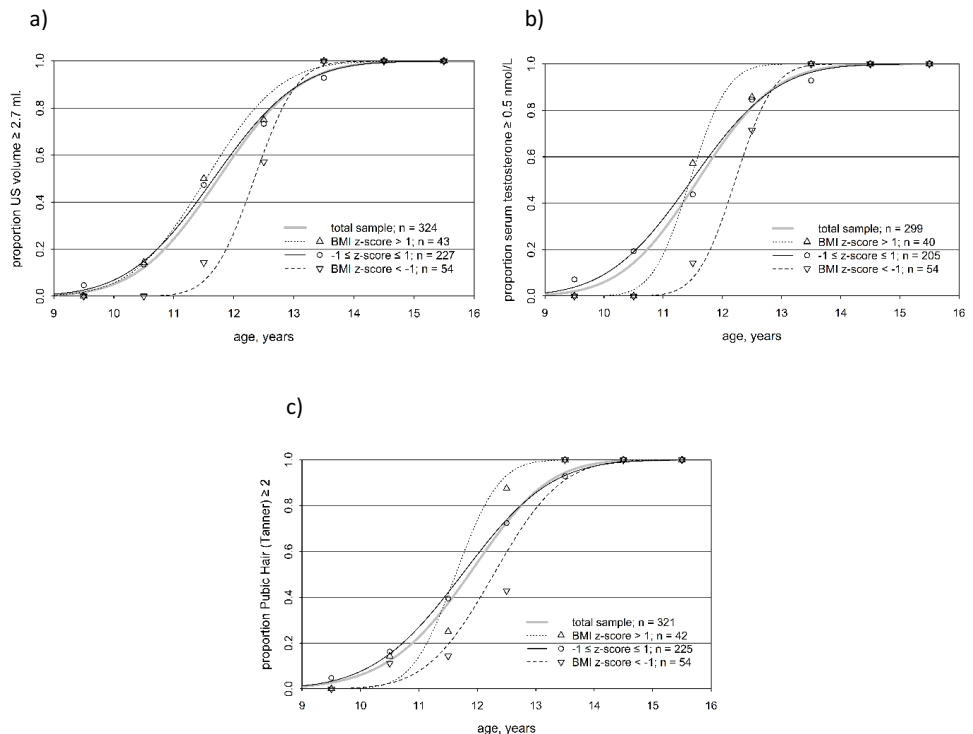

a)

b)

c)

Figure 2 a-c: Proportion of boys having attained a) a pubertal testicular volume (USTV ≥ 2.7 mL, n=324), b) a pubertal testosterone level (≥ 0.5 nmol/L, n=299) and c) Tanner stage 2 for pubic hair (PH2, n=321) in each of the three BMI z-groups in boys aged 9-16 years. A generalized additive model with probit link was used to estimate the cumulative distribution curve in each BMI group. The mean ages of reaching a pubertal marker in boys with a low, average, and high BMI for age was 12.34, 11.66 and 11.54 years for testicular volume (USTV ≥ 2.7 mL), 12.22, 11.48 and 11.46 years for serum testosterone level (≥ 0.5 nmol/L) and 12.28, 11.74 and 11.63 years for Tanner PH2. BMI z-scores were calculated using references from the Bergen Growth Study 1[31], USTV z-scores were calculated using references from the Bergen Growth Study 2 [26]

Discussion

In the current study, we examined the association between the timing of sexual maturation and a low or high weight status in a cross-sectional cohort of healthy boys. We found that boys with a low BMIz and a low WCz reached puberty almost eight months later than those with an average BMIz or WCz and were delayed over the whole pubertal age range as demonstrated by the smaller testicular volume by age. On the other hand, neither a high BMI nor high WC for age were associated with earlier maturity as originally anticipated. These results were confirmed for puberty onset according to the level of serum testosterone.

Our endpoints for male puberty status included measurements of testicular volume with ultrasound, a pubertal level of serum testosterone, and the development of pubic hair as described by Marshall and Tanner [29]. Indisputably, the best and most objective clinical marker of male puberty is the assessment of testicular volume [37]. The size of the testicle is traditionally assessed by Prader orchidometry, but measurements of testicular dimensions with ultrasound have been shown to be the preferred method when accuracy of testicular volume is important [38]. In addition, the ultrasound volume is a continuous variable which facilitated the development of testicular volume-for-age reference charts [26, 39]. Age adjusted testicular volume z-scores calculated with the Norwegian references [26] allowed us to stratify boys into tertiles of pubertal progress, with the 33rd and 67th percentiles as cutoffs for late, average, and early maturation. Sørensen and Juul previously used a similar approach based on the discrete testicular volume measured with a Prader orchidometer [40], while Ribeiro et al. divided the boys into quartiles based on age and Tanner G stage [41].

To assess the association of adiposity and body composition on the timing of puberty and degree of maturation we stratified boys into three groups according to their BMI, WC, SSF and %BF for age z-score. Boys with a z-score below -1 were considered as low, and those with a z-score above 1 as high. The effect of having a low "weight status" was analyzed separately since previous studies revealed effects of low vs. average values for anthropometric variables that were independent from the high values [14, 42]. For instance, Tomova et al. studied more than 4 000 boys between 7 and 19 years of age[42]. They observed that boys with a low BMI (< 12th percentile) were delayed at every stage of pubertal development, while boys with a high BMI (> 85th percentile) started puberty at an earlier age and reached the final stage of puberty ahead of their normal weight peers[42]. But most previous studies have compared pubertal development in overweight versus non-overweight subjects without considering low weight class as a separate group [15, 43].

It is well known that energy homeostasis is an important factor for the timing of puberty and that adequate nutrition is key for normal puberty [44]. The satiety hormone leptin produced in fat cells has

been suggested as a possible link between weight status and pubertal timing [45]. Our finding that boys with a low BMI and WC for age were delayed is therefore not surprising and is supported by others [42, 46]. The finding that boys with a high BMIz did not significantly differ from normal weight boys and thus not achieve pubertal milestones at an earlier age was more surprising given the numerous studies reporting an association between adiposity and earlier puberty onset [4, 15, 19, 20, 41, 47-49]. However, even though we did not find an association for a high BMIz, we cannot exclude that this is due to the limited number of boys with overweight, and even lower number with obesity.

Busch et al. recently demonstrated that boys with obesity (defined as BMIz > 2) experienced earlier timing of testicular enlargement (mean age 11.3 years), as compared to control group with a BMIz < 2 (mean age 11.7 years) [15]. However, all boys with a BMI z-score of 0 to 1, 1 to 2, and 2 to 3 entered puberty at the same mean age of 11.4 years, while boys with a BMIz 0 to -1 entered puberty at a mean age of 11.9 years and those with a BMIz below -1 at 12.4 years. Their conclusion of an advancement in boys with obesity could thus also be interpreted as a delay in boys with a low BMIz in line with our current findings. Another Danish study using self-reported pubertal data also concluded that overweight boys reached Tanner G2 almost three months earlier than normal weight boys [43], but a normal weight was defined as any BMI below the 85th percentile. Further scrutiny of the tabulated results confirmed that boys with low BMI (< 16 kg/m^2) appeared to reach Tanner G2 at an older age than those with a higher weight.

In the current study, WC, a proxy for abdominal fat that has shown a stronger association with cardiovascular risk than BMI [50], followed that for BMI, in that boys with lower WC for age had lower probability of being more mature than their peers, while having a larger WCz was not associated with earlier maturation. This contrasts with a recent study from Brazil showing that boys with early pubertal development presented higher prevalence of central adiposity, which was defined as increased WC [51].

No significant differences were found between SSFz and %BFz and early or late maturing boys in the present study. SSF is a direct measure of subcutaneous (trunk) fat, and the %BF measured with BIA, is generally considered to be more sensitive and specific for grading adiposity than anthropometric indices such as the BMI [52]. Vizmanos and colleagues measured skinfolds and %BF in a longitudinal study of 282 boys [53]. They found that the BMI increased with age at onset of puberty in boys, but since the amount of body fat mass was constant, it was concluded that puberty onset initiates with a characteristic accumulation of subcutaneous body fat mass that is independent of the age of puberty onset. In contrast to this, Biro et al. found that boys with more advanced maturation at age 12 had

lower sum of skin folds, and that boys who arrived at any given maturation stage at a younger age had lower BMI and lower adiposity [54].

We did not find an association between high BMI and pubertal timing as anticipated. Because of the cross-sectional design, we can only describe the associations, but not causality between weight class and pubertal timing. Conclusions drawn from cross-sectional studies are vulnerable to potential confounding by reverse causality, i.e. that children could be assigned to wrong weight classes due to early or late puberty onset, or due to differential tempo of growth [55]. Sørensen and Juul found that early pubertal timing was not associated with a degree of higher adiposity, measured with BIA, and that BMIz tended to overestimate adiposity and more readily classified children as overweight in early versus late maturing children [40]. Considering the associations found for BMI and WC, but not for SSF and %BF, may imply that BMI is a marker of maturity more than adiposity.

The conflicting results in association studies between weight class and pubertal timing are striking, however, it is plausible that differences in methods to assess pubertal development and different definitions of obesity have contributed to a diverging range of conclusions. Moreover, the lack of longitudinal studies limits the possibility of defining the causal relationship between obesity and pubertal maturation. These inconsistencies warrant further investigations using a longitudinal design and consensus endpoints to determine puberty onset to solve the effect of adiposity on pubertal timing.

In addition to the cross-sectional design, another limitation of the current study is the potential of selection bias. Only 37% of the invited boys agreed to participate, potentially making very early or late maturing boys, less inclined to participate. In addition, non-significant findings should be interpreted cautiously since the relatively small number of boys with a high (> 1) or low (< -1) z-score for anthropometric measurements (the expected prevalence is 16%) may have impacted the statistical power of our analysis.

A major strength of our study is the use of ultrasound, which facilitated measurements of the testicular volume on a continuous scale, without the interference of the surrounding scrotal tissue. This, in turn enabled the calculation of age-adjusted z-scores for each study participant in accordance with our previously published reference chart [26]. We have previously shown that the USTV of 2.7 mL immediately precedes a drastic surge in testosterone levels [35] and our current findings for the associations between testicular volume and anthropometric measurements were corroborated by equivalent findings with regard to serum testosterone. This highlights the co-occurrence of testicular enlargement and testosterone production. Another strength is that we not only included BMI, but also WC, SSF and %BF in addition to blood tests in a quite large cohort of healthy boys.

Conclusion

A good understanding of the relationship between sexual maturation and weight status has many important clinical and public health implications. We have demonstrated an association between a low BMI for age and pubertal timing, but no association for high BMI for age was found. Boys with a low BMIz entered puberty with a delay of eight months. We found that variables related to shape (BMI and WC) were significant in relation to pubertal timing, whereas variables related to composition (SSF and %BF) were not. Weight status should therefore always be taken into consideration when assessing pubertal status in children and adolescents.

Author Contributions:

Dr. Oehme coordinated and supervised data collection carried out initial analyses and interpretation and drafted the initial manuscript and reviewed and revised the manuscript.

Dr. Roelants carried out initial analysis, substantial statistical work and critically reviewed the manuscript.

Mrs. Bruserud coordinated, supervised, and collected data, and reviewed the manuscript.

Dr. Madsen contributed with statistical analysis and interpretation of data and reviewed the manuscript.

Prof. Bjerknes contributed with conceptualization and design of the study and reviewed the manuscript.

Prof. Rosendahl contributed with the design of the study, supervision and collection of data, and revision of the manuscript.

Prof. Júlíusson conceptualized and designed the study, supervised data collection, and critically reviewed the manuscript.

All authors approved the final manuscript as submitted and agree to be accountable for all the aspects of the work.

References:

1. Aksglaede L, Sorensen K, Petersen JH, Skakkebaek NE, Juul A. Recent decline in age at breast development: the Copenhagen Puberty Study. *Pediatrics*. 2009;123(5):e932-939.
2. Herman-Giddens ME, Slora EJ, Wasserman RC, Bourdony CJ, Bhapkar MV, Koch GG, et al. Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings network. *Pediatrics*. 1997;99(4):505-512.
3. Herman-Giddens ME, Wang L, Koch G. Secondary sexual characteristics in boys: estimates from the national health and nutrition examination survey III, 1988-1994. *Arch Pediatr Adolesc Med*. 2001;155(9):1022-1028.
4. Sorensen K, Aksglaede L, Petersen JH, Juul A. Recent changes in pubertal timing in healthy Danish boys: associations with body mass index. *The Journal of clinical endocrinology and metabolism*. 2010;95(1):263-270.
5. Goede J, Hack WW, Sijstermans K, van der Voort-Doedens LM, Van der Ploeg T, Meij-de Vries A, et al. Normative values for testicular volume measured by ultrasonography in a normal population from infancy to adolescence. *Hormone research in paediatrics*. 2011;76(1):56-64.
6. Day FR, Elks CE, Murray A, Ong KK, Perry JR. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. *Scientific reports*. 2015;5:11208.
7. Golub MS, Collman GW, Foster PM, Kimmel CA, Rajpert-De Meyts E, Reiter EO, et al. Public health implications of altered puberty timing. *Pediatrics*. 2008;121 Suppl 3:S218-230.
8. Burt Solorzano CM, McCartney CR. Obesity and the pubertal transition in girls and boys. *Reproduction*. 2010;140(3):399-410.
9. Reinehr T, Roth CL. Is there a causal relationship between obesity and puberty? *The Lancet Child & adolescent health*. 2019;3(1):44-54.
10. Kaplowitz PB, Slora EJ, Wasserman RC, Pedlow SE, Herman-Giddens ME. Earlier onset of puberty in girls: relation to increased body mass index and race. *Pediatrics*. 2001;108(2):347-353.
11. Wang Y. Is obesity associated with early sexual maturation? A comparison of the association in American boys versus girls. *Pediatrics*. 2002;110(5):903-910.
12. Rosenfield RL, Lipton RB, Drum ML. Thelarche, pubarche, and menarche attainment in children with normal and elevated body mass index. *Pediatrics*. 2009;123(1):84-88.
13. Currie C, Ahluwalia N, Godeau E, Nic Gabhainn S, Due P, Currie DB. Is obesity at individual and national level associated with lower age at menarche? Evidence from 34 countries in the Health Behaviour in School-aged Children Study. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*. 2012;50(6):621-626.
14. Bratke H, Bruserud IS, Brannsether B, Assmus J, Bjerknes R, Roelants M, et al. Timing of menarche in Norwegian girls: associations with body mass index, waist circumference and skinfold thickness. *BMC pediatrics*. 2017;17(1):138.
15. Busch AS, Hojgaard B, Hagen CP, Teilmann G. Obesity is associated with earlier pubertal onset in boys. *The Journal of clinical endocrinology and metabolism*. 2019.
16. Kleber M, Schwarz A, Reinehr T. Obesity in children and adolescents: relationship to growth, pubarche, menarche, and voice break. *Journal of pediatric endocrinology & metabolism : JPEM*. 2011;24(3-4):125-130.
17. Lee JM, Wasserman R, Kaciroti N, Gebremariam A, Steffes J, Dowshen S, et al. Timing of Puberty in Overweight Versus Obese Boys. *Pediatrics*. 2016;137(2):e20150164.
18. Herman-Giddens ME, Steffes J, Harris D, Slora E, Hussey M, Dowshen SA, et al. Secondary sexual characteristics in boys: data from the Pediatric Research in Office Settings Network. *Pediatrics*. 2012;130(5):e1058-1068.
19. He Q, Karlberg J. Bmi in childhood and its association with height gain, timing of puberty, and final height. *Pediatric research*. 2001;49(2):244-251.

20. Juul A, Magnusdottir S, Scheike T, Prytz S, Skakkebaek NE. Age at voice break in Danish boys: effects of pre-pubertal body mass index and secular trend. *International journal of andrology*. 2007;30(6):537-542.
21. Abreu AP, Kaiser UB. Pubertal development and regulation. *The lancet Diabetes & endocrinology*. 2016;4(3):254-264.
22. Diamond DA, Paltiel HJ, DiCanzio J, Zurakowski D, Bauer SB, Atala A, et al. Comparative assessment of pediatric testicular volume: orchidometer versus ultrasound. *The Journal of urology*. 2000;164(3 Pt 2):1111-1114.
23. Paltiel HJ, Diamond DA, Di Canzio J, Zurakowski D, Borer JG, Atala A. Testicular volume: comparison of orchidometer and US measurements in dogs. *Radiology*. 2002;222(1):114-119.
24. Rivkees SA, Hall DA, Boepple PA, Crawford JD. Accuracy and reproducibility of clinical measures of testicular volume. *J Pediatr*. 1987;110.
25. Fuse H, Takahara M, Ishii H, Sumiya H, Shimazaki J. Measurement of testicular volume by ultrasonography. *International journal of andrology*. 1990;13(4):267-272.
26. Oehme NHB, Roelants M, Saervold Bruserud I, Madsen A, Eide GE, Bjerknes R, et al. Reference data for testicular volume measured with ultrasound and pubic hair in Norwegian boys are comparable with Northern European populations. *Acta paediatrica (Oslo, Norway : 1992)*. 2020.
27. Oehme NHB, Roelants M, Bruserud IS, Eide GE, Bjerknes R, Rosendahl K, et al. Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys - intra- and interobserver agreement and comparison with Prader orchidometry. *Pediatric radiology*. 2018;48(12):1771-1778.
28. Lambert B. The frequency of mumps and of mumps orchitis and the consequences for sexuality and fertility. *Acta Genet Stat Med*. 1951;2.
29. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Archives of disease in childhood*. 1970;45(239):13-23.
30. Juliusson PB, Roelants M, Eide GE, Hauspie R, Waaler PE, Bjerknes R. Overweight and obesity in Norwegian children: secular trends in weight-for-height and skinfolds. *Acta paediatrica (Oslo, Norway : 1992)*. 2007;96(9):1333-1337.
31. Juliusson PB, Roelants M, Nordal E, Furevik L, Eide GE, Moster D, et al. Growth references for 0-19 year-old Norwegian children for length/height, weight, body mass index and head circumference. *Annals of human biology*. 2013;40(3):220-227.
32. Brannsether B, Roelants M, Bjerknes R, Juliusson PB. Waist circumference and waist-to-height ratio in Norwegian children 4-18 years of age: reference values and cut-off levels. *Acta paediatrica (Oslo, Norway : 1992)*. 2011;100(12):1576-1582.
33. Brannsether B, Roelants M, Bjerknes R, Juliusson PB. References and cutoffs for triceps and subscapular skinfolds in Norwegian children 4-16 years of age. *European journal of clinical nutrition*. 2013;67(9):928-933.
34. McCarthy HD, Cole TJ, Fry T, Jebb SA, Prentice AM. Body fat reference curves for children. *International journal of obesity (2005)*. 2006;30(4):598-602.
35. Madsen A, Oehme NB, Roelants M, Bruserud IS, Eide GE, Viste K, et al. Testicular ultrasound to stratify hormone references in a cross-sectional Norwegian study of male puberty. *The Journal of clinical endocrinology and metabolism*. 2019.
36. Methlie P, Hustad SS, Kellmann R, Almås B, Erichsen MM, Husebye E, et al. Multisteroid LC-MS/MS assay for glucocorticoids and androgens, and its application in Addison's disease. *Endocr Connect*. 2013;2(3):125-136.
37. Biro FM, Lucky AW, Huster GA, Morrison JA. Pubertal staging in boys. *J Pediatr*. 1995;127(1):100-102.
38. Kuijper EA, van Kooten J, Verbeke JI, van Rooijen M, Lambalk CB. Ultrasonographically measured testicular volumes in 0- to 6-year-old boys. *Human reproduction (Oxford, England)*. 2008;23.

39.   Joustra SD, van der Plas EM, Goede J, Oostdijk W, Delemarre-van de Waal HA, Hack WW, et al. New reference charts for testicular volume in Dutch children and adolescents allow the calculation of standard deviation scores. *Acta paediatrica (Oslo, Norway : 1992)*. 2015;104(6):e271-278.

40.   Sorensen K, Juul A. BMI percentile-for-age overestimates adiposity in early compared with late maturing pubertal children. *European journal of endocrinology*. 2015;173(2):227-235.

41.   Ribeiro J, Santos P, Duarte J, Mota J. Association between overweight and early sexual maturation in Portuguese boys and girls. *Annals of human biology*. 2006;33(1):55-63.

42.   Tomova A, Robeva R, Kumanov P. Influence of the body weight on the onset and progression of puberty in boys. *Journal of pediatric endocrinology & metabolism : JPEM*. 2015;28(7-8):859-865.

43.   Brix N, Ernst A, Lauridsen LLB, Parner ET, Arah OA, Olsen J, et al. Childhood overweight and obesity and timing of puberty in boys and girls: cohort and sibling-matched analyses. *International journal of epidemiology*. 2020.

44.   Muñoz-Calvo MT, Argente J. Nutritional and Pubertal Disorders. *Endocrine development*. 2016;29:153-173.

45.   Kiess W, Reich A, Meyer K, Glasow A, Deutscher J, Klammt J, et al. A role for leptin in sexual maturation and puberty? *Hormone research*. 1999;51 Suppl 3:55-63.

46.   Heger S, Korner A, Meigen C, Gausche R, Keller A, Keller E, et al. Impact of weight status on the onset and parameters of puberty: analysis of three representative cohorts from central Europe. *Journal of pediatric endocrinology & metabolism : JPEM*. 2008;21(9):865-877.

47.   Aksglaede L, Juul A, Olsen LW, Sorensen TI. Age at puberty and the emerging obesity epidemic. *PloS one*. 2009;4(12):e8450.

48.   Mamun AA, Hayatbakhsh MR, O'Callaghan M, Williams G, Najman J. Early overweight and pubertal maturation--pathways of association with young adults' overweight: a longitudinal study. *International journal of obesity (2005)*. 2009;33(1):14-20.

49.   Sandhu J, Ben-Shlomo Y, Cole TJ, Holly J, Davey Smith G. The impact of childhood body mass index on timing of puberty, adult stature and obesity: a follow-up study based on adolescent anthropometry recorded at Christ's Hospital (1936-1964). *International journal of obesity (2005)*. 2006;30(1):14-22.

50.   Maffeis C, Corciulo N, Livieri C, Rabbone I, Trifirò G, Falorni A, et al. Waist circumference as a predictor of cardiovascular and metabolic risk factors in obese girls. *European journal of clinical nutrition*. 2003;57(4):566-572.

51.   Adami F, Benedet J, Takahashi LAR, da Silva Lopes A, da Silva Paiva L, de Vasconcelos FAG. Association between pubertal development stages and body adiposity in children and adolescents. *Health Qual Life Outcomes*. 2020;18(1):93.

52.   Houtkooper LB, Lohman TG, Going SB, Howell WH. Why bioelectrical impedance analysis should be used for estimating adiposity. *The American journal of clinical nutrition*. 1996;64(3 Suppl):436s-448s.

53.   Vizmanos B, Marti-Henneberg C. Puberty begins with a characteristic subcutaneous body fat mass in each sex. *European journal of clinical nutrition*. 2000;54(3):203-208.

54.   Biro FM, Khoury P, Morrison JA. Influence of obesity on timing of puberty. *International journal of andrology*. 2006;29(1):272-277; discussion 286-290.

55.   Ong KK, Ahmed ML, Dunger DB. Lessons from large population studies on timing and tempo of puberty (secular trends and relation to body size): the European trend. *Molecular and cellular endocrinology*. 2006;254-255:8-12.

uib.no