NOVEL WAYS OF ASSESSING PUBERTY

Findings from the Adolescent Rural Cohort Study of Hormones, Health, Education, Environments and Relationships

Ben W.R. Balzer
B.Med.Sc., M.B., B.S. (Syd.)

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Discipline of Child and Adolescent Health,
Sydney Medical School,
Faculty of Medicine and Health,
The University of Sydney

Primary Supervisor: Professor Katharine S. Steinbeck
Medical Foundation Chair in Adolescent Medicine,
The University of Sydney

2019
STATEMENT OF AUTHENTICATION

This thesis is submitted to the University of Sydney in fulfilment of the requirement is for the Degree of Doctor of Philosophy.

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

Signature:

Ben W.R. Balzer

Date: 3 June 2019
ACKNOWLEDGEMENTS

I would like to thank Professor Kate Steinbeck for her role as my primary supervisor. I have worked with Kate since 2010, when I first joined the Academic Department of Adolescent Medicine as a Summer Research Scholar. Her expertise, supervision and support have been inspiring and have brought this thesis to fruition. Thank you also to my associate supervisors, Professor David Handelsman and Associate Professor Catherine Hawke, whose time and input have benefitted me greatly and been vital to this project’s success.

I wish to acknowledge the National Health and Medical Research Council Project Grant awarded to Professor Steinbeck and I am grateful for personal support from the University of Sydney Faculty of Medicine through the Postgraduate Research Support Scheme and the Dean’s Scholarship Fund Conference Grant.

Additionally, I am thankful for the assistance of Tina Cunningham, Dr. Hoi Lun (Helen) Cheng, Karen Paxton and Mazen Amatoury, as well as the contributions of my co-authors: Associate Professor Patrick Kelly, Professor Philip Hazell, Dr. Gurmeet Surindar Singh, Dr. Frances Garden and Dr. Georgina Luscombe. I would also like to acknowledge the participants of the Adolescent Rural Cohort, Hormones and Health, Education, Environments and Relationships (ARCHER) study.

Finally, to my parents Ben and Sandra, sisters Bianca and Isabella, my grandparents and extended family for their unwavering love and continual encouragement. Also, Con and Maria James for their support and originally suggesting I apply to work with Kate in 2010. Above all to my wife Anthea – without your love, patience and support, none of this would have been possible.
ETHICS APPROVAL

The Adolescent Rural Cohort study of Hormones, health, Education, environments and Relationships (ARCHER) has ethical approval from the Human Research Ethics Committee, University of Sydney (HREC 13094) within the National Health and Medical Research Council Guidelines for Human Experimentation, which is consistent with the Declaration of Helsinki. Prior to commencing the study, all participants assented, and a parent/guardian provided written informed consent.
AUTHORSHIP ATTRIBUTION STATEMENT


I designed and performed the literature search, extracted the data and wrote the drafts of the manuscript. Dr Duke and A/Prof Hawke provided critical revisions of the manuscript. Prof Steinbeck also assisted with data analysis and providing critical revision of the manuscript.


I assisted with data analysis, study data extraction and writing the drafts of the manuscript. Dr Duke was the lead author and responsible for the literature search and principal data extraction. Prof Steinbeck also assisted with data analysis and providing critical revision of the manuscript.


I designed the analysis plan, analysed the collected data and wrote the drafts of the manuscript. A/Prof Kelly assisted with data analysis. Prof Hazell, Ms Paxton and A/Prof Hawke made substantial contributions to the overall ARCHER study. Prof
Steinbeck is the principal investigator of the ARCHER study and has overseen its conception, design and progress. I wrote the drafts of the manuscript and all authors provided critical reviews of the drafts prior to submission. A poster based on this work was also presented at the National Youth Health Conference, Fremantle, Australia in November 2013.


Dr Singh and I are equal first authors on this manuscript. Dr Singh designed and performed the hormone assays and assisted with drafting the manuscript. I was responsible for data analysis and was the principal author of the manuscript. I conducted the data analysis with the help of A/Prof Kelly. Ms Paxton and A/Prof Hawke made substantial contributions to the overall ARCHER study. Prof Handelsman was substantially involved in the conception and design of the ARCHER study and designed and performed steroid hormone assays with Dr Singh. Prof Steinbeck is the principal investigator of the ARCHER study and has overseen its conception, design and progress. I wrote the drafts of the manuscript and all authors provided critical reviews of the drafts prior to submission. A poster based on this work was also presented at the International Congress of Endocrinology/Endocrine Society of America Annual Meeting, Chicago, USA in June 2014.

I co-designed the study with the co-authors, interpreted the analysis done with Dr Garden and Dr Luscombe. Mr Amatoury helped with data collection by overseeing specimen transport and storage and provided reviews of manuscript drafts. Ms Paxton and A/Prof Hawke made substantial contributions to the overall ARCHER study. Prof Handelsman was substantially involved in the conception and design of the ARCHER study and designed and performed steroid hormone assays. Prof Steinbeck is the principal investigator of the ARCHER study and has overseen its conception, design and progress. I wrote the drafts of the manuscript and all authors provided critical reviews of the drafts prior to submission. Part of this research was presented as an oral presentation at the National Youth Health Conference, Melbourne, Australia, in November 2015.


I designed the analysis plan and performed the data analysis with Dr Cheng, Dr Garden and Dr Luscombe. Ms Paxton and A/Prof Hawke made substantial contributions to the overall ARCHER study. Prof Handelsman was substantially involved in the conception and design of the ARCHER study and designed and
performed steroid hormone assays. Prof Steinbeck is the principal investigator of the ARCHER study and has overseen its conception, design and progress. I wrote the drafts of the manuscript and all authors provided critical reviews. A poster based on this work was also presented at the European Society for Paediatric Endocrinology Annual Meeting, Athens, Greece, in September 2018.

Ben William Robert Balzer Date: 3 June 2019

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Professor Katharine S. Steinbeck Date: 3 June 2019

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.
# TABLE OF CONTENTS

Statement of Authentication ......................................................................................... i
Acknowledgments ........................................................................................................ ii
Ethics Clearance .......................................................................................................... iii
Authorship Attribution Statement .............................................................................. iv
Contents ...................................................................................................................... viii
List of Tables ............................................................................................................... xii
List of Figures ............................................................................................................. xiv
Abstract ..................................................................................................................... xv

1. Puberty and Adolescence in Context ................................................................. 1
   1.1 Introduction ........................................................................................................ 1
   1.2 Puberty ................................................................................................................ 2
   1.3 Currently Used Methods of Assessment of Puberty in Research Studies ....... 15
   1.4 Adolescence ....................................................................................................... 34
   1.5 The Effects of Pubertal Changes on Adolescent Well-Being ......................... 40
   1.6 Concluding Remarks ....................................................................................... 49
   1.7 References ......................................................................................................... 51

2. Methodology .......................................................................................................... 63
   2.1 Aims and Hypotheses of this Thesis ............................................................... 63
   2.2 Subjects and Setting ......................................................................................... 66
   2.3 Surveys, Anthropometry and Specimen Collection ........................................ 69
   2.4 Liquid Chromatography-Tandem Mass Spectrometry ................................... 71
2.5 Statistical Analyses .......................................................... 72
2.6 Concluding Remarks ...................................................... 76
2.7 References ..................................................................... 76

3. The Effects of Oestradiol on Mood and Behaviour in Human Female

Adolescents: A Systematic Review ............................................. 78
3.1 Abstract ........................................................................ 79
3.2 List of Abbreviations ..................................................... 80
3.3 Introduction ................................................................... 80
3.4 Methods ......................................................................... 83
3.5 Results .......................................................................... 86
3.6 Discussion ...................................................................... 97
3.7 References ..................................................................... 101

4. Testosterone and Its Effects on Human Male Adolescent Mood and

Behaviour: A Systematic Review ............................................. 104
4.1 Abstract ........................................................................ 105
4.2 Introduction ................................................................... 107
4.3 Methods ......................................................................... 109
4.4 Results .......................................................................... 113
4.5 Discussion ..................................................................... 124
4.6 References ..................................................................... 128

5. Text Messaging is a Useful Research Tool ............................. 131
5.1 Abstract ........................................................................ 132
5.2 Introduction ................................................................... 133
5.3 Methods ......................................................................... 133
8.5 Discussion ........................................................................................................200
8.6 References ........................................................................................................203

9. Thesis Discussion .................................................................................................204
   9.1 Key Findings ......................................................................................................204
   9.2 Future Directions ...............................................................................................216
   9.3 References ........................................................................................................217

Appendices ..............................................................................................................

   Appendix A: ARCHER Study Methodology ..........................................................220
   Appendix B: ARCHER Study Cohort Description .................................................245
   Appendix C: Requirement for Specific Gravity and Creatinine Adjustments for
   Urinary Steroids and Luteinizing Hormone Concentrations in Adolescents
   ...............................................................................................................................277
   Appendix D: Supplementary Material for Chapter 3 .............................................294
   Appendix E: Supplementary Material for Chapter 4 .............................................298

Note: Appendices A, B and C comprise publications which add context to work in
the thesis but do not form part of its examinable content.
TABLES

1.1 Descriptions of male Tanner staging .................................................................32
1.2 Descriptions of female Tanner staging ............................................................33
2.1 ARCHER baseline demographics compared with the comparison local
government areas (LGAs) and New South Wales (NSW) ..................................68
3.1 Study characteristics .........................................................................................89
3.2 Extracted study data .........................................................................................100
3.3 Quality assessment of included studies ............................................................106
4.1 Longitudinal studies .........................................................................................115
4.2 Cross-sectional studies or longitudinal studies with cross-sectional analysis ....116
4.3 Aggressive/disruptive behaviour studies .........................................................118
4.4 Other behaviours, moods and affect .................................................................120
5.1 Time to SMS reply and urine collection scheduled data by gender ...............136
6.1 Baseline and 12-month follow-up anthropometry measurements .....................150
6.2 Self-Rated Tanner staging at baseline and 12-month follow-up .........................150
6.3 Mean urine E$_2$, T and LH over time for males (M) and females (F) .............154
6.4 Pearson’s Correlation Coefficients for Serum and Urinary Hormones using 0 and
12 month data ........................................................................................................156
6.5 Baseline and 12-month follow-up serum E$_2$, T and LH .................................156
6.6 Regression results for serum (left) and urine (right) hormones and anthropometric
markers of puberty ...............................................................................................158
7.1 Self-rated Tanner stage, puberty progress and questionnaire responses ..........173
7.2 Odds ratio (95% confidence intervals) of hormone increase based on self-rated Tanner stage, relative to Tanner stage 1. For each Tanner stage, the odds ratio refers to the multiplicative factor increase in hormone levels relative to stage 1.

7.3 Odds ratio (95% confidence interval) of hormone levels based on adolescents who rated their pubertal development as ‘behind’ or ‘ahead’ of their peers, relative to those who had ‘same’ pubertal development.

7.4 Percent agreement, κ score, sensitivity and specificity of subjective markers (survey) of puberty compared with objective markers (hormone levels, males testosterone < 0.5nmol/L and females oestradiol < 40pmol/L).

8.1 Participant anthropometry, self-rated Tanner stage and menarcheal status.

8.2 Differences in anthropometry for males and females who were pre-pubertal at baseline compared to pubertal, and pre- and post-menarcheal (at any timepoint) females.

8.3 Mixed-model regression analyses of longitudinal changes in anthropometry and Tanner stage with changes in serum hormones.
FIGURES

1.1 A proposed model of gonadotrophin release based upon studies in humans. ........6
1.2 Tanner staging for male genital development.........................................................26
1.3 Tanner staging for breast development in females..............................................27
1.4 Tanner staging for pubic hair distributions in males and females.........................28
1.5 Tanner staging line drawings for female breast and pubic hair development ......29
1.6 Tanner staging line drawings for male genital and pubic hair development.......30
1.7 Sequence of events at puberty. ...........................................................................31
3.1 PRISMA diagram for study selection...................................................................87
4.1 PRISMA flow diagram for study selection.............................................................114
5.1 Plot of urine specimen collection time and SMS reply time..............................137
6.1 Plots of mean and SD of urinary E$_2$, T and LH concentrations derived from 3, 6, 
9 and 12 months collection changes from baseline (0)........................................153
6.2 Plots of mean and SD E$_2$ (left), T (centre) and LH (right) in urine (top) and serum 
(bottom) specimens, by Tanner stage........................................................................159
7.1 Box plots of Tanner stages at each year by sex......................................................172
7.2 Mean serum testosterone (left) and oestradiol (right) with 95% confidence 
interval (bars) by self-rated Tanner staging in males and females.........................176
8.1 Mean annual % change in foot length over three years’ follow up, separated by 
baseline pubertal status...............................................................................................196
ABSTRACT

Puberty involves profound anthropometric and hormonal changes that have been associated with changes in mood, behaviour and other health risks that emerge in adolescence. However, there is little longitudinal research in the effects of puberty on adolescents. The overarching aim of this thesis was to demonstrate novel ways of studying puberty and how these improve our understanding of the longitudinal changes that occur over this period.

Oestradiol is popularly associated with changes in mood and depression and testosterone with aggression and risk taking. Chapters 3 and 4 reviewed the effects of oestradiol on female, and testosterone on male adolescent mood and behaviour. In both reviews there were inconsistent cross-sectional data, and limited longitudinal data, supporting such associations.

In Chapter 5, text messaging reminders for study compliance was studied. Specimen collection correlated with text message reply time.

Liquid chromatography-tandem mass spectrometry assays on urine and serum specimens quantified oestradiol, testosterone and luteinising hormone levels in 97% of urine and 95% of serum specimens (Chapter 6). Hormone changes over one year were not linear, with frequent urine sampling offering a more nuanced description of puberty hormone change.

Self-rated Tanner stage and other subjective measures of puberty (adolescent- and parent-rated) were compared with longitudinal changes in serum hormones in
Chapter 7. Positive longitudinal associations were observed between these subjective measures and hormone changes.

Chapter 8 identifies foot length growth as a novel marker for early puberty changes. Positive longitudinal relationships were observed between foot length, height, weight, Tanner stage and serum sex steroids. Foot length offers a practical, novel and cost-effective marker of early puberty.

This thesis provides novel insights into puberty and adolescence, particularly how this life transition is studied. The importance of longitudinal research to determine the true effects of puberty hormone changes on adolescents is highlighted, and data are provided to show how this can be feasibly achieved.
1. PUBERTY AND ADOLESCENCE IN CONTEXT

1.1. Introduction

Puberty and adolescence involve profound biological and psychosocial changes. Despite the known importance of these changes, there are persisting gaps in our understanding of the impact of puberty on adolescence, and of both transitions on health and wellbeing across the lifespan. Many studies of pubertal hormones are cross-sectional, or, if longitudinal, lack sufficient sample size or adequate and appropriate specimen collection to capture the dynamic changes over this period. Additionally, limitations in hormone assay technology mean that it is only with the recent advent of accessible mass spectrometry that measuring low circulating hormone levels in early puberty with accuracy is now feasible. Puberty occurs over a relatively short period of time with subtle indications of its onset. Community-based adolescents are needed to provide sufficient samples to capture normal hormone changes.

In Australia, the Adolescent Rural Cohort, Hormones and Health, Education, Environments and Relationships (ARCHER) study (1), was conducted to investigate the effects of pubertal hormone changes on the health and wellbeing in a cohort of 342 adolescents living in rural NSW. This thesis is based on data from the ARCHER cohort and plans to systematically evaluate the literature for hormone effects on adolescent mood and behaviour, as well as describe the methodology and logistics necessary to complete the ARCHER study. Further, this thesis will consider how we might relate these novel pubertal hormone results to our established understanding of
Puberty in terms of Tanner staging and the other anthropometric changes which occur.

Puberty generally occurs during adolescence, the second decade of life (2), though adolescence is a life phase that extends beyond the bounds of puberty (3). Physiologically, puberty signifies a time of accelerated somatic growth and sexual development, which is promoted by the synergistic effects of rapidly increasing gonadal hormones and growth hormone (4). Much controversy remains as to the effects of these somatic and hormone changes on health and wellbeing, for example the effects of hormone changes on mood and behaviour. Adolescent changes include not just puberty’s biological transitions (which prepare the individual for reproductive capacity), but also a variety of other cognitive, social and psychological changes necessary for independent adult life (4). There are many popular ideas and beliefs about the effects of these life changes on the individual. This introductory chapter will consider the changes that occur during both puberty and adolescence, and how these may affect the individual, as well as identifying some of the unanswered questions about this crucial life transition.

1.2. Puberty

Puberty is the biological process by which an individual sexually matures and gains the physiological and psychological capacity to reproduce (3-7). This process involves activation of the hypothalamic-pituitary-gonadal (HPG) axis and the concomitant growth of the gonads, secretion of sex hormones (mainly testosterone for males and oestradiol for females (4)), the development of secondary sex characteristics, skeletal growth and finally fusion of epiphyseal growth plates to
achieve adult stature (3, 5). Other physical and psychological changes may occur that prepare the adolescent for reproductive capacity and general maturity (6, 7).

1.2.1. The Onset of Puberty

The mechanisms underpinning the timing of pubertal onset are still being elucidated, though a variety of factors are known to play a modulating role. These include genetics, nutrition, sex hormone levels, chronic illness and general health (4, 5, 7). Many of these factors will also affect the somatic growth of the individual during the pubertal transition (8). Body weight and adiposity have been studied extensively due to observations that overweight and obese children tend to commence puberty earlier (9). Many of these studies focus on females, while data on boys are less clear. In a cross-sectional Bulgarian cohort, Tomova found that heavier boys commenced puberty earlier, with a positive association between body mass index (BMI), body weight and onset of puberty (10). It is thought that leptin (see below) plays a significant role as well as insulin resistance, which creates compensatory increases in insulin levels and decreased sex hormone-binding globulin, thereby increasing the amount of bioavailable sex steroids (10). Similarly females with increased BMI at seven years of age were found to have earlier puberty in a prospective Hong Kong cohort (11) and in a longitudinal study of Greek girls, Pantsiotou reports earlier onset of puberty (Tanner breast stage 2) for girls with greater BMI, but they did not observe a difference in age at menarche (12). This latter finding is controversial, as other studies have shown earlier menarche in girls with greater BMI than age-matched peers (13).
The genetic heritability of pubertal onset has also been studied. Genetic polymorphisms in a variety of regulatory networks have been found to influence pubertal timing (14-16), and genome wide association studies have identified several unique loci whose polymorphisms influence timing of pubertal onset. For example, \textit{LIN28B} single nucleotide polymorphisms are associated with earlier age at menarche and breast development in females, earlier pubic hair development in males and faster height growth (17-20). Girls homozygous for the \textit{FSHR}-29AA allele (associated with decreased follicle stimulating hormone receptor expression) enter puberty 7.4 months later than other common variants (\textit{FSHR}-29GG or \textit{FSHR}-29GA) (21). The overall heritability of pubertal timing has been considered in several studies. A mono- and dizygotic twin study by Silventoinen estimated 91\% of variability of the pubertal growth spurt was explained by genetic heritability (9) and a large British cohort study found that familial timing of menarche was significantly associated with adolescent’s timing of menarche, with approximately 57\% heritability and the remaining 43\% of variability explained by non-shared environmental factors (22). Recent work by Wohlfahrt-Veje and colleagues has shown that pubertal timing in adolescents is influenced by the timing of their parents’ pubertal milestones, which suggests that at least some of the variation in pubertal timing is genetic (23). Parents were asked if their pubertal timing was early, average or late compared to peers and adolescents underwent repeated clinician examination for Tanner staging and girls were asked when they underwent menarche. For girls, the age at menarche was better associated with both maternal and paternal pubertal timing compared to thelarche or pubarche, indicating the role of other factors in the timing of the latter two milestones. The timing of menarche in girls was significantly earlier than their mothers’. The timing of boys’ milestones was also associated with
1. Puberty and Adolescence in Context

parental milestones. Early parental pubertal timing in parents was associated with earlier pubarche in males and menarche in girls (23).

1.2.2. Initiation of Puberty

We have a clearer picture of the initiation of puberty. Gonadal development and the increase in physical growth during puberty both begin with reactivation of the HPG axis by the nocturnal pulsatile secretion of gonadotrophin-releasing hormone (GnRH) from the hypothalamus (4-7). The neuropeptides kisspeptin (4, 24, 25), neurokinin B (which is encoded by TAC3) and dynorphin, leptin (26) and the recently described Makorin Ring Finger 3 (MKRN3) gene and its products (27-29) are all involved in promoter or brake roles (Figure 1.1). Inactivating neurokinin B (via TAC3 missense loss-of-function mutations or deletion) or its receptor is associated with a lack of progression into puberty (25, 29, 30) and mutations in MKRN3 have been associated with central precocious puberty (27, 28). These findings have been used to infer the contributions of these factors to normal pubertal initiation. A decrease in serum MKRN3 concentration precedes puberty onset in healthy girls, while MKRN3 concentration was negatively correlated with LH and FSH (31). Additionally, leptin has been implicated in the initiation of puberty through an indirect regulatory mechanism on GnRH neurons (26).
In this model, kisspeptin neurons in the pre-optic area and kisspeptin-neurokinin B-dynorphin neurons (KNDy) in the arcuate nucleus of the hypothalamus secrete kisspeptin to stimulate KISS1 receptors (KISS1R) gonadotrophin-releasing hormone (GnRH) neurons in the pre-optic area and infundibular nucleus of the hypothalamus. GnRH stimulates the GnRH receptor (GnRHR) on gonadotrophs in the anterior pituitary, leading to the secretion of LH and FSH which act on the gonads. KDNy neurons have complex activating and inhibiting stimuli which are not fully elucidated. Auto- and paracrine secretion of Neurokinin B (NKB) stimulates the NK3R receptor to enhance kisspeptin secretion by KNDy neurons and dynorphin acts on κ-opioid receptors (KOR) to inhibit kisspeptin secretion. Leptin is another proposed stimulator of kisspeptin secretion from KNDy neurons. MKRN3 and ghrelin both act as brakes on kisspeptin secretion, and sex steroids (oestradiol (E2), testosterone (T) and progesterone (P) are thought to act on KNDy oestrogen receptor-α (ERα), progesterone receptor (PR) and androgen receptor (AR) as a negative feedback loop.
The HPG axis is initially active during the foetal period, potentiating sexual differentiation (4-6), but is largely suppressed after birth (4), apart from in the first six months of life, when both sexes experience a mini-puberty. Mini-puberty refers to the transient activation of the HPG axis in the first months of life (39, 40). There is a brief surge in luteinising hormone (LH) and follicle stimulating hormone (FSH) from the first week of life to approximately six months of age (peaking at one month), which helps in gonadal development (40). In males, there is evidence that testosterone secreted in mini-puberty is critical for normal penile growth in infancy, as well as Sertoli cell proliferation and differentiation of germ cells into adult spermatogonia, and is thus presumed to be associated with later male fertility (40, 41). The relationship between oestradiol and mini-puberty in females with respect to later fertility is less clear (40, 42, 43). In both males and females, mini-puberty is thought to play critical roles in childhood growth and development. The sex differences in growth are noted to begin around the time of mini-puberty (44) and have demonstrable associations with growth up to six years of age (42). In a longitudinal study of thirty-five infants (seventeen male) followed from one month to six years of age, there was a significant negative relationship between testosterone and LH at eight weeks of age and body weight and body mass index at three and six years (42). A similar negative relationship was seen for those outcomes and oestradiol levels at twenty weeks of age (42). Development of sex-based toy preferences and sex-specific behaviours in early childhood have also been consistently associated with changes in testosterone levels during mini-puberty (45, 46). After the period of childhood suppression, the pubertal reactivation of the HPG axis has been described as part of the continuum of sexual development, beginning
with sexual differentiation in utero and the mini-puberty in the neonate, childhood quiescence, followed by puberty and then sexual maturity and reproductive capacity, and concluded by senescence (6, 7).

In puberty, increased GnRH secretion causes the release of LH and FSH from the anterior pituitary gland. LH and FSH cause gonadal growth and sex steroid production through a negative feedback loop (4-7). The gonads are the most important site of sex steroid formation, with the ovaries producing over 90% of female oestradiol (and 50% of testosterone) and the testes producing over 90% of testosterone and 50% of oestradiol in males (4). LH acts upon gonadal interstitial cells (ovarian theca cells and testicular Leydig cells) to promote androgen secretion. In females, these androgens are precursors of oestrogens (oestrone, oestradiol, oestriol, the most potent and prevalent of which is oestradiol) (4). FSH stimulates testicular Sertoli cells in males to support the production of gametes, and ovarian granulosa cells in females to form oestrogens from androgenic precursors, such as androstenedione and testosterone. Inhibins A and B, secreted by testicular Sertoli cells and ovarian granulosa cells, play an important negative feedback role in gonadotrophin secretion. Inhibin B can serve as a marker of spermatogenesis in males and its increase in late pre-pubertal females suggests increasing follicular activity (47).

Sex steroids also help develop secondary sexual characteristics, such as breast and uterine growth by oestrogens in females and penile growth, body hair and deepening voice in males as a result of laryngeal cartilage growth by androgens (3-5). These also affect changes in body composition during puberty. The increase in sex steroids (particularly oestrogens (48)) increases pituitary secretion of growth hormone (GH),
which in turn increases the levels of insulin-like growth factor 1 (IGF-1). GH and IGF-1 increase the rate of physical growth in a synergistic manner with sex steroids (especially oestradiol), and GH also optimises LH and FSH action on gonadal growth (4, 5, 7). The completion of physical growth in both genders is due to high-level oestradiol-induced fusion of epiphyseal plates (4).

The clinical changes of puberty normally commence between 8 and 13 years in girls and between 9 and 14 years in males (49), with abnormally early pubertal onset defined as precocious and later puberty being delayed (although time differentiates between simple delay and pathological absence). Puberty is usually completed within 2-4 years of onset (5, 6). Bone age is a more accurate predictor of maturation than chronological age, with less variation between skeletal age compared to their chronological age in 170 adolescents in the longitudinal Harpenden Growth Study (4, 50). The age of menarche (as an indicator of pubertal development) has decreased during the mid-late 20th century (7) and similar trends have been seen using breast development as a marker (51-53). However, conflicting data have recently been found in white girls in the USA (54) and accordingly there is some controversy whether secular trends for earlier puberty are persistent (7). In males, recent studies have provided evidence for continued reduction in the timing of onset of puberty, using direct measurement of testicular size (55, 56) and peak height velocity (PHV), though PHV as a marker of pubertal progress has limitations (see Section 1.2.3). Tena-Sempere describes puberty as a “biological sensor”, the timing of which (and trends thereof) reflecting a possible change in environmental influences on reproductive health (26), such as intrauterine nutrition, changing dietary patterns and endocrine-disrupting chemicals (53).
Two separate endocrine processes contribute to puberty (57): the onset of adrenal androgen secretion (adrenarche) usually between 6-8 years, and the re-activation of the HPG axis and onset of sex steroid secretion from the gonads (gonadarche) approximately two years after this. The temporal relationship between adrenarche and gonadarche may suggest adrenal development might influence future gonadal development, however this remains unproven and there is no obvious mechanism. Androgens secreted by the adrenal glands, such as androstenedione and dehydroepiandrosterone (DHEA and its sulphate DHEAS), cause the development of axillary and pubic hair, as well as the development of acne, apocrine sweat glands and body odour (5, 58, 59). Those with primary adrenal insufficiency, however, still undergo gonadarche within the normal age range (6). A specific pituitary stimulating hormone for adrenarche has never been identified, but it is generally assumed that adrenocorticotropic hormone (ACTH) is responsible. Anecdotally, children with central pan-hypopituitarism including adrenal insufficiency frequently have little adrenarchal response.

1.2.3. Physical Growth during Puberty

One of the most striking features of puberty is the growth spurt, which transforms the individual to final adult height. This growth is dynamic and sexually dimorphic (60), and is generally independent of prior growth (61, 62). Epiphyseal fusion occurs during this time as well, together with increasing bone density. Between 15% and 25% of adult height is gained during puberty – up to 25 cm in females and up to 30 cm in males (63). Weight gain occurs too, whether by enhanced musculature (more marked in males; under the influence of testosterone action) or increased fat
deposition (more marked in females; an action of oestradiol). Male growth begins and concludes later than female growth.

Bone mass doubles from the onset of puberty to young adulthood, usually peaking in the third decade (64), with concomitant changes in trabecular and cortical bone mass. This is modulated by activation of the GH-IGF-1 axis by the sex hormones (in particular oestrogens in both sexes), which leads to the pubertal growth spurt (64-67). Epiphyseal chondrocyte fusion finalises bone growth and is primarily an action of high-level oestradiol in both sexes (65). Sex hormones also enhance calcium and phosphate metabolism to increase bone mass and mineral density (64), assuming adequate protein and calcium intake and sufficient Vitamin D.

Peak height velocity describes the time when the height change over time is maximal. In males, average PHV is 10.5 cm per year (range 7-12 cm/year) and for females it is 9.0 cm/year (range 6-11 cm/year). From the age of two years to the onset of puberty, both sexes have approximately the same growth velocities: average 5-6 cm/year height increase and average 2.5 kg/year weight increase, dependent on factors such as nutrition, GH and thyroid hormone levels (60). According to Marshall and Tanner’s studies in British children, PHV in males occurred at a mean age of 14.1 years, at around the time that Tanner stage 4 was reached (68). Female PHV was approximately two years before boys, with the average PHV occurring at 12.1 years during Tanner stage 3 (69). These are illustrated in Figure 1.7. Given the secular decrease in age at puberty onset, these findings are likely to have been altered, although the patterns of growth and development remain unchanged.
Recent data from German adolescents noted age ranges for PHV in females (between 10.0 and 13.0 years), and for males PHV between 11.9 and 14.7 years (62). A retrospective analysis in Swedish males born between 1946 and 1991 found that male PHV was reached at a mean age of 14.0 ± 1.1 years and had been decreasing by 1.3 months for each decade of the study (56). Tanner found that adolescents of both sexes could still reach the same final height whether their age of pubertal onset was 2 standard deviations earlier or later than the mean – the main difference was their PHV, which was more rapid and of higher magnitude in early maturing adolescents and more prolonged and lower magnitude for those maturing later (70). Recent work has also shown this in girls: those maturing earlier (as defined by age at menarche) had greater PHV and gained more height relative to later-maturing girls, although at the end of growth they were shorter and had greater adiposity (71). Measuring height over several visits to calculate PHV allows longitudinal trends to be observed and is less invasive than other measures of pubertal development. Recent work by Cole demonstrated that annual measurement of height generates a sufficiently accurate PHV as measuring more frequently (72). However, height and PHV are too indirect to use as standalone pubertal assessments. While sex steroids exert an effect on the GH-IGF-1 axis, growth is dependent on many factors, including genetics, nutrition and general health.

Foot length also increases during the pubertal growth phase. The change in foot length is held to be among the first anthropometric changes to occur, and conclude, during the adolescent growth phase (73-76). Foot length is a marker of interest in puberty as it may offer a non-invasive and cost-effective means to assess pubertal development (77, 78). Indeed, measuring feet may not even be necessary as shoe
sizes are directly proportionate to foot length. Foot length is convenient as parents may recall changes in foot length (as measured by shoe size changes) more easily than recalling the beginnings of breast changes or body hair development. The latter two may be difficult to recall due to subtlety of change and to adolescent privacy around their body image, whereas foot length changes are memorable (74), possibly because these necessitate the purchase of larger shoes.

In a large Indian cross-sectional study (973 participants, 489 male), Mitra (78) observed feet reached adult length at onset of puberty – that is, there was a significant increase in foot length between clinician-rated Tanner stage 1 and Tanner stage 2, but thereafter no differences were observed. Ford (77) followed 86 girls from six years of age with biannual follow-up (of unknown duration), comparing the age of foot size increase (maximum foot velocity) with the age of development of secondary sex characteristics (clinician assessed Tanner staging). This work found that foot growth peaked at a mean of 8.40 years, which was before the age of occurrence of secondary sex characteristics (8.79 years). Additionally, Knudtzon (79) retrospectively surveyed 106 female members of a Norwegian association for people with large feet and found that an earlier menarche was associated with larger adult foot size, although there is likely significant selection bias. At present, there is no evidence linking changes in foot length with changes in sex hormones, the direct measure of pubertal progress. Similarly, there are no longitudinal data in males, nor any large-scale longitudinal data. While there are ostensible temporal associations with indirect markers of puberty, one would need to demonstrate a relationship with longitudinal hormone changes to properly conclude this.
Weight gain also occurs during puberty, with the average increase for females 17.5 kg and for males 23.7 kg. (63), with peak velocities of 9.1 kg/year in males and 8.3 kg/year in females. The peak weight velocity for females occurs at 12.9 years, and 14.3 years for males (80). These occur slightly after peak height gain, so some weight increase can be attributed to increased bone mass. Two other factors add to weight gain during this time: increased musculature and increased adipose tissue. Increased skeletal muscle is common to both sexes, though the velocity of muscle growth is much greater in males (61). Acquisition of skeletal muscle mass peaks at the time of PHV and is associated with increased physical strength, though muscle bulk and strength may continue to change long after PHV is reached (61). As children, boys have greater muscle mass than girls, though this sex difference lessens when girls commence puberty (generally earlier). Once male puberty commences however, the difference is re-established, with male musculature increasing faster and with greater magnitude than in females, and males having greater muscle mass than females (61). Growth hormone stimulates muscle mass and strength (64), as well as regulating fat mass (81). The androgens (especially testosterone), GH and IGF-1 are anabolic for protein (60). Peak GH-IGF-1 levels occur in females around Tanner stages 3 and 4, and in males around Tanner stage 4, with acquisition in both sexes declining in Tanner stage 5 (60). The peak in IGF-1 is thus, as expected, around the time of PHV and then it decreases (82).

Similarly, body adiposity changes with puberty. Android fat deposition is predominately abdominal, and gynoid deposition is gluteo-femoral. Females have almost double the adipose tissue than males after puberty (60) as oestrogens enhance fat deposition, whereas androgens are lipolytic (60). GH is also lipolytic (83). Boys
also lose more limb fat than girls at puberty, though truncal fat loss is less in both sexes (63). Pubertal girls’ fat mass increases at a rate of 1.14 kg/year, whereas pubertal boys decrease by 1.15 kg/year (8).

It must be recognised that much of the data regarding growth changes do not take into account recent trends of increasing overweight and obesity (84) and earlier onset of puberty (7, 55, 85); however, the trend of increased linear growth, musculature and weight changes remain. A fifty-year longitudinal study suggested that secular changes in body mass index (weight divided by height squared; kg/m²) may be independent of any trend towards earlier menarche (86). More contemporaneous growth velocity data are needed to accurately account for secular increases in overweight and obesity prevalence, especially in adolescent populations. While these changes have been described with respect to the timing and the tempo of the pubertal transition (measured through Tanner staging), the relationship between hormonal changes and the physical manifestations is less clear.

1.3. Currently Used Methods of Assessment of Puberty in Research Studies

The assessment of puberty in research studies tends to consist of clinical assessment or hormonal assessment (57, 87, 88). Clinical assessment may be in the form of pubertal staging, or measurement of anthropometry such as PHV, foot length (both discussed in Section 1.2.4) or voice maturation. Sampling the hormonal milieu is another method and assessing reproductive milestones such as menarche or spermarche offer alternative measures. Below, the most widely used methods of pubertal assessment are considered.
1.3.1. Clinical Assessment

Two methods are regularly used to clinically assess pubertal development: the staging system developed by Tanner and Marshall (68, 69, 89) and the Pubertal Development Scale (PDS) (90). Tanner staging (see Figures 1.2, 1.3, 1.4, 1.5, 1.6 and Tables 1.1 and 1.2) is generally seen as the standard measure of pubertal status (57), and divides progress through puberty into five stages, with Stage 1 being pre-pubertal and Stage 5 full adult development. These changes occur in a continuum, though discrete stages are used to indicate progression (91). Pubic hair development is categorised for both genders, while females have their breast development assessed and males their genitalia (68, 69). As oestradiol will stimulate breast development and testosterone the development of axillary and pubic hair as well as the male genitalia (4, 92), it is assumed that there is a direct and causative relationship between increase in sex hormones during puberty and changes in these characteristics.

Dorn’s review of measures of pubertal status (57) notes that the original research used to create Marshall and Tanner’s stages was limited by a relatively small number of subjects (228 males and 192 females (68, 69)), all of whom were Caucasian. This may hinder its application in ethnically diverse settings and with increasing prevalence of obesity since the original stages were devised. However, it remains the most widely used system for assessment of pubertal status, in which the subject’s development is compared (ideally by a trained clinician) with standard photographs (such as Figures 1.2, 1.3 and 1.4) or line drawings (Figures 1.5 and 1.6) reflecting the development at each stage. Many studies continue to use Tanner’s black-and-white photographs, which may be less acceptable to adolescents than standardised line
drawings (87), and indeed colour drawings have shown to have higher accuracy when compared with self-rating (93). Figure 1.7 shows the temporal relationships between Tanner stages and other maturational milestones during this period.

In research studies, there are possible ethical issues in clinician assessment of Tanner staging, particularly where such clinician assessment would not have been performed were the adolescent not enrolled in the study. Principal concerns include privacy and modesty for adolescents and parents regarding examination, for whom such assessment may be intrusive and limit involvement in the study. Whereas one can justify a trained examiner assessing a child’s Tanner stage in a clinical context, for example to assess early or delayed puberty, it may be less justifiable or indeed unacceptable to the adolescent enrolled in a community-based research study, who may not have concerns about their development, and for whom it could constitute an assault. Therefore, clinician examination can pose challenges for the researcher, in terms of ethics review board approval and gaining parental and adolescent consent. Additionally, the logistics of having trained examiners assess large cohorts provides further challenges. These mean that alternative pubertal assessment tools may be required. Self-rating by the adolescent therefore offers a less intrusive and safer means by which Tanner staging can be provided.

Self-report of Tanner stage has been shown to be adequate when compared to clinician assessment (94-96), though some studies disagree (97, 98). A systematic review by Walker found that agreement between self- and clinician-assessed Tanner stage ranged from 43% to 81%, though one issue in accuracy is that some studies use the photographs used in Tanner’s original staging (Figures 1.2, 1.3, and 1.4) and others use line drawings based upon these photographs (such as the drawings in
Figures 1.5 and 1.6) (87). Duke’s study in 66 adolescents (43 girls) showed excellent agreement between self-rated and clinician-rated Tanner stage (female breast development 81% agreement (κ=0.81), female pubic hair 93% agreement (κ=0.91) and male genital stage 91% agreement (κ=0.88)) (94). Similar results in the 1980s and 1990s showed good correlation or agreement (κ coefficient range 0.59-0.91) between self-reported Tanner stage and clinician assessment in American boys and girls (95, 99). However, these findings have been challenged in more recent studies in larger and more diverse cohorts (96, 98, 100). A reason for declining accuracy in more contemporary studies may reflect secular trends in increasing obesity prevalence, with poor agreement particularly shown in studies of overweight or obese adolescents (93, 97, 101). Recent work by Rasmussen provides the largest sample size to date (n=898) to answer this question, and found that self-rated Tanner staging was not sufficiently accurate (κ=0.28-0.55) to be used in clinical practice, where accurate staging is important, for example to diagnose disorders of pubertal onset (102). However, despite this lack of accuracy, results are concordant enough to continue the use of self-rating in research studies where the emphasis on precise Tanner staging may not be as high, or not possible due to the reasons described. Recent work has shown self-rating is useful in research settings to differentiate pre-pubertal and pubertal development as assessed by adolescents or their parents (102-105). In Terry’s longitudinal study (104), maternal and adolescent rating of development was relatively sensitive and specific compared to clinician rating. Chavarro’s cross-sectional cohort showed that self-assessment, while inferior to clinician assessment, correlated to a composite index ranking of adolescent hormonal development derived from hormone levels measured with radioimmunoassay (105).
This form of hormonal assessment is not routinely used and does not have a clear physiological correlate.

The PDS was developed in the 1980s and involves self- or parental report in several domains: body hair, growth spurt and skin changes (both genders); voice changes and facial hair (males only); and breast change and menarche (females only) (90). These all reflect the effects of hormonal changes on the body and reduce the need for direct visualisation. These domains are scored from one to four, with increasing scores reflecting more development (90). The PDS is used less commonly than Tanner staging in adolescent research (88). The greatest contrast between the Tanner scales and the PDS is that the PDS lacks pictorial standards of sexual development (90). While this makes it more appropriate in sensitive settings, it also may enhance subjectivity biases, as subjects using the PDS must perceive and rate their development in comparison to their peers.

Another method of non-invasive clinical assessment of puberty is voice maturation assessment, characterised by changing pitch, loudness and tone due to laryngeal growth (87, 106, 107). This is often a late pubertal event in males (107, 108). Ong’s study of voice changes in a birth cohort of 2008 males found that more advanced voice-breaking at age 14 years was associated with higher adult body mass index and fat mass at 60-64 years (107). Harries’ study of 26 teenage boys (aged 13-14 years at recruitment) showed Cooksey stages of voice development correlated with Tanner stage and testis volume, though not salivary testosterone over a one-year follow up (106). Expert opinion in Walker’s review was that this method has promise but requires further research and adolescent opinion was that this non-invasive method would be an acceptable method of pubertal assessment (87).
An interesting concept is that, despite a lack of accuracy when compared to clinician assessment, self-rating of puberty is actually more important for adolescent researchers because it reflects the individual’s self-perception (109). While neuro-endocrine changes are probably better associated with the clinician’s rating (as the gold standard), the adolescent’s perceived maturity is critical to understand as they may behave differently, seeking experiences or environments more consistent with their presumed development (109), which may naturally have more public health importance as these behavioural changes most often lead to adverse health outcomes in adolescence (see Sections 1.4 and 1.5). Moore and colleagues found that perceived pubertal timing was important in predicting engagement in sexual and romantic behaviour (dating, romantic sex and non-romantic sex) whereas objective pubertal timing (age at menarche) only predicted the onset of sexual behaviour. The perception of pubertal timing is therefore thought to alter the adolescent’s interpretations and subsequently the emotional and cognitive processing of puberty-related cues (110). So, despite a perceived lack of accuracy with clinician-rating, self-rating of puberty may actually be the more important research outcome in some circumstances.

1.3.2. Reproductive Milestones for Assessing Pubertal Development

As one of puberty’s core biological functions is to mature an individual’s reproductive system, the use of reproductive milestones may be of use in assessing progress through puberty. For females, this is primarily the assessment of menarche, and for males the assessment of testicular volume or determination of the onset spermatozoa production (spermarche) or as is generally reported, semenarche (first ejaculation). Self-report of menarche has been shown to be relatively reliable, even
into later life and when reported by mothers for their daughters (111-113). It occurs later in puberty, at approximately Tanner breast Stage 4 (69) and is thought to be memorable because of its salience. Menarche itself, however, does not strictly infer reproductive capacity, as cycles may remain anovulatory for some time after the onset of menstruation (114, 115). Regardless, menarche is consistently seen as the best marker of female sexual development. Regarding ovulatory cycles, a rise in plasma progesterone in the luteal phase is the clearest indicator of ovulation, while changes in cervical mucus and cyclical temperature changes have been used clinically. Similarly, the report of dysmenorrhoea as an indication of post-ovulatory prostaglandin activation (116) is another means of assessing ovulation.

For males, there is less consistency in the literature with regards to markers for the attainment of reproductive capacity such as spermarche, as male development does not have such a clearly defined event as menarche (117, 118). The transition of testicular volume from 3 mL to 4mL correlates to pubertal onset (119). Marshall and Tanner used the development of the testes, scrotum and penis to indicate androgen action and infer reproductive capacity (68, 89). The use of a Prader orchidometer to measure testicular volume has shown that the visible physical effects of testosterone occur around 7 mL volume and ejaculatory capacity correlates with a volume of approximately 10 mL (120). Marshall and Tanner used the development of the testes, scrotum and penis to indicate androgen action and infer reproductive capacity (68, 89). The use of a Prader orchidometer to measure testicular volume has shown that the visible physical effects of testosterone occur around 7 mL volume and ejaculatory capacity correlates with a volume of approximately 10 mL (120). Ultrasound measurement of testicular volume shows it is accurate when compared with orchidometer assessment (121-123), though both methods are operator-dependent, and invasive (122), with associated costs of ultrasound limiting widespread use in research (57). Measurement of testicular size is less likely to be acceptable to young people in a recent consultation paper, with males feeling they would prefer to self-examine to determine testicular size (87). The
The average age of spermarche is approximately 13 years 5 months (124-126), which relates to Tanner pubic hair stages 2 and 3 (127, 128). Direct assessment of semenarche may be done through ascertaining ejaculatory capacity, whether by obtaining sperm samples or self-report of behaviours such as nocturnal emission, sexual debut or masturbation. Spermarche may be established through examining urine samples for retrograde ejaculation. (118, 125). However, there have not been studies comparing laboratory findings of spermarche with parental or self-report (57) and few studies have used this as a pubertal assessment method. In fact, males are more willing to report non-sexual problem behaviour than they are to report masturbation (129), which means these measures are probably unsuitable for epidemiological studies. In Walker’s systematic review and consultation with young people on methods of assessing puberty, there was consensus among the consulted boys that discussion of nocturnal emissions was both embarrassing and may be incorrectly recalled (87).

1.3.3. Hormonal Assessment

The concentrations of sex steroid hormones, particularly testosterone for males and oestradiol for females, provides information as to the individual’s reproductive maturity. Accordingly, these are the most direct measure of pubertal progress, as these are the primary drivers of the development of secondary sexual characteristics measured in Tanner staging (along with adrenal androgens for hair development).

*The longitudinal changes in gonadal steroids, while rarely undertaken, would be the most accurate study of an individual’s puberty.*
Other hormones commonly assayed to assess pubertal development are adrenal androgen hormones, gonadotropins and IGF-1, which reflect different aspects of development and growth. Heightened activity of the HPG axis, reactivated at the very start of puberty, increases levels of gonadotrophins (130). Hormone levels do not exactly correlate with a specific pubertal ‘stage’ (as defined by clinical staging), as there are overlaps between hormone levels at different stages (57, 131-133). Dorn outlines a variety of considerations which must be factored in to studies involving hormonal assays, the most important being that studies control for the diurnal variations in hormones by ensuring collections occur at approximately the same time of day, and for post-menarcheal females, the same time in their menstrual cycle (57).

Immunoassays were the first convenient method for measuring circulating hormone levels, though with commercial assay development, accuracy was traded for affordability and throughput efficiency (134). The poor analytic specificity of immunoassays for steroid hormones at low circulating levels, such as in childhood and early puberty, as well as significant cross-reactivity between antibody and epitopes from the hormone, its precursor, metabolite or conjugate, limits their accuracy and therefore utility for high quality research (134). Several studies have demonstrated such limitations. For example, Sikaris analysed testosterone levels in 124 healthy men (21-35 years) and found that immunoassays diverged from gold-standard mass spectrometry techniques such that diagnostic utility was impaired (137). Similarly, oestradiol assays with radioimmunoassay and fluoroimmunoassay in pre- and pubertal children by Ankarberg-Lindgren’s group demonstrated insufficient sensitivity for clinically useful results, unless further labour-intensive purification steps were undertaken (138), with attendant cost and time limitations.
Taieb also compared analytic methods using enzyme immunoassays and radioimmunoassays with gas-chromatography mass spectrometry and showed that none of the immunoassay techniques were sufficiently reliable for detecting testosterone levels in women and children compared with mass spectrometry (139). In response to Taieb’s paper, Herold used a random number generator and compared his data to Taieb’s and found that in some instances the random number generator produced comparable data (compared to mass spectrometry) as did Taieb’s immunoassay data (140), again casting doubt on the utility of immunoassays for high-quality analytic research, particularly at the low levels of circulating sex steroids seen in early puberty. With the recent development of lower cost, highly sensitive and specific liquid chromatography-tandem mass spectrometry (LC-MS/MS) assays, the very low levels of sex hormones present in pre- and early puberty can be accurately measured. This was a limitation of previous radioimmunoassay and immunoassay techniques that lacked sufficient accuracy at such low circulating levels and thus affected estimates in hormone research at this life stage (134-136).

The first step in LC-MS/MS is liquid chromatography, which separates compounds of interest from other analytes in the specimen (141). Following this, the eluate is nebulised and ionised, and specimen analysis is achieved using triple quadrupole mass spectrometry (142). Atmospheric pressure photo-ionisation is used to generate ions of steroid hormones in the ARCHER study (136), though other methods can also do this, such as electrospray ionisation or atmospheric pressure chemical ionisation (142). LC-MS/MS is able to provide the following data: retention time,
mass: charge ratio (m/z) of both the precursor and product ions, and these data give this technique its high degree of analytical specificity (142).

A study of twenty adolescents followed every six months for five years found LC-MS/MS could detect serum testosterone concentrations approximately 1.5 years earlier than detectable by immunoassay (143). This study was limited by its small sample size and only measuring testosterone but is promising in showing the analytical strength of LC-MS/MS in early puberty. Another possible improvement to these methodologies is to use urine samples. These are advantages in that venepuncture is avoided and that urine provides a more integrated measurement of sex steroids and LH, which are secreted diurnally and in a pulsatile fashion respectively and which can be collected more frequently than the 6-12 monthly serum samples usually used. A major drawback is that it is more time- and resource-intensive, requiring extra steps in specimen processing that lengthen assay time by a day more than for serum samples (136). LC-MS/MS can also be applied to saliva samples but contamination by blood limits the accuracy (144).
Figure 1.2. Tanner staging for the development of male genitalia.

Figure 1.3. Tanner staging for breast development in females.

Figure 1.4. Tanner staging for pubic hair distributions in males (top, A) and females (bottom, B).

Note: Tanner stage 1 (no pubic hair) is not shown in either image.

Figure 1.5. Tanner staging line drawings for female breast and pubic hair development. Note: Tanner stage 1 pubic hair development is not shown.

Figure 1.6. Tanner stage line drawings for male pubic hair and genitalia for (from left): stage 2, 3, 4 and 5. Stage 1 for pubic hair (no pubic hair) and genitalia (same size as in childhood) are not shown.

Reproduced from Australian Paediatric Endocrine Group Growth Chart for Boys 2-18 years: https://apeg.org.au/clinical-resources-links/growth-growth-charts/
Figure 1.7. Sequence of events at puberty for females (top) and males (bottom).

Table 1.1. Descriptions of male Tanner staging (as shown in Figures 1.2 and 1.6).

<table>
<thead>
<tr>
<th>Tanner Stage</th>
<th>Genitalia</th>
<th>Pubic Hair</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pre-adolescent. Testes, scrotum, and penis are of about the same size and proportion as in early childhood.</td>
<td>Pre-adolescent. The vellus over the pubes is no further developed than that over the abdominal wall, i.e. no pubic hair.</td>
</tr>
<tr>
<td>2</td>
<td>The scrotum and testes have enlarged and there is a change in the texture of the scrotal skin. There is also some reddening of the scrotal skin.</td>
<td>Sparse growth of long, slightly pigmented, downy hair, straight or only slightly curled, appearing chiefly at the base of the penis.</td>
</tr>
<tr>
<td>3</td>
<td>Growth of the penis has occurred, at first mainly in length but with some increase in breadth. There has been further growth of testes and scrotum.</td>
<td>Considerably darker, coarser, and more curled. The hair spreads sparsely over the junction of the pubes.</td>
</tr>
<tr>
<td>4</td>
<td>Penis further enlarged in length and breadth with development of glans. Testes and scrotum further enlarged. There is also further darkening of the scrotal skin.</td>
<td>Hair is now adult in type, but the area covered by it is still considerably smaller than in most adults. There is no spread to the medial surface of the thighs.</td>
</tr>
<tr>
<td>5</td>
<td>Genitalia adult in size and shape. No further enlargement takes place after stage 5 is reached.</td>
<td>Adult in quantity and type, distributed as an inverse triangle of the classically feminine pattern. Spread to the medial surface of the thighs, but not up the linea alba or elsewhere above the base of the inverse triangle*</td>
</tr>
</tbody>
</table>

*80% of males progress to “Stage 6” with spread of pubic hair beyond this triangle

Table 1.2. Descriptions of female Tanner staging (as shown in Figures 1.3, 1.4 and 1.5).

<table>
<thead>
<tr>
<th>Tanner Stage</th>
<th>Breast</th>
<th>Pubic Hair</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pre-adolescent; elevation of papilla only.</td>
<td>Pre-adolescent; the vellus over the pubes is not further developed than that over the anterior abdominal wall, i.e. no pubic hair.</td>
</tr>
<tr>
<td>2</td>
<td>Breast bud stage; elevation of breast and papilla as small mound, enlargement of areola diameter.</td>
<td>Sparse growth of long, slightly pigmented, downy hair, straight or only slightly curled, appearing chiefly along the labia.</td>
</tr>
<tr>
<td>3</td>
<td>Further enlargement of breast and areola, with no separation of their contours.</td>
<td>Considerably darker, coarser, and more curled. The hair spreads sparsely over the junction of the pubes.</td>
</tr>
<tr>
<td>4</td>
<td>Projection of areola and papilla to form a secondary mound above the level of the breast.</td>
<td>Hair is now adult in type, but the area covered by it is still considerably smaller than in most adults. There is no spread to the medial surface of the thighs.</td>
</tr>
<tr>
<td>5</td>
<td>Mature stage; projection of papilla only, due to recession of the areola to the general contour of the breast.</td>
<td>Adult in quantity and type, distributed as an inverse triangle of the classically feminine pattern. Spread to the medial surface of the thighs, but not up the linea alba or elsewhere above the base of the inverse triangle.</td>
</tr>
</tbody>
</table>

1.4. Adolescence

Adolescence is the life phase occurring in the second decade of life (10-19 years) (2). There are over 1.8 billion adolescents and youth (12-24 years) globally (the largest age demographic in the world) (2), the majority of whom live in low- and middle-income countries (145). Adolescence has long been recognised as a defining life stage, with rites and rituals, both formal and informal, indicating the passage from childhood to adulthood. Adolescence also involves profound changes in the individual’s life (146), such as increasing legal and political rights and the transitions from schooling to higher education and employment. Other cardinal events include sexual debut, first consumption of alcohol and beginning to drive, each of which are important social steps in the adolescent’s life and identity and are associated with a variety of risk factors. In 1904, G. Stanley Hall described adolescence as a time of ‘storm and stress’ for the individual (147), a perception that still persists within the community, though this is not necessarily true for all (49, 148).

Several theories were posited during the 20th Century with regards to the development of the adolescent. Three will be considered here: Havighurst’s developmental task theory (an ‘educational’ theory) (149), Piaget’s ‘cognitive development’ theory (150, 151) and Erikson’s Stages of Psychosocial Development (a ‘psychoanalytic’ theory) (152). Though these theories differ in the nature of their focus, a common theme is the development of independence and transition to a more mature form.

Havighurst’s task theory outlines a series of ‘tasks’ or ‘challenges’ that the individual faces during their development. Tasks specific for adolescence include developing
relationships with peers; accepting of sexuality and gender identities; becoming independent of parents and other adults; acquiring a value set and ideological and ethical systems; developing socially responsible behaviour and accepting the physique and effective use of the body (149).

From a cognitive viewpoint, Piaget theorised the adolescent moves into the “formal operational” stage, in which there is a development of abstract thought, creativity and hypothesis formation (150). As such, “reality becomes secondary to possibility” (151), with the individual thinking beyond their world and belief system through the use of propositions (153). This reflects the final stage of cognitive development (153).

Erikson’s theory reflects Hall’s postulation of adolescence being a time of crisis for the individual (147, 152). Through questioning themselves in a range of domains, Erikson argues that the adolescent develops their sense of identity (153). Additionally, the peer group becomes more important for the adolescent’s relationships than the family, including the formation of their own ideological perspectives independent of the family group (152).

1.4.1. Adolescent Brain Development

During adolescence, the brain continues to develop and mature, allowing a greater range of cognitive, perceptive and executive functions (154-159). This allows the individual to better master their behaviours and motivations (160). Neurological changes include neurogenesis, dendritic growth, increased myelination, synaptic formation and elimination, neuronal apoptosis and changes in receptor sensitivity, and are in regions associated with behaviour and emotional regulation (5, 161-163).
Rapid cognitive changes and increased abstract reasoning (163) reflect these neural architectural changes. Following this, different connective changes in the brain may potentiate the onset of such behavioural or affective alterations as depression, eating disorders and anxiety, all of which become more prevalent during adolescence. Evidence from animal studies suggests at least some of these changes are influenced by sex hormone changes during puberty (164). The true effect of pubertal hormones on human brain development is yet to be fully determined. There are, however, well recognised structural differences between male and female brains (165).

1.4.1.1. Theoretical Effects of Oestradiol and Testosterone on the Brain

Pubertal increases in testosterone and oestradiol occur over a time of important psychosocial change. As such, it is difficult to discern the exact contribution of each factor to the onset of changes in mood, behavioural and other psychological or psychiatric conditions.

It is postulated that neural changes in girls commence 1-2 years before boys, suggesting that sex hormones influence brain structure in a way that relies upon sex and the predominant sex hormone for the sex of the individual (163). An analysis of brain imaging studies on pubertal subjects found that the sex steroids as well as LH are all associated with dynamic brain changes (though oestradiol and testosterone’s effects were specific to girls and boys, respectively). White matter increases correlate with LH in both sexes and testosterone in males. Grey matter changes were observed such that increased levels of sex-specific sex hormones were associated with decreases in grey matter in the pre-frontal, parietal and temporal cortices in both sexes (163). Issues with this review included the cross-sectional nature of the data,
limiting the ability to draw conclusions with regards to causality and directionality of the effects. Additionally, not all studies reviewed found a relationship between sex steroid levels and neural changes (163).

Oestradiol and testosterone appear to induce synaptogenesis (166, 167). These are associated with the organisational and activational effects hypothesis, wherein steroids act during foetal brain development to alter neuroarchitecture (organisation), limiting the repertoire of the brain’s responses when acted upon by that same hormone later in life (activation) (162, 168). An example of such activation is the commencement of reproductive behaviour in adulthood, which is thought to be programmed earlier during the organisation phase of neural development (162).

Various neuronal and glial populations express aromatase (169), therefore evidence of associations between oestradiol concentrations and behaviour or affective changes may be confounded by variable enzyme activity and substrate availability. A review of aromatase activity in human brains suggested it plays an important in role in modulating neural plasticity and behaviour (170). It appears that aromatisation to oestradiol is the predominant mechanism by which testosterone exerts its effects on the brain, and thus aromatisation is suggested to be important in such events as prenatal masculinisation of the brain (171-174).

Oestradiol has been shown to modulate genetic expression in neural monoamine systems (such as the dopaminergic, serotonergic and noradrenergic systems) that are commonly implicated in psychiatric conditions, including depression, anxiety and psychoses, such as schizophrenia (175-178). The presence of the oestrogen receptor-beta (ER-β) in areas that regulate mood (179) may also indicate an influential role of
oestradiol on behaviour and affect modulation. For example, a post-mortem study observed decreased ER-β expression in the locus coeruleus (a noradrenergic region of the brain) of suicide victims (176), which may indicate a role of oestradiol (and its receptor activity) in depression and suicidality.

1.4.2. The Emergence of Risk Factors for Future Health and Well-Being during Adolescence

Though childhood mortality has drastically decreased in the past century, adolescent health has not improved at rates commensurate with progress in general child health and mortality rates in this population have only marginally decreased (180, 181). Ascertaining the health risk behaviours that emerge during adolescence and the mechanisms which drive this increase, as well as how these behaviours can be mitigated, are a public health imperative.

A large number of health risk behaviours and their associated health issues emerge during adolescence. Many of these have the potential to have an impact on future as well as current health and wellbeing. It is estimated that 70% of adult premature deaths are due to risk factors that emerge during adolescence (182), such as smoking, alcohol consumption and unsafe sexual practices, together with decreasing physical activity (183, 184) and increasing adiposity for many populations. One cannot assume pubertal changes, either in body size or hormone elevation, are the sole contributor to the emergence of these problems. Other factors like individual susceptibility, family or peer relations and environment will also contribute (49). However, unlike puberty, these are potentially modifiable. It would also be
unfortunate if puberty received excessive attribution and modifiable factors were ignored.

Social, biological and emotional changes during adolescence play a major role in the increased burden of health risk behaviours during adolescence and into adulthood (185). These behaviours include increased rates of unintentional injury, suicides and homicides, depression, anxiety, eating disorders, and substance and alcohol abuse, as well as sexually transmitted infections and unwanted pregnancies. Many of these can be viewed, in part, as the result of the adolescent’s difficulty in controlling emotions and behaviour (179, 185-189). General trends show that females are more predisposed to problems such as depression and eating disorders than males, and these differences first emerge during the adolescent life phase (187). For males, conduct disorders increase during puberty in a similar fashion, with a lesser preponderance in females (190). However, emotional and behavioural lability is sometimes considered typical of the life phase and as such may not be recognised as a clinical disorder (191). These changes may be due to reorganisation of the adolescent brain, which could alter judgement and behavioural patterns, as discussed previously (185, 192, 193).

Despite these observations, the cross-sectional nature of many studies in puberty and adolescence means that the aetiology and duration of these mood and behavioural changes are unclear (194). It has also been asserted that while affective expression does change during adolescence, extremes of moodiness and emotionality are relatively uncommon (195). Though the adolescent is faced with a variety of potential health risks and challenges, each of which may contribute to future health
and wellbeing, the contribution of puberty’s biological changes in this regard remains contentious.

1.5. The Effects of Pubertal Changes on Adolescent Well-Being

1.5.1. Pubertal Timing

There are numerous endogenous and exogenous influences on the timing of the onset of puberty. There are also known associations between the timing of puberty and adolescent psychosocial factors, including emotional and behavioural problems (166, 196). Pubertal timing generally refers to the time of onset (such as reaching Tanner stage 2), but it can also be used to indicate the congruency of an individual’s pubertal stage relative to their age-matched peers (91).

For both sexes, it has been observed that off-timed, especially early, puberty may potentiate depression (197). Other potential associations between early but normal puberty and adolescent health and wellbeing have been observed. For girls, there exists well-documented evidence of a link between timing of puberty and problem behaviours, including earlier sexual debut, although a variety of cofactors are implicit (198). Pubertal timing also has implications for future health, with early-maturing girls at higher risk of obesity, cardiovascular disease and breast cancer in later life (199-202). Early puberty for girls is associated with greater popularity, though it also increases their risk of emotional problems, impairs self-image and increases susceptibility to depression, anxiety, delinquency (disorders of conduct), substance use and eating disorders (49, 188, 197, 198, 203-205). Recent work in a large Hong Kong Chinese cohort (n=5795) found that earlier breast development (as assessed by a trained clinician) was associated with a higher risk of the presence of
depression in girls, though no associations were identified for boys (11). While this risk is highest for early maturing girls, on-time puberty is also associated with increasing rates of delinquency (206).

There is evidence for earlier puberty being both beneficial (associated with greater career success and marital satisfaction (205, 207) and detrimental (associated with substance misuse and aggressive or anti-social conduct (49, 205)) in boys (49, 205). Conversely, late maturation has also been shown to predispose males to anxiety and higher rates of excess alcohol consumption, with one investigator proposing these behaviours as a coping strategy for diminished social dominance due to later maturation and achievement of adult stature (208). The lack of consistency indicates a need for further research to better clarify any associations or causative factors.

Early maturing adolescents are often more popular, possibly due to achieving adult bodies (and the effect of popular media on perceptions of this as being superior). However, females face challenges to their wellbeing and health from early puberty. These may result from associating with chronologically older peers of similar physical development, or personal dissatisfaction with the rate of changes to their body, in particular changes in adiposity. For males, the earlier achievement of adult stature and enhanced musculature may promote physical prowess, with attendant benefits to self-esteem. However in one North American study, it was found that regardless of early or late timing (as measured by two standard deviations from the mean) of peak height velocity, adolescent males and females both reached similar final heights – the main difference between the groups was the pace and timing at which they achieved these heights, indicating physical status is not necessarily determined solely by timing of pubertal onset (70).
1.5.2. Pubertal Tempo

While the degree of hormonal change during the course of puberty is well documented, the effects of pubertal tempo (time from onset to completion of puberty (91)) have been less well studied. Further, the effects of different pubertal tempos on adolescent health and wellbeing outcomes have not been studied in as much detail as have the effects of variations in timing of pubertal onset. The duration of puberty is thought to be typically four years (range one to seven years) (55, 109).

Longitudinal studies in healthy female European adolescents have shown that the tempo of pubertal change (time from Tanner breast stage 2 to menarche) was inversely related to age at Tanner breast stage 2, with longer duration of puberty (and thus slower maturational tempo of change) in girls who entered Tanner breast stage 2 at an earlier age (12, 209, 210). In contrast, girls entering Tanner stage 2 at a later age underwent menarche at a similar age to their counterparts, therefore having a greater tempo of change. The girls with earlier onset (Tanner stage 2 at 10 years) were initially shorter than the later maturing girls (Tanner stage 2 at 13 years), though their final heights were not significantly different, indicating different growth patterns (210). In an American study of white and African-American females, earlier-maturing girls had greater height velocity and absolute height increase than those who matured later, and early-maturing girls were more likely to be shorter at the conclusion of their pubertal growth spurt (71). One study that looked at boys showed that the age of puberty commencement does not significantly affect the final height; rather the period of growth changes to compensate, meaning that height growth tempo increases with later onset to reach an adult height similar to early-
maturing boys (211), even though early-maturing adolescents started growth from a shorter height, meaning their absolute height gain was greater.

There is no universally accepted definition of puberty, or for its measurement, which may account for some of the differences in findings. Further, none of the above studies performed any direct assessment of hormonal levels, or their change over the course of maturation. Hormonal assessment may offer a better explanation to these observed phenomena and provide a concrete biological explanation as to why growth rates change. Work by Marceau showed that males with increasing levels of testosterone and dihydrotestosterone at a younger age had a slower pubertal tempo (212).

The relationship between pubertal tempo and psychological outcomes has been described in terms of a “maturational compression” hypothesis (109, 213), in which a faster development is thought to exacerbate the challenges that accompany puberty as the adolescent must adapt more rapidly to changes. However, Beltz found that tempo was not clearly correlated with psychological outcomes, whereas timing was (213). Marceau notes that there is a dearth of evidence regarding the effects of pubertal tempo (using Tanner staging, not hormone levels) on adolescents’ psychological wellbeing (91). For those undergoing a rapid pubertal transition, arising psychological issues may be due to the inability to meet the new challenges arising from faster biological and social changes than their peers, whereas a normal maturational tempo may allow one to accustom themselves with such changes in a constructive way and at a manageable pace (91). More rapid tempo has been associated with depressive symptoms, social difficulties and substance use in males (193, 214, 215) and psychological problems in females (91). However, this
hypothesis is yet to be confirmed: as an example, separate studies on pubertal tempo and depression showed opposite effects: while one found a rapid tempo (as measured by rate of change in PDS score) was linked to greater risk of depression (216), another two found tempo (as measured by rate of change in Tanner stage) protective (217), including at ten years post-puberty (218). Additionally, Marceau and colleagues found tempo (time rate of change in Tanner stage between Tanner stages 1 and 5) was positively associated with male externalising behaviours (aggressive and delinquent) and female internalising (anxious or depressive) and externalising behaviours (91).

1.5.3. Hormone Changes

The contributions of sex hormones to physical growth were described previously in this review (section 1.2.3). Studying the associations between pubertal sex steroid changes and mood and behaviour in normal adolescents only began in the 1980s, with the advent of radioimmunoassay (219), though these measurements have now been superseded by the modern, highly accurate LC-MS/MS.

Evidence for the contributions of the primary sex steroids testosterone and oestradiol to adolescent affect and behaviour is limited. The majority of the literature addressing this question involves cross-sectional analyses, which limits the ability to form conclusions. Causality cannot be inferred from cross-sectional studies, as the data may simply show an association that could be a persistent or short-term change or even coincidence (208). Likewise, the direction and progression of the effect can only be postulated, and moreover the lack of a statistically significant association may be due to confounding factors not otherwise accounted for. A longitudinal
cohort study with sufficient power to draw conclusions would offer the best confirmation of any effect of gonadal hormone changes on adolescent mood and behaviour. The contributions of hormones to sexual behaviours will not be examined in this literature review, as pubertal development is implicitly necessary universally for sexual behaviour. Further, the long-term implications of off-timed puberty hormone changes are outside the scope of this discussion.

1.5.3.1. Testosterone

Owing to evidence that testosterone levels correlate with aggressive behaviours in animals (220, 221) and the recognised effects of excessive anabolic steroids in humans (222), there has been much study into the effects of endogenous testosterone in male humans. Higher testosterone has been found to correlate with greater aggression in several cross-sectional studies in adolescent males (221, 223, 224) and more recently in cross-sectional investigations of both sexes (225). However, there are numerous studies that contradict such findings (189, 204, 226, 227). Added to this is evidence of testosterone’s relationship with social dominance. Vermeersch found females with higher testosterone levels were more likely to have dominant personality traits (as scored by the California Personality Inventory – Dominance Scale). This result was not observed in the males in their cohort. (228). Since there is at least a tenfold difference between testosterone levels in males and females by the end of puberty, there is difficulty interpreting the effects of the one hormone without accounting for changes in other hormones, including oestradiol for females.

Aggression can also be manifested in behavioural problems. Some cross-sectional male studies (229, 230) observed positive correlations between testosterone and anti-
social and problem behaviours, while other such studies did not (223, 231). A significant association has also been demonstrated between testosterone levels and assaultive behaviour in boys who live in “adverse” neighbourhoods (232). However, there exist numerous confounding factors, including the nature of the neighbourhood – Tarter states that in such socioeconomically depressed settings, notions of male honour are often demonstrated physically (232), which may prompt violent and assaultive behaviour, regardless of hormone levels.

In two cross-sectional studies of Belgian youth, one male and the other female, Vermeersch and colleagues found boys had significant positive associations between free testosterone and non-aggressive risk taking (sensation seeking and novel behaviour that is not aggressive or violent in nature). No association in this study was observed between free testosterone and aggressive risk taking (233). In contrast, no relationships were observed for either free or total testosterone and either aggressive or non-aggressive risk taking in females (234).

Not all risk taking involves physicality. Substance use is a common risk behaviour in adolescents, and a study by Foshee has observed that testosterone levels negatively correlate with cigarette and alcohol consumption for both sexes, with peer context playing an important role in mediating this association (235). In contrast to Foshee, Udry found that higher testosterone levels were related to alcohol abuse in adolescent boys (236). The conflicting results observed between studies may be due to the communities in which they are set, sample sizes and participant ages. Additionally, there are few large-scale longitudinal studies considering causative relationships between hormone levels and any of these behaviours. Furthermore, none of these
1. Puberty and Adolescence in Context

studies used mass spectrometry to measure hormone levels and accuracy of the data may have played a part.

Depressive affect and anxiety disorders increase in prevalence during adolescence. Angold’s longitudinal Great Smoky Mountains Study found higher testosterone was associated with an increased odds ratio for depression for girls (237). Associations between testosterone and depression were not found in cross-sectional studies (189, 204, 238), nor did these observe any significant relationships between testosterone and measures of anxiety. Booth has noted that there was an important modifying effect of family interactions with testosterone in its association with depression: positive family relations led to non-significant associations with depression, whereas negative family relations interacted with testosterone to positively correlate with depressive symptomatology (239). This latter finding suggests the complexity of the interactions between hormonal changes and environmental factors in the onset of adolescent depression, amongst a variety of other conditions that emerge during this life phase.

1.5.3.2. Oestradiol

Studies involving oestradiol tend to focus on female adolescents, though some also involve males. In adult males, oestradiol levels are similar to females in their menstrual/early follicular phase. For the remainder of the menstrual cycle, oestradiol levels in females are up to ten times higher than in males. Additionally, studies do not account for the potential confounding effects of testosterone, which in adult males is consistently at least ten times higher than in females.
Oestradiol was associated with depression most strongly during the transition from pre-pubertal to adult hormone levels in two cross-sectional analyses from Brooks-Gunn and Susman (191, 240). However, more recent work from Angold and Graber in this field has found linear associations between oestradiol and female depressive affect (237, 241). Angold’s study is the largest longitudinal study to date (339 adolescents followed for three years) to consider whether there is a relationship between hormone changes and depression (237), though there is a limitation to the applicability of their findings as specimen collection time was not strictly controlled.

As for anxiety, Rapkin observed a negative correlation between this mood state and oestradiol in females (242).

Adolescent behavioural changes have also been associated with oestradiol levels. These studies are all cross-sectional in their analyses and therefore only provide preliminary associations of interest, which require longitudinal studies to confirm whether the hormone-outcome association is of a causative nature. For example, analyses observed that lower oestradiol correlated with delinquent behaviours (204, 243) and another showed that higher oestradiol correlated with positive behaviour (238). Risk taking behaviours in both adolescent males and females positively correlate with oestradiol levels, though peer associations affected this outcome (233, 234). Additionally, alcohol consumption was associated with a higher mean oestradiol level in one study on females (244).

In summary, there is insufficient evidence to suggest a clear role for either testosterone or oestradiol in adolescent behavioural, mood or affect changes, although it remains likely that there will be some role demonstrated in appropriately devised longitudinal studies. The longitudinal studies discussed have had either small
sample sizes or insufficient time points that limit the power from which their conclusions can be drawn. The cross-sectional evidence that has been discussed suggests that a relationship might exist between hormonal changes and behaviour/affect. However, it is likely that this would be heavily modified by a variety of other factors which interact during this phase of the adolescent’s life, including changing peer and family relations, education and their general social environment. Any future study on hormonal changes and behaviour or affect changes should be longitudinal in its design, with sufficient participants for adequate power. A variety of behavioural and affect measures should be considered to better typify these states during adolescence and how these change with the changing hormonal milieu, and to account for all possible confounders which may interact and influence any changes.

1.6. Concluding Remarks

Puberty and adolescence represent universal transitions for the individual, though numerous gaps in the evidence remain regarding their implications for health and wellbeing. Adolescence, while allowing the development of peak physical and reproductive capacity for the individual, involves and introduces health risks that may affect the individual both during and after this life phase.

Using clinical staging or cross-sectional hormone analyses offer some insights into the effects of the pubertal transition upon the adolescent. However, the evidence is often inconsistent and does not adequately account for the full effect of the hormonal transition as would a longitudinal cohort study. It is clear that profound physical and neurological changes occur during the second decade. The physical changes create
an adult body and allow the individual reproductive capability. Neurological changes improve cognitive ability and executive function, although may predispose to increased risk taking and sensation seeking, both causes for concern given their associations with increased adolescent morbidity and mortality. The contribution of pubertal hormone changes to these are not entirely elucidated, although given the contemporaneous nature of hormone and physical/neurological changes, consideration must be given to how differing hormonal milieu and the interaction with an adolescent’s environment may alter physical, mood or behavioural systems.

1.6.1. The ARCHER Study

The Adolescent Rural Cohort, Hormones and Health, Education, Environments and Relationships (ARCHER) study (1) is a three-year longitudinal study of 342 adolescents (153 females; ages 9-14 at baseline) and their parents (n=272; 239 female), living in rural New South Wales, Australia, centred on the towns of Dubbo and Orange. It was designed to be the most intensive biological study of puberty hormones ever undertaken, with three-monthly urine collection and annual blood collection for hormone analysis (and other biological variables of interest). Novel assay methodologies have been developed, using liquid chromatography-tandem mass spectrometry (LC-MS/MS) on urine and serum specimens, which will allow the detection of sex steroids at much lower and accurate levels than previously achieved (135, 136, 245). As well as collection of biological samples (serum and urine), participants undergo anthropometric assessment and complete a lengthy psychosocial survey annually. Their parents also complete a similar survey about their child as well as proving data about family, home and community context. The principal aim of the ARCHER study is to determine how
the timing and tempo of hormone changes affect adolescent behavior, health and wellbeing. In so doing, many of the questions raised in this literature review should be answered.

As regular biological sample collection is intrinsic to achieving the aims of the ARCHER study, it is important for adolescents (and their parents) to comply with study protocols, particularly the three-monthly home-based urine sample collection. Given the large study catchment area, participants need to be easily contacted for specimen collection. Text messaging is being used in the ARCHER study, however the utility of this as a research tool is unclear (246, 247). Potential benefits of text messaging include its relatively low costs and the ubiquity of mobile telephones (248-250) and adolescents’ familiarity with text messaging as a communication tool (251, 252). Indeed, the utility of text messaging in healthcare settings has been demonstrated previously (249) and also in child and adolescent populations in research as a data collection tool (253, 254). However, its role in enhancing compliance to protocols in adolescent research has not yet been established.

The published methodology of the ARCHER study is provided for reference in Appendix A.

1.7. References

1. Puberty and Adolescence in Context

1. Puberty and Adolescence in Context


64. Saggese G, Barongelli GI, Bertelloni S. Puberty and bone development. Best Pract Res Clin Endocrinol Metab. 2002;16(1):53-64.


73. Anderson M, Blais M, Green WT. Growth of the normal foot during childhood and adolescence; length of the foot and interrelations of foot, stature, and lower extremity as seen
1. Puberty and Adolescence in Context


1. Puberty and Adolescence in Context


1. Puberty and Adolescence in Context


1. Puberty and Adolescence in Context


1. Puberty and Adolescence in Context

1. Puberty and Adolescence in Context


1. Puberty and Adolescence in Context


2. GENERAL METHODS

The detailed protocol for the Adolescent Rural Cohort, Hormones and Health, Education, Environments and Relationships (ARCHER) study, upon which this thesis is based, is provided in Appendix A. Note that non-published chapters use British spelling while the published chapters (Chapters 3-8) use the requisite spelling for the journal’s requirements.

2.1. Aims and Hypotheses of this Thesis

The overall purpose of this thesis is to examine relevant research methodologies that are needed to answer important questions regarding puberty hormone changes and the relationship between these and other changes during adolescence. This thesis has five aims, each of which will be addressed in chapters 3-8. The research conducted to meet these aims will provide a multi-dimensional demonstration of novel techniques in adolescent research, through which the ARCHER study can provide a deeper understanding of the longitudinal effects of puberty on adolescent health and wellbeing.

The aims of this thesis are:

1) Summarise, in greater depth, the present body of knowledge regarding the effect of oestradiol levels on female adolescent mood and behaviour and the effects of testosterone on male adolescent mood and behaviour, so as to identify gaps in the evidence. This is a way to provide context to the ARCHER study going forward and indicate ways that a longitudinal study of adolescence and pubertal hormones can add to our understanding of the complexities of this life phase (see Section 1.5.3).
2. General Methods

2) Demonstrate the feasibility of the ARCHER study methodology with respect to specimen collection. Investigating the role of text messaging as a novel research tool to enhance compliance with the study protocol will supplement this, as regular repeated sample collection is crucial to the success of the ARCHER study, and indeed to future studies involving adolescents where frequent biological specimen collection is required.

3) Demonstrate the feasibility of the ARCHER study methodology to analyse collected specimens and describe preliminary outcomes on relationships between urine puberty hormones measured by liquid chromatography-tandem mass spectrometry and more traditional research descriptors of puberty.

4) Investigate the longitudinal relationship between serum hormone levels and adolescent and parental subjective reports of pubertal progress in the ARCHER study cohort to determine the applicability of these assessments in longitudinal studies where direct physical examination of pubertal status is not possible. This would also determine how subjective assessments are related to hormone levels and whether these have sufficient accuracy to identify the onset and progress of puberty, a finding of utility to both clinical and research settings.

5) Investigate the use of foot length change as a possible marker of early puberty. Some data exist regarding foot length in puberty (see Section 1.2.4) but no studies have considered the relationship between this possible anthropometric marker and hormone levels.

For each of the above aims, and based upon the literature review presented in Chapter 1, it is hypothesised that:
2. General Methods

1) There are insufficient longitudinal studies of high-quality design to support popular assertions of a causative link between oestradiol or testosterone changes and alterations in mood and behaviour for female or male adolescents, respectively.

2) Text messaging will be a useful reminder tool for adolescents participating in the ARCHER study, which would be indicated by a significant positive correlation between message response time and specimen collection time. This in turn is hypothesised to lead to an increased specimen collection rate and compliance with study protocol.

3) The more frequent sampling and LC-MS/MS methodology for urine hormone analyses will provide novel insights into hormone changes over one year of the ARCHER study.

4) The changes in urine and serum hormones will be associated with longitudinal changes in traditional markers of puberty, specifically anthropometry (height, weight, BMI) and self-rated Tanner stage.

5) Adolescents will be able to self-rate their Tanner staging with some accuracy (1-6) and this will be significantly and positively associated with hormone levels. However, parental assessment of puberty is hypothesised to be less accurate, possibly due to increased adolescent privacy as their bodies change. Both ratings will be sufficiently acceptable, when compared to the true markers of pubertal progress (hormone levels) to be used in research settings where direct clinician examination is not possible.
6) Foot length is a novel marker of early puberty and foot growth and is related to the hormonal drivers of pubertal changes. It could provide a useful, cost-effective and novel way to assess early puberty. Studies looking at Tanner stage change and peak height velocity have shown foot growth is an earlier event in puberty, occurring between Tanner stage 1 and 2 and prior to peak height velocity (7, 8). Therefore, it is hypothesised that foot growth will be greatest in adolescents whose hormone levels are relatively lower than their peers.

As this thesis is being completed by publication, the specific methodologies for each chapter are included in each publication. The rest of this chapter provides details of broader methods employed across the thesis, such as study setting, hormone assay techniques and statistical analysis methodology.

2.2. Subjects and Setting

The ARCHER study enrolled 342 adolescents (153 females; ages 9-14 at baseline) and their parents (n=272; 239 female), with some sibling recruitment, and followed them over three consecutive years. The study was set in Central Western New South Wales (NSW), specifically around the towns of Dubbo and Orange where research offices were established for annual participant visits. These are large regional towns whose local hospitals are affiliated with the University of Sydney, the overseeing institution for the ARCHER study (9). As demonstrated in Table 2.1, there were no significant differences in gender proportions in the recruited cohort of ARCHER study compared to either the local government area (LGA) or the state of NSW for the age range included in the study (10). There was no significant difference in the proportion of adolescents identifying as Indigenous recruited as compared to living
in the LGA, though the proportion of Indigenous participants in the ARCHER study is over double the percentage of Indigenous adolescents in NSW (10). Recruited parents were more likely to be employed full time, speak English as a first language and have a tertiary or post-graduate qualification than those in the LGA or the rest of NSW (10). Lack of parental competence English was an exclusion criterion for the ARCHER study (9). A detailed description of the ARCHER study cohort is provided in Appendix B.

By focusing on adolescents in regional and rural Australia, the ARCHER study will give added insights into growing up and transitioning through puberty in a relatively under-researched geographic population. Rural adolescents, both in Australia and abroad, have their own unique health risk profile (11, 12). Feasibility studies were conducted before the rolling recruitment of the ARCHER study proper commenced in 2012. Given the number of participants, timeframe of the study and enormity of the data collected and its complex longitudinal analysis (discussed further in this chapter), this thesis addresses a subset of data from within the longitudinal study.

Recruitment occurred through schools, local community groups and both traditional (newsletters and advertisements) and social (Facebook page) media (for parents) (9, 13). Adolescents provided informed assent and their parents informed consent. The ARCHER study has full ethics approval through the University of Sydney (HREC 13094) and within the Australian National Health and Medical Research Council Guidelines for Human Experimentation, which are consistent with the Declaration of Helsinki.
Table 2.1. ARCHER baseline demographics compared with the comparison local government areas (LGAs) and New South Wales (NSW).

<table>
<thead>
<tr>
<th></th>
<th>ARCHER(^1)</th>
<th>LGA(^2)</th>
<th>NSW</th>
<th>Statistics</th>
<th>ARCHER vs Region</th>
<th>ARCHER vs NSW</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adolescents</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>55.3 (189)</td>
<td>51.2 (2,947)</td>
<td>51.3 (179,821)</td>
<td>p value = 0.147</td>
<td></td>
<td>0.158</td>
</tr>
<tr>
<td>Females</td>
<td>44.7 (153)</td>
<td>48.8 (2,807)</td>
<td>48.7 (170,500)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Indigenous status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indigenous</td>
<td>11.3 (38)</td>
<td>16.6 (907)</td>
<td>5.0 (16,546)</td>
<td>p value = 0.011</td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Non-Indigenous</td>
<td>88.7 (299)</td>
<td>83.4 (4,571)</td>
<td>95.0 (316,951)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) Some Adolescent data are missing so not all categories total to N = 342.
\(^2\) The regional comparison area includes LGAs of Cabonne and Wellington areas and the cities of Dubbo and Orange.

Data generated from ABS 2011 Census of Population and Housing.
2. General Methods

2.3. Surveys, Anthropometry and Specimen Collection

Adolescents responded to a survey and underwent anthropometric assessment annually. Parents answered survey questions at the same time as the adolescent. Adolescent and parent surveys focused on adolescent health and well-being, mood and behaviour, as well as questions regarding living situation, family relations, education, risk taking behaviours and sexuality, among others. Adolescents provided self-rated Tanner stage based on line drawings, and both parent and adolescent were asked simple subjective questions about the adolescent’s pubertal development. Close to the time of the annual survey, adolescents provided fasted, first-morning blood collections for analysis and a first morning urine sample. Every three months between annual visits, a collection kit was sent to the adolescent for further first-morning urine samples. A text message was sent to the adolescent (or their parent) on the morning of collection to prompt their collection. The message also asked the adolescent to provide a rating of their mood on a linear scale from 0 (worst) to 9 (best ever). See Section 5.3 for further details.

At the time of the annual clinic visit, adolescents underwent anthropometry assessment. Height (to 0.1 cm) was measured using a portable stadiometer (Seca GmbH, Hamburg, Germany). Weight (to 0.1 kg) was measured in light clothing using a Tanita TBF-300 Pro Body Composition Analyzer (Tanita, Tokyo, Japan). Foot length (to 0.1 cm) was measured using a Ritz stick foot measure (Footer, Wisconsin, USA), modified with a millimetre-graduated measuring tape (to 0.1 cm). The foot measure was placed under the sole of the foot and length measured from the heel to the end of the distal phalanx of the hallux. The right foot was measured
2. General Methods

unless this was inaccessible, in which case the left foot was used (3 of 1316 measurements, in three separate adolescents). As anthropometric assessment occurred annually, there was concern growth curves and peak height velocity could not be adequately determined given the relative infrequency of anthropometry. Additionally, over three years’ follow-up, the full extent of pubertal growth may be missed. Recent work by Cole, using superimposition by translation and rotation (SITAR) modelling, compared longitudinal height data (128679 measurements of 3172 boys) with intervals of two, three, four, six, twelve and twenty-four months, and found that measuring at intervals between two to twelve months provided essentially identical growth curve models for generating peak growth data and therefore concluded that annual anthropometric assessment is just as informative as more frequent measurement (14). This was reassuring for the ARCHER study, as the sheer geography of the study catchment area and the human resources available would not facilitate more frequent clinic visits, particularly outside school holiday periods.

Serum and urine specimens were frozen and transported to Sydney for storage in alarmed -80 °C freezers. Urine samples underwent hormone assay using liquid chromatography-tandem mass spectrometry (LC-MS/MS) for sex steroids: oestrone, oestradiol, testosterone, dihydrotestosterone, androstenedione, and dihydroepiandrosterone (DHEA). The ARCHER study has ethics approval for future analysis of other analytes of interest related to pubertal development and metabolic health. This thesis will principally look at the oestradiol and testosterone data. A subset of urine samples was analysed with immunoassay for both luteinising hormone (LH) and follicle stimulating hormone (FSH) using an Immulite analyser.
2. General Methods

Only LH satisfied the requirements for specificity and FSH analysis in urine was therefore abandoned. Collection date was controlled for post-menarcheal females to be in the early-mid-follicular phase (menstrual cycle days 7-10). While this may not control for pre-menarcheal ovarian activity, it ensured standard collection time in girls who were post-menarcheal.

2.4. Liquid Chromatography-Tandem Mass Spectrometry

As described in Section 1.3.3, LC-MS/MS involves liquid chromatography, which separates compounds of interest from other analytes in the specimen (15) and then mass spectrometry, which nebulises and ionises the chromatographic eluate (16). In the ARCHER study, atmospheric pressure photo-ionisation is used to generate ions of steroid hormones (17). The composition of a mixture (such as urine or serum) is determined by analysis of both the precursor and product ions, with minimal cross-reactivity (compared to immunoassay), which gives LC-MS/MS a high level of sensitivity and specificity. This is useful in early puberty, as testosterone and oestradiol levels are too low for other assay methodologies to accurately determine (such as radioimmunoassay and enzyme-linked immunosorbent assay, unless the laboratory uses in-house assays specifically developed for lower range levels; see Section 1.3.3) (16). It should be noted that many peer reviewed journals will now only accept gonadal hormone results obtained by mass spectrometry (18).

Urine and serum specimens were separated by liquid chromatography using a Shimadzu Nexera UHPLC system (Shimadzu Scientific Instruments, Columbia, USA). Following this, tandem-mass spectrometry analysis was performed on samples using an API-5000 triple-quadrupole mass spectrometer (Applied
2. General Methods

Biosystem/MDS SCIEX, Ontario, Canada) (17). Both urine and serum analyses followed previously described methodologies (17, 19). A thorough validation was carried out for the LC-MS/MS method to measure oestradiol and testosterone according to standard analytical validity criteria (20). Urine analysis requires an extra extraction step that lengthens the procedure by a day longer than for serum extraction for LC-MS/MS analysis, making urine analysis currently more expensive and resource intensive (17). Urine specimens were adjusted for urine creatinine or urine specific gravity to correct for the influence of hydration status on urine analysis, but such an adjustment did not appear to be necessary, at least in adolescents, in a preliminary methodological paper by our group (Appendix C). Deming regression and Bland-Altman comparison of the unadjusted and adjusted measured concentrations both showed a lack of bias, whether uncorrected or corrected for specific gravity or urine creatinine (17). This possibly reflects the standardisation of urine collection as a first morning specimen which ensures a more controlled hydration state (17). For further detail, refer to Section 6.3.3.

2.5. Statistical Analyses

This thesis required multiple techniques to be used for the analysis of our longitudinal dataset, with multiple time-based variables. The repeated measures allowed the use of longitudinal regression techniques such as mixed-models linear regression. These will aid our understanding of puberty in general by providing results using longitudinal sets of data that provide more robust analysis than that possible with cross-sectional data. For this thesis specifically, using the collected data will provide novel findings to demonstrate the feasibility of longitudinal research in adolescents, both through determining the compliance to an intensive
specimen collection regimen (Aim 2) and that using such intensive sampling can be associated with anthropometric changes (Aim 3) and parental and adolescent rating of pubertal progress (Aim 4) as well as foot length change providing a novel marker of early puberty (Aim 5). This will provide a platform from which further longitudinal analytical models can be developed. The data and their analysis will ultimately allow translation of these findings to answer questions about the effects of puberty on adolescent health and well-being (as demonstrated through Aim 1). These are findings which we hypothesise will be related to onset of hormone change in part, as well as to tempo of change and stability of that hormone change.

Collected data were divided into binary, ordinal and continuous variables. Binary variables included gender and the achievement of menarche for females. Ordinal data were self-rated Tanner stage and text message mood scores. Continuous variables were age, height (in m), weight (in kg), body mass index (BMI; kg/m²), foot length (cm), serum and urine testosterone (nmol/L), oestradiol (pmol/L) and luteinising hormone (IU/L) concentrations, time to reply to text message and time to provide urine sample. Differences in continuous variables were compared using Student’s t-test (significance p<0.05).

In Chapter 5, Pearson’s correlation coefficient was used to test the relationship between time until specimen collection and time to reply to text message. In Chapter 6, linear regression was used to determine relationships between changes in anthropometry (height, weight, BMI) and self-rated Tanner stage with changes in urine and serum hormones between baseline and one-year follow-up. For these, the models were adjusted for the baseline anthropometric measurement and baseline hormone concentration. Mixed-models linear regression was used to assess
longitudinal changes in serial measures of hormones with gender and collection time from baseline as covariates, with models including a random effect for each child.

In Chapter 7, the relationship between serum hormone concentration and survey response questions (self-rated Tanner stage, parent and adolescent questions regarding pubertal development) were assessed using mixed-models linear regression (hormone levels as outcome) or mixed-models logistic regression (dichotomous survey responses as outcome) with models including a random effect for each child. Cohen’s kappa (κ) was used to assess agreement between adolescent and parent questionnaire results. Hormone data were logarithmically transformed as these were skewed, so regression results were presented as odds ratios. Chapter 8 uses mixed-models linear regression to determine the relationship between hormone level changes and changes in anthropometry (foot length, height and weight) and Tanner stage with random slopes for each individual. In Chapters 7 and 8, adolescents were classified as pre-pubertal if serum testosterone was < 0.5 nmol/L (males) or serum oestradiol < 40 pmol/L (females). These definitions were based upon a large cross-sectional clinical research study defining hormone confidence intervals using LC-MS/MS in adolescents and adults (21, 22) and conservative cut-offs were set.

Mixed-models linear regression allows for time-dependent covariates, which enables its use in longitudinal research that involves the repeated measurement of the same variables over time. In general, it estimates longitudinal changes in an outcome (23) and allows for missing data points which are to be expected in studies of free living humans, which gives an advantage over analysis of variance methods that do not provide individual-specific analysis and are less suited to missing data points (24). Longitudinal models allow for a clearer interpretation of changes in outcomes over
time. In contrast, cross-sectional studies allow less time-intensive (and thus cheaper) data collection as only one collection point is required, which may simplify research with adolescent populations, an age group that is under-researched, and which offers both unique research challenges and opportunities (25). The cost of cross-sectional approaches is that analyses only offer a “snapshot” which may mean associations are under- or overestimated, or even caused by chance, with longitudinal analyses allowing a truer estimation of the relationship between exposure and outcome over time, and thus indicating any causative relationships. The implications of this are explained further in Chapters 3 and 4.

This analytical strategy will offer a better understanding of the changes that occur in puberty by providing, for the first time, frequent repeated samples collected from a large cohort of community adolescents who are commencing or in early puberty. These data can be used to demonstrate the nature of hormone changes during this time, in addition to changes in anthropometric and psychosocial outcomes to give a holistic perspective of this transition. Using the aforementioned statistical methods, longitudinal regression models can be developed to demonstrate how pubertal hormone changes may truly affect the adolescent over the study duration. These could then be translated to provide population-based curves that may be used to predict those at risk of adverse outcomes during puberty, such as risk-taking, mood or behavioural changes, so that targeted interventions can be developed. If such identification of at-risk youth occurred, the approach would clearly not be aimed at altering the hormonal trajectory of physiological puberty, rather it would provide other interventions and support that may ameliorate adverse risks. For the
preliminary data in this thesis, we aim to show how the full study data could be used for such longitudinal analysis at the completion of the ARCHER study.

2.6. Concluding Remarks

In undertaking this thesis, I have learnt about many of the challenges of research in adolescents, and why they remain an under-researched population with respect to longitudinal studies with biological data collection. The attribution of pubertal hormone effects on adolescent mood, behaviour and wellbeing are yet to be determined. The ARCHER study is adequately powered to determine that attribution (9). The findings of this thesis are novel and represent the first of many exciting and explanatory findings which go some way to elucidate, for the first time, the true effects of puberty hormones.

2.7. References

2. General Methods


3. THE EFFECTS OF OESTRADIOL ON MOOD AND BEHAVIOUR IN HUMAN FEMALE ADOLESCENTS
A SYSTEMATIC REVIEW

Ben W.R. Balzer¹,², Sally-Anne Duke³, Catherine I. Hawke⁴, Katharine S. Steinbeck¹,²

Affiliations: ¹ Academic Department of Adolescent Medicine, The Children’s Hospital at Westmead, Westmead, NSW, Australia; ² Discipline of Paediatrics and Child Health, Sydney Medical School, The University of Sydney, NSW, Australia; ³ Department of Endocrinology, Royal North Shore Hospital, St. Leonards, NSW, Australia; ⁴ School of Rural Health, Sydney Medical School, The University of Sydney, Orange, NSW, Australia

3. Oestradiol and Female Adolescent Mood and Behaviour

3.1. Abstract

Mood disorders and health risk behaviours increase in adolescence. Puberty is considered to contribute to these events. However, the precise impact of pubertal hormone changes to the emergence of mood disorders and risk behaviours is relatively unclear. It is important that inappropriate attribution is not made. Our aim was to determine what is known about the effect of endogenous oestradiol on human adolescent girls’ mood and behaviour. The databases searched were MEDLINE, Embase, PsycINFO, ERIC, Pre-MEDLINE, Web of Science and Scopus for all dates to October 2014. For inclusion, contemporaneous hormone and mood or behavioural assessment was required. Data were extracted following a template created by the authors. Fourteen studies met our inclusion criteria. There was some consistency in findings for mood and oestradiol levels, with associations between oestradiol and depression, and emotional tone and risk taking. Results were less consistent for studies assessing other mood and behavioural outcomes. Most studies were cross-sectional in design; assay methodologies used in older studies may lack the precision to detect early-pubertal hormone levels.

Conclusion: Three longitudinal and several cross-sectional studies indicate potential associations between oestradiol and certain mood or affective states, especially depression and mood variability though there are insufficient data to confirm that the rise in oestradiol during puberty is causative. We believe that it is important for health professionals to take care when attributing adolescent psychopathology to puberty hormones, as the current data supporting these assertions are limited.
3.2. List of Abbreviations:

AQ Aggression Questionnaire; BESAA Body Esteem Scale for Adolescents and Adults; CAPA = Child and Adolescent Psychiatric Assessment; CBC Child Behaviour Checklist; CDI Children’s Depression Inventory; CPA Children’s Physical Activity Scale; CPI California Psychological Inventory; CSI Children’s Somatization Inventory; DISC Diagnostic interview Schedule; DSM-IV Diagnostic and Statistical Manual of Mental Disorders, 4th Edition; ERIC Education Resources Information Centre (ERIC); MAACL Multiple Affect Adjective Checklist; MASC Multi-dimensional Anxiety Scale for Children; MEBS Minnesota Eating Behaviours Survey; MTFS Monitoring the Future Survey; OSIQ Offer Self-Image Questionnaire; PACES Physical Activity Enjoyment Scale; PDS Pubertal Development Scale; PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses; SIQYA Self-Image Questionnaire for Young Adolescents; SPP Self Perception Profile for Children; SSAS Sensation Seeking and Anxiety States Test; YBP Youth Behaviour Profile; YSR Youth Self Report

3.3. Introduction

Adolescence is a formative time for an individual’s identity, development and functioning (1). During this time, long term behavioural and affective patterns emerge, as well as psychopathologies (2). Puberty is the universal biological event of adolescence, the primary purpose of which is to achieve adult reproductive capacity (3, 4). Puberty involves dramatic and well-remembered physical changes (5) and it is
perhaps unsurprising that intuitively these changes are allocated importance in both physical and psychosocial contexts.

The popular assertions of adolescence as a time of “puberty blues”, “storm and stress” or other behavioural changes (6, 7) are palpably not correct (8) for all adolescents. Certain problem behaviours and mental health issues do emerge during adolescence (1, 9), and deterioration in health may be linked to such behavioural changes (10). These include the appearance of sex differences in the prevalence of depression (the prevalence of depression in females is twice that of males) (11), increased sensation seeking (2, 12), substance use (13) and disordered eating (14). For girls, depressive illnesses, self-image and eating disorders are of particular concern. Gender differences and the coincidental increase in pubertal hormones at this time of increased risk implicate some role for sex hormones. However there exists no systematic analysis of the literature.

Given the importance of non-biological factors such as social, family and peer relations in the development of adolescent behaviours and affect (15), it could be postulated that hormones influence adolescent behaviour indirectly, either by modulating internal status reactivity to modify affect, and/or by the effect on others of phenotypic development (6). While the hormone levels remain high after puberty, behavioural and mood disorders often ameliorate, giving rise to popular notions of “puberty blues” being confined to this period alone. Why some of these mood and behavioural changes fail to abate in certain individuals is unclear. It is important to understand the true effect and duration of that effect of sex hormones on adolescent mood and behaviour, so that clinicians can target established, evidence-based interventions to those most at risk (5). Additionally, while diagnosable psychiatric
conditions increase in prevalence, many parents will describe mood and behaviour changes despite absence of an identifiable mental illness. While these changes fall within a normal spectrum of mood or behaviour, these can still be confronting and challenging to the parent and adolescent alike.

The age of onset of puberty in many Western countries has declined over the recent decades, (3, 16), with earlier exposure to rising levels of oestradiol, the primary puberty hormone in females. This is an additional reason to better understand how oestradiol affects the mood and behaviour in the pubertal transition, and which girls might be especially at risk for any negative impacts of their puberty hormones.

One specific mechanism for any effect of oestradiol might be through central neural monoamine systems, which have been implicated in a wide range of mood and behavioural disorders (17, 18). Oestradiol may also induce the formation of new synapses (19). Such a neuroarchitectural effect is part of the organizational and activational effects hypothesis, wherein steroids act during brain development to alter neuroarchitecture, which limits the repertoire of the brain’s responses when acted upon by that hormone later in life (activation) (17, 20). An example of such activation is the commencement of reproductive behaviour in adulthood, which is programmed earlier during the organization phase of neural development (20).

Animal studies, especially those in rodents (21, 22), have informed the role of oestradiol in adolescent behaviour. For example, the importance of oestradiol in sexually differentiated behaviour patterns in rodents (23) and other animals (24) may provide hypotheses as to sex-based behavioural differences in humans. While these studies proffer putative roles for oestradiol in human mood and behaviour, we know
that variation between different animal species is significant (25) and it is difficult to adequately model the complexities of human behaviour and mood in animal settings.

The aim of this systematic review was to determine what evidence exists for the true effect of the endogenous puberty hormone, oestradiol, on adolescent girls’ mood and behaviour.

3.4. Methods

3.4.1. Search Strategy

A systematic search was conducted to identify publications on the effect of endogenous oestradiol on mood and behaviour in healthy adolescent girls (10-19 years) using the terms as laid out in Appendix D. The following databases were searched: MEDLINE, Embase, PsycINFO, Education Resources Information Centre (ERIC), Pre-MEDLINE, Web of Science and Scopus from the date of database inception to October 2014. No language limits were set. The search strategy for MEDLINE is included in Appendix D, with search terms for the other databases modified to their requirements. Where relevant, reference lists were hand searched for further records. No initial restrictions were placed upon publication type.

3.4.2. Inclusion Criteria

To be considered for this review, study participants were female adolescents (10-19 years) from community samples, with no specified diseases. Institutionalized or incarcerated populations were excluded, unless a control group was reported separately, due to potential confounders for behavioural outcomes. Study participants must have undergone, or be undergoing spontaneous puberty. Thus studies involving
exogenous oestradiol or other oestrogens were excluded. Oral contraceptive pill use was also a criterion for exclusion, as these preparations contain synthetic and biologically potent oestrogens that suppress endogenous oestradiol production. In the event that a study did not explicitly define oral contraceptive use, authors were contacted for further information.

Oestradiol measurement in blood, saliva or urine was required, with the laboratory methodology provided. Though assay methodology and quality has improved markedly over the past decades, we did not limit studies by assay type. It should be noted that even now assays might not be able to detect very low (i.e. pre-and early pubertal) oestradiol levels with adequate sensitivity (26). We considered whether to include only studies that used a standardized time of biological data collection, in order to control for known diurnal physiological variation in oestradiol in early puberty (3). If we had adhered to this criterion the only study to be excluded would have been a large, longitudinal study in which the collection time for the majority of samples was relatively constant (27). Only one study in females controlled for cycle time (28). In the others there was no stratification for pre- or post-menarchal status, and hence no explicit control for menstrual cycle. We decided to retain these (all cross-sectional) studies, but also to address the inherent limitations of this approach in the discussion. Studies were included if outcomes were mood and/or behaviour measured by a recognized, validated tool and with the mood/behavioural measurement concurrent with hormone measurements.

Studies on the effects of oestradiol generally do not account for the potential confounding effects of testosterone, and thus testosterone was not addressed in the review. The mood or behavioural effects of progesterone were considered beyond the
scope of the review, as we were interested in the pubertal transition, rather than the mature adult ovulatory cycle where progesterone induced mood variation is possible.

Specific moods and behaviours sought included depression, anxiety, eating disorders and self-image disturbance, social interactions, aggressive, disruptive or conduct-disordered behaviour; risk taking including substance abuse. Studies were excluded if these addressed primarily sexual behaviours, as such behaviours are essential for reproduction which is the key biological function of puberty (3).

3.4.3. Data Collection and Analysis

3.4.3.1. Selection of Studies

Once irrelevant studies and duplicates were removed, one reviewer (BB) scanned the title, abstract and keywords of the remaining articles. Where a reference seemed suitable for the review, the full text was retrieved for further analysis by two reviewers (BB, KS) and was either included or excluded on the above criteria.

3.4.3.2. Data Extraction

Information was extracted from each study into a template developed by the authors. The following information was recorded: participants (sample size, sex and age range), affect or behaviour measured (type, measurement tool and its validity), oestradiol measurement (assay type, time of day and time of cycle where available), assessment of pubertal status (method used; examination, self-report or parental report), study outcomes and discussion of limitations. These data are described in Tables 3.1 and 3.2.
3.4.3.3. **Quality Assessment**

A general methodology for quality assessment was followed, based upon checklists for the evaluation of studies (30, 31) and is reported in Table 3.3 in the Results section.

3.4.3.4. **Statistical Analyses**

Given the heterogeneity of the outcomes and outcome measures in this systematic review, no further analyses (such as meta-analysis) could be performed and the results are presented as descriptive data.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (32) was followed in the writing of this review.

3.5. **Results**

3.5.1. **Search Results**

The result of the searches is displayed in Figure 3.1. Of the initial 9904 results obtained from database searches, 9574 were removed. Three additional studies were identified from manual reference list searching. After exclusion based upon abstracts, the remaining 76 citations were inspected in full text. Fourteen of these met the inclusion criteria. The reviewers had full consensus in their findings. All included publications were written in English.

The 62 full text articles which were excluded, as well as the reasons for their exclusion, are detailed in Appendix D. Participant age, lack of oestradiol measurement and oral contraceptive use were the main reasons for exclusion.
Figure 3.1. PRISMA Diagram for Study Selection.
3.5.2. **Included Studies**

Fourteen studies met the inclusion criteria. These studies are summarized in Tables 3.1 and 3.2. Thirteen studies used standard radioimmunoassay measurement of oestradiol for blood samples. One used an enzyme-linked immunoassay in saliva (33). Three studies (27, 34, 35) were longitudinal and the remainder was cross-sectional, whether by study design or analysis. Four of the studies were based upon the same participant sample (35-38). Nine of 14 studies examined mood and affect alone, and 11 of 14 of the studies examined aggression, delinquency and behavioural or conduct disorders. Six of the nine studies on aggression also included an assessment of mood and affect.

For clarity of interpretation, the variables of interest were grouped under four categories: A: mood and affect; B: self-image and social competency and related behaviours; C: risk taking, sensation seeking and substance use; and D: aggression, behaviour/conduct disorder and delinquency as shown in Table 3.1. The groupings describe similar outcomes so that a degree of inter study comparison can be made in the absence of formal meta-analysis. Most of the included studies considered more than one categorical outcome, and in order to reduce repetition we have not looked at the four outcome categories separately. In addition most studies were unable to demonstrate an association, so repetition of negative findings is also reduced. As previously stated, it was not possible, due to heterogeneous outcomes, to perform meta-analysis.
Table 3.1. Study Characteristics (males are excluded from further analysis).

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>N</th>
<th>Sex</th>
<th>Setting</th>
<th>Age (mean (SD) or range)</th>
<th>Behaviour or affect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susman et al.</td>
<td>1991</td>
<td>108</td>
<td>56 Male, 52 Female</td>
<td>USA, community based</td>
<td>10-14 (M); 9-14 (F)</td>
<td>A, D</td>
</tr>
<tr>
<td>Slap et al.</td>
<td>1994</td>
<td>54</td>
<td>Female</td>
<td>USA, community based, high schools</td>
<td>10-14</td>
<td>A, B</td>
</tr>
<tr>
<td>Angold et al.</td>
<td>1999</td>
<td>339</td>
<td>Female</td>
<td>USA, community based</td>
<td>9-15</td>
<td>A</td>
</tr>
<tr>
<td>Susman et al.</td>
<td>1985</td>
<td>108</td>
<td>56 Male, 52 Female</td>
<td>USA, community based</td>
<td>10-14 (M); 9-14 (F)</td>
<td>A, D</td>
</tr>
<tr>
<td>Nettelmann et al.*</td>
<td>1987</td>
<td>108</td>
<td>56 Male, 52 Female</td>
<td>USA, community based</td>
<td>10-14 (M); 9-14 (F)</td>
<td>B, D</td>
</tr>
<tr>
<td>Susman et al.*</td>
<td>1987</td>
<td>108</td>
<td>56 Male, 52 Female</td>
<td>USA, community based</td>
<td>10-14 (M); 9-14 (F)</td>
<td>A, D</td>
</tr>
<tr>
<td>Brooks-Gunn et al.</td>
<td>1989</td>
<td>103</td>
<td>Female</td>
<td>USA, community based, private schools</td>
<td>12.1 (0.8)</td>
<td>A, D</td>
</tr>
<tr>
<td>Warren et al.</td>
<td>1989</td>
<td>100</td>
<td>Female</td>
<td>USA, community based</td>
<td>10.6-13.3</td>
<td>A, B, D</td>
</tr>
<tr>
<td>Martin et al.</td>
<td>1999</td>
<td>94</td>
<td>Female</td>
<td>USA, community based, high schools</td>
<td>16.6 (1.0)</td>
<td>C</td>
</tr>
<tr>
<td>Graber et al.</td>
<td>2006</td>
<td>100</td>
<td>Female</td>
<td>USA, community based, urban high schools</td>
<td>12.1 (0.8)</td>
<td>A, D</td>
</tr>
<tr>
<td>Rapkin et al.</td>
<td>2006</td>
<td>106</td>
<td>Female</td>
<td>USA, community based</td>
<td>13.0 (3.0)</td>
<td>A</td>
</tr>
<tr>
<td>Davison et al.</td>
<td>2007</td>
<td>178</td>
<td>Female</td>
<td>USA, community based</td>
<td>11.3 (0.28)</td>
<td>B, D</td>
</tr>
<tr>
<td>Vermeersch et al.</td>
<td>2008</td>
<td>298</td>
<td>Female</td>
<td>Belgium, community based, high schools</td>
<td>14.3 (0.59)</td>
<td>C, D</td>
</tr>
<tr>
<td>Klump et al.</td>
<td>2010</td>
<td>258</td>
<td>Female</td>
<td>USA, twin study, community based</td>
<td>12.0 (1.40)</td>
<td>B</td>
</tr>
</tbody>
</table>

† Longitudinal analysis
* These studies are based upon the same data set
# In studies which included males, only the female data are reported
A = mood and affect; B = self-image and related behaviours; C = risk taking, sensation seeking and substance use; D = aggression, behaviour/conduct disorder and delinquency
Table 3.2. Extracted Study Data.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>n</th>
<th>Relevant Scale(s)</th>
<th>Assessment of Pubertal Status</th>
<th>Assessor of Pubertal Status</th>
<th>Control for Cycle</th>
<th>Behaviour or affect</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susman et al.</td>
<td>1985</td>
<td>52</td>
<td>OSIQ, CBC (maternal)</td>
<td>Tanner</td>
<td>Physician/ Nurse</td>
<td>No</td>
<td>A, D</td>
<td>No significant relationships observed for girls.</td>
</tr>
<tr>
<td>Nottelmann et al.</td>
<td>1987</td>
<td>52</td>
<td>OSIQ, CBC (maternal)</td>
<td>Tanner</td>
<td>Physician/ Nurse</td>
<td>No</td>
<td>B, D</td>
<td>No significant relationships between E2 and behavioural or self-image issues for girls.</td>
</tr>
<tr>
<td>Susman et al.</td>
<td>1987</td>
<td>52</td>
<td>OSIQ, CBC (maternal), Self-ratings, MAACL (maternal)</td>
<td>Tanner</td>
<td>Physician/ Nurse</td>
<td>No</td>
<td>A, D</td>
<td>No significant relation between E2 and affect in girls</td>
</tr>
<tr>
<td>Brooks-Gunn et al.</td>
<td>1989</td>
<td>103</td>
<td>YBP</td>
<td>Tanner</td>
<td>Physician/ Nurse</td>
<td>No</td>
<td>A, D</td>
<td>Non-linear effect of E2 on depressive affect: most depression during rapid increase in E2. No significant relation between E2 and aggression. Negative life events interacted significantly with hormonal changes for depression.</td>
</tr>
<tr>
<td>Warren et al.</td>
<td>1989</td>
<td>100</td>
<td>YBP, SIQYA, Maternal Depression Questionnaire</td>
<td>Tanner</td>
<td>Physician/ Nurse</td>
<td>No</td>
<td>A, B, D</td>
<td>Significant curvilinear relationship between oestradiol level and depression, impulse control and psychopathology with depression highest for E2 184-275 pmol/L. Impulse control was lowest and psychopathology highest at E2 92-184 pmol/L. (F value for depression relationship 6.78 (p=0.01); impulse control 4.45 (p=0.04); psychopathology 4.73 (p=0.03))</td>
</tr>
<tr>
<td>Susman et al.†</td>
<td>1991</td>
<td>52</td>
<td>DISC, SIQYA, CBC</td>
<td>Tanner</td>
<td>Physician/ Nurse</td>
<td>No</td>
<td>A, D</td>
<td>No significant cross-sectional relation between E2 and emotional tone, internalizing behaviour problems or symptoms of depression/anxiety in girls. Longitudinal changes in E2 positively associated with higher emotional tone in girls (r=0.26; p&lt;0.05)</td>
</tr>
</tbody>
</table>
Table 3.3. (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>n (Female)</th>
<th>Relevant Scale(s)</th>
<th>Assessment of Pubertal Status</th>
<th>Assessor of Pubertal Status</th>
<th>Control for Cycle</th>
<th>Behaviour or affect</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slap et al.†</td>
<td>1994</td>
<td>54</td>
<td>SIQYA</td>
<td>Tanner</td>
<td>Physician/ Nurse</td>
<td>No</td>
<td>A, B</td>
<td>No association between change in $E_2$ and body image/mood variability</td>
</tr>
<tr>
<td>Angold et al. †</td>
<td>1999</td>
<td>339</td>
<td>CAPA</td>
<td>Tanner</td>
<td>Self-rated</td>
<td>No</td>
<td>A</td>
<td>$E_2$ has a linear relationship to depression</td>
</tr>
<tr>
<td>Martin et al.</td>
<td>1999</td>
<td>94</td>
<td>MTFS</td>
<td>Nil</td>
<td>Nil</td>
<td>Not stated</td>
<td>C</td>
<td>Mean $E_2$ significantly higher in those who used alcohol recently ($F$-statistic 69.81, $p&lt;0.001$). No significant difference in $E_2$ for other drug use.</td>
</tr>
<tr>
<td>Graber et al.</td>
<td>2006</td>
<td>100</td>
<td>YSR</td>
<td>Tanner</td>
<td>Physician/ Nurse</td>
<td>Not stated</td>
<td>A, D</td>
<td>ANOVA showed alcohol use was highest mid-cycle (when $E_2$ highest)</td>
</tr>
<tr>
<td>Rapkin et al.</td>
<td>2006</td>
<td>106</td>
<td>MASC, CDI, CSI</td>
<td>Tanner</td>
<td>Self-rated</td>
<td>No</td>
<td>A</td>
<td>Higher $E_2$ correlated with depressive affect ($r=0.27$; $p&lt;0.01$). Negative life events mediated effect of $E_2$ on aggression</td>
</tr>
<tr>
<td>Davison et al.</td>
<td>2007</td>
<td>178</td>
<td>PACES, CPA, CDI, SPP, BESAA</td>
<td>Tanner, PDS</td>
<td>Nurse (Tanner), Mother (PDS)</td>
<td>Not stated</td>
<td>B, D</td>
<td>Anxiety was inversely correlated with trichotomized $E_2$ ($r=-0.202$; $p=0.038$). No relationship observed for depression or somatization.</td>
</tr>
<tr>
<td>Vermeersch et al.</td>
<td>2008</td>
<td>298</td>
<td>Self-derived questionnaire</td>
<td>Tanner</td>
<td>Physician</td>
<td>Yes</td>
<td>C, D</td>
<td>Negative correlation between $E_2$ and body esteem at 11 years ($r=-0.19; p&lt;0.05$)</td>
</tr>
<tr>
<td>Klump et al.</td>
<td>2010</td>
<td>258</td>
<td>MEBS</td>
<td>PDS</td>
<td>Self-rated</td>
<td>Not stated</td>
<td>B</td>
<td>Total and free $E_2$ positively related with both aggressive and non-aggressive risk taking ($r=0.21; p &lt; 0.001$ for both), controlling for and independent of age and stage. Effects were most evident mid-cycle for aggressive risk taking (but not non-aggressive). Differential association was an important mediator. Higher $E_2$ plays a role in the onset of disordered eating attitudes and behaviours in a twin study.</td>
</tr>
</tbody>
</table>

†Longitudinal analysis; ‡ Differential association refers to the theory that risk-taking behaviour is learnt through interactions with risk taking peers (Vermeersch et al.)

A = mood and affect; B = self-image and related behaviours; C = risk taking, sensation seeking and substance use; D = aggression, behaviour/conduct disorder and delinquency
3. Oestradiol and Female Adolescent Mood and Behaviour

3.5.3. **Study Data**

The studies characterized in Table 3.1 are further detailed in Table 3.2. To briefly summarize the most important findings:

3.5.3.1. **Mood and affect (Category A)**

Most of the studies considered mood and affect. The studies can be summarized as follows: female sample size varied from 52 to 339, with a mean of 122 and median 100. The age ranges of these subjects are listed in Table 3.1. In the nine publications considering this domain, 13 different measures were used for participant assessment. All of these studies were based on community samples from the United States of America. Mood was investigated in different ways, such as DSM-IV depression (27), emotional tone (variability in mood) (34, 35, 38, 39), depressive affect (10, 40), and anxiety (41). While these studies do not all examine the same aspects of mood or affect, all do consider what can be identified as pathological or maladaptive states.

Susman’s longitudinal study in 52 girls over one year considered the roles played by several hormones (luteinizing hormone, follicle stimulating hormone, testosterone, oestradiol dehydroepiandrosterone and its sulfate, androstenedione and cortisol) in a variety of negative affective states: emotional tone, internalizing behaviour problems, symptoms of depression and anxiety. Longitudinal data were given for emotional tone and internalizing behaviour problems, though only cross-sectional data were provided for depression and anxiety symptoms. For girls, no significant cross-sectional relationships were observed between oestradiol and any of the study outcomes, though longitudinal analyses showed a significant positive relationship between oestradiol and increased emotional tone (35). Slap *et al.* found no
longitudinal association between oestradiol and mood (34). This study followed girls for one year, focusing on changes in self-image throughout puberty. Emotional tone was one scale of interest for their questionnaire. Regression analyses did not find a relationship between changes in oestradiol and changes in emotional tone over one year. It should be noted that there was a significant decline in emotional tone scoring (indicating a more variable mood) though changes in oestradiol were not significant between baseline and follow-up. In contrast, Angold’s larger (n=339) three-year longitudinal study found an approximately linear relationship between oestradiol and depression, which was diagnosed by interview using DSM-IV criteria (27). This study found that the odds ratio of oestradiol changes being related to depression was significant (odds ratio 2.5, 95% confidence interval 1.5 to 4.3). Additionally, Angold’s study found that oestradiol was more strongly related to depression than Tanner stage, supporting, not unexpectedly, the primacy of hormone change over morphological change in providing an etiological basis for depression in their cohort. When oestradiol levels were divided into quintiles, the percentage of girls with DSM-IV depression was approximately linear as hormone levels increased.

Two cross-sectional studies observed that oestradiol had a non-linear relationship with depressive affect. When oestradiol concentration was reported as quartiles, depression was lowest in the pre- and post-pubertal quarters (bottom 25% and uppermost 25% respectively), and highest in the transition between these two (10, 40). For Brooks-Gunn’s data, there were significant interactions between negative life events and hormonal levels in depression rating (40). In Rapkin’s study, anxiety was inversely correlated with oestradiol when hormone levels were tertiled. Without
categorization, oestradiol was not significantly correlated with anxiety, however the authors note a trend to this correlation (p=0.051) between the two variables.

Of the seven cross-sectional analyses, emotional tone was considered in three (35, 38, 39). Susman’s results from 1991 are reported above, as both longitudinal and cross-sectional analyses were provided. An earlier paper in the same data set did not find any significant cross-sectional relationship with emotional tone and oestradiol (38). Both of these studies used different mood assessment tools. In contrast to Susman’s studies, more recent work in a larger study population (n=100) observed a correlation between oestradiol and emotional tone (39).

3.5.3.2. *Self-image and related behaviours (Category B)*

Slap *et al.* (34) found no longitudinal association between changes in oestradiol and self-image. In contrast, cross-sectional studies observed correlations between oestradiol and body esteem (42) and with the appearance of genetically influenced disordered eating attitudes and behaviours (43). In Davison *et al* a negative correlation between oestradiol and body esteem was observed at 11 years in girls, though not at any other time point (42). Nottelmann *et al.* found no association between oestradiol and body image (36).

3.5.3.3. *Risk taking, sensation seeking and substance use (Category C)*

A significant correlation was found between the level of oestradiol and alcohol use in Martin *et al.* (28). No relationship between oestradiol and other substance use was identified. In a large Belgian population-based study, Vermeersch *et al.* found positive relationships between total and free oestradiol in girls and risk taking (44).
Aggressive risk taking was most evident at mid-cycle (defined by the investigators as an oestradiol of greater than 60 pg/mL (220 pmol/L) (28)), with non-aggressive risk taking showing no such cycle effect.

3.5.3.4. Aggression, behaviour/conduct disorder and delinquency (Category D)

There were limited data in this Category. Brooks-Gunn et al. and Nottelmann et al. both observed no correlation between oestradiol and aggression in their studies (36, 40). Graber et al. found that aggression only correlated with oestradiol in females when negative life events were included as a mediating effect (39).

3.5.4. Quality Assessment

Table 3.3 provides a detailed description of quality assessment. No study included a power analysis, nor were there adequate descriptions of sampling methods. Four studies specified exclusion criteria. Of the fourteen studies, only Martin et al. controlled for menstrual cycle (28). Though Angold’s study did not use exact timing of sample collection, approximately three-quarters of samples were collected at consistent times which the authors report as minimizing diurnal variation effect (27).

With respect to confounding factors, four studies accounted for or controlled for factors such as age and socio-economic status.
<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Was the study setting and population adequately described? (Age, gender, menarcheal status)</th>
<th>Did papers consider a power analysis?</th>
<th>Exclusion criteria described?</th>
<th>Adequate control or adjustment for menstrual cycle?</th>
<th>Appropriate methodology described for variable measures? (Including oestradiol assay limits)</th>
<th>Relevant confounders (including age, SES) accounted for or controlled?</th>
<th>All primary outcomes reported in results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susman et al., 1991</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No (assay limits not stated)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Slap et al., 1994</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Angold et al., 1999</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Susman et al., 1985</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No (assay limits not stated)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Nottelmann et al., 1987</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No (assay limits not stated)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Susman et al., 1987</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No (assay limits not stated)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Brooks-Gunn et al., 1989</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Warren et al., 1989</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Martin et al., 1999</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Graber et al., 2006</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Rapkin et al., 2006</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Davison et al., 2007</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Vermeersch et al., 2008</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Klump et al., 2010</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
3.6. Discussion

This systematic review is the first to examine the effect of oestradiol levels on mood and behaviour in adolescent girls. It is timely because long-standing assumptions and the evidence regarding sex hormone effects on adolescent mood and behaviour have never been comprehensively explored. With the growing awareness of mood and behavioural changes in adolescents, confirming or challenging these assumptions is important to develop or refine new paradigms in the diagnosis and management. Given the earlier onset of female puberty (3, 16) it is important to be aware of when any oestradiol effects might be expected to occur, especially as there are established links between early onset and later female mood and behavioural problems (14, 36, 45, 46).

The review demonstrated reasonably consistent findings on mood and affect, with depression and increased mood variability being positively correlated with oestradiol for at least some stages of puberty. Associations were most consistently observed between oestradiol concentrations and depressive states during the transition from pre-pubertal to adult oestradiol levels (10, 27, 39, 40), suggesting that tempo or rapidity of hormone change might play a part. A positive correlation was generally observed between oestradiol and affective variability (35, 38, 39). While mood variability more likely reflects “puberty blues” than does DSM-IV diagnosed depression (27), both benign and pathological changes in mood and affect are important causes of parental and adolescent concern during puberty. We have shown both of these have some association with oestradiol levels. There was no clear consistency of oestradiol effect for behaviours such as aggression, behaviour/conduct...
disorder and delinquency (see 3.3.4). These behaviours are more often (falsely) associated with testosterone, as another review by our group has shown (47).

The cross-sectional nature of all but three studies limits the conclusions of this review. Thus, reported outcomes can only show associations between oestradiol concentrations and the affect or behaviour of interest. Causality cannot be inferred, as an association could be a persistent relationship, short-term change or mere coincidence (48); likewise the direction of the effect can only be postulated. Additionally, four of the studies were from the same cohort (35-38), and three of these described similar outcomes (all assessed with different scales), which risks over-representation of their data. Only four of the twelve cross-sectional studies controlled for potential confounders such as socio-economic status.

Brooks-Gunn et al. estimated only 1% of variance in negative emotional expression was due to oestradiol (40) in their study. Environmental and social factors or determinants in an individual may mediate or amplify susceptibility to behavioural or affective changes as a result of pubertal hormone change (48). Angold et al. postulated that hormonal influences on mood and behaviour are less causative than sensitizing – that is, pubertal increases in sex hormones surpass a threshold that would render one more likely to alter affect or behaviour (27). That one-fifth of samples had hormone levels that fell below assay limits of detection and timing of sample collection was not as rigorous as other studies should be viewed as specific study limitations.

The studies appraised in this systematic review are both observational and primarily cross-sectional. Longitudinal cohort studies would provide the best evidence for an
3. Oestradiol and Female Adolescent Mood and Behaviour

effect, if any, of oestradiol on adolescent mood and behaviour and this was the finding in the large longitudinal study from Angold et al (27).

None of the studies in the systematic review addressed mechanisms for a putative relationship between oestradiol and behaviour and/or mood. Such mechanisms are likely primarily central. Oestrogens, especially oestradiol, have been shown to modulate genetic expression in neural monoamine systems (such as the dopaminergic, serotonergic and noradrenergic systems) (17, 18). These systems are commonly targeted by psychotropic drugs such as monoamine oxidase inhibitors and selective serotonin re-uptake inhibitors, which support at least a role for oestradiol in the onset or progression of mood or behavioural disorders. Pubertal hormone changes are often thought to modulate mood and behaviour by activating previously organized neuroarchitecture (see Introduction). It would be anticipated that these changes persist into adulthood, but given the limited follow-up in the studies reviewed, we cannot postulate as to whether oestradiol’s effects in this regard are persistent. Regardless, it is important to consider how sex hormone changes might affect adolescent mood and behaviour, especially given the salient increase in psychopathology during and after adolescence and the importance of earlier intervention.

This systematic review has several strengths. Our methodology included comprehensive searches of many databases of potential relevance to this review, as well as detailed data extraction and quality analysis. Of the fourteen included studies, seven included over 100 girls.
3. Oestradiol and Female Adolescent Mood and Behaviour

There are a number of limitations. The studies included in the systematic review used a range of measurement tools to assess behaviour and affect, and meta-analysis or quantitative analysis of the collected study data was not possible. There may be a bias in that only published manuscripts were included in our final analysis, but as a counter to this argument, published studies had mainly negative findings.

Information about the quality of the environment and early-life stresses experienced by the participants were not offered in the majority of the studies. These factors might well contribute to any aberrant moods or behaviours, and are thus important confounders.

The irregular nature of menstrual cycles for some time after menarche makes control for menstrual cycle difficult in any study, and was only done in one of the retained studies. Failing to control for menstrual cycle may be a limitation of the review. Alternatively selecting a specific phase of the cycle (follicular, mid or luteal) may also skew the findings – if samples were not collected in the mid-follicular phase, progesterone may be a confounding factor. The most important methodological limitation of every study is the method of oestradiol measurement. The American Endocrine Society’s consensus statement on oestradiol measurement (26) concludes that even current assays are inadequate for pre-pubertal oestradiol measurement, indicating that the methodologies of the reviewed studies may lack sufficient sensitivity to detect oestradiol, especially at low levels or indeed to measure with enough specificity mid puberty levels. In older, less sensitive assays especially oestradiol-mediated effects will be underestimated if too many oestradiol levels are below limits of quantification.
In conclusion, this systematic review has found that there are insufficient longitudinal data of high methodological quality to confirm that the rise in oestradiol during puberty plays a causative role in adolescent mood and behavioural changes, though the current evidence clearly suggests such a relationship exists. Future studies would require sufficient duration and frequency of sampling of biological markers, adequate statistical power and the use of mass spectrometry techniques to clarify the role of oestradiol. However, given the intra- and inter- individual variability of oestradiol levels from the beginning of puberty onwards, and the variation in timing and tempo of puberty it is possible that even with careful longitudinal studies, the assignation of a causal role may prove elusive.

We believe that it is important for health professionals to take care when attributing adolescent psychopathology to puberty hormones, as the current data supporting these assertions are limited. Both timing and tempo of puberty may contribute to vulnerability in certain adolescents, particularly if other adverse psychosocial circumstances exist. Such adolescents may require supportive intervention during their adolescence in order to ensure that health trajectories are optimized. To this end, further understanding on how the dramatic rise in oestradiol during puberty affects adolescent girls’ mood and behaviour has important public health repercussions.

3.7. References

3. Oestradiol and Female Adolescent Mood and Behaviour

4. TESTOSTERONE AND ITS EFFECTS ON HUMAN MALE ADOLESCENT MOOD AND BEHAVIOUR

A SYSTEMATIC REVIEW

Sally-Anne Duke¹, Ben W.R. Balzer²,³, Katharine S. Steinbeck²,³

Affiliations: ¹Department of Endocrinology, Royal North Shore Hospital, St. Leonards, NSW, Australia ²Academic Department of Adolescent Medicine, The Children’s Hospital at Westmead, Westmead, NSW, Australia; ³Discipline of Paediatrics and Child Health, Sydney Medical School, The University of Sydney, NSW, Australia

4. Testosterone and Male Adolescent Mood and Behaviour

4.1. Abstract

Purpose: During human puberty, there is an approximate 30-fold increase in testosterone production in boys. This increase is often linked to changes in mood and behaviour in adolescence such as aggression, an increase in risk taking, and depression. The aim of this systematic review was to determine what evidence exists on the effects of endogenous testosterone on behaviour and mood in males during adolescence.

Methods: The following databases were searched: MEDLINE, Pre-MEDLINE, Education Resources Information Centre, PsycINFO, EMBASE, Scopus, and Web of Science. Only human studies were included. The study is community based, and the participants were healthy male adolescents within the age range of 9-18 years. Studies were required to have a validated mood and/or behaviour assessment contemporaneous with a timed testosterone measurement.

Results: A total of 27 studies met the inclusion criteria of which only one was a longitudinal study. The remaining 26 studies were cross sectional in their analysis. As a variety of measurement tools were used, no meta-analysis was possible. Most studies focused on aggression. The one longitudinal study looking at testosterone and aggression showed little relationship with concurrent changes in aggression. Most of the cross-sectional studies of adolescent males observed relationships between aggression and testosterone levels. With respect to other behaviours and moods and/or affect, no consistent relationships with testosterone were observed in cross-sectional studies.
Conclusions: This systematic review concludes that there are insufficient longitudinal data of high methodological quality to currently confirm that changing testosterone levels during puberty are significantly associated with mood and behaviour in adolescent males. To discount these findings is to risk apportioning blame inappropriately and missing other important diagnoses in adolescent males.
4.2. Introduction

During human puberty, there is a 20- to 30-fold increase in endogenous testosterone production in males (1). Many assumptions are made about the nonphysical effects of this dramatic rise in testosterone; in particular that testosterone is responsible for changes in mood and behaviour. These changes are often considered to be undesirable or concerning and include aggression, an increase in risk taking, and depression (2, 3).

Aggression is the most commonly cited behaviour associated with testosterone. Aggression is defined as a range of behaviours that can result in either physical or psychological harm to oneself, others, or objects in the environment. The common measurement for aggression in children is within the Youth Self-Report, and its parent-report equivalent, the Child Behavioural Check List (CBCL). The Youth Self-Report and CBCL assess several types of aggressive behaviours, including bragging, arguing, screaming, showing off, attention seeking, teasing, being demanding, threatening behaviour, and displaying a temper.

The implied association of testosterone with these behaviours is based on related but indirect evidenced normal developmental events, conditions of abnormal androgen excess and exogenous testosterone use. The second trimester surge in testosterone production in utero in males (4) is presumed to play a role in neural modifications that allow expression of behaviours more consistently seen in boys, such as physically rough-and- tumble play. This play pattern is also observed in females with congenital adrenal hyperplasia who are exposed to high levels of their own adrenal androgens, including testosterone, in utero (4). In males who fail to progress through
puberty because of abnormalities in the hypothalamic-pituitary-gonadal axis, the initiation of puberty by exogenous androgen therapy has demonstrated increases in physically aggressive behaviour (5). The behavioural changes induced by the abuse of exogenous androgens in testosterone-replete individuals (6) have been used to support assertions that testosterone has an important role in the etiology of mood and conduct disorders. However, there is no certainty that these latter data can be extrapolated to include the normal physiological rise in testosterone observed in puberty. For example, at normal physiological levels of testosterone, androgen receptors are saturated and therefore some of the effects of androgenic anabolic steroids are likely to be through different mechanisms such as high oestrogen levels or effects on the GH-IGF₁ axis (7).

There is indeed a concern that behaviour and mood change may be falsely attributed to puberty hormones, which are unalterable, and the true diagnosis ignored or not excluded. Although adolescence is often viewed as a period of storm and stress, most adolescents negotiate the period without short- or long-term detrimental behavioural outcomes, and the changes in mood and behaviours often ameliorate with time, despite the testosterone levels remaining largely unchanged relative to the changes during puberty.

There have been a large number of animal studies (including rat, mice, hamster, and monkey) that have informed the role of testosterone in adolescent behaviour. This includes the activational and organizational effects of testosterone on neural circuitry and neurohormones (8), the effects of androgenic anabolic steroids (8), and the influence of prenatal testosterone on behaviour (9, 10). Although these models can inform human studies, we know that variation between strains and species is
significant (11) and may or may not translate to the complexities of human behaviour.

The aim of this systematic review, therefore, was to determine what evidence exists on the effects of endogenous testosterone on behaviour and mood in human males during adolescence.

4.3. Methods

4.3.1. Search strategy

A systematic search was conducted to determine the effect of endogenous testosterone on behaviour, affect, and mood in healthy male adolescents. Only human studies were sought. There is a large body of animal, primarily rodent, data using non-physiological testosterone manipulation. These are well beyond the scope of this review, even in discussion. The following databases were searched from the data on which records began up until the first week of March 2013: MEDLINE, Pre-MEDLINE, Education Resources Information Centre, PsycINFO, EMBASE, Scopus, and Web of Science. Studies published in any language were included. The MEDLINE search strategy can be viewed in Appendix E. The other databases were searched with a similar strategy adjusted for the particular database.
4.3.2. Inclusion criteria

4.3.2.1. Study type.

The study can be a published prospective or cross-sectional study in any language, despite the inability to draw conclusions on causality from onetime point data (addressed in the Discussion section).

4.3.2.2. Participants.

Study participants are male adolescents. Most participants are within the age range of 9-18 years as this age range captures the extremes of early (non-precocious) puberty and of constitutional delay. Participants are not to have a specified disorder or disability, or if a specified disorder or disability is present, there is a control group with results reported separately. The study is community based and not from an institutionalized or incarcerated population (unless community-based controls are reported for comparison) as these may confound behavioural outcomes. Participants are not to be taking exogenous testosterone, which implies nonspontaneous puberty.

4.3.2.3. Hormone studies.

All studies must have a measure of testosterone from serum or saliva, by a validated immunoassay or mass spectrometry (MS). The Endocrine Society position statement on testosterone measurement supports the use of MS as a more accurate way of measuring sex steroids especially at low levels (12). However, MS was not commonly available when most of these studies were performed, and therefore, studies that used a validated serum and/or plasma or saliva immunoassay are
4. Testosterone and Male Adolescent Mood and Behaviour

included. It is not possible to comment on how comparable or otherwise the various assays are.

The time of the testosterone collection must be standardized to account for diurnal variation in testosterone production. Testosterone levels fall approximately 40% from peak to nadir (13). In this review, studies must include a documented time of collection or adjustments for timing.

4.3.2.4. Outcome measure.

The outcomes are those behaviours and affect disorders that are known to have a clear increased incidence and prevalence in the second decade of life and the behaviours commonly inferred to be the result of puberty hormones: aggressive/disruptive/conduct disordered behaviour, substance abuse, social competency and social interactions, depression and anxiety, and self-image concerns. Contextual moderators such as family or peer relationships and social context were sought because of their known influence on behaviour and mood.

Because most published studies address a single behaviour or mood, the extracted manuscripts were grouped together based for clarity on common themes: (1) externalizing behaviours, including aggression; (2) alcohol and other drug use; (3) self-image and social behaviours; and (4) mood and affect. The behaviours or affect disturbance needed to have been measured using recognized and validated tools. Those studies that focused primarily on sexual behaviours were excluded. This was because testosterone’s role in the development of male secondary sexual development and sexual behaviour is universally acknowledged, and the focus of this review is on behaviours that may indicate psychopathology or at the least troubling
behaviours. The assessment of the specified outcome needed to be contemporaneous with testosterone measure. A physical examination for puberty status was not required but was extracted if included.

4.3.3. Data collection and analysis

4.3.3.1. Selection of studies.

After the removal of obviously irrelevant studies or duplicates, a single reviewer (S.A.D.) scanned the title, abstract, and keywords of every record and retrieved the full text for further assessment when the reference suggested that the study might fit the inclusion criteria. The full text of these studies was analyzed by two reviewers (S.A.D. and K.S.S.) and identified as included or excluded based on the previously mentioned criteria.

4.3.3.2. Data extraction.

Information was extracted from each study into a specific template developed by the authors. The following information was extracted: study design, description of participants which included sample size, age range, the outcome measured and the tool used, testosterone level(s) including sample type, time collected and laboratory assay used, physical pubertal status (stage, scale, and by whom), results obtained with analysis type, and whether co-variables or confounders were assessed.

4.3.3.3. Statistical analyses.

Given the heterogeneity of both the measures and the outcomes, no further analyses could be performed and the results are presented as descriptive data.
4. Testosterone and Male Adolescent Mood and Behaviour

4.4. Results

4.4.1. Search results

From the initial search, 8,731 citations were identified from the following databases: MEDLINE (2,151), Pre-MEDLINE (31), Education Resources Information Centre (45), PsycINFO (756), EMBASE (1,536), Scopus (3,352), and Web of Science (860). One reviewer (S.A.D.) scanned these references, and 7,981 were removed as being either not on topic, reviews, or duplicates. The abstracts (and full reference if abstract unavailable) were studied in detail of the remaining 750, with 117 citations identified as potentially relevant. Two reviewers (S.A.D. and K.S.S.) studied the full references of the 117 citations and 27 studies completely fulfilled the inclusion criteria. There was consensus between the two reviewers. All 27 final publications were in English (Figure 4.1).

The main reasons for exclusion were age of participants, if the subjects were selected for a particular disease or disability, the time of testosterone collection not documented or standardized, the behaviour measured not contemporaneous with testosterone collection, testosterone was the dependent variable (the effect of behaviour on testosterone levels), or sexual behaviours only were studied (see Appendix E for more details).
Records identified through database searching (n=8731)
- MEDLINE (2151)
- Pre-MEDLINE (31)
- ERIC (45)
- PsycINFO (756)
- EMBASE (1536)
- Scopus (3352)
- Web of Science (860)

Exclude obviously irrelevant, duplicates, reviews (n=7981)

Abstracts reviewed in detail and full references when necessary (n=750)

Did not fulfil inclusion criteria (n=633)

Full references of potentially relevant articles studied in detail by 2 reviewers (n=117)

Finally excluded studies (n=90)

Studies completely fulfilling inclusion criteria (n=27)

**Figure 4.1.** PRISMA flow diagram for study selection (14). ERIC = Education Resources Information Centre.
Table 4.1. Longitudinal Studies.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>No.</th>
<th>Sex</th>
<th>Setting</th>
<th>Age</th>
<th>Behaviour*</th>
<th>Testosterone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halpern, CT, et al. (15)</td>
<td>1993</td>
<td>82</td>
<td>M</td>
<td>US, school students</td>
<td>12-13</td>
<td>A</td>
<td>Plasma 3-9pm, Diagnostic Production Corporation, I 125 count-a-coat kit</td>
</tr>
</tbody>
</table>

*Behaviour* indicates the specific type of behavior being measured.
Table 4.2. Cross-sectional studies or longitudinal studies with cross-sectional analysis.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Sex</th>
<th>No. of Males</th>
<th>Setting</th>
<th>Age range</th>
<th>Behaviour</th>
<th>Testosterone measurement</th>
<th>Assessment of pubertal stage*</th>
</tr>
</thead>
<tbody>
<tr>
<td>De Water E et al. (16)</td>
<td>2013</td>
<td>M/F</td>
<td>86</td>
<td>Netherlands, school students</td>
<td>12-17</td>
<td>B</td>
<td>Saliva on waking, isotope dilution-LC-T-MS</td>
<td>PPDS</td>
</tr>
<tr>
<td>Booth A, et al. (17)</td>
<td>2003</td>
<td>M/F</td>
<td>315</td>
<td>US, school students</td>
<td>6 – 18</td>
<td>A/D</td>
<td>Saliva 7am RIA</td>
<td>PPDS</td>
</tr>
<tr>
<td>Dawes MA, et al. (18)</td>
<td>1999</td>
<td>M</td>
<td>297</td>
<td>US, community, targeted</td>
<td>10 – 12</td>
<td>A</td>
<td>Plasma 7am RIA (Amersham’s)</td>
<td>T (p)</td>
</tr>
<tr>
<td>Eriksson CI et al. (19)</td>
<td>2005</td>
<td>M</td>
<td>464</td>
<td>Finland, Community twin study</td>
<td>14</td>
<td>B</td>
<td>Saliva before noon RIA (Orion diagnostics, Finland)</td>
<td>PPDS</td>
</tr>
<tr>
<td>Fang CY, et al. (20)</td>
<td>2009</td>
<td>M/F</td>
<td>164</td>
<td>US, 6 clinical centers, targeted</td>
<td>11-14</td>
<td>A</td>
<td>Serum RIA (PerkinElmer, Waltham, MA)</td>
<td>T (p)</td>
</tr>
<tr>
<td>Foshee VA et al. (21)</td>
<td>2007</td>
<td>M/F</td>
<td>209</td>
<td>US, school students</td>
<td>6th and 8th grade</td>
<td>B</td>
<td>Saliva 8-10am RIA</td>
<td>PPDS</td>
</tr>
<tr>
<td>Gerra G et al. (22)</td>
<td>1998</td>
<td>M</td>
<td>30</td>
<td>Italy, school students</td>
<td>12</td>
<td>A</td>
<td>Serum 430pm RIA IFO-CLONE (Saluzia, Italy)</td>
<td>T (p)</td>
</tr>
<tr>
<td>Gerra G et al. (23)</td>
<td>2000</td>
<td>M</td>
<td>40</td>
<td>Italy, school students</td>
<td>12</td>
<td>D</td>
<td>Serum 430pm, IFO-CLONE (Saluzia, Italy)</td>
<td>T (p)</td>
</tr>
<tr>
<td>Granger DA et al. (24)</td>
<td>2003</td>
<td>M/F</td>
<td>106</td>
<td>US, community, targeted</td>
<td>11 -17</td>
<td>A/D</td>
<td>Saliva 7.30am¹</td>
<td>PPDS</td>
</tr>
<tr>
<td>Kirillova GP et al. (25)</td>
<td>2008</td>
<td>M</td>
<td>478</td>
<td>US, community, targeted</td>
<td>9 - 20</td>
<td>A/B</td>
<td>Serum 830 (DSL Webster, TX)</td>
<td>T (p)</td>
</tr>
<tr>
<td>Nottelmann ED, et al. (1)</td>
<td>1987</td>
<td>M/F</td>
<td>56</td>
<td>US, community</td>
<td>9-14</td>
<td>A/C</td>
<td>Plasma 8-10am RIA not stated</td>
<td>T (p)</td>
</tr>
<tr>
<td>Olweus D, et al. (26)</td>
<td>1980</td>
<td>M</td>
<td>58</td>
<td>Sweden, school students</td>
<td>15-17</td>
<td>A</td>
<td>Plasma 8-10 RIA not stated</td>
<td>T (p)</td>
</tr>
<tr>
<td>Olweus D, et al. (27)</td>
<td>1988</td>
<td>M</td>
<td>58</td>
<td>Sweden, school students</td>
<td>15-17</td>
<td>A</td>
<td>Plasma 8-10 RIA not stated</td>
<td>T (p)</td>
</tr>
<tr>
<td>Sánchez-Martín JR, et al. (28)</td>
<td>2011</td>
<td>M/F</td>
<td>44</td>
<td>Spain, school students</td>
<td>9</td>
<td>A</td>
<td>Saliva 9am ELISA (Salimetrics, State College, USA)</td>
<td>None</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Sex</td>
<td>No. of Males</td>
<td>Setting</td>
<td>Age range</td>
<td>Behaviour</td>
<td>Testosterone measurement</td>
<td>Assessment of pubertal stage</td>
</tr>
<tr>
<td>----------------------</td>
<td>------</td>
<td>------</td>
<td>--------------</td>
<td>--------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>--------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Susman EJ, et al.</td>
<td>1985</td>
<td>M/F</td>
<td>56</td>
<td>US, community</td>
<td>9-14</td>
<td>D</td>
<td>Plasma 8-10am RIA (technique Nieschlag &amp; Loriaux)</td>
<td>T (p)</td>
</tr>
<tr>
<td>Susman EJ, et al.</td>
<td>1987</td>
<td>M/F</td>
<td>56</td>
<td>US, community</td>
<td>9-14</td>
<td>A/D</td>
<td>Plasma 8-10am RIA</td>
<td>T (p)</td>
</tr>
<tr>
<td>Susman EJ, et al.</td>
<td>1991</td>
<td>M/F</td>
<td>56</td>
<td>US, community</td>
<td>9-14</td>
<td>D</td>
<td>Plasma 8-10am RIA</td>
<td>T (p)</td>
</tr>
<tr>
<td>Tarter RE, et al.</td>
<td>2007</td>
<td>M</td>
<td>179</td>
<td>US, community, targeted</td>
<td>16</td>
<td>A</td>
<td>Plasma 830 RIA (DSL, Webster, Texas)</td>
<td>T (p)</td>
</tr>
<tr>
<td>Udry JR (35)</td>
<td>1990</td>
<td>M</td>
<td>101</td>
<td>US, school students</td>
<td>13-16</td>
<td>A</td>
<td>Serum 3-7pm</td>
<td>T (s)</td>
</tr>
<tr>
<td>Vermeersch H, et al.</td>
<td>2008</td>
<td>M</td>
<td>301</td>
<td>Belgium, school students</td>
<td>13-18</td>
<td>A</td>
<td>Serum 9-12am RIA (Orion Diagnostica, Finland)</td>
<td>T (p)</td>
</tr>
<tr>
<td>Vermeersch, H, et al.</td>
<td>2009</td>
<td>M/F</td>
<td>301</td>
<td>Belgium, school students</td>
<td>13-18</td>
<td>C</td>
<td>Serum 9-12am RIA (Orion Diagnostica, Finland)</td>
<td>T (p)</td>
</tr>
<tr>
<td>Vermeersch, H, et al.</td>
<td>2010</td>
<td>M/F</td>
<td>301</td>
<td>Belgium, school students</td>
<td>13-18</td>
<td>C</td>
<td>Serum 9-12am RIA (Orion Diagnostica, Finland)</td>
<td>T (p)</td>
</tr>
<tr>
<td>Victoroff, J, et al.</td>
<td>2011</td>
<td>M</td>
<td>52</td>
<td>Gaza, refugee Muslims</td>
<td>14</td>
<td>A</td>
<td>Saliva 9am RIA (Salimetrics, PA)</td>
<td>None</td>
</tr>
<tr>
<td>Yu, YZ, et al.</td>
<td>2009</td>
<td>M/F</td>
<td>20</td>
<td>China, school students</td>
<td>11-16</td>
<td>A</td>
<td>Saliva 1.30-2.30pm RIA (CHEMCLIN of Beijing)</td>
<td>T (p)</td>
</tr>
</tbody>
</table>

* A=aggressive/disruptive behaviour B= Substance abuse C= Self-image/social interaction and competency D= Depression/Anxiety

# T =Tanner staging (practitioner (p) or self (s)); PPDS – Petersen Pubertal Development Scale (Petersen, Crockett, Richards, & Boxer, 1988) – this is a self report scale
### Table 4.3. Aggressive/Disruptive Behaviour Studies.

<table>
<thead>
<tr>
<th>Author</th>
<th>Behaviour scale</th>
<th>Statistics</th>
<th>Co-variables</th>
<th>Outcomes in males</th>
<th>Funding stated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halpern 1993</td>
<td>Self-ratings of aggression, OMFAI</td>
<td>ANOVA</td>
<td>Pubertal development (PD)</td>
<td>Longitudinal study: 0,6,12,18,24,36. Few and inconsistent relationship between testosterone and aggression</td>
<td>Y</td>
</tr>
<tr>
<td>Booth 2003</td>
<td>RBS, measures of parent child relationship</td>
<td>Regression</td>
<td>Age, parents T</td>
<td>T no relationship to risk behaviour, but modified by parent child relationship</td>
<td>Y</td>
</tr>
<tr>
<td>Dawes 1999</td>
<td>DFI, CAPI, PDS, BRS</td>
<td>Regression</td>
<td>PD, peer, family</td>
<td>T not significant in regression model of behaviour self regulation</td>
<td>Y</td>
</tr>
<tr>
<td>Fang 2009</td>
<td>CBCL, FES</td>
<td>Correlation, Regression</td>
<td>PD, family cohesions, PD, BMI</td>
<td>T positively associated with delinquent behaviour (moderated by cohesion), no significant relationship with aggression</td>
<td>Y</td>
</tr>
<tr>
<td>Gerra 1998</td>
<td>CDI, TIA, CPQ</td>
<td>ANOVA</td>
<td>Nil</td>
<td>T significantly higher in group with high aggression, than low or normal aggression</td>
<td>Y</td>
</tr>
<tr>
<td>Granger 2003</td>
<td>CBCL, YSR, DISC</td>
<td>Regression</td>
<td>Age, PD</td>
<td>T no relationship with disruptive behaviour</td>
<td>Y</td>
</tr>
<tr>
<td>Kirillova 2008</td>
<td>K-SADS-E, PDS</td>
<td>Regression, Cox Hazard Path analysis</td>
<td>No of affected parents, PD</td>
<td>T level at times 1 and 2 significantly related to rate of conduct disorder development, with a trend only at time 3</td>
<td>Y</td>
</tr>
<tr>
<td>Nottelmann 1987</td>
<td>OSIQ, CBCL, OMQI, personality</td>
<td>Regression</td>
<td>Age, PD</td>
<td>T alone no relationship with aggression or delinquent behaviour Elements of aggression (physical, verbal) positively correlated with T; no association with anti social behaviour</td>
<td>N</td>
</tr>
<tr>
<td>Olweus 1980</td>
<td>OMFAI, OQI, personality</td>
<td>Correlation</td>
<td>PD, parent</td>
<td>as above when other variables controlled for</td>
<td>Y</td>
</tr>
<tr>
<td>Olweus 1988</td>
<td>OMFAI, OQI, personality</td>
<td>Regression</td>
<td>PD, parent</td>
<td>as above when other variables controlled for</td>
<td>Y</td>
</tr>
<tr>
<td>Sánchez-Martin 2011</td>
<td>DIAS (peer), anger and impulsivity</td>
<td>Correlation; Regression</td>
<td>Impulsivity, anger</td>
<td>Pearson’s correlation – no significant relationship between T and any forms of aggression or anger; general linear model - after sex, T found to be the best predictor (positive sense) of all three types of aggression</td>
<td>Y</td>
</tr>
<tr>
<td>Susman 1987</td>
<td>CBCL (DQB, A), OSIQ</td>
<td>Regression</td>
<td>PD, age</td>
<td>No relationship between T and any behavioural measures</td>
<td>N</td>
</tr>
<tr>
<td>Tarter 2007</td>
<td>ND, SPS, PPBS, PDS</td>
<td>Path analysis</td>
<td></td>
<td>T at 16 significantly predicts social dominance/norm violating behaviour at 16 (p&lt;0.05) but not neurobehavioural disinhibition</td>
<td>Y</td>
</tr>
</tbody>
</table>
### Table 4.3. (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Behaviour scale</th>
<th>Statistics</th>
<th>Co-variables</th>
<th>Outcomes in males</th>
<th>Funding Stated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tarter 2009</td>
<td>Assaultiveness Scale of the Hostility-</td>
<td>Path</td>
<td>T at 10-12</td>
<td>T level at age 12-14 significantly correlated with assaultive behaviour at that age</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Guilt Inventory</td>
<td>analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tremblay 1998</td>
<td>Teacher- SBQ, Peer- PEI, &amp; self-DQB</td>
<td>Regression</td>
<td>BMI, T at age 12</td>
<td>Testosterone levels at 12 were not significantly correlated to physical aggression or antisocial behaviour; T did not significantly add to either model of aggression or social dominance at age 13 after BMI and T at age 12</td>
<td>Y</td>
</tr>
<tr>
<td>Udry 1990</td>
<td>14 problem behaviours</td>
<td>Regression</td>
<td>Age, PD, others e.g. SES, school</td>
<td>T significantly predicts problem behaviour -moderated by good child and grades</td>
<td>Y</td>
</tr>
<tr>
<td>Vermeersch 2008</td>
<td>Dutch self-report</td>
<td>Regression</td>
<td>Age, PD, DA</td>
<td>Significant but modest relationship between T &amp; non aggressive risk taking, no relationship between T &amp; aggressive risk taking, moderated by differential association</td>
<td>Y</td>
</tr>
<tr>
<td>Victoroff 2011</td>
<td>AQ (self)</td>
<td>Correlation</td>
<td>None</td>
<td>No significant correlation between T and aggression</td>
<td>Y</td>
</tr>
<tr>
<td>Yu 2009</td>
<td>CBCL</td>
<td>Student t</td>
<td>Other hormones</td>
<td>Corr: T level significantly higher in aggressive students; RA: -no relationship between T and aggression</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Regression</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Behaviour Scales/Questionnaires:** AQ - Aggression Questionnaire; BISBRS - Behaviour Regulation Scale; CAPI - Child Abuse Potential Inventory; CBCL – Child Behaviour Checklist; CDI - Children’s Depression Index; CES- Center for Epidemiological Studies Depression Scale; CPI-D - California Psychological Inventory-Dominance scale; CPQ – Children Personality Questionnaire; C-SSAGA- A - Finnish translation of the adolescent Semi-Structured Assessment of Genetics of Alcoholism; DFI - Dysfunctional Family Index; DIAS - Direct and Indirect Aggression Scale; DISC - Diagnostic Interview Schedule for Children; FES-Family Environment Scale; K-SADS-E - Kiddi-e-Schedule for Affective Disorders and Schizophrenia for School Age Children; Neurobehavioural disinhibition trait; OMFAQ - Olweus Multi-faceted Aggression Inventory for Boys; OQI – Olweus Q Inventory; OSIQ- Offer Self-Image Questionnaire for Adolescents; PDS- Peer Delinquency Scale; PEI - The Pupil Evaluation Inventory; PPBS- Perception of Problem Behaviours Scale; PPDS - Petersen Development Scale; RBS - Risky Behaviour Scale; SBQ- Social Behaviour Questionnaire; self-DQB- Self-reported Delinquency; SIQYA - Self Image Questionnaire for Adolescents; SPS - Social Potency Scale of the Differential Personality Questionnaire; TIA-Test Anxiety Inventory; YBP – Youth Behaviour Profile; YSR - Youth Self Report

**Other abbreviations:** PA- path analysis; Corr- correlations; RA – regression analysis; PD – pubertal development; PT – pubertal timing
<table>
<thead>
<tr>
<th>Author</th>
<th>Scale/Measure</th>
<th>Statistical Analysis</th>
<th>Co-variables</th>
<th>Outcomes</th>
<th>Funding stated</th>
</tr>
</thead>
<tbody>
<tr>
<td>De Water 2013</td>
<td>Alcohol use last 30 days and Alcohol questionnaire – frequency of drinking and intoxication</td>
<td>Regression</td>
<td>Age</td>
<td>T was not associated with recent alcohol use. No association was found between T levels and frequency of drinking. Frequency of intoxication was associated with T levels (p = 0.02), but not after adjustment for PD (p = 0.125).</td>
<td>Y</td>
</tr>
<tr>
<td>Eriksson 2005</td>
<td>Tobacco and alcohol use, family, peer, neighbor and school context</td>
<td>Regression</td>
<td>PD</td>
<td>No association was found between T levels and frequency of drinking. Frequency of intoxication was associated with T levels (p = 0.02), but not after adjustment for PD (p = 0.125).</td>
<td>Y</td>
</tr>
<tr>
<td>Foshee 2007</td>
<td>K-SADS-E; PDS</td>
<td>Regression, correlation, regression</td>
<td>PD, PT, age, context</td>
<td>Association between T and drug use became positive and stronger as the context more harmful. The relationship between T and substance use disorder were variable with a trend towards association at visit 1, significant association at visit 2 and no association at visit 3.</td>
<td>Y</td>
</tr>
<tr>
<td>Kirillova 2008</td>
<td>K-SADS-E; PDS</td>
<td>Regression (Cox CHR, &amp; Path Analysis)</td>
<td>No of affected parents, PD</td>
<td>The relationship between T and substance use disorder were variable with a trend towards association at visit 1, significant association at visit 2 and no association at visit 3.</td>
<td>Y</td>
</tr>
<tr>
<td>Nottelmann 1987</td>
<td>OSIQ, CBCL (maternal)</td>
<td>Regression</td>
<td>Age, PD</td>
<td>No relationship between T and self image</td>
<td>N</td>
</tr>
<tr>
<td>Vermeersch 2009</td>
<td>Dutch translation of Carver and White's behavioural scale activation/inhibition</td>
<td>Bivariate &amp; Regression</td>
<td>Age, PD</td>
<td>No direct relationship between T and behavioural inhibition or behavioural activation</td>
<td>N</td>
</tr>
<tr>
<td>Vermeersch 2010</td>
<td>CPI-D</td>
<td>Regression</td>
<td>Age, PD, stature, weight</td>
<td>No association between T and dominance</td>
<td>N</td>
</tr>
<tr>
<td>Booth 2003</td>
<td>CES (adol), p-c relation quality</td>
<td>Regression</td>
<td>Age, parents T p-c relation</td>
<td>No significant relationship between T and depression moderated by quality of p-c relationship</td>
<td>Y</td>
</tr>
<tr>
<td>Gerra 2000</td>
<td>Teacher VAS, CPQ, CDI and TAI</td>
<td>Student t</td>
<td>None</td>
<td>Basal T level not different in group with and group without anxiety</td>
<td>Y</td>
</tr>
<tr>
<td>Granger 2003</td>
<td>CBCL, YSR, DISC</td>
<td>Regression, Path Analysis</td>
<td>Age, PD</td>
<td>Anxiety and depression inversely associated with T</td>
<td>Y</td>
</tr>
<tr>
<td>Nottelmann 1987</td>
<td>OSIQ, CBCL</td>
<td>Regression</td>
<td>Age, PD</td>
<td>No significant association between T alone and anxiety or depression</td>
<td>N</td>
</tr>
<tr>
<td>Susman 1985</td>
<td>OSIQ</td>
<td>Regression</td>
<td>Age</td>
<td>No significant correlations seen between T and anxiety or depression</td>
<td>N</td>
</tr>
<tr>
<td>Susman 1991</td>
<td>OSIQ, CBC and DISC (time 3)</td>
<td>Regression</td>
<td>Age, PD</td>
<td>Mixed findings of relationship between T and anxiety/depression at different times and dependent if age and PD in equation</td>
<td>N</td>
</tr>
</tbody>
</table>
4. Testosterone and Male Adolescent Mood and Behaviour

4.4.2. Included studies

A total of 27 studies met the inclusion criteria. These are summarized in Tables 4.1 and 4.2. There was one longitudinal study and 26 cross-sectional studies or longitudinal studies with cross-sectional analysis.

The extraction data are reported in behavioural type subgroup for clarity.

4.4.2.1. Aggressive and/or disruptive behaviour.

Most of the studies focused on this behaviour type. There was only one longitudinal study that investigated the association between testosterone and aggressive behaviours (15). Halpern et al. reported data from a 3-year study of 100 boys in early adolescence with the aim of determining whether pubertal changes in testosterone predicted changes in aggressive behaviour. This was a random sample of white boys in grade 7 and 8 with a 50% participation rate. Analyses considered first whether 6-month changes in testosterone predicted 6-month changes in aggression and second whether changes in testosterone over 1-year periods predicted changes in aggression over the concurrent 1-year periods. In this study, changes in testosterone were not significant over 6-month periods. Testosterone change over a 1-year period was significantly and positively associated with concurrent 1-year changes in one of the six measures of aggression (fights) and one of the two aggression scales (ACL). However, the direction of the relation-ship varied over time, and the authors concluded that change in testosterone showed little relationship with concurrent changes in aggression. There was no evidence of a systematic increase in aggression during the 3 years of the study.
The cross-sectional studies are detailed in Table 4.3 and can be summarized as follows: sample size varied from 20 to 478 with a mean of 143 and median 93.5. The ages of the subjects are given in Table 4.2 with a variation in age range between studies. The scale used to measure the behaviour differed greatly from study to study. There were at least 16 different scales used in the 18 studies. The separation of aggression, aggressive risk taking and risk taking or delinquency, and other nuances of aggressive behaviour did not allow for comparisons between studies.

The most common scale, but only used in five studies, was the CBCL. The CBCL is completed by a parent or teacher and covers a number of different aspects of behaviour including aggression and delinquent behaviour. None of the five studies using CBCL found a significant relationship between testosterone and aggression, and only one of four studies (16) found a positive relationship between testosterone and delinquent behaviour which was moderated by family cohesion.

The three largest studies (17, 25, 36) focused on delinquent or risk behaviour and each used a different scale. Aggression may have been a component but was not separated out in all. A study by Vermeersch et al. (36) used a Dutch self-report scale designed to measure both nonaggressive and aggressive risk taking. This study found a significant but modest relationship between testosterone and nonaggressive risk taking, moderated by peers and no relationship between testosterone and aggressive risk taking. Booth et al. (17) used the Risky Behaviour Scale which is a self-report questionnaire of 18 risky behaviours in the last year and found no relationship between testosterone and risk behaviour but found modification of behaviour by parent-child relationships. The study by Kirillova et al. (25) was the largest study in this group and did find a relationship between testosterone and the development of
4. Testosterone and Male Adolescent Mood and Behaviour

conduct disorder in two of the three periods studied. However, their focus was on conduct disorder, the more extreme degree of aggressive behaviour, and a psychiatric diagnosis measured with the K-SADS-E (Kiddie-Schedule for Affective Disorders and Schizophrenia for School Age Children).

4.4.2.2. Substance use, self-image or social interaction and competency, depression, and anxiety.

There were very few studies looking at other behaviours, mood, or affect. These are summarized in Table 4.4: alcohol and other drug use (group B), self-image and self-competency (group C), and depression and anxiety (group D), and all were cross sectional. Four investigated alcohol and other drug use, three considered aspects of self-image and social competency, and six included depression and anxiety in their analysis.

In group B, only one study (21) found a positive association between testosterone and drug use, which increased as the family, peer, neighbor, and school context became more harmful. The other studies did not find associations between testosterone and drug use but did not consider any peer or family context or other potential modifiers of behaviour.

The three studies in group C focused on different aspects of self-image and competency and included two from Vermeersch’s group that reported on the same study sample (37, 38). None found significant associations between testosterone and measures of self-image or competence.
Of the studies that looked at testosterone and depression or anxiety, sample size varied from 40 (23) to 315 (17). Granger (24) demonstrated an inverse association between testosterone level and anxiety or depression. The remaining five studies demonstrated no significant relationship between testosterone and anxiety or depression.

4.5. Discussion

To the best of the authors’ knowledge, this systematic review is the first to examine the effect of endogenous testosterone on behaviours, mood, or affect in male adolescents. Twenty-seven studies were included in the data extraction: only one was a longitudinal analyses with the other studies either cross sectional in their design or longitudinal cohorts with cross-sectional analyses. As a variety of measurement tools were used for the specified outcomes, no quantitative meta-analysis was possible.

To summarize the findings of this review, there was only one longitudinal study that satisfied the inclusion criteria. This study by Halpern et al. (15) found a positive association between male testosterone and aggression but only in one of six measures of aggression and with variation of the direction of the relationship over time. The findings from 18 cross-sectional studies on the relationship between testosterone and aggression and risk taking were equivocal and conflicting. Significant relationships, when described, often related to only a subset of measures. Most of the 13 cross-sectional studies focusing on other behaviours or mood demonstrated no relationship between testosterone and the behaviour measured.
Contextual modification was only analyzed in a minority of studies. However, there was significant modification of aggressive behaviour by parent-child relationship (17) and modification of substance use through peer, family, and school context (21). Vermeersch et al. (36) found that peer association moderated the relationship between testosterone and risk taking. These moderating external influences are not surprising and should inform future studies which aim to describe the role of puberty hormones on adolescent behaviour and mood.

Our findings are limited by the cross-sectional nature of most of the studies reviewed. Reported outcomes can only show associations between testosterone levels and the behaviour or affect of interest. As such, the conclusions of this review are simply indicative of potential relationships between testosterone and adolescent behaviour and affect. Accordingly, causality cannot be inferred, as an association could be a persistent relationship, short-term change, or mere coincidence. Likewise, the direction of the effect can only be postulated.

There is also a limitation in the acceptance of all testosterone assays, when in fact these may be highly variable in their assessment of the true biological effect of testosterone.

The Greater Smoky Mountains Study (3) of children and adolescents remains one of the largest cohort studies to date that has addressed the relationship between pubertal hormonal and psychological changes, as a part of this major study on the prevalence and causes of psychiatric illness. It was excluded on the criterion of sample timing as the study used only annual single-point measures of testosterone uncontrolled for the
time of day. The study showed an association between depression and testosterone over the pubertal transition.

It is important to consider how the timing of the initial testosterone rise in puberty (onset) and the rate of testosterone change (tempo) might influence behaviour. Because many undesirable adolescent behaviours ameliorate over time, despite hormone levels remaining high, it is possible that the speed of hormone change may be more important than the absolute hormone levels. It is also likely, as suggested by cross-sectional studies, that any effect of testosterone is ameliorated or modified by exogenous factors. As an example, there is some evidence in males that delayed puberty is associated with increased anxiety, decreased self-esteem, and lower social competency (41). This might be affected by timing and tempo of testosterone rise. However, such a situation could also be affected by dissonance from peers in sexual maturation. The descriptors of male pubertal timing are less well characterized when compared with females because of the lack of a clearly defined and easy-to-remember pubertal event like menarche and emphasize the importance of measurement of biological markers of puberty. The knowledge of neural change and cognitive development over adolescence has dramatically advanced over the last decade (42-45). None of the studies in the systematic review addressed neurocognitive change and its association with testosterone, a topic that is currently under investigation by a number of groups. Human studies cannot investigate central neurotransmitter changes (although the peripheral androgen receptor can be studied), which are the likely pathways for testosterone effects on mood and behaviour, either directly or via aromatization to oestradiol (46).
The strength of this study is the comprehensive search strategy and strict inclusion criteria. The weaknesses are the lack of quality longitudinal data and the validity of the testosterone assays used, particularly in older studies. Thus, the authors conclude that insufficient evidence exists to answer the original question about the nonphysical effects of endogenous testosterone on behaviour and mood in adolescent males. This is nevertheless an important outcome, given the assumptions about testosterone’s effect on male behaviour that are often made both by clinicians and the general population and appear in the lay press, including parenting self-help publications. These assumptions, which particularly relate to undesirable behaviours in both adolescent and younger male children, need to be challenged. This is both because of the lack of evidence currently and the fact that the etiology of such behaviours is likely multifactorial and with strong psychosocial interactions. These predictors are likely to be both significant and modifiable, unlike a normal pubertal testosterone rise, which could never be a target for intervention. The optimal approach to demonstrate a relationship between testosterone and any behavioural or mood/affective patterns would be a longitudinal study in a well-described, sufficiently powered cohort that accounted for possible confounders, and with concurrent, frequent, and accurate measurement of testosterone levels and contextual modifiers.

**Summary**

This systematic review has shown that there are insufficient longitudinal data of high methodological quality to currently confirm that the changing testosterone levels during puberty significantly affect male adolescent behaviours and mood.
4. Testosterone and Male Adolescent Mood and Behaviour

4.6. References


4. Testosterone and Male Adolescent Mood and Behaviour

5. TEXT MESSAGING IS A USEFUL RESEARCH TOOL

Ben W.R. Balzer¹,², Patrick J. Kelly³, Philip Hazell⁴, Karen Paxton⁵,
Catherine Hawke⁶, Katharine S. Steinbeck¹,²

Affiliations: ¹Academic Department of Adolescent Medicine, The Children’s
Hospital at Westmead, Sydney, NSW, Australia; ²Discipline of Paediatrics and
Child Health, Sydney Medical School, The University of Sydney, Sydney, NSW,
Australia; ³Sydney School of Public Health, The University of Sydney, Sydney,
NSW, Australia; ⁴Discipline of Psychiatry, Sydney Medical School, The University
of Sydney, Sydney, NSW, Australia; ⁵School of Rural Health, The University of
Sydney, Orange, NSW, Australia

Citation: Balzer, B.W.R., Kelly, P.J., Hazell, P., Paxton, K., Hawke, C.I., Steinbeck,
(DOI: 10.1136/archdischild-2014-306075)


5.1. Abstract

Introduction: Longitudinal studies of adolescents must be ‘adolescent-friendly’, both to collect data and encourage maintenance in the study cohort. Text messaging may offer a feasible means to do both.

Methods: Adolescents in the Adolescent Rural Cohort, Hormones and Health, Education, Environments and Relationships (ARCHER) study (n=342) are sent automated text messages every three months, prompting biological specimen collection.

Results: A total of 99.2% of participants (or their parents) owned a mobile phone, of which 89.1% of participants responded to text messages and 97.3% of intended urine samples were collected. The average time to provide a urine sample after prompting correlated with time to reply to SMS.

Conclusions: This study shows SMS can be used effectively in longitudinal research involving adolescents, and is both feasible and useful as a reminder tool for regular biological specimen collection.
5.2. Introduction

Longitudinal studies in adolescents have their own particular challenges, especially when the collection of biological samples is included. (1) In this context, ‘adolescent-friendly’ research tools are required both for the collection of longitudinal data and to encourage retention in the study cohort. One potential tool is text messaging via Short Message Service (SMS), (2) a ubiquitous communication means among adolescents. SMS has been shown to be effective in a variety of healthcare contexts, (3) though its usefulness in research involving adolescents is unclear. The aim of this study was to determine the usefulness of SMS as a reminder tool in a longitudinal study of adolescents.

5.3. Methods

The Adolescent Rural Cohort, Hormones and Health, Education, Environments and Relationships (ARCHER) study (2) is a three-year longitudinal cohort study of adolescents, which requires the regular collection of urine samples for puberty hormone assay. The principal aim of the ARCHER study is to explore how the onset and tempo of changes in gonadal hormones during puberty affect adolescent behavior, health and wellbeing.

Participants are sent urine collection kits every three months. SMS messages are sent the morning of the scheduled urine collection day. Depending on family cellular phone ownership, participants (or parents) are instructed to reply when the urine collection is completed. Participants without cellular phones were provided with a sticker on which they wrote their mood score number (part of the broader study and data not presented here). The standard text was: “Hi <First Name><Last Name>. On
5. Text Messaging is a Useful Research Tool

a scale from 0 to 9 (with 9 = best ever). What is your mood now? Thanks
<Investigator Name>.” This reply is in a form of a linear analog self-assessment scale of mood (focusing on the adolescent). Prior research has shown that any request for action increased the chances of a response to a text message (4) and we hypothesized our request would increase the likelihood of specimen collection. If participants did not provide an SMS response or urine specimen within two weeks, they were followed up by telephone call.

The ARCHER Study has full ethical approval from the Human Research Ethics Committee, University of Sydney (HREC 13094) within the Australian National Health and Medical Research Council Guidelines for Human Experimentation, which are consistent with the Declaration of Helsinki. All participants and their parents provided written informed consent prior to commencing the study.

5.4. Results

The total study recruitment was 342 adolescents (153 female). Three hundred and thirty-nine participants (151 female) in the ARCHER study had provided at least one SMS mood indicator response by June 30, 2013. The average age (SD) for adolescent participants at enrollment was 11.8 (0.99) years for males and 11.7 (0.98) years for females. Three hundred and thirty-seven (99.4%) of these participants or their parents owned a cellular phone and were able to provide responses by SMS. The two participants without a mobile phone provided their responses written on stickers provided with the urine collection kit. A total of 1555 SMS were sent to the 339 participants between 1 June 2011 and 30 June 2013, with 1386 replies (89.1% response rate).
From 1555 intended urine collections, 42 were not provided (2.7%). Of these, 36 (85.7%) did not provide an SMS response. Over two-thirds of SMS replies and urine samples were obtained within seven days of the scheduled date (see Table 5.1). Males and females did not differ in time to reply to SMS (mean difference = 1.75 days; 95% CI = -0.431 – 3.933 days; p=0.116), though females took longer than males to provide a urine specimen (mean difference = 2.01 days; 95% CI 0.526 – 3.50 days; p=0.008).

Time of urine specimen collection was correlated with the time of SMS reply, as shown in Figure 1 (r = 0.606, p<0.001). Some SMS replies and urine collections occurred before scheduled dates if participants were unavailable then, giving negative times relative to scheduling (Figure 5.1).
Table 5.1. Time to SMS reply and urine collection scheduled date by gender.

<table>
<thead>
<tr>
<th></th>
<th>Overall N (%)</th>
<th>Male N (%)</th>
<th>Female N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SMS Reply</strong>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time to reply (days) (Median/IQR)</td>
<td>3 (0-10)</td>
<td>2 (0-8)</td>
<td>4 (0-12)</td>
</tr>
<tr>
<td>Reply within 24 hours</td>
<td>599 (39)</td>
<td>333 (40)</td>
<td>266 (36)</td>
</tr>
<tr>
<td>Reply between 2 to 7 days</td>
<td>365 (23)</td>
<td>208 (25)</td>
<td>157 (21)</td>
</tr>
<tr>
<td>Reply between 8 to 28 days</td>
<td>332 (21)</td>
<td>159 (19)</td>
<td>173 (24)</td>
</tr>
<tr>
<td>Reply between 29 to 90 days</td>
<td>90 (6)</td>
<td>38 (5)</td>
<td>52 (7)</td>
</tr>
<tr>
<td>Did not reply (no reply after 90 days)</td>
<td>169 (11)</td>
<td>83 (10)</td>
<td>86 (12)</td>
</tr>
<tr>
<td><strong>Urine Collection</strong>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time to collection (days) (Median/IQR)</td>
<td>3 (0-10)</td>
<td>3 (0-8)</td>
<td>4 (0-12)</td>
</tr>
<tr>
<td>Collection within 24 hours</td>
<td>643 (41)</td>
<td>354 (43)</td>
<td>289 (39)</td>
</tr>
<tr>
<td>Collection between 2 to 7 days</td>
<td>409 (26)</td>
<td>239 (29)</td>
<td>170 (23)</td>
</tr>
<tr>
<td>Collection between 8 to 28 days</td>
<td>373 (24)</td>
<td>177 (21)</td>
<td>196 (27)</td>
</tr>
<tr>
<td>Collection between 29 to 90 days</td>
<td>88 (6)</td>
<td>42 (5)</td>
<td>56 (8)</td>
</tr>
<tr>
<td>Did not collect (no specimen after 90 days)</td>
<td>42 (3)</td>
<td>20 (2)</td>
<td>22 (3)</td>
</tr>
</tbody>
</table>

* Time in days to SMS reply or urine collection from scheduled date
Figure 5.1. Plot of urine specimen collection time and SMS reply time. Line shows $y=x$. Dots on the line indicate SMS reply and urine sample conducted on same day, dots above line indicate urine sample occurred before SMS reply and dots below the line indicates urine sample conducted post SMS reply.
5.5. Discussion

In this study, SMS reminders were used to increase compliance and retention. The high cellular phone ownership and high response rates to the text messages in this cohort indicate that contact via SMS is both feasible and acceptable in a longitudinal cohort study of adolescents.

The time to reply to the SMS strongly correlated with the time to supply a urine sample, and less than 10% of collections were missed. While this does not imply SMS was the sole reason for remembering to collect a sample, the timing of the message with scheduled collection is likely to have an influence. The mean time from scheduled date to reply SMS receipt and the provision of a urine sample were both approximately one week. This is primarily due to scheduling and logistics between research staff and participants: dates had to be organized for specimen kit delivery, timing for follicular phase of menstrual cycle and participant preference. Menstrual cycle adjustments may explain the difference in urine specimen collection between sexes. The negative times seen in Figure 1 did not affect the correlation between SMS reply and collection time.

There are several limitations to our study. There was no control group so we cannot be certain the SMS prompt contributed to compliance and may limit the conclusions which we can draw. Likewise, the SMS component was used once every three months so results could differ if contact occurred at different frequencies, though weekly SMS contact with children and adolescents has been shown to have similar response rates (5). We could not tell whether the parent or the adolescent was replying to the text message. As the SMS was one aspect of the reminder system
(along with the delivery of the specimen collection kit), it may not be the sole reason for collection reminder, though the strong correlation between reply time and collection rate shows participants were compliant in replying to the SMS as requested. The use of smartphone applications, such as “SnapChat” or “What’s App” may appeal more to adolescents and possibly increase reply rate, however SMS is a feature common to all cellular phones so avoids limitations imposed by operating systems, data access and application costs.

In conclusion, SMS has been shown to be an acceptable and feasible tool for use as part of a reminder system in longitudinal research involving adolescents. The current analysis suggests that SMS may be used in research studies involving multiple home-based data collection points and replies are in line with instructions (regarding timing and content). Young people readily reply to a simple and short SMS. Our findings suggest that SMS may be a potential data collection tool for other information, provided the question does not require a complex response. Clinically, SMS may be useful in monitoring and compliance with adolescent patients who require multiple specimen collections over a period of time.

5.6. References

6. URINARY SEX STEROIDS AND ANTHROPOMETRIC MARKERS OF PUBERTY

A NOVEL APPROACH TO CHARACTERISING WITHIN-PERSON CHANGES OF PUBERTY HORMONES

Gurmeet K.S. Singh¹,²,*, Ben W.R. Balzer³,⁴,*, Patrick J. Kelly⁵, Karen Paxton⁶, Catherine I. Hawke⁶, David J. Handelsman¹, Katharine S. Steinbeck³,⁴,*

* Equal First Authorship

Affiliations: ¹ ANZAC Research Institute, University of Sydney, NSW, Australia; ² Faculty of Pharmacy, Universiti Teknologi MARA, Selangor, Malaysia; ³ Academic Department of Adolescent Medicine, The Children’s Hospital at Westmead, Westmead, NSW, Australia; ⁴ Discipline of Paediatrics and Child Health, Sydney Medical School, University of Sydney, NSW, Australia; ⁵ Sydney School of Public Health, Sydney Medical School, University of Sydney, NSW, Australia; ⁶ School of Rural Health, Sydney Medical School, University of Sydney, Dubbo and Orange, NSW, Australia

6.1. Abstract

Background/Aims: The longitudinal relationships of within-individual hormone and anthropometric changes during puberty have not ever been fully described. The objectives of this study were to demonstrate that 3 monthly urine collection was feasible in young adolescents and to utilise liquid chromatography-tandem mass spectrometry assay methods for serum and urine testosterone (T), oestradiol (E₂) and luteinizing hormone (LH) in adolescents by relating temporal changes in urine and serum hormones over 12 months to standard measures of pubertal development.

Methods: A community sample of 104 adolescents (57 female) was studied over 12 months with annual anthropometric assessment, blood sampling and self-rated Tanner staging and urine collected every 3 months. Serum and urine sex steroids (T, E₂) were measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) and LH by immunoassay.

Results: A high proportion (92%) of scheduled samples were obtained with low attrition rate of 6.7% over the 12 months. Urine hormone measurements correlated cross-sectionally and longitudinally with age, anthropometry and Tanner stage.

Conclusion: We have developed a feasible and valid sampling methodology and measurements for puberty hormones in urine, which allows a sampling frequency by which individual pubertal progression in adolescents can be described in depth.
6.2. **Introduction**

The circulating gonadotropin, testosterone (T), and oestradiol (E2) changes that drive the external manifestations of puberty are well described from cross-sectional studies according to chronological age or Tanner staging interpreted quasi-longitudinally (1, 2). Such cross-sectional analysis artificially smooths longitudinal data due to a low resolution in temporal sampling, markedly underestimating the underlying within-subject variability. Combined with this variation in individual hormones is the normal variability in both time of onset and tempo of completing puberty. Both timing and tempo of hormone change might be an important intermediate factor in the marked behavioural and psychological changes of adolescence, but presents challenges in its study. Previous studies that have considered how the individual variability in puberty hormone change might influence the dramatic psycho-bio-behavioural changes in adolescence relied upon older methods of often direct (unextracted) sex steroid immunoassays. This is a less accurate technology especially at low circulating steroid levels concentrations (3-5), and is now being supplanted by more sensitive and specific liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based steroid assays (3). Similar more sensitive LC-MS/MS -based methods have been recently described for serum sex steroids in pre-pubertal children (6).

In clinical settings the usual methods to appraise pubertal development and, by implication, its variability, comprise hormone measurements and anthropometry, with emphasis on timing of the height growth spurt, clinical inspection and rating of secondary sexual characteristics and bone age (7, 8). In epidemiological studies and community studies the definition of puberty has to be simplified to be based on
adolescent self-report or parental report against either Tanner stage line drawings (9, 10) or Petersen’s Pubertal Development Scale (1). Self-rated Tanner staging is less intrusive and often the only Ethics Board approved method to assess puberty as defined by secondary sexual characteristics for epidemiological and community studies involving adolescents who remain legally minors. This has to sacrifice some reliability, especially in early puberty, and with the limited number of developmental stages available to select (11-14). Recent work has demonstrated self-rating is of insufficient accuracy to be of use in the clinical setting, however may be acceptable in research settings where clinician assessment is not possible (15). Menarche (a late pubertal event), spermarche (16, 17) which is difficult to evaluate (18), and voice breakage have all been described in relation to age and/or Tanner stage (19, 20). However, as singular time points these provide minimal information on either timing of onset or tempo of puberty. To date in the literature there have been no reports describing individual puberty hormone change with a sufficient measurement frequency to adequately capture individual variation. Unless this is achieved, it is not possible to determine the true effects of puberty hormones on adolescent mood, wellbeing and behaviour.

This study aims to examine the feasibility of frequent three-monthly urine collection in a large Australian adolescent cohort, as well as the utility of LC-MS/MS assays urine and serum sex hormones. Additionally, we will describe changes in E₂, T (using these LC-MS/MS) assays and luteinizing hormone (LH) over 12 months in adolescents to compare results for urine and serum hormones. Finally, we aim to examine the association between the changes in anthropometry and self-reported Tanner stage with changes in hormones in both urine and serum.
6. Materials and Methods

6.3. Setting and Participants

The study was set in two regional towns in the state of New South Wales (NSW), Australia. Adolescents between the ages of 10 and 12 years were recruited from local schools. Fasting morning blood samples were collected at 0 and 12 months for the measurement of E$_2$, T and LH and first morning (fasting) urine collected three-monthly for the same measures. No participants had an endocrine disorder or were on any type of gonadal steroid hormone therapy.

6.3.2. Anthropometry and pubertal staging

Height was measured using a portable stadiometer (to 0.1 cm). Weight was measured in light clothing using a Tanita TBF-300 Pro Body Composition Analyzer (21). Body mass index (BMI) was calculated using these measures. The adolescents provided a self-rating of puberty using line drawings based on the Tanner stages (9, 10). Females provided self-rating of breast stage and males provided self-rating of genitalia stage. Self-report of Tanner stage (11-14) was the only feasible and ethically acceptable measure of pubertal staging available to the investigators. This situation is now a common limitation, as both Institutional Review Boards and parents are reluctant to permit direct physical examination of undressed healthy children for a research study of normal puberty.

6.3.3. Hormone measurements

Following a 12-hour fast urine was collected at home before blood samples were collected between 7:00 am and 8:30 am in order to minimise the effects of diurnal
hormone variation (1, 22). The urine sample was collected as a first morning void into a lidded container and immediately placed into an insulated carry bag on an ice brick. This was refrigerated in a freezer at -20 °C within 3 hours. Blood and urine samples were then transported on dry ice to the research laboratory where these were kept at -80 °C until analysed – approximately 12 months later. Post-menarcheal girls provided urine and blood specimens in the mid-follicular phase (Day 7-10) of their menstrual cycle in order to standardise collection time. Urine and serum steroids were measured by liquid chromatography, tandem mass spectrometry (LC-MS/MS) as modified from a previously described method for serum (23) and adapted for urine specimens following enzymatic deconjugation. The conjugated steroid underwent hydrolysis using β-glucuronidase enzyme from *Escherichia coli* K12 (Roche Diagnostic, Mannheim, Germany) that deconjugates the glucuronides moiety from steroids. The developed LC-MS/MS method measures unconjugated steroids (i.e. originally unconjugated plus deconjugated). The details of the novel urine assays have been published (24). Briefly, urine and serum specimens were separated by liquid chromatography using a Shimadzu Nexera UHPLC system (Shimadzu Scientific Instruments, Columbia, MD). Following this, tandem-mass spectrometry analysis was performed on samples using an API-5000 triple-quadrupole mass spectrometer (Applied Biosystem/MDS SCIEX, Ontario, Canada) (24). A thorough validation was carried out for the LC-MS/MS method to measure E₂ and T according to standard FDA/EMEA analytical validity criteria.

The calibration curves of the urinary steroid LC-MS/MS method ranged from 0.025 and 32 ng/mL for T and 0.05 and 32 ng/mL for E₂, and fitted quadratic functions with r of 0.999 or better. Within-day and between-day accuracies and precision at all
levels of quality control ranged from 95-105% and 2.6-9.7%, respectively. Using a 500 µL urine sample the limits of detection (LOD) and limits of quantitation (LOQ), respectively, were 5 pg/mL and 25 pg/mL for T and 25 pg/mL and 50 pg/mL for E2. Matrix effects were negligible with no significant ion suppression or enhancement for either analyte with recovery values between 102-108%. Extraction recovery and process efficiency were between 93-103% for both the analytes at all levels of QC. The β-glucuronidase enzyme hydrolysis (deconjugation) efficiency was 100-102% after an overnight incubation at room temperature.

Specificity of both the analytes was tested against structurally related compounds that potentially may interfere with the method, including oestrone, epitestosterone, androsterone, etiocholanolone, epitiocholanolone, 3-α androstanediol, 3-β androstanediol, and dehydroepiandrosterone. All the steroids listed did not interfere with the studied analytes. Blood and urine samples with values less than the lower limit of quantitation (LLOQ) for E2 and T were taken as half the LLOQ. Total of 27 urine (E2: 16, T: 11) and 2 serum (T) samples were below the LLOQ. Serum and urine LH were measured by Immulite 1000 LH (Siemens) which detects intact LH and LHβ subunit and provided reproducible measurements in frozen stored urine(25). The within-assay coefficients of variation were <10%. Serum and urine LH values below the detection limit (0.1 IU/L) were set at zero (10 serum and 5 urine samples). Urine FSH assays (Immulaite, Delfia) did not pass validity tests (dilutional linearity, quantitative spike recovery) and were therefore not used in this study. All urine hormone concentrations were adjusted for urine specific gravity (SG) measured by reagent strip (ChoiceLine 10, Roche Diagnostics) to a standard SG of 1.020.
6.3.4. **Statistical analyses**

Anthropometric measurements and hormones levels (blood and urine) were summarised by gender over time using means and standard deviations for continuous variables and frequencies and proportions for categorical variables. Linear mixed effects models were used to assess longitudinal changes in hormones, with gender and collection time as covariates. A random effect for child was included in these models. Interaction between gender and collection time was tested for each hormone. Linear regression was used to analyse the association between changes in anthropometry (height, weight, BMI) and self-rated Tanner stage and changes in urine and serum hormones. For these, the models were adjusted for the baseline anthropometric measurement and baseline hormone concentration. All p-values were calculated using Wald tests. Statistical analyses were conducted using Stata 12.1 (StataCorp, Texas, USA). Statistical significance was set at 0.05.

6.3.5. **Ethical considerations**

The study has ethical approval from the Human Research Ethics Committee, University of Sydney (HREC 13094) within the National Health and Medical Research Council Guidelines for Human Experimentation, which is consistent with the Declaration of Helsinki. All participants assented, and a parent provided written informed consent prior to commencing the study.
6.4. **Results**

6.4.1. **Cohort Characteristics**

One hundred and four participants were recruited. The mean ages (SD) for the study participants at baseline were 12.5 (0.93) years for males and 11.8 (0.98) years for females. At follow-up, the ages were 13.5 (0.94) years for males and 12.9 (0.97) years for females. Anthropometric characteristics for this cohort are shown in Table 6.1. For the females, 22 (39%) had menarche prior to the study and one additional girl experienced menarche during the follow-up year. Post-menarcheal girls were significantly older than their pre-menarcheal counterparts (12.9 years vs. 11.9 years, p<0.001). A high proportion of scheduled samples were collected for urine (484, 92%) and serum (194, 93%). There was a low loss to follow-up (7, 6.7%). For those not lost to follow-up, 16 females and 12 males did not provide at least one urine specimen. There was no difference in age between those who provided a specimen and those who did not (mean difference 0.35 years; 95% CI -0.21-0.92; p=0.23), nor was there a difference in self-rated Tanner stage (p=0.53).

Mean anthropometric measurements (Table 6.1) increased significantly over 12 months whereas age- and gender-standardized z-scores did not change over the 12 months follow-up. Pre-menarcheal girls significantly increased their age-standardized weight (p<0.001) and height z-scores (p = 0.002), whereas these z-scores did not change for post-menarcheal girls. Rate of change in anthropometric measurements were similar between genders (interaction p>0.05).

Self-rated Tanner staging increased with fewer in stage 1 and more in stage 5 at 12-month follow-up (Table 6.2). During the year, 14 participants (13.5%) progressed.
two Tanner stages, 43 (41.3%) progressed one stage and 41 (39.4%) did not change in their self-rated Tanner stage. Six (5.8%) participants (three boys) provided a lower self-rated Tanner stage at follow-up than baseline. One participant (1%, 1 boy) did not provide baseline Tanner staging and seven (6.7%, 1 boy, 6 girls) did not provide follow-up Tanner staging.
Table 6.1. Baseline and 12-month follow-up anthropometry measurements.

<table>
<thead>
<tr>
<th></th>
<th>Baseline Mean (SD)</th>
<th>12 Months Mean (SD)</th>
<th>p-value</th>
<th>Gender*</th>
<th>Time*</th>
<th>Gender x Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (n=47)</td>
<td>Female (n=57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female (n=57)</td>
<td>Male (n=47)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female (n=57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Height (cm)</strong></td>
<td>156.8 (9.9)</td>
<td>163.8 (10.1)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>150.8 (8.5)</td>
<td>157.1 (7.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Weight (kg)</strong></td>
<td>49.3 (12.0)</td>
<td>55.2 (13.2)</td>
<td>0.026</td>
<td>&lt;0.001</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44.1 (10.9)</td>
<td>50.1 (11.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BMI (kg/m²)</strong></td>
<td>19.7 (3.5)</td>
<td>20.4 (3.8)</td>
<td>0.56</td>
<td>&lt;0.001</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19.3 (3.7)</td>
<td>20.3 (4.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Self-rated Tanner stage</strong></td>
<td>3 (1.2)</td>
<td>4 (1.0)</td>
<td>0.001</td>
<td>&lt;0.001</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 (1.1)</td>
<td>3 (1.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* these p-values were calculated with both gender and time in the model, but with no interaction term

Table 6.2. Self-Rated Tanner staging at baseline and 12-month follow-up.

<table>
<thead>
<tr>
<th>Tanner</th>
<th>Baseline Male (%)</th>
<th>Baseline Female (%)</th>
<th>12 Months Male (%)</th>
<th>12 Months Female (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6 (13)</td>
<td>14 (25)</td>
<td>0 (0)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>2</td>
<td>8 (17)</td>
<td>17 (30)</td>
<td>7 (16)</td>
<td>13 (25)</td>
</tr>
<tr>
<td>3</td>
<td>9 (20)</td>
<td>16 (28)</td>
<td>6 (13)</td>
<td>15 (29)</td>
</tr>
<tr>
<td>4</td>
<td>18 (39)</td>
<td>7 (12)</td>
<td>18 (40)</td>
<td>12 (24)</td>
</tr>
<tr>
<td>5</td>
<td>5 (11)</td>
<td>3 (5)</td>
<td>14 (31)</td>
<td>8 (16)</td>
</tr>
<tr>
<td>Total</td>
<td>46</td>
<td>57</td>
<td>45</td>
<td>51</td>
</tr>
</tbody>
</table>

6. Urinary Sex Steroids and Anthropometric Markers of Puberty
6. Urinary Sex Steroids and Anthropometric Markers of Puberty

6.4.2. **Hormone measurements**

6.4.2.1. **Urine hormone levels versus time**

For the 484 urine collections, 14 (3%) E₂, nine (2%) T and five (1%) LH assays were below the LLOQ. For LH, these participants were younger than the rest of the cohort (mean difference 1.35 years; 95% CI 0.41-2.30; p=0.005), but there was no age difference between those with urine E₂ or T samples below or above the limits of detection. In five such cases (one urine E₂, one urine E₂ and T, one urine T, one urine and serum and one urine LH) the participant provided a Tanner stage 1 at baseline.

Figure 6.1 shows within-person changes from baseline over the 12-month collection period and Table 6.3 shows the mean three monthly values for urinary E₂, T and LH over 12 months. There is a significant increase and clear upward pattern for all subjects in mean urine T and E₂ plotted over time from baseline to follow-up in E₂, T and LH. The same pattern is seen in females when stratified by menarcheal status (data not shown). Urine hormone levels were not strictly progressive and in some instances decreased over time, though there was an overall increase in mean levels for all three hormones for males and females. For example, there was a 22.2% decrease in mean male and a 40.0% decrease in mean female E₂ between six and nine months, a 14.6% decrease in male T between baseline and three months, a 24.6% decrease for male and 16.7% decrease for female T between six and nine months. For LH, there was a decrease for males between baseline and three months of 3.4% and between six and nine months a 14.3% decrease and a 23.3% decrease for males and females, respectively. At all other times there was an increase in urine hormones. For serum hormones, a decline from baseline to follow-up was observed.
6. Urinary Sex Steroids and Anthropometric Markers of Puberty

22 adolescents (11 female) for E₂, 19 (13 female) for T and in 26 adolescents (18 female) for LH. Overall declines were observed in 23 (9 female) for urinary E₂, 24 (14 female) for urinary T and in 48 adolescents (30 female) for urinary LH.
Figure 6.1. Plots of mean and SD of urinary E₂, T and LH concentrations derived from 3-, 6-, 9- and 12-months collection changes from baseline (0) for urine (top) and serum (bottom)
Table 6.3. Mean urine E2, T and LH over time for males (M) and females (F).

<table>
<thead>
<tr>
<th>Mean</th>
<th>Baseline</th>
<th>3 Months</th>
<th>6 Months</th>
<th>9 Months</th>
<th>12 Months (Follow-Up)</th>
<th>p-value</th>
<th>Gender*</th>
<th>Time*</th>
<th>Gender x Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2 (ng/mL)</td>
<td>M 0.7 (0.9) F 1.8 (2.0)</td>
<td>M 0.7 (0.9) F 2.1 (2.8)</td>
<td>M 1.1 (1.6) F 2.5 (3.1)</td>
<td>M 0.9 (1.1) F 3.5 (4.8)</td>
<td>M 1.0 (1.0) F 2.7 (2.8)</td>
<td>0.002</td>
<td>0.002</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>T (ng/mL)</td>
<td>M 23.3 (35) F 3.9 (3.8)</td>
<td>M 19.9 (26.1) F 5.0 (6.2)</td>
<td>M 35.1 (80) F 5.0 (4.1)</td>
<td>M 26.4 (4.1) F 6.0 (34)</td>
<td>M 26.2 (28) F 5.4 (5.1)</td>
<td>0.024</td>
<td>&lt;0.001</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>LH (IU/L)</td>
<td>M 8.7 (7.4) F 9.6 (10.0)</td>
<td>M 8.4 (5.9) F 10.2 (10.1)</td>
<td>M 9.6 (7.4) F 11.5 (11.4)</td>
<td>M 9.8 (6.1) F 13.3 (6.1)</td>
<td>M 10.1 (16.6) F 10.1 (7.9)</td>
<td>0.232</td>
<td>0.001</td>
<td>0.82</td>
<td></td>
</tr>
</tbody>
</table>

* these p-values were calculated with both gender and time in the model, but with no interaction term
6.4.2.2. Urine versus Serum

Urinary E₂, T and LH all correlated with their serum hormone measurements (p<0.001 for all; see Table 6.4), and increased in each Tanner stage.

6.4.2.3. Serum hormone levels versus time

Mean serum hormone levels at baseline and one year follow-up are shown in Table 6.5. Of the 194 serum collections, two (1%) T and nine (5%) LH assays (all in separate individuals) were below the LLOQ. No E₂ samples were below LLOQ. Those with LH samples below the limits of detection were significantly younger than their peers (mean difference = 1.47 years 95% CI 0.81-2.14 p<0.001). No difference in age was observed for those with serum E₂ or T samples below limits of detection.

Hormone values all significantly increased over the 12-month period and were significantly different between genders, but rate of change was not statistically significant between genders (interaction p>0.05), except for T.
Table 6.4. Pearson’s Correlation Coefficients for Serum and Urinary Hormones using 0- and 12-month data.

<table>
<thead>
<tr>
<th></th>
<th>Urine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
</tr>
<tr>
<td>Serum E2</td>
<td>0.7089</td>
</tr>
<tr>
<td>Serum T</td>
<td>–0.7864</td>
</tr>
<tr>
<td>Serum LH</td>
<td>–0.5513</td>
</tr>
<tr>
<td><strong>Follow-Up</strong></td>
<td></td>
</tr>
<tr>
<td>Serum E2</td>
<td>0.7245</td>
</tr>
<tr>
<td>Serum T</td>
<td>0.7918</td>
</tr>
<tr>
<td>Serum LH</td>
<td>–0.4206</td>
</tr>
<tr>
<td><strong>Pooled</strong></td>
<td></td>
</tr>
<tr>
<td>Serum E2</td>
<td>0.7187</td>
</tr>
<tr>
<td>Serum T</td>
<td>0.7746</td>
</tr>
<tr>
<td>Serum LH</td>
<td>–0.4735</td>
</tr>
</tbody>
</table>

p<0.001 for all

Table 6.5. Baseline and 12-month follow-up serum E₂, T and LH.

<table>
<thead>
<tr>
<th>Mean (SD)</th>
<th>Baseline</th>
<th>12 months</th>
<th>p-value Gender</th>
<th>Time</th>
<th>Gender x Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td></td>
</tr>
<tr>
<td>E₂ (pg/mL)</td>
<td>19.1 (8.5)</td>
<td>43.4 (29)</td>
<td>24.9 (9)</td>
<td>54.7 (30)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>T (ng/mL)</td>
<td>2.4 (2.3)</td>
<td>0.16 (0.10)</td>
<td>3.4 (2.55)</td>
<td>0.2 (0.12)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>LH (IU/L)</td>
<td>1.7 (1.3)</td>
<td>2.3 (2.4)</td>
<td>2.5 (1.8)</td>
<td>3.6 (2.7)</td>
<td>0.23</td>
</tr>
</tbody>
</table>

* these p-values were calculated with both gender and time in the model, but with no interaction term
6.4.3. **Association between Anthropometry Measurements and Urine and Serum Hormones**

Urinary E2, T and LH all positively correlated with Tanner staging at baseline and 12-month follow-up (p<0.001 for all). Table 6.6 shows the associations between anthropometry changes and changes in sex hormones. Change in height was associated with changes in serum T, and serum and urine LH in females and both serum and urine T in males. Change in weight was associated with changes in urine E2, serum T and serum LH in females. Serum T and LH was associated with self-rated Tanner stage in males, but this was not statistically significant based on the urine samples. No other significant associations were observed between serum or urinary hormones and changes in anthropometry over 12 months.

Figure 6.2 shows urine and serum E2, T and LH stratified by Tanner stages. Data stratified by chronological age were similar. Hormone concentrations increased through each Tanner stage and each year of age.
Table 6.6. Regression results for serum (left) and urine (right) hormones and anthropometric markers of puberty.

<table>
<thead>
<tr>
<th></th>
<th>Serum R²</th>
<th>β</th>
<th>p-value</th>
<th>Urine R²</th>
<th>β</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Change in Height</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td>0.15</td>
<td>-0.001</td>
<td></td>
<td>E₂</td>
<td>0.25</td>
<td>0.26</td>
</tr>
<tr>
<td>T</td>
<td>0.22</td>
<td>10.19</td>
<td></td>
<td>T</td>
<td>0.22</td>
<td>0.13</td>
</tr>
<tr>
<td>LH</td>
<td>0.32</td>
<td>0.52</td>
<td>0.001</td>
<td>LH</td>
<td>0.24</td>
<td>0.09</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td>0.15</td>
<td>0.09</td>
<td></td>
<td>E₂</td>
<td>0.16</td>
<td>0.61</td>
</tr>
<tr>
<td>T</td>
<td>0.34</td>
<td>0.89</td>
<td>&lt;0.001</td>
<td>T</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>LH</td>
<td>0.02</td>
<td>-0.16</td>
<td>0.68</td>
<td>LH</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Change in Weight</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.66</td>
<td>E₂</td>
<td>0.15</td>
<td>0.55</td>
</tr>
<tr>
<td>T</td>
<td>0.08</td>
<td>11.81</td>
<td>0.031</td>
<td>T</td>
<td>0.12</td>
<td>0.21</td>
</tr>
<tr>
<td>LH</td>
<td>0.26</td>
<td>0.45</td>
<td>0.013</td>
<td>LH</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td>0.03</td>
<td>0.07</td>
<td>0.58</td>
<td>E₂</td>
<td>0.06</td>
<td>1.56</td>
</tr>
<tr>
<td>T</td>
<td>0.04</td>
<td>0.21</td>
<td>0.69</td>
<td>T</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>LH</td>
<td>0.01</td>
<td>-0.10</td>
<td>0.90</td>
<td>LH</td>
<td>0.003</td>
<td>-0.01</td>
</tr>
<tr>
<td><strong>Change in BMI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td>0.04</td>
<td>0.001</td>
<td>0.92</td>
<td>E₂</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td>T</td>
<td>0.07</td>
<td>2.18</td>
<td>0.27</td>
<td>T</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>LH</td>
<td>0.16</td>
<td>0.03</td>
<td>0.63</td>
<td>LH</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td>0.04</td>
<td>-0.02</td>
<td>0.41</td>
<td>E₂</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>T</td>
<td>0.46</td>
<td>-0.01</td>
<td>0.96</td>
<td>T</td>
<td>0.01</td>
<td>0.003</td>
</tr>
<tr>
<td>LH</td>
<td>0.02</td>
<td>-0.10</td>
<td>0.58</td>
<td>LH</td>
<td>0.22</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Change in Self-Rated Tanner Stage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td>0.19</td>
<td>0.005</td>
<td>0.40</td>
<td>E₂</td>
<td>0.23</td>
<td>0.10</td>
</tr>
<tr>
<td>T</td>
<td>0.23</td>
<td>2.18</td>
<td>0.15</td>
<td>T</td>
<td>0.19</td>
<td>0.03</td>
</tr>
<tr>
<td>LH</td>
<td>0.26</td>
<td>0.10</td>
<td>0.06</td>
<td>LH</td>
<td>0.20</td>
<td>0.02</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td>0.39</td>
<td>0.04</td>
<td>0.042</td>
<td>E₂</td>
<td>0.31</td>
<td>0.20</td>
</tr>
<tr>
<td>T</td>
<td>0.46</td>
<td>0.18</td>
<td>0.008</td>
<td>T</td>
<td>0.35</td>
<td>0.01</td>
</tr>
<tr>
<td>LH</td>
<td>0.29</td>
<td>-0.06</td>
<td>0.62</td>
<td>LH</td>
<td>0.36</td>
<td>0.02</td>
</tr>
</tbody>
</table>
6. Urinary Sex Steroids and Anthropometric Markers of Puberty

**Figure 6.2.** Plots of mean and SD E\(_2\) (left), T (centre) and LH (right) in urine (top) and serum (bottom) specimens, by Tanner stage.
6.5. Discussion

There are two major findings of this study. First is the demonstration that it is feasible to collect urine samples from a community-based adolescent cohort at three-monthly intervals, with high compliance (93% serum and 92% urine collections completed) and low follow-up attrition rate. Second is the finding that urine levels both correlate with serum testosterone and oestradiol and have the potential to provide a more subtle and nuanced description of individual puberty hormone progression.

In order to demonstrate the specific effects of puberty hormones on any biological aspect of adolescent development and health, a methodology that allows more frequent biological sampling than has ever been previously reported for community-based field studies is essential. A single reading of testosterone or oestradiol in whatever biological sample is of limited use. It is change that is relevant and hence the importance of longitudinal samples with frequent sampling. Urine samples have the advantage over blood samples that these are more acceptable to both research ethics and adolescents. Each overnight sample also provides a more time-integrated hormone measure. This is particularly true in early puberty when pubertal hormones commence pulsatile secretion nocturnally (1), so that a morning overnight urine sample may be more informative than a serum sample at any single time point. Urine collections also allow for more frequent collection than repeated venepuncture would be tolerated (26). Salivary samples are potentially easier to collect, but blood contamination and influence of flow-rate on measurements seriously limits validity and accuracy and these may be subject to non-compliance by children (27).
Urine steroid measurements have been dramatically improved upon by introduction of new LC-MS/MS technology. Urine steroid LC-MS/MS measurements have been used exclusively as the basis of for anti-doping tests worldwide for the last few decades (3) and were required to have high level specificity. As a result this led to the recognition of the lower specificity of steroid immunoassays and resulted in major changes to clinical endocrinology research (3). Steroid mass spectrometry ensures the accuracy and specificity necessary for the measurement of low levels of E₂ and T and the detection of subtle changes in these gonadal steroids over time.

Anthropometric and serum hormone changes in our study revealed the anticipated increases over the course of one year in a cohort of young adolescents. The urine data also revealed anticipated hormone increases over the 12-months of observation. However, mean and individual urinary hormone changes were not strictly progressive, suggesting within-subject variability in early and mid-pubertal hormone levels, which may contain novel and hitherto unexploited information on determinants of biological aspects of pubertal progression. As these urine data are novel we are unable to compare with any previously published data, notably the degree of individual variability. Mouritsen and colleagues have shown that serum testosterone is variable in a longitudinal study of 20 adolescents (10 male) with sample collection biannually for five years (28). Individual hormone curves showed testosterone levels fluctuated during the study, though the overall pattern demonstrated an increase in serum hormones testosterone (as assayed by LC-MS/MS and immunoassay) from recruitment to the end of follow-up. Biro’s work in 252 females followed every six months for six years similarly demonstrated androstenedione, oestrone, E₂ and T by LC-MS, with changes which increased over
6. Urinary Sex Steroids and Anthropometric Markers of Puberty

the transition from pre-puberty to six months after reaching Tanner stage 2 (29); however, neither provide inter-individual variability or mean inter-individual hormone levels. Both previous studies measured only serum steroids in a single gender. We also identify the need for studies of longer duration in order to comment on what implications this variability (or what could be interpreted as instability) of hormone change may have on more distal responses to gonadal hormone change patterns in puberty be these physical, neurobehavioural or mood. This finding also suggests that three monthly urine collections over an extended period have the capacity to provide new insights into the biology of puberty.

We have shown considerable overlap between hormone levels at each age and self-reported Tanner stage, which emphasises the need for better descriptors of individual puberty hormone changes. Such overlap is consistent with other data using LC-MS/MS and clinician rating (30-32) and indicate the complexity and dynamic nature of puberty. Both Tanner stage and anthropometric change lag behind hormonal change. Using the former as surrogates for puberty hormone change will less accurately describe the relationship between hormone changes and the resultant physical changes, as well as any other adolescent health or developmental change of interest, such as mood or behaviour. More broadly still, this study has potential importance to the understanding of normal puberty because there is a paucity of frequently sampled longitudinal studies of the hormonal changes during puberty. This situation leads to over-interpretation of how adolescent mood, behaviour and wellbeing relate to puberty hormones, going well beyond available data (33, 34). Our study methodology with frequent sampling of an easily accessible biological fluid is likely provide new insights into the dynamics of pubertal hormonal changes,
especially in the tempo and stability of puberty change. These data in turn may allow an understanding of how puberty hormone change relates to the behavioural aspects of adolescence.

Previous work has questioned the validity of self-rated Tanner staging (1, 11-14). In this study a high proportion of adolescents completed self-assessment with findings of stable or an advance in Tanner stage at 12 months follow-up in all but 6% of adolescents (who went backwards in a Stage) Self-rated Tanner staging also corresponded well with conventional anthropometric measures of puberty change.

In conclusion, our work has established a feasible method for intensive urine sampling of community-dwelling adolescents and used a robust methodology of urine sex steroid hormone measurement, using liquid chromatography-tandem mass spectrometry measurements for urine sex steroids (3), which display the high sensitivity and specificity to detect lower levels of sex hormones, a particular challenge in the study of pubertal progression (6, 35). Based on previously recorded longitudinal growth data (36, 37), it is anticipated that frequently measured urine samples over the two to three year window of normal puberty might well provide a firmer biological basis for clinically observed patterns of puberty, such as rapid, slow or variable tempo, which may over the longer term support some of the observed differences in adolescent mood and behaviour.

**Acknowledgements**

We are grateful for the technical assistance of Ron Newman in assay development.
6.6. References


6. Urinary Sex Steroids and Anthropometric Markers of Puberty


7. SELF-RATED TANNER STAGE AND SUBJECTIVE MEASURES OF PUBERTY ARE ASSOCIATED WITH LONGITUDINAL GONADAL HORMONE CHANGES

Ben W.R. Balzer1,2,3, Frances L. Garden4,5, Mazen Amatoury2, Georgina M. Luscombe6, Karen Paxton6, Catherine I. Hawke6, David J. Handelsman7, Katharine S. Steinbeck1,2

Affiliations: 1 University of Sydney, Discipline of Child and Adolescent Health, Westmead, NSW, Australia; 2 Academic Department of Adolescent Medicine, The Children’s Hospital at Westmead, Westmead, NSW, Australia; 3 University of New South Wales, School of Women’s and Children’s Health, Randwick, NSW, Australia; 4 University of New South Wales, South-Western Sydney Clinical School, Liverpool, NSW, Australia; 5 Ingham Institute of Applied Medical Research, Liverpool, NSW, Australia; 6 University of Sydney, School of Rural Health, Orange and Dubbo, NSW, Australia; 7 ANZAC Research Institute, University of Sydney, Concord, NSW, Australia

7. Self-Rated Tanner Stage and Subjective Measures of Puberty

7.1. **Abstract**

**Introduction:** In large community-based studies of puberty, Tanner staging by a clinician is often not possible. We compared self-rated Tanner staging and other subjective ratings of pubertal development with serum hormone levels measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) to reassess the utility of self-rated pubertal stage using highly sensitive and specific hormone analysis.

**Methods:** Adolescents and their parents enrolled in the Adolescent Rural Cohort study of Hormones and Health, Education, Environments and Relationships (ARCHER) answered annual survey questions on pubertal development. Annually, adolescents provided blood samples for serum testosterone and oestradiol measured by LC-MS/MS.

**Results:** Longitudinally, self-rated Tanner stage was positively associated with serum testosterone and oestradiol levels in both sexes. Confirmation by adolescent and parent that puberty had commenced was associated with higher gonadal hormone levels in both sexes. Parent and adolescent responses demonstrated ‘fair’ to ‘moderate’ agreement.

**Conclusions:** Over a three-year follow-up, self-rated Tanner staging and simple questions regarding pubertal onset and development are positively associated with adolescent gonadal hormone concentrations in serum measured by mass spectrometry. Thus, self-report of puberty stage still has a role in large community-based studies where physical examination is not feasible.
7.2. Introduction

Identifying the onset and progression of puberty is a significant developmental milestone of interest to clinicians, adolescents and parents. This also has important public health significance given the known associations between pubertal timing and long-term health outcomes (1, 2). Multiple markers of puberty onset and progress have been used, such as Tanner stage, peak height velocity and circulating sex steroid levels (3). Tanner staging or similar physical change assessment by the Pubertal Development Scale (PDS) are the most widely used pubertal assessment methods (4, 5) and describe a continuum of change. These represent an assessment of secondary sex characteristic changes induced primarily by gonadal steroids. While direct physical examination remains the gold standard for Tanner staging, it is applicable to patients but not feasible in large community-based research of healthy children due to ethical and privacy concerns (4). Self-examination and/or comparison to images of Tanner stages offer a substitute method by which pubertal staging can be determined, removing the discomfort or concerns with physical examination clinical research involving healthy children.

The validity of self-rated Tanner staging for research is debated (4). It appears to differentiate children prior to and during puberty as assessed by adolescents or their parents (6-10) but findings using this method vary. As examples, in one longitudinal study (8), maternal and adolescent rating of development was slightly less sensitive and specific compared to clinician rating. In another cross-sectional cohort, self-assessment was inferior to clinician assessment, but still correlated with a composite hormone rating score derived from puberty hormone levels measured with radioimmunoassay (9). Two recent cross-sectional community-based studies from
one group have recently used clinician-based Tanner staging to deduce pubertal reference ranges for serum testosterone and oestradiol measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (11, 12). The present study is the first to assess longitudinally self-rated Tanner stage and other pubertal rating by adolescents and parents against mass spectrometry-measured circulating puberty hormones.

7.3. Methods

Subjects were enrolled in the Adolescent Rural Cohort study of Hormones and health, Education, environments and Relationships (ARCHER), a three-year longitudinal study of adolescents living in regional New South Wales, Australia investigating the effects of puberty hormone changes on mood, well-being and behaviour (13). Adolescents provided self-rated Tanner staging by comparison to line drawings each year. Females matched breast and pubic hair, whereas males matched genitalia and pubic hair. At each assessment visit over three years (four in total) adolescents (and parent or guardians) nominated whether they thought their puberty had started and indicated whether they felt their pubertal progress was behind, equal to, or more advanced than their age- and sex-matched peers. Female adolescents were asked if and when they had reached menarche at each time point (13). Surveys were piloted before use in the study.

Adolescents provided fasting, first-morning blood samples annually (around the time of questionnaire completion). Collection times were standardised only for the follicular phase of the menstrual cycle in post-menarcheal females. Serum testosterone and oestradiol were measured by LC-MS/MS (13), which is sensitive
7. Self-Rated Tanner Stage and Subjective Measures of Puberty

and specific for low circulating sex steroids in early puberty (14). Adolescents were classified as pre-pubertal if serum testosterone was < 0.5 nmol/L (males) or serum oestradiol < 40 pmol/L (females).

Means of pooled continuous data were compared by t-test or analysis of variance as appropriate. Where serum hormone levels were the outcome, the relationship with survey answers (Tanner stage, relative pubertal development compared to peers) was modelled using mixed-models linear regression with random intercepts for each individual. Where dichotomous outcomes were assessed (pubertal commencement), mixed-models logistic regression was used with random intercepts for each adolescent. For analysis of differing degrees of pubertal development, Tanner stage was dichotomised three ways: pre-pubertal (stage 1) vs early pubertal (stages 2-3); pre-pubertal (stage 1) vs late pubertal (stages 4-5) and pre-pubertal (stage 1) vs pubertal (stages 2-5). Hormone levels were logarithmically transformed due to skewed data so regression results are presented as odds ratios when back-transformed. Odds ratios were relative to Tanner stage 1 (i.e. pre-puberty) or a rating of ‘same’ compared to ‘behind’ or ‘ahead’ for relative development questions.

Cross-tabulations between survey questions and dichotomised Tanner stage outcomes or pubertal commencement (based upon serum hormones) were assessed using chi-squared ($\chi^2$) tests. Cochran’s Q test was used to assess if Tanner stage increased with time. Cohen’s weighted kappa ($\kappa$) was used to assess agreement between adolescent and parent questionnaire results and between these questionnaire results and dichotomised Tanner staging as well as pre-puberty vs puberty based on serum hormones. Interpretation of $\kappa$ scores was based on Landis’ criteria: ‘poor’ ($\kappa$ <0.20), ‘fair’ ($\kappa$ 0.21-0.40), ‘moderate’ ($\kappa$ 0.41-0.60), ‘substantial’ ($\kappa$ 0.61-0.80),
‘almost perfect’ (κ 0.81-1.00) (15). Sensitivity and specificity of the questionnaire results was calculated using the pre-pubertal definition above as reference standard. Statistical analyses were conducted using Stata 12.1 (StataCorp, College Station, Texas, USA) with significance set at p<0.05.

The ARCHER study had ethics approval from the Human Research Ethics Committee, University of Sydney (HREC 13094) consistent with the Declaration of Helsinki. Participants provided full informed written assent and their parents provided full written informed consent prior to commencing the study.

7.4. Results

There were 329 adolescents (185 male) who provided an initial survey response and 1193 survey responses were provided over the course of the study. Mean (standard deviation) age at baseline was 11.3 (1.0) years for males and 11.2 (1.0) years for females (p=0.43). Table 7.1 shows a progression through pubertal stages from baseline to three years follow-up. Most adolescents commenced puberty and self-rated higher pubertal stages over the duration of the study (p<0.001 for males and females). Based on the hormone cut off for pubertal onset, 57.0% (n=94/165) of males and 86.7% (n=111/128) females had commenced puberty at baseline and 95.9% of males and 92.8% of females had commenced puberty by three years’ follow up (p<0.001). Self-rated Tanner stage increased through the study (p<0.001 for males and females) and most adolescents (84.9% males and 75.9% females) self-rated as Tanner stages 4 or 5 at the end of the study, as shown in Figure 7.1 and Table 7.1.
Figure 7.1. Box plots of Tanner stages at each year by sex. Median (bold line) and dots indicate outliers.
Table 7.1. Self-rated Tanner stage, puberty progress and questionnaire responses.

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time from baseline (years)</td>
<td>Time from baseline (years)</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>185</td>
<td>169</td>
</tr>
<tr>
<td>Age (mean (SD))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.3 (1.0)</td>
<td>12.4 (1.1)</td>
</tr>
<tr>
<td>Age (median (IQR))</td>
<td>11 (11-12)</td>
<td>12 (11-13)</td>
</tr>
<tr>
<td>Serum hormone levels, (median (IQR))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testosterone (nmol/L)</td>
<td>1.0 (6.6)</td>
<td>8.5 (14.6)</td>
</tr>
<tr>
<td>Oestradiol (pmol/L)</td>
<td>40.4 (40.4)</td>
<td>62.4 (69.8)</td>
</tr>
<tr>
<td>Self-rated Tanner stage, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>26 (14.2)</td>
<td>13 (7.7)</td>
</tr>
<tr>
<td>2</td>
<td>63 (34.4)</td>
<td>38 (22.5)</td>
</tr>
<tr>
<td>3</td>
<td>54 (29.5)</td>
<td>42 (24.9)</td>
</tr>
<tr>
<td>4</td>
<td>27 (14.8)</td>
<td>53 (31.4)</td>
</tr>
<tr>
<td>5</td>
<td>13 (7.1)</td>
<td>23 (13.6)</td>
</tr>
<tr>
<td>Commenced Puberty, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adolescent report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>111 (60.3)</td>
<td>132 (80.0)</td>
</tr>
<tr>
<td>No</td>
<td>73 (39.7)</td>
<td>33 (20.0)</td>
</tr>
<tr>
<td>Parent/guardian report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>87 (47.5)</td>
<td>119 (71.3)</td>
</tr>
<tr>
<td>No</td>
<td>96 (52.5)</td>
<td>48 (28.7)</td>
</tr>
</tbody>
</table>
### Table 7.1. (Continued)

<table>
<thead>
<tr>
<th>Time from baseline (years)</th>
<th>Males</th>
<th></th>
<th></th>
<th></th>
<th>Females</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Post-menarche, n (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>35 (28.5)</td>
<td>58 (50.9)</td>
<td>79 (71.8)</td>
<td>89 (85.6)</td>
</tr>
<tr>
<td>No</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>88 (71.5)</td>
<td>56 (49.1)</td>
<td>31 (28.2)</td>
<td>15 (14.4)</td>
</tr>
<tr>
<td>Development Relative to Peers, n (%)&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adolescent report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behind</td>
<td>19 (10.3)</td>
<td>19 (11.2)</td>
<td>20 (12.1)</td>
<td>20 (12.8)</td>
<td>34 (23.9)</td>
<td>23 (17.4)</td>
<td>25 (20.0)</td>
<td>22 (18.5)</td>
</tr>
<tr>
<td>Same</td>
<td>131 (71.2)</td>
<td>123 (72.4)</td>
<td>120 (72.3)</td>
<td>105 (67.3)</td>
<td>83 (58.5)</td>
<td>91 (68.0)</td>
<td>80 (64.0)</td>
<td>78 (65.6)</td>
</tr>
<tr>
<td>Ahead</td>
<td>34 (18.5)</td>
<td>28 (16.5)</td>
<td>25 (16.7)</td>
<td>31 (19.9)</td>
<td>25 (17.6)</td>
<td>18 (13.6)</td>
<td>20 (16.0)</td>
<td>19 (16.0)</td>
</tr>
<tr>
<td>Parent/guardian report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behind</td>
<td>21 (11.5)</td>
<td>22 (13.2)</td>
<td>22 (13.3)</td>
<td>17 (11.5)</td>
<td>28 (19.7)</td>
<td>27 (20.9)</td>
<td>20 (15.9)</td>
<td>18 (17.1)</td>
</tr>
<tr>
<td>Same</td>
<td>138 (74.5)</td>
<td>119 (71.3)</td>
<td>114 (68.7)</td>
<td>105 (71.0)</td>
<td>88 (62.0)</td>
<td>83 (64.3)</td>
<td>87 (69.1)</td>
<td>72 (68.6)</td>
</tr>
<tr>
<td>Ahead</td>
<td>24 (13.1)</td>
<td>26 (15.6)</td>
<td>30 (18.1)</td>
<td>26 (17.8)</td>
<td>26 (18.3)</td>
<td>19 (13.6)</td>
<td>19 (15.1)</td>
<td>15 (14.3)</td>
</tr>
</tbody>
</table>

Note: some survey data are missing where participants or parents/guardians did not answer a question. Percentages are based on the total number of responses to a question.

<sup>a</sup> “Do you think you have/your child has started puberty?”

<sup>b</sup> Female adolescents asked if they have had their first period

<sup>c</sup> “Compared to other kids your (your child’s) age and sex, what is your (your child’s) stage of puberty?”
There were fifty-three occasions where adolescents rated their Tanner stage as lower than the previous year (of the 869 ratings performed after baseline – 6.1%). Thirty-one males and 21 females provided a lower rating on one occasion, and one male and one female provided lower ratings on two occasions. The most common decrease in stage was from stage 3 to stage 2 and stage 4 to stage 3 in females (eight instances each) and from stages 3 to 2 and 5 to 4 in males (eight instances each). One male rated himself stage 5 at baseline and stage 1 at one year follow up. There was no significant difference in age between adolescents who rated their Tanner stage as lower (mean (SD) 12.8 (1.4) years) compared to those who reported either no change or an increased stage (12.7 (1.5) years), p=0.77. When pooled across the study, no difference between mean testosterone or oestradiol levels was observed for adolescents with a lower subsequent rating at any point compared to those who reported no change or an increased stage, p=0.65 for males, p=0.36 for females.

Mean hormone levels with each Tanner stage, separated by gender, are shown in Figure 7.2. Overall, serum testosterone and oestradiol levels were strongly related to self-rated Tanner stage for both males and females (p<0.001). Table 7.2 provides the odds ratios of hormone level increase for each Tanner stage, relative to Tanner stage 1.
Figure 7.2. Mean serum testosterone (left) and oestradiol (right) with 95% confidence interval (bars) by self-rated Tanner staging in males and females.
Table 7.2. Odds ratio (95% confidence intervals) of hormone increase based on self-rated Tanner stage, relative to Tanner stage 1. For each Tanner stage, the odds ratio refers to the multiplicative factor increase in hormone levels relative to stage 1.

<table>
<thead>
<tr>
<th>Self-rated Tanner Stage</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Testosterone</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>2.31 (1.34-3.99)*</td>
<td>1.73 (1.40-2.15)*</td>
</tr>
<tr>
<td>3</td>
<td>4.47 (2.57-7.78)*</td>
<td>2.46 (1.97-3.07)*</td>
</tr>
<tr>
<td>4</td>
<td>22.53 (12.94-39.24)*</td>
<td>2.50 (1.99-3.14)*</td>
</tr>
<tr>
<td>5</td>
<td>30.47 (17.90-51.85)*</td>
<td>2.31 (1.80-2.96)*</td>
</tr>
<tr>
<td></td>
<td>Oestradiol</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>1.13 (0.85-1.51)</td>
<td>1.72 (1.29-2.28)*</td>
</tr>
<tr>
<td>3</td>
<td>1.56 (1.16-2.08)*</td>
<td>2.48 (1.86-3.31)*</td>
</tr>
<tr>
<td>4</td>
<td>2.19 (1.63-2.93)*</td>
<td>2.67 (1.99-3.59)*</td>
</tr>
<tr>
<td>5</td>
<td>2.29 (1.74-3.03)*</td>
<td>2.18 (1.58-2.99)*</td>
</tr>
</tbody>
</table>

*p<0.05 (compared to Tanner stage 1)
As shown in Table 7.1, the percentage of adolescents who answered that they commenced puberty increased over time for adolescent (males p<0.001; females p<0.001) and parent (males p<0.001; females p<0.001) surveys. The majority of adolescents (64.1% of females, 70.9% of males for all surveys) and their parents (65.7% female, 71.7% male for all surveys) felt that pubertal development was the same as their age- and sex-matched peers. There was no significant change in the proportions of adolescent or parent ratings of relative pubertal development (behind/same/ahead) over the three years’ follow-up (p>0.05 for both). However, males were more likely to rate their development as ahead of their peers and females more likely to rate as behind (p<0.001). There were no significant differences in age between the relative rating groups (p=0.42 for males, p=0.79 for females). Self-rated Tanner stage was higher in adolescents who rated their development as greater than their peers’ (mean stage (SD) males: 4.1 (1.1); females 3.9 (1.1); p<0.001), and the converse was seen for adolescents who rated their development as behind (males: 2.9 (1.1); females: 2.7 (1.2); p<0.001).

Table 7.3 provides the relative (odds ratio) hormone levels for adolescents who rated themselves ‘behind’ or ‘ahead’ of peers, relative to those who rated themselves as having ‘same’ pubertal development.
Table 7.3. Odds ratio (95% confidence interval) of hormone levels based on adolescents who rated their pubertal development as ‘behind’ or ‘ahead’ of their peers, relative to those who had ‘same’ pubertal development.

<table>
<thead>
<tr>
<th>Relative pubertal rating</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Testosterone</td>
<td></td>
</tr>
<tr>
<td>Behind</td>
<td>0.55 (0.31-0.96)*</td>
<td>0.79 (0.66-0.95)*</td>
</tr>
<tr>
<td>Same</td>
<td>0.88 (0.70-1.10)</td>
<td></td>
</tr>
<tr>
<td>Ahead</td>
<td>1.63 (1.04-2.56)*</td>
<td>0.98 (0.80-1.19)</td>
</tr>
<tr>
<td></td>
<td>Oestradiol</td>
<td></td>
</tr>
<tr>
<td>Behind</td>
<td>0.95 (0.73-1.24)</td>
<td>0.88 (0.70-1.10)</td>
</tr>
<tr>
<td>Same</td>
<td>0.88 (0.70-1.10)</td>
<td></td>
</tr>
<tr>
<td>Ahead</td>
<td>1.21 (1.00-1.47)</td>
<td>1.05 (0.82-1.34)</td>
</tr>
</tbody>
</table>

*p<0.05 (compared to ‘same’ rating)
There were 1170 parental survey responses over three years’ follow up, with 1040 (88.9%) completed by a female parent/guardian. Where a sex discrepancy existed (such as female parent/guardian rating male adolescent), parents/guardians were less likely to rate their child as having started puberty (77.5% for sex discrepancy compared to 82.4% for like-sex pairs; p=0.03). Similarly, they were more likely to rate their child’s relative development as ‘same’ or ‘ahead’ (88.0% compared to 80.8% for like-sex pairs; p=0.02).

The accuracy, sensitivity and specificity of the subjective markers of puberty onset and progress compared to the objective assessment (testosterone >0.5 nmol/L for males; oestradiol >40 pmol/L for females) are provided in Table 7.4. The ratings ranged between κ 0.13-0.66 (Table 7.4). All subjective markers demonstrated high sensitivity and poor specificity for objective puberty onset. Adolescents in early puberty (Tanner 2-3) showed ‘poor’ to ‘fair’ agreement (κ 0.10-0.29) with objective assessment of puberty onset, whereas adolescents in later puberty (Tanner 4-5) had ‘moderate’ to ‘substantial’ (κ 0.15-0.66) agreement based upon κ scores. Separating females by menarcheal status showed post-menarcheal females’ answers had greater sensitivity and specificity for survey responses compared to hormone levels, with better percent agreement compared to pre-menarcheal females. Agreement was ‘poor’ for survey answers whether females were post-menarcheal or not.
Table 7.4. Percent agreement, κ score, sensitivity and specificity of subjective markers (survey) of puberty compared with objective markers (hormone levels, males <0.5nmol/L and females < 40pmol/L).

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females (all)</th>
<th>Females (pre-menarche)</th>
<th>Females (post-menarche)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Commencement of Puberty (adolescent reported)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreement (%)</td>
<td>81.4</td>
<td>74.3</td>
<td>63.6</td>
<td>86.0</td>
</tr>
<tr>
<td>κ</td>
<td>0.45*</td>
<td>0.13*</td>
<td>0.05</td>
<td>0.13</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>90.3</td>
<td>78.2</td>
<td>89.6</td>
<td>95.1</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>52.3</td>
<td>41.4</td>
<td>14.6</td>
<td>15.4</td>
</tr>
<tr>
<td><strong>Commencement of Puberty (parent/guardian reported)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreement (%)</td>
<td>80.9</td>
<td>79.3</td>
<td>74.4</td>
<td>89.0</td>
</tr>
<tr>
<td>κ</td>
<td>0.52*</td>
<td>0.19*</td>
<td>0.11</td>
<td>0.19*</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>82.8</td>
<td>83.9</td>
<td>90.1</td>
<td>95.0</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>74.7</td>
<td>41.4</td>
<td>19.2</td>
<td>22.2</td>
</tr>
<tr>
<td><strong>Self-rated Tanner Stage: Pre-Puberty vs Early Puberty</strong>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreement (%)</td>
<td>60.4</td>
<td>80.0</td>
<td>77.9</td>
<td>90.5</td>
</tr>
<tr>
<td>κ</td>
<td>0.13*</td>
<td>0.26*</td>
<td>0.10</td>
<td>0.29*</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>60.0</td>
<td>91.0</td>
<td>88.8</td>
<td>97.4</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>63.0</td>
<td>32.3</td>
<td>20.0</td>
<td>25.0</td>
</tr>
<tr>
<td><strong>Self-rated Tanner Stage: Pre-Puberty vs Late Puberty</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreement (%)</td>
<td>93.1</td>
<td>80.4</td>
<td>64.9</td>
<td>89.6</td>
</tr>
<tr>
<td>κ</td>
<td>0.66*</td>
<td>0.32*</td>
<td>0.18</td>
<td>0.15</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>97.4</td>
<td>94.4</td>
<td>95.5</td>
<td>93.2</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>63.0</td>
<td>32.3</td>
<td>20.0</td>
<td>25.0</td>
</tr>
<tr>
<td><strong>Self-rated Tanner Stage: Pre-Puberty vs Puberty</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreement (%)</td>
<td>79.1</td>
<td>85.7</td>
<td>81.2</td>
<td>90.9</td>
</tr>
<tr>
<td>κ</td>
<td>0.22*</td>
<td>0.26*</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>96.6</td>
<td>91.4</td>
<td>90.2</td>
<td>94.6</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>19.8</td>
<td>35.7</td>
<td>20.0</td>
<td>25.0</td>
</tr>
</tbody>
</table>

*p<0.05

* Self-rated Tanner stage 1 vs 2-3
** Self-rated Tanner stage 1 vs 4-5
*** Self-rated Tanner stage 1 vs 2-5
7.5. Discussion

Our investigation has firstly shown that mean serum gonadal hormone levels, measured by LC-MS/MS, are strongly associated with increasing Tanner stage self-ratings by adolescent males and females. Secondly, serum gonadal hormone levels verify that both adolescents and their parents can identify puberty onset and progression with reasonable accuracy. Given that earlier and later puberty onset are both associated with impaired wellbeing in later adolescence and beyond (1, 2), simple subjective questions such as those used in this study may be useful to identify puberty onset. Certainly, in the context of large longitudinal community-based research studies of healthy children or where puberty may not be the main focus, these findings are reassuring. While we did not include text with Tanner line drawings (6, 7) because of varying age-related literacy, our cohort (aged between 9 and 14 years at baseline) demonstrated similar levels of agreement between objective (gonadal hormones) and subjective pubertal assessments to those statistics seen in an older cross-sectional cohorts with over 75% Tanner stage 4 using clinician compared to self-rated Tanner stage (10). This finding differs from a report that younger children are less accurate in self-rating their development compared to older peers (7).

Our longitudinal study investigated subjective measures of pubertal development and their associations with gonadal hormone levels measured by sensitive and specific LC-MS/MS, in order to further define the utility of self-rating tools in longitudinal adolescent health research. We have combined these data with additional information on pubertal progression, such as whether an adolescent or their parent felt they had started puberty and relative pubertal progress compared to age- and sex-matched
peers. Longitudinal regression showed a significant positive relationship between both serum hormones and self-rated Tanner stage in both males and females, consistent with findings reported by a cross-sectional reference-range study using LC-MS/MS and clinician-based Tanner staging (11, 12). While a positive relationship was seen between testosterone levels and male and female subjective ratings of pubertal commencement and progress relative to peers, a relationship was seen with oestradiol for males only. It is unclear why oestradiol did not have a significant relationship with relative pubertal progress in females.

Females generally had poorer agreement and lower κ scores between survey answers and hormone measures than males. In later puberty (Tanner stage 4 or 5) the sensitivity and specificity were increased in both sexes, indicating more accurate identification at later puberty. Nevertheless, agreement, sensitivity and specificity were largely unchanged for self-rated Tanner staging in earlier puberty, the exception being for males in early puberty who displayed lesser agreement and sensitivity. Collection in the early follicular stage, where oestradiol levels are relatively low, may explain the poorer agreement observed in females compared to males. Other explanations could be the variability in ovarian follicular activity even prior to menarche and menarche may be perceived by adolescents or parents as the ‘true’ marker of pubertal onset, rather than the late event it is. The latter explanation is supported by our analysis by menarcheal status; post-menarcheal females had better agreement with subjective pubertal questions than pre-menarcheal. In contrast, testosterone potentiates more obvious physical changes which act as clearer indicators of puberty in males.
The increasing demand for physical privacy in adolescence is the likely reason for the overall lack of agreement between parent and adolescent responses. This rather obvious explanation does however highlight the importance of who is asked questions about puberty development and using youth-friendly protocols and questions to engage adolescents in such research (16). Given the low number of fathers performing the annual surveys, we are unable to comment on the reporting effect of same sex versus opposite sex parent. We were however able to provide data on how uncommonly adolescents rate their Tanner stage as lower with follow-up, which is useful to refute criticism about the reliability of self-report. A similar number of males and females self-rated their pubertal development as lower than the preceding year, though no differences in age or hormone levels were observed to explain the lower ratings, which would be anticipated given adolescents cannot go ‘backwards’ in pubertal development. This may reflect better familiarity with the survey questions and Tanner stage diagrams though they were only surveyed annually, making such a memory effect less likely. It may also reflect a more mature understanding of their bodies.

Our study also found no difference in age between adolescents who rated their development as behind or ahead of their peers, though males were more likely to answer that they were ahead and females were more likely to describe themselves as behind their peers in development. This suggests that this type of subjective question may not be useful in the current format, or that numbers are too low to detect extremes of development. The reason for the discrepancy between males and females’ developmental ratings is unclear, though may be related to body image. Further, assumptions that menarche as the ‘true’ marker of puberty in females may
make pre-menarcheal females more likely to rate their development as behind their peers. Alternatively, differences in breast size (but not development) may be interpretation by females as indicative of their pubertal status and thus alter their perceptions. This discrepancy may also explain why the known difference in timing of puberty between boys and girls is less apparent than expected in the self-rating results.

Our study’s strengths include the large cohort of adolescents with variable progress through puberty, inclusion of parental perspectives about puberty, and the use of sensitive LC-MS/MS gonadal hormone assays to detect the low circulating hormone levels of early puberty, something of importance to our understanding of this life phase. The lack of any clinician-performed staging which was not appropriate in a community study, or indeed allowable by our Ethics Committee for non-clinical populations is a limitation. More females (86.7%) were pubertal at baseline (based on hormone values) than males (57.0%), which means recruitment did not adequately capture early puberty in many, which may also explain the differential results between males and females. Recruiting females at a younger age than males would have enabled us to better capture their pubertal onset, however, the ARCHER study’s overall focus was not the onset of puberty, but rather how puberty hormone changes influence adolescent health, mood and behaviour (13).

7.6. Conclusions

Longitudinal changes in self-rated Tanner stage are strongly associated with changes in circulating serum testosterone and oestradiol concentrations measured by LC-MS/MS. Adolescents and their parents can reasonably assess whether puberty has
commenced by using Tanner stage or simple yes/no questions. Our study is the first longitudinal study of self-rated Tanner stage against mass spectrometry-based measurement of puberty hormones. While self-rating is probably poor on an individual level, its utility at a population level has value in longitudinal research studies where clinician assessment and/or gonadal hormone measures are either not feasible or required, and especially where repeated measures are required.

7.7. References

7. Self-Rated Tanner Stage and Subjective Measures of Puberty


8. FOOT LENGTH GROWTH AS A NOVEL MARKER OF EARLY PUBERTY

Ben W.R. Balzer1,2,3, Hoi Lun Cheng1,2, Frances Garden4,5, Georgina M. Luscombe6, Karen T. Paxton6, Catherine I. Hawke6, David J. Handelsman7, Katharine S. Steinbeck1,2

Affiliations: 1 University of Sydney, Discipline of Child and Adolescent Health, Westmead, NSW, Australia; 2 Academic Department of Adolescent Medicine, The Children’s Hospital at Westmead, Westmead, NSW, Australia; 3 University of New South Wales, School of Women’s and Children’s Health, Randwick, NSW, Australia; 4 University of New South Wales, South-Western Sydney Clinical School, Liverpool, NSW, Australia; 5 Ingham Institute of Applied Medical Research, Liverpool, NSW, Australia; 6 University of Sydney, School of Rural Health, Orange and Dubbo, NSW, Australia; 7 ANZAC Research Institute, University of Sydney, Concord, NSW, Australia

8. Foot Length Growth as a Novel Marker of Early Puberty

8.1. Abstract

Foot growth is part of overall pubertal growth but its relation to other anthropometric and hormonal changes is unclear. Our objective was to determine how foot length changes relate to changes in other growth parameters (height and weight), Tanner stage and serum hormones. Adolescents (n=342) were recruited to a three-year longitudinal cohort study, underwent annual anthropometric assessments (height, weight, foot length) and provided self-rated Tanner staging. They also provided blood samples that were analysed using liquid chromatography-tandem mass spectrometry for serum testosterone and oestradiol and classified as pre-pubertal or pubertal based on circulating hormone levels. Average annual percent increase in foot length was greater for pre-pubertal adolescents compared to pubertal. Increased foot length was associated with increases in height, weight and Tanner stage and serum hormones in males and pre-menarcheal females but not post-menarcheal females. Foot length offers a novel, non-invasive, cost-effective and easily demonstrable marker of early pubertal changes.
8. Foot Length Growth as a Novel Marker of Early Puberty

8.2. Introduction

It is important to know when the pubertal transition is occurring because of the profound biological changes puberty brings (1). There are also known associations between the pubertal timing and adolescent psychosocial factors in both early- and late-maturing girls and boys compared to their on-time counterparts. These include emotional and behavioural problems (1) as well as poorer health outcomes in adulthood such as hypertension, type two diabetes mellitus, depression and anxiety (1-3).

Commonly used indicators of puberty all have limitations that restrict their interpretation. Tanner stage, the gold standard clinical assessment of pubertal development, requires undressing and trained clinicians (though self-rating can be used with reduced accuracy) (4, 5). Physical inspection may not be acceptable in non-clinical environments. The accuracy of developmental self-rating by adolescents may be slanted according to subjective factors such as body image and understanding. Peak height velocity requires serial calibrated measurements with sufficient frequency to determine a clear peak. Sex steroid measurement directly samples the hormonal milieu that drives the observed physical changes of puberty, but is limited due to acceptability of repeat biological sampling with adequate frequency (6) and the assay cost making this approach less useful for population-based assessment of puberty. Thus, a marker that is simple and easily measured, cost-effective, correlates with established markers of puberty (especially sex steroids) and which additionally may indicate the early stages of puberty is likely to be of significant research and clinical utility. Foot growth may be such a marker.
Current cross-sectional evidence suggests that change in foot length is an early pubertal event, preceding peak height growth (7, 8) and that peak change in foot length is related to the transition from Tanner stage 1 (pre-puberty) to Tanner stage 2 (9). Limited longitudinal data in girls suggest foot length change precedes peak height velocity (7). One retrospective study in women with large feet showed a negative correlation between age at menarche and final foot size (10). While none of these studies comprehensively examine the relationship of foot growth to other markers of growth and development in adolescence, the current literature suggests it could be a novel objective marker of early pubertal development. Importantly, there are no data on relationships with hormonal drivers of pubertal growth (7-9).

The aim of this study is to describe how longitudinal changes in foot length relate to changes in height, weight, Tanner stage and serum sex steroid levels (testosterone (T) and oestradiol (E2) in order to determine the utility of foot length change as a suitable marker of pubertal stage. This might offer a practical, novel and cost-effective marker of puberty, something of interest to adolescents, parents, researchers and clinicians alike.

8.3. Methods

The Adolescent Rural Cohort, Health, Hormones, Environment, Education and Relationships (ARCHER) study enrolled 342 healthy community-dwelling adolescents (153 females) and their parents (n=272; 239 mothers) living in Central Western New South Wales, Australia, specifically around the towns of Dubbo and Orange (11). The primary aim of the ARCHER study was to determine the attributable impact of puberty hormones on adolescent mood, behaviour and
wellbeing. Recruitment occurred through schools and the local media. Adolescents provided informed assent and their parents informed written consent. The ARCHER study has full ethics approval through the University of Sydney (HREC 13094) within the National Health and Medical Research Council Guidelines for Human Experimentation, which are consistent with the Declaration of Helsinki.

Participants were followed for three years, with annual clinic visits where adolescents underwent anthropometric assessment by trained staff and provided a self-rating of puberty using line drawings based on the Tanner stages (12, 13). Adolescents were contacted annually to arrange clinic follow-up (11). Females provided self-rating of breast stage and males provided self-rating of genital stage. Females were also asked if they had undergone menarche and, if so, when. Adolescents provided annual blood samples. Foot length (to 0.1 cm) was measured using a Ritz stick foot measure (Footer, Wisconsin, USA), modified with a millimetre-graduated measuring tape. The foot measure was placed under the sole of the foot and length measured from the heel to the end of the distal phalanx of the hallux. The right foot was measured unless this was inaccessible, in which case the left foot was used (3 of 1316 measurements, in three separate adolescents). Height (to 0.1 cm) was measured using a portable stadiometer (Seca GmBH, Hamburg, Germany). Weight (to 0.1 kg) was measured in light clothing using a Tanita TBF-300 Pro Body Composition Analyzer (Tanita, Tokyo, Japan).

8.3.1. Hormone Assays

Blood samples collected by venepuncture at a study collection site were frozen at -80 °C until assay in a single batch. Early morning serum E2 (collected in follicular phase
for post-menarcheal participants) and T were measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using previously described methods (6, 14). Briefly, organic solvent extracts of serum were separated by liquid chromatography using a Shimadzu Nexera UHPLC system (Shimadzu Scientific Instruments, Maryland, USA). Following this, tandem-mass spectrometry analysis was performed using an API-5000 triple-quadrupole mass spectrometer (Applied Biosystem/MDS SCIEX, Ontario, Canada) (6, 14). This method can quantify low levels of circulating E₂ and T in 99% of serum specimens from adolescents (6).

8.3.2. Statistical Analysis

Anthropometric and hormone outcomes by sex were summarised over time using means and standard deviations (SD). A t-test was used to assess the differences in age at baseline between the sexes. Adolescents were classified as being ‘pre-pubertal’ if their baseline serum T was less than 14 ng/dL (0.5 nmol/L) for males or serum E₂ less than 10 pg/mL (40 pmol/L) for females based on cross-sectional data using LC-MS/MS and clinician-rated Tanner stage (15, 16). Levels above these cut-offs were defined as ‘pubertal’. Differences in the mean annual percent change in foot length, height and weight over the study between pre-pubertal and pubertal males and females were analysed using t-tests. Similarly, differences in the mean percent change in foot length, height and weight over the study between pre- and post-menarcheal females, were analysed using t-tests. Mixed-model linear regression was used to assess the relationship between longitudinal changes in the exposure variable foot length with changes in outcome variables height, weight and Tanner stage, by sex. Mixed-model linear regression was also used to assess the relationship between changes in the outcome variables height, weight, foot length and Tanner
stage, respectively, with changes in the exposure variables serum hormone levels, by sex and menarche status in females. Random intercepts for each subject were used in all mixed models to account for the repeated measures per subject. Analyses were conducted using Stata 12.1 (StataCorp, Texas, USA) and statistical significance was set at 0.05.

8.4. Results

8.4.1. Cohort Characteristics

There was no difference in age between sexes: mean age (SD) was 11.7 (1.0) years for females and 11.8 (1.0) years for males (p=0.27) at baseline. Table 8.1 describes the basic demographics and anthropometric characteristics for each year of the study. Serum specimens were provided by 293 adolescents at recruitment, with 17 of 128 (13.3%) females and 71 of 165 (43.0%) males classified as pre-pubertal at baseline by their serum E₂ and T levels, respectively.

8.4.2. Anthropometry

Foot length, height, weight, and self-rated Tanner stage, all increased over the study period (Table 8.1) as did the prevalence of menarche. There were significant differences in mean percent annual change in foot length, height, and weight between adolescents who were pre-pubertal at baseline and those who were already pubertal (Table 8.2).

Mean annual percent change in foot length over the course of the study was significantly greater if the adolescent was pre-pubertal at baseline (females: 2.7% vs 1.5% p<0.001; males: 4.2% vs 2.9%, p<0.001; see Figure 8.1). Overall percent foot
growth (from baseline to final follow-up) was greater if the adolescent was pre-pubertal at baseline (females 8.4% vs 5.0% p<0.001; males 12.5% vs 8.8% p<0.001). Pre-menarcheal females had greater mean percent foot length, height and weight change than girls who were post-menarcheal at baseline (foot length: 1.8% vs 0.9% p<0.01; height: 3.3% vs 2.1% p<0.001; weight: 12.1% vs 8.7% p<0.001).
Figure 8.1. Mean annual % change in foot length over three years’ follow up, separated by baseline pubertal status.
8. Foot Length Growth as a Novel Marker of Early Puberty

The longitudinal changes in foot length were associated with statistically significant increases in height, weight and Tanner stage. On average, a one-centimetre increase in foot length was associated with a 6.7 (95% confidence interval (CI) 6.5-6.9) cm increase in height and 6.2 kg (95% CI 5.9-6.5) weight gain in males, and the respective changes for females were 5.8 cm (95% CI 5.3-6.2) height growth and 6.6 kg (95% CI 5.9-7.4) weight gain (p<0.001 for all analyses).

8.4.3. Hormone Measures

Mean serum hormone levels increased for both males and females over the study period. Serum T increased from mean (SD) 121.0 (175.8) ng/dL (4.2 (6.1) nmol/L) at baseline to 279.5 (129.7) ng/dL (9.7 (4.5) nmol/L) at three years in males. Serum E2 was mean (SD) 39 (28.8) pg/mL (144 (105.8) pmol/L) at baseline and 37 (21.3) pg/mL (134 (78.1) pmol/L) at three years for females. Positive longitudinal relationships were seen for changes in foot length and changes in serum T and E2 for males and females (except post-menarcheal females) (p<0.001 for all). Self-rated Tanner stage progression was also positively associated with changes in serum T and E2 for both sexes (p<0.001). These results are shown in Table 8.3. In pre-menarcheal females, serum T and E2 were significantly positively associated with increasing foot length and increasing self-rated Tanner stage.
Table 8.1. Participant anthropometry, self-rated Tanner stage and menarcheal status

|                      | Males |           |           |           |           |           |           |           |
|----------------------|-------|-----------|-----------|-----------|-----------|-----------|-----------|
|                      | Time from baseline (years) | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 |
| N                    |       | 185 | 169 | 166 | 158 | 142 | 133 | 126 | 120 |
| Age (mean (SD))      |       | 11.3 (1.0) | 12.4 (1.1) | 13.4 (1.1) | 14.3 (1.1) | 11.2 (1.0) | 12.2 (1.0) | 13.2 (01.0) | 14.3 (1.1) |
| Anthropometry (mean (SD)) |       |   |   |   |   |   |   |   |
| Foot length (cm)     |       | 23.8 (1.7) | 24.7 (1.7) | 25.6 (1.5) | 26.2 (1.4) | 23.1 (1.4) | 23.5 (1.2) | 23.9 (1.2) | 24.1 (1.1) |
| Height (cm)          |       | 151.9 (10.0) | 158.9 (10.8) | 165.4 (10.4) | 171.0 (9.2) | 151.0 (8.5) | 156.7 (7.8) | 161.1 (6.9) | 163.6 (6.3) |
| Weight (kg)          |       | 44.9 (12.2) | 50.6 (13.4) | 57.1 (14.6) | 63.3 (15.8) | 46.7 (12.9) | 51.8 (13.1) | 57.0 (13.4) | 60.3 (13.2) |
| Self-rated Tanner stage, n (%) |       |   |   |   |   |   |   |   |
| 1                    |       | 26 (14.2) | 13 (7.7) | 1 (0.6) | 1 (0.6) | 28 (19.9) | 9 (6.8) | 2 (1.6) | 0 (0) |
| 2                    |       | 63 (34.4) | 38 (22.5) | 26 (15.8) | 8 (5.1) | 50 (35.5) | 29 (21.8) | 12 (9.7) | 3 (2.5) |
| 3                    |       | 54 (29.5) | 42 (24.9) | 29 (17.6) | 15 (9.5) | 38 (27.0) | 49 (36.8) | 38 (30.7) | 26 (21.7) |
| 4                    |       | 27 (14.8) | 53 (31.4) | 53 (32.1) | 32 (20.3) | 20 (14.2) | 31 (23.3) | 44 (35.4) | 41 (34.2) |
| 5                    |       | 13 (7.1) | 23 (13.6) | 56 (33.9) | 102 (64.6) | 5 (3.6) | 15 (11.2) | 28 (22.3) | 50 (41.7) |
| Post-menarche, n (%)  |       |   |   |   |   |   |   |   |
| Yes                  |       |   |   |   |   |   |   |   |
| No                   |       |   |   |   |   |   |   |   |

\[a\] Menarche based upon participant survey response
Table 8.2. Differences in anthropometry for males and females who were pre-pubertal at baseline compared to pubertal \(^a\), and pre- and post-menarcheal \(^b\) (at any timepoint) females.

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Pre-Pubertal</td>
<td>Pubertal</td>
<td>p-value</td>
<td>Pre-Pubertal</td>
<td>Pubertal</td>
<td>p-value</td>
</tr>
<tr>
<td>Change in foot length (%)</td>
<td>4.2 (2.0)</td>
<td>2.9 (1.6)</td>
<td>&lt;0.001</td>
<td>2.7 (1.1)</td>
<td>1.5 (1.7)</td>
<td>&lt;0.001</td>
<td>2.1 (1.6)</td>
</tr>
<tr>
<td>Change in height (%)</td>
<td>4.1 (0.8)</td>
<td>4.2 (1.5)</td>
<td>0.39</td>
<td>3.9 (0.6)</td>
<td>2.9 (1.5)</td>
<td>&lt;0.001</td>
<td>3.6 (1.0)</td>
</tr>
<tr>
<td>Change in weight (%)</td>
<td>12.3 (3.2)</td>
<td>12.9 (4.9)</td>
<td>0.06</td>
<td>13.7 (3.8)</td>
<td>10.6 (4.7)</td>
<td>&lt;0.001</td>
<td>12.6 (3.9)</td>
</tr>
</tbody>
</table>

\(^a\) Pre-puberty defined as serum T<14ng/dL (0.5nmol/L) for males; serum E2<10pg/mL (40pmol/L) for females

\(^b\) Menarche based on self-report

Table 8.3. Mixed-model regression analyses of longitudinal changes in anthropometry and Tanner stage with changes in serum hormones.

<table>
<thead>
<tr>
<th>β (95% CI)</th>
<th>Foot length (cm)</th>
<th>Height (cm)</th>
<th>Weight (kg)</th>
<th>Tanner stage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testosterone</td>
<td>0.02 (0.01-0.04)*</td>
<td>0.05 (0.04-0.05)*</td>
<td>0.04 (0.03-0.04)*</td>
<td>0.005 (0.004-0.005)*</td>
</tr>
<tr>
<td>Oestradiol</td>
<td>0.06 (0.04-0.08)*</td>
<td>0.08 (0.04-0.1)*</td>
<td>0.4 (0.3-0.6)*</td>
<td>0.06 (0.04-0.07)*</td>
</tr>
</tbody>
</table>

**Females (all)**

| Testosterone | 0.006 (0.005-0.007)* | 0.2 (0.1-0.3)* | 0.2 (0.06-0.3)* | 0.04 (0.03-0.06)* |
| Oestradiol   | 0.01 (0.005-0.02)*  | 0.5 (0.4-0.7)* | 0.03 (-0.02-0.08) | 0.01 (0.006-0.02)* |

**Females (pre-menarcheal)**

| Testosterone | 0.03 (0.003-0.05)* | 0.2 (-0.004-0.3) | 0.2 (-0.02-0.4) | 0.04 (0.03-0.06)* |
| Oestradiol   | 0.01 (0.003-0.02)* | 0.08 (0.01-0.1)* | 0.03 (-0.05-0.1) | 0.009 (0.001-0.02)* |

**Females (post-menarcheal)**

| Testosterone | 0.007 (-0.006-0.02) | 0.00007 (-0.1-0.1) | -0.05 (-0.2-0.1) | -0.002 (-0.02-0.02) |
| Oestradiol   | 0.002 (-0.003-0.007) | -0.02 (-0.07-0.03) | -0.07 (-0.1-0.004) | -0.003 (-0.01-0.004) |

*p<0.05
8. Foot Length Growth as a Novel Marker of Early Puberty

8.5. Discussion

This study demonstrates a positive longitudinal relationship between foot length change and pubertal hormone change which provides a biological basis for prior findings of a temporal relationship between foot length change and other pubertal anthropometric changes such as height velocity or Tanner stage (7-9). In earlier studies on foot growth (7-10), peak height velocity (PHV), Tanner staging and menarche timing have all been used as surrogates for pubertal hormone changes. In this study we further characterised this relationship, first, by showing that the longitudinal relationships between foot length change and other anthropometric changes are more marked in the adolescents who were younger at baseline. Secondly, a significant longitudinal relationship exists between gonadal hormone changes and foot length changes.

Mean percent change in foot length was greater in pre-pubertal than pubertal adolescents in this study indicating this variable may be a useful indicator of early pubertal stage. Additionally, foot growth was consistently associated with serum hormones in pre-menarcheal but not post-menarcheal females. Both these observations combined with anthropometry and Tanner stage data support the concept that an increase in foot length velocity is an indicator of early puberty with increases in foot growth indicating that puberty has started. These findings likely come as no surprise to parents, as increasing frequency of changes in shoe size necessitates the more frequent purchase of larger and more costly shoes – concrete indicators more easily determined or recalled than the breast change or the appearance of body hair. The latter two features are visually subtle and often hidden
by adolescent privacy concerns. Foot length changes are both memorable and open (8).

In our cohort, both male and female foot growth was greater in pre-pubertal compared to pubertal adolescents. In comparison, height and weight demonstrated different sex-based patterns. From an evolutionary viewpoint, humans as exclusively bipedal mammals need a stable platform to support the somatic growth spurt of puberty; foot growth preceding height growth would naturally facilitate this.

Assessing foot length in the clinical setting is a low-cost, non-invasive measure, and as shoe sizes are directly proportionate to foot length, asking about shoe size changes is a readily accessible and non-intrusive measure. Like all measures the advantages and limitations of foot measurement need to be recognised. Shoe sizes are linearly related to foot length – there is an 8.47 mm (one third of an inch) increase in length for each full US shoe size. This utility is reflected in shoe sales data which show increases in the time preceding age at peak height velocity, based on population data (8). The practical applications of our results would be that serial increase in foot length or asking parents or adolescents about increasing shoe sizes may indicate the earliest phases of the pubertal transition and could offer a useful screening question. Further, in adolescents for whom accurate height assessment is not possible or population-based non-clinical studies using recalled development, serial foot measurements may be a convenient surrogate for other anthropometric assessment. While the Brannock device has been used in other studies (7, 9), it is only accurate to the nearest half shoe size (i.e. 4.3 mm or one sixth of an inch) so did not offer sufficient precision for our investigation. We were interested in how foot length changed, as opposed to shoe size. However, shoe size offers a practical correlate for
parents and adolescents. The translational implication is that regardless of manufacturers’ producing differing lengths for the same advertised size, a history of increasing shoe purchase frequency due to growing feet would sufficiently demonstrate increasing foot length and therefore suggest early pubertal changes. Additionally, foot length may be a useful measure in children for whom height measurement is not possible.

The strengths of this study are its longitudinal design with a large cohort given the intensity of follow-up and the use of liquid chromatography-tandem mass spectrometry which is capable of measuring low levels of circulating sex steroids in early puberty (6). The annual anthropometry is a limitation, as more frequent measures would have provided further detail about foot growth velocity. Recent work has shown that, using the Super-Imposition by Translation And Rotation (SITAR) model, annual anthropometric measurements (such as those performed in our study) of sufficient duration are precise enough to estimate peak growth velocity during puberty (17). However, with only three years of sampling we did not have the requisite data to confirm that peak height velocity was preceded by peak foot growth velocity as described in Ford et al (2009) (7). As expected, due to the earlier puberty in females, more girls were pubertal at baseline compared to males. Some of these girls may have experienced epiphyseal fusion by the end of the study, with cessation of all bone growth, which may explain the decline in growth velocities being more than for males. As blood collection was in the early follicular phase for females, we observed consistent oestradiol levels over the course of the study.

In conclusion, we provide longitudinal evidence that the pubertal increase in foot length is related to growth (height and weight) as well as secondary sex characteristic
changes (self-rated Tanner stage) and serum sex steroids. Foot growth is greatest in adolescents with the lowest gonadal hormone levels and is therefore a practical marker of other early pubertal changes.

8.6. References

9. THESIS DISCUSSION

9.1. Key Findings

9.1.1. Introduction

The primary aim of this thesis was to demonstrate novel ways of studying puberty using preliminary data from the longitudinal ARCHER study. Chapters 3 and 4 reviewed the current evidence base for oestradiol’s effect on mood and behaviour in female adolescents and testosterone’s effect on male adolescent mood and behaviour respectively. The feasibility of the ARCHER study’s methodology with respect to specimen collection and analysis was considered (Chapters 5 and 6, respectively), with Chapter 6 also investigating the relationship between hormone changes and other anthropometric markers of puberty. Self-rated Tanner staging, a research tool for pubertal assessment, has been used in the ARCHER study, and its usefulness in this context was investigated in Chapter 7. The use of foot length growth as a novel marker of early puberty was demonstrated in Chapter 8. Each of these studies has demonstrated that it is both possible and necessary to perform intensive, complex longitudinal studies in adolescent research. There are important gaps in the evidence as to how the dynamic changes occurring during puberty over a relatively short timeframe affect adolescent health and well-being, although this latter question has not been examined in this thesis. This thesis’ studies have also identified ways we can identify early puberty and the means by which cost-, time- and labour-efficient research can be conducted in this important and under-researched population.
Here, the preceding chapters will be briefly summarised to provide an interpretation and discussion of their results in the greater context of the thesis. The limitations of the work will be discussed, and future research directions identified by this study will be considered.

9.1.2. The Effects of Oestradiol on Mood and Behaviour in Human Female Adolescents and the Effects of Testosterone on Mood and Behaviour in Human Male Adolescents

The systematic reviews presented in Chapters 3 and 4 summarised and synthesised the evidence base for the effects of oestradiol (Chapter 3) and testosterone (Chapter 4) on female and male adolescents, respectively. We found that despite there being long-standing assumptions regarding sex hormone effects on female adolescents, the evidence to support these was not convincing. Nine studies on mood and affect in adolescent females were identified, with one longitudinal study showing an approximately linear relationship between oestradiol and depression (1), though cross-sectional data were less consistent. There was little consistency in studies regarding self-image and related behaviours with oestradiol, though a large Belgian study identified a positive correlation between oestradiol levels and risk taking (2).

There were 27 studies focusing on testosterone’s effects on male adolescents. The majority of these were cross-sectional, and aggression was generally the outcome of interest. While one longitudinal study failed to demonstrate a clear relationship between testosterone levels and changes in aggression, several cross-sectional studies showed significant correlations between testosterone levels and conduct disorder, risk-taking and aggression. Similar to oestradiol, some of the reviewed studies
identified relationships between testosterone and anxiety or depression, though the relationship between testosterone and substance use was only demonstrated in one study. As with the oestradiol systematic review, issues with cross-sectional analyses and lack of consistency between studies limited the ability of this review to demonstrate a clear causal relationship between changes in testosterone and changes in mood and behaviour in male adolescents. Of note, most of these studies used the less reliable method of immunoassay to assess sex steroid levels, with our study moving to the much more reliable (especially at low levels) LC-MS/MS technique.

As there are demonstrable effects of oestradiol on the adolescent brain (3) and in animal studies (4), one might anticipate these data to be translated in clinical research. We were surprised that there was little consistency in studies with positive results, though it is likely that confounding factors, failure to control for menstrual cycle, social factors, and the quality of oestradiol or testosterone assays played a role in this. For example, no studies used gold standard LC-MS/MS assays, which can detect the low levels of circulating oestradiol or testosterone in early puberty, as demonstrated later in Chapter 6.

Our findings in these chapters will enable clinicians, policymakers, parents and adolescents alike, for whom the increasing awareness of mental health issues in the adolescent population indicates the need for greater research. Despite popular assumptions that mood and behaviour changes in puberty are a consequence of “raging hormones” or are simply “puberty blues”, we have demonstrated little systematic evidence for this conclusion. One of the main goals of the ARCHER study is to provide hormone data (using high quality LC-MS/MS assays) over three years of puberty. This, combined with several validated survey tools (Youth Self

206
Report/Child Behaviour Checklist and the Short Moods and Feelings Questionnaire), will provide longitudinal evidence for the true and attributable role of sex hormones in adolescent mood and behaviour. Surveys have also been designed to ask adolescents and their parents about a variety of other factors which may act as confounders or interaction factors, such as demographic data, education, socio-economic status and sleep patterns (5). Accordingly, the ARCHER study data will provide more precise estimates than offered in the studies reviewed in Chapters 3 and 4. We would anticipate that oestradiol and testosterone may contribute to a certain degree to mood and behavioural changes at adolescence, however it is likely that a variety of other environmental factors will play a larger role. For example, Brooks-Gunn identified 1% of variance in mood in their study was due to oestradiol level (6), highlighting the importance of other factors. Since the publication of these chapters, no work has been performed on a longitudinal cohort of sufficient size and intensity of hormone sampling to demonstrate whether a relationship truly exists between pubertal hormones and mood and behaviour in adolescence. The Avon Longitudinal Study of Parents and Children, a large longitudinal UK birth cohort (n=15656 children recruited), involves annual follow-up and will likely provide supportive data for future findings from the ARCHER study.

9.1.3. Text Messaging is a Useful Research Tool

Text messaging and mobile phones are nearly ubiquitous in Australia, particularly amongst adolescents (7). Indeed, it is now the most used communication tool globally (8, 9). Therefore, it is something to which adolescents are accustomed and can be effectively utilised by them and their parents. Text messaging has been used effectively in many clinical contexts, such as appointment reminders (10), and for
data collection (8, 11-15). Thus, our text message, requesting a self-rating of mood at
the time a urine specimen was due for collection, was intended to utilise the benefits
seen for researchers using text messaging for both goals. We were able to achieve a
very high level of both specimen collection (primary aim; 97.3% of intended
samples) and a high rate of text message response (89.1% response rate), with a
strong correlation (r=0.606) and approximately linear relationship between these (see
Figure 5.1).

The Framingham Children’s Study (16), a longitudinal study in the United States
(n=187), identified several aspects that benefitted their cohort’s retention and study
compliance. These included flexibility in scheduling, reminders and minimal
response burden (16). Our SMS reminder tool provides participants with a non-
intrusive reminder that also allows data collection and is cost-effective. This
reminder tool has translated to a low level of attrition (2.6% per annum) compared to
other longitudinal adolescent cohorts with intensive data collection (such as Koenis’
study of 224 adolescent twins, which had an attrition rate of 6.8% per annum (17)).
Further, the SMS reminder tool was associated with a majority of specimens
collected. This is especially important in the context of the ARCHER study, whose
recruitment catchment is spread over 45000 km², meaning effective, timely
communication is important.

The practical implication of this study is that text messaging can be feasibly
employed to enhance compliance with a study protocol (18). While there was no
control arm to demonstrate text messaging alone was the cause for this, our results
show it is an economical and efficacious reminder and data collection tool. Further
work by the ARCHER study group will identify how these SMS mood scores could
act as a marker for mood disorders such as depression or anxiety, which would then offer an adolescent-friendly, easily accessible and affordable tool for clinicians to detect these disorders. Future research could benefit from this finding by implementing text messaging for collection of other data from study participants through the prompting question.

9.1.4. **Urinary Sex Steroids and Anthropometric Markers of Puberty – A Novel Approach to Describing Puberty**

As in Chapter 5, this chapter demonstrated the feasibility of part of the ARCHER study’s methodology, namely whether LC-MS/MS could determine hormone levels in pre-pubertal fasting urine and serum samples and thus provide an earlier indication of pubertal progress than previous assay techniques have allowed. Further, we demonstrated the relationship between hormone measures and anthropometric markers of development such as height, weight and self-rated Tanner stage.

Our methodology, which is more frequent and intensive than other recent work with a large study cohort, will provide the most comprehensive evidence for the effects of pubertal hormone changes on the adolescent than has been previously available. For example, the Copenhagen Puberty Study collected serum samples for LC-MS/MS hormone assays from twenty adolescents every six months for five years (19), whereas ARCHER followed 342 adolescents every three months for three years (5). In so doing, the ARCHER study aims to identify the role hormones play longitudinally in modifying growth, mood, behaviour and other aspects of adolescent health.
Mouritsen’s work in the Copenhagen Puberty Study also used LC-MS/MS and showed it to be superior to immunoassay in measuring serum testosterone levels in early adolescence. Our LC-MS/MS serum assay was unable to quantify testosterone in only two (1%) of our first 104 participants and luteinising hormone levels could not be detected by immunoassay in eleven specimens (of 194 collected; 6%). Similarly, only 28 of 484 urine collections (6%) were below the limits of quantification for any hormone (23 were oestradiol or testosterone LC-MS/MS assays). Given an early-pubertal study cohort, our methodology has demonstrated, as with Mouritsen’s, that LC-MS/MS able to quantify most hormone levels at this life stage, which we could not previously determine with sufficient precision. Our data have added to this by including LC-MS/MS data on oestradiol and further demonstrated the utility of this assay methodology on urine assays for the same hormones. First-morning urine samples can be feasibly collected (Chapter 5) and are intuitively more acceptable than repeated venepuncture and blood collection for adolescents and their parents. Collection time is also important given the initially nocturnal pulsatile secretion of pubertal hormones in early puberty (20, 21).

This work showed urine and serum hormone levels are correlated, and also identified that urine hormone changes were not strictly progressive, instead fluctuate, suggesting a greater level of complexity to the pubertal transition than previously known. The top panel of Figure 6.1 showed this, and it is not clear why hormone levels demonstrate this pattern. It should be noted that there were overall increases in urinary hormones over one year, however the apparent mid-year peak cannot be adequately interpreted until the final analyses of all ARCHER study urine samples. We could not demonstrate patterns for serum hormones as these are collected
annually and this Chapter only looked at one year of collection. Interestingly, Figure 6.2 showed that there were increases in hormones at each self-rated Tanner stage, with the exception of urine and serum luteinising hormone that showed less clear progression and for females there was an apparent decrease in luteinising hormone between Tanner stage 4 and Tanner stage 5 which we cannot explain. Though both sexes showed increases in sex hormones over each Tanner stage, females demonstrated greater increases in oestradiol and males in testosterone, as anticipated physiologically. As these are preliminary data, one must bear in mind the significant limitation this poses upon the interpretation of these findings. Our data follow anticipated trends and identify new ones. However, these data are from approximately one-third of the cohort over one year (of three) for the ARCHER study. Therefore, it is not possible to say whether such trends will continue for the larger cohort or in their longer-term follow-up. However, we do demonstrate the feasibility of the methodology, which has implications for the larger study and provides promising data suggesting the success of the ARCHER study as a whole in achieving its aims.

9.1.5. Adolescent and Parental Ratings of Puberty: Associations with Serum Hormones

The use of clinician-rated Tanner staging may pose ethical issues in studies involving community (non-clinical) adolescents, as it requires exposing their genitalia. The logistical issues requiring trained examiners added further challenges given the wide catchment area of the ARCHER study, and subjecting adolescents to this examination may have further deterred them from participating in ongoing research. For this reason, the ARCHER study relied on self-rating of Tanner stage and other
more basic assessments (such as “Do you think you have started puberty?”). In this chapter, we compared adolescent self-rating and adolescent and parent reports with LC-MS/MS assays of serum testosterone and oestradiol. We found significant relationships between self-rated Tanner staging and serum hormones and that those who affirmed the onset of puberty (or parental report of same) had significantly higher hormone levels than those who said puberty had not commenced. To this end, we have shown what Rasmussen and colleagues identified in their work (22) – self-rated Tanner staging is useful in a broad epidemiological context where clinician assessment and/or gonadal hormone measures are not feasible, though may not be useful in clinical situations requiring accurate assessment (where clinician assessment is the gold standard). We have demonstrated that adolescents and their parents are able to assess whether puberty has commenced to a reasonable degree. Asking parents and adolescents whether puberty has commenced may thus offer a useful screening tool in non-clinical situations. However, it remains unclear whether an adolescent or parent can assess puberty progress relative to other age- and sex-matched adolescents.

We also identified that statistical agreement between adolescent and parent report is generally poor. One of the principal reasons for this is that adolescents may become more private about their bodies, and thus less likely to divulge changes or allow parental inspection. Despite this, both parental and adolescent reporting was associated with hormone changes. Additionally, more adolescent males said they had started puberty compared to their parents, and opposite trends were observed for females. This is an interesting finding, that may relate to the clinical manifestations of pubertal development. For males, events such as spermarche and penile
development may be seen as developmental milestones, but these are often kept private. In contrast, female breast development and menarche may be less so, due to the visibility of the former and salience of the latter, both necessitating purchases (bras or hygiene products). However, it is unclear why female adolescents are less likely to report the commencement of puberty than their parents given this. It is possible that parents base their observation on behavioural change and infer this change is hormone related.

Without clinician assessment (gold standard of Tanner staging) we are unable to determine the accuracy of self-rating in this cohort, so our work relies on previous findings that self-rating has an adequate accuracy to be used as a tool in epidemiological research (22). Confidence intervals for hormone levels at each self-rated stage fall within clinician-rated Tanner stage reference ranges established for LC-MS/MS (23, 24). We did not provide adolescents or parents with any information about signs of puberty by which they could answer questions which creates inherent issues in the definition of “pubertal onset”, which would likely differ between individuals in this context. Additionally, we asked adolescents and parents whether they felt puberty has commenced and to rank their development as behind, same or ahead of their peers. As discussed in Chapter 1, the subjective assessment of puberty by an adolescent is an important piece of data, regardless of accuracy, as it reflects the individual’s self-perception (25). The perceived maturity may therefore lead to the adolescent behaving in ways they feel more congruous with their idealised development, as opposed to the clinician rating or neuro-endocrine drivers of these changes. Moore and colleagues found that perceived pubertal timing was important in predicting engagement in sexual and romantic behaviour whereas objective
pubertal timing (age at menarche) only predicted the onset of sexual behaviour (26). The adolescent’s (and their parent’s) perception is therefore critical to understanding some of the changes observed in pubertal behaviour. When using this frame of reference, future work in the ARCHER study could look at how self- or parental-rating (Tanner stage or brief ranking questions) is associated with changes in behaviour, relationships (familial, friendships and romantic or sexual) and mood. For example, whether boys who rate their development as ‘ahead’ of their peers are more likely to engage in delinquent behaviour (27), could provide useful information about the intersection of puberty, self-image and mood and behavioural outcomes in this group and allow better targeted interventions.

9.1.6. Foot Length Growth as a Novel Marker of Early Puberty

In this chapter, foot growth was greatest over the three-year study follow-up period for adolescents whose hormone levels were pre-pubertal (at baseline) compared to those with pubertal hormone levels. There are four studies which have looked at foot length as a pubertal marker (28-31). Ford found foot growth preceded peak height velocity in a small longitudinal study of girls (29). In a large Indian cross-sectional study, Mitra demonstrated foot growth was greatest between clinician-rated Tanner stages 1 and 2 (30). Using population growth data and local shoe sale data, Busscher showed an increase in shoe sales in the time preceding peak height velocity (31). These studies showed foot length preceded other anthropometric changes in puberty, but our work is the first to demonstrate a relationship with hormone changes.

In our work, foot length was longitudinally related to changes in serum testosterone and oestradiol, as well as height, weight and self-rated Tanner stage. Percent annual
foot growth was significantly greater in male and female adolescents with pre-
pubertal hormone levels. There are only three years’ follow-up data so further work could be focused at studying foot length from before puberty until completion of growth, in concert with height and weight growth velocities. Additionally, annual foot measurements may not provide sufficient detail to define adequate growth curves, though, using the Super-Imposition by Translation And Rotation (SITAR) model, recent work found that annual anthropometric measurements (such as those performed in our study) may estimate peak growth velocity during puberty (32). However, with only three years of sampling we could not determine if peak growth velocities were reached to delineate whether peak foot growth velocity preceded peak height velocity as it did in Ford’s work (29). Using longitudinal regression, we showed that foot growth was positively associated with self-rated Tanner stage, a further finding of utility.

Foot length therefore offers a novel and cost-effective way of assessing early puberty. It can be used by adolescents, parents and clinicians to broadly determine if an adolescent is in the early stages of puberty. Formal pubertal assessment, whether using clinician-rated Tanner stage or hormone assessment, would of course be needed if there were clinical concerns about puberty. Foot length is an obvious and public change that may necessitate new shoe purchases, which is therefore memorable to parents and adolescents, whereas other growth may not be noticed as early, whether due to subtlety of changes (such as initial height or weight change) or due to adolescent privacy or body image concerns. Foot growth could thus be used as a screening question to identify the early stages of the pubertal transition.
9.2. Future Directions

The data presented here have some important translational implications for adolescent research. We have shown the feasibility of our novel study methodology, particularly in terms of meeting the technical challenges of measuring lower hormone concentrations than previously achieved through novel LC-MS/MS techniques, and the logistical challenges of a longitudinal study in rural New South Wales, where the study’s catchment area (over 45000 km²) and frequent study specimen collection, require novel research tools for communication and ways to minimise study attrition. Our findings can be utilised by other researchers as we have shown how successful regular specimen collection is in adolescents using SMS reminders, alternative ways to obtain puberty staging data and most importantly that novel urine LC-MS/MS techniques are feasible in this setting to measure previously unquantifiable early pubertal hormone levels. Our simple tools for reminders and data collection in early puberty are relevant and useful for those considering adolescent research, particularly longitudinal work.

Findings regarding the true effects of pubertal hormones will inform policy and public health decisions and enable better education of parents and adolescents about puberty. It may also allow early intervention for those at risk of mood or behavioural changes. Future research guided at understanding how longitudinal changes in sex hormones are associated with mood and behaviour will provide stronger evidence than that found in two systematic reviews on the topic (33, 34). We will also be able to identify the role sex hormones play in moderating growth and health during adolescence, with further scope to understand how these affect adolescents in later life with planned work to follow the ARCHER study through data linkage.
All of these findings will significantly advance our understanding of puberty and its role in adolescence, either by confirming previous work with longitudinal data consisting of more accurate hormone assays, or providing novel results that question previous hypotheses and form the basis for a deeper, more rigorous understanding of this crucial life phase. The second decade of life is under-researched and yet is a key time in understanding how long-term health trajectories are established.

9.3. References


APPENDICES

Appendix A: ARCHER Study Methodology ............................................................ 220
Appendix B: ARCHER Study Cohort Description ................................................... 245
Appendix C: Requirement for Specific Gravity and Creatinine Adjustments for Urinary Steroids and Luteinizing Hormone Concentrations in Adolescents ...... 277
Appendix D: Supplementary Material for Chapter 3 ................................................ 294
Appendix E: Supplementary Material for Chapter 4 ................................................ 298

Note: Appendices A, B and C comprise publications which add context to work in the thesis but do not form part of its examinable content.
APPENDIX A – ARCHER STUDY METHODOLOGY

1.1. **Abstract**

1.1.1. **Background**

Adolescence is characterized by marked psychosocial, behavioural and biological changes and represents a critical life transition through which adult health and well-being are established. Substantial research confirms the role of psycho-social and environmental influences on this transition, but objective research examining the role of puberty hormones, testosterone in males and oestradiol in females (as biomarkers of puberty) on adolescent events is lacking. Neither has the tempo of puberty, the time from onset to completion of puberty within an individual been studied, nor the interaction between age of onset and tempo. This study has been designed to provide evidence on the relationship between reproductive hormones and the tempo of their rise to adult levels, and adolescent behaviour, health and wellbeing.

1.1.2. **Methods/Design**

The ARCHER study is a multidisciplinary, prospective, longitudinal cohort study in 400 adolescents to be conducted in two centres in regional Australia in the State of New South Wales. The overall aim is to determine how changes over time in puberty hormones independently affect the study endpoints which describe universal and risk behaviours, mental health and physical status in adolescents. Recruitment will commence in school grades 5, 6 and 7 (10–12 years of age). Data collection includes participant and parent questionnaires, anthropometry, blood and urine collection and geocoding. Data analysis will include testing the reliability and validity of the chosen measures of puberty for subsequent statistical modeling to assess the impact over time of tempo and onset of puberty (and their interaction) and mean-level repeated
measures analyses to explore for significant upward and downward shifts on target outcomes as a function of main effects.

1.1.3. Discussion

The strengths of this study include enrollment starting in the earliest stages of puberty, the use of frequent urine samples in addition to annual blood samples to measure puberty hormones, and the simultaneous use of parental questionnaires.
1.2. **Background**

Adolescence is a time when interventions have the capacity to make changes in individual health trajectories (1-3). The aim of this study is to understand the effects of longitudinal changes in puberty hormones, especially the onset and tempo of change, on adolescent health and well-being. The present paper presents the research protocol for the study, with an emphasis on the newer technologies which have made this type of study possible.

The two prime hormones, or biological measures, of puberty are testosterone (in boys) and oestradiol (in girls) (4), Figure 1.1. The dramatic hormone changes of puberty are a universal and unforgettable experience. Inevitably such a phenomenon prompts many assumptions about the effects of puberty hormones on wellbeing and health. However, the authentic, longitudinal effects of the puberty hormones on human health and wellbeing in adolescence and in later adult life are not well understood, and remain under-researched. Previous research has relied on proxy measures of puberty hormonal changes including self-report of physical change and physical growth (5).
Figure 1.1. Puberty hormone change and tempo. Shows the pattern of testosterone and oestradiol change during puberty, the puberty milestones of menarche and Spermarche in relation to puberty stage and how the tempo of puberty may vary between individuals.
The major and rapid increases in puberty hormones mean that repeated hormone measurement is essential to appraise the relationships between biological variation in puberty hormones and the concomitant adolescent health problems and to allow exploration of the biological determinants of individual, social and physical environmental factors that are prominent in adolescence (6). This is the first challenge in research of this nature. The onset of puberty occurs when testosterone and oestradiol begin to rise above the levels of childhood. This event occurs anywhere between the ages of 8 and 13 years (7, 8) but chronological age is an unreliable indicator of pubertal stage. The tempo, or time, to complete the pubertal increase in reproductive hormones is comparatively rapid, between 18 months and four years, which is too rapid to be characterized by annual or less frequent blood sampling. Although the impact of variable tempo of puberty on health and wellbeing has not been well studied (or reported) (9), we propose that tempo is a key variable mediating the clinical effects of dynamically changing puberty hormones on adolescent health and behaviour (10). It is notable that hormonal events of puberty are as dramatic in scope and amplitude as more well-known examples of rapid hormonal change that cause defined physical, behavioural and mental effects such as the menstrual cycle, pregnancy, castration, anabolic steroid use and the menopause (11-13).

The additional challenges to research on puberty hormones are to create a framework where factors can be repeatedly studied with minimum reporter burden, and to ensure that the interactions between factors are understood and accounted for in order to satisfactorily interpret outcomes. The factors that have been chosen for study are those that are objectively measurable as well as being clinically important and
modifiable, thus forming the basis of subsequent interventional studies. These factors are listed below and a brief review of the factors, how these are interrelated and their roles as possible aetiological factors, co-variates or predictors follows (Figure 1.2).

The Study Variables are:

- Universal aspects of adolescent health and wellbeing which include social and emotional wellbeing, educational engagement and motivation, and sleep.
- Health risk behaviours which include unintentional injury, alcohol and other drugs, sexual activity and conduct disorder.
- Mental health which includes depression, anxiety and self harm.
- Physical health which includes chronic illness and disability and cardiovascular risk factors – overweight, physical inactivity and tobacco use.
Figure 1.2. Depicts how the hormonal environment of puberty, **INCLUDING TEMPO AND ONSET** as defined by blood and urine testosterone & oestradiol may be related to the events of adolescence.

**Internal hormone environment of PUBERTY,**

**AETIOLOGICAL FACTORS**
e.g. genetic background, body weight and composition & physical illness

**HAS AN EFFECT ON EVENTS IN ADOLESCENCE**
e.g. learning, disruptive behaviour, unintentional injury, depression

**HAS AN EFFECT ON EVENTS IN ADOLESCENCE**
e.g. sleep cycle alteration, depression, blood pressure

**External influences on adolescence health and wellbeing which act as confounders include socioeconomic status, family and broader environments and external stressors**

**Figure 1.2.** Depicts how the hormonal environment of puberty as defined by blood and urine testosterone and oestradiol may be related to the events of adolescence.
1.2.1. **Universal aspects of adolescent health and wellbeing**

Socially competent adolescents are more likely to do well in education and less likely to have mental health problems (14). Learning and engagement in education in adolescence is critical to social and emotional development and well-being, but the influence of the onset of and tempo of pubertal hormone rise on motivation and engagement in education, particularly when compared with the effects of socio-economic status and mental health, is as yet unknown (15). Sleep disturbances in adolescents are common and may affect school performance and mood. The most common adolescent sleep disorder is delayed sleep phase syndrome (5-10%) (16). What is not known is how the onset and tempo of pubertal hormone changes relate to sleep alterations and disturbances (17).

1.2.2. **Health risk behaviours**

Unintentional injury is the most common cause of death in adolescents (18), mainly as a result of road traffic accidents (19), poisoning and drowning. Previous work has highlighted possible associations between pubertal stage as estimated by direct physical examination and physical injury incidence (20), but the onset of and the rate of the rise in pubertal hormones have yet to be studied in this context. Drug and alcohol use in adolescence increases the likelihood for reduced educational attainment, co-morbid mental health disorders, substance abuse and dependence, criminality, and psychiatric disorders in adulthood (20). We propose that puberty onset and tempo also play an important role in these associations (22). Sexual activity in teenagers is occurring at younger ages and teenagers are more likely to report multiple partners than in previous generations (23). The relative contribution...
of changing pubertal hormone levels, specifically tempo, to sexual activity has not been confirmed and the mechanism of their influence and modifiers are completely unknown. Conduct disorders appear in pre-puberty or first emerge in puberty (24). There are conflicting data about these disorders and their relationship to early or late maturation as measured by self-report and single measurements of blood testosterone levels.

1.2.3. Mental health

Depression and anxiety are the most common mental health problems in young people. At any time point, up to five percent of adolescents experience depression severe enough to warrant treatment, and around 20% of adolescents will have experienced significant depressive symptoms by adulthood (25, 26). On self-reported Tanner Stages, depression occurs most commonly at Stage 3 in girls and before Stage 3 in boys. There is conflicting evidence about whether age of onset of puberty influences depression (27-29). Higher rates of self-harm are found among adolescents in self-reported late puberty compared to early puberty (30). The association is mediated by depressive symptoms, sexual activity and alcohol intake, and often attenuates with increasing age (31).

1.2.4. Physical health

Fifteen percent of adolescents have a chronic physical illness or disability (32). If serious this may retard the onset of puberty and potentially alter tempo (33). Overweight, physical inactivity and tobacco use, which are the top three contributors to the burden of disease in developed countries (34), all increase during adolescence (35). Twenty percent of adolescents smoke by the age of 17 years (36). There are no
longitudinal studies on how the age of onset and rate of change or tempo of puberty hormones influence the cardiovascular risk factors of overweight, physical inactivity and tobacco use.

1.3. **Methods/Design**

1.3.1. **Introduction**

The ARCHER study is a multidisciplinary, three year, longitudinal cohort study, using a convenience sample. The research hypothesis of the study is that onset and/or tempo of the rise in puberty hormones play a significant role in adolescent health and wellbeing. The ARCHER study commenced recruitment in June 2011 and has enrolled 200 adolescent participants (with parent/guardian) as of March 2012.

A feasibility study for all aspects of baseline data collection was completed in 28 students in years 5&6 at a rural primary school. Sample collection, labeling, transport and storage were successful. The plasma testosterone and oestradiol levels demonstrated that the participants (except 2 -1 M, 1 F) had levels consistent with pre- or very early puberty. On-line questionnaires were revised and successfully re-piloted in August 2009 in a rural primary school with 26 students and their parents. In 2010 focus group discussions (single and mixed gender with 10–12 year olds and 13–15 year olds) were held with 58 participants. Topics included what would positively influence recruitment and retention, how peers might participate in the study and the collection of biological samples. The young people expressed strong preferences in relation to the collection process and the focus group outcomes are now published (37). The methodology was further tested in 2010 in a funded pilot study on puberty, sleep and depression, with included the use of the Actiwatch. This
is a digitally integrated recording of wrist activity which provides more detailed information on sleep wake/times (38).

1.3.2. Setting and study sample

The study is being conducted in Dubbo and Orange, two large towns in regional Australia in the State of New South Wales and surrounding areas. Enrollment is in school Grades 5, 6 and 7 (10–12 years of age) to ensure that both the transition into early puberty and the end of puberty are captured in adequate numbers. Each student will require a parent or guardian to also participate in the study. Lack of parental competence in English or adolescent intellectual disability are exclusion criteria. The least common primary outcome that we measure is depression, which we estimate will affect 5% of children in mid- to late-puberty (39). We estimate that 160 boys and 160 girls at Tanner Stage 3 (the pubertal stage where clinical onset of depression is most likely) are required to detect a one standard deviation (SD) difference in testosterone nmol/l (SD = 5) (40) or oestradiol pmol/l (SD = 250) (41) levels between those with and without depression. The anticipated loss to follow up is 5% per annum in this setting. Assuming a mean age of puberty onset of 11.4 years (42), 400 subjects will ensure adequate numbers of adolescents at Tanner Stage 3 in the study, as a few younger subjects may not reach this Stage at the end of the study (age 14 years) and some older children may have already entered Stage 3 at the beginning of the study (age 12 years).

1.3.3. Recruitment procedures and follow up rates

Recruitment is primarily through schools, as well as through the media (targeted at parents) and through established links with local community groups. We have
permission from the state Department of Education and Communities - Western Region and the Catholic Education Office – Bathurst Diocese to recruit through their schools. Data collection does not take place in the schools. Information sessions are held in the community and at consenting schools. Parents are asked to provide two additional adult contact names with the written consent, to reduce loss to follow up. Participants who agree are sent birthday cards, updates and study newsletters as reminders by their preferred mode of contact (email, letter or Short Message Service (SMS)). These practices have been shown to increase retention rates in school age children (43). The young people requested ‘get togethers’ in the focus groups to celebrate their involvement in the study and these will be held annually.

1.3.4. Data collection

Adolescents will complete questionnaires at baseline and annually for the following three years. We will use a computerized on-line questionnaire with branched algorithmic structure, to enhance confidentiality and accuracy of responses and to reduce exposure to sensitive questions (particularly related to sexual activity). This has been created and trialed. The total number of questions for adolescents is 256. These are completed in approximately one hour (shorter for the older participants and closer to 75 minutes for younger participants). Adolescents complete the questionnaire under supervision of research staff. Questionnaires will not be performed in the fasted state and, as with all other data and biological sample collection, are done outside school hours. Physical examination occurs at baseline and annually for three years. Blood is collected at baseline and annually for three years, in the fasting state. A first morning urine sample is collected at baseline and every three months for three years. A single indicator question providing information
on mood fluctuation is sent to all participants with access to mobile phones every 3 months and functions as a reminder to complete urine collection.

A parent or guardian completes a questionnaire at baseline and annually for three years, and in the same month as their child. The questionnaires are available online or as hard copy and contain approximately 180 questions. This questionnaire reduces reporter burden on adolescents for demographic data, as well as providing other relevant family and environmental data.

1.3.5. Ethics

Ethical approval has been granted for the pilot questionnaires, anthropometry, biological sample collection (HREC 10612), the pilot SMS and sleep studies (HREC 12502) and for the full study described in this protocol (HREC 13094) from the Human Research Ethics Committee, University of Sydney.

1.3.6. Instruments

1.3.6.1. Questionnaires - adolescent

The Child Behaviour Check List (CBCL) in the version validated for ages 11–18 years as the Youth Self Report (YSR) (44) is the main questionnaire measurement instrument. It contains 20 social competence items that measure participation in hobbies, games, sports, jobs, chores, friendship, and activities and 8 sub-scales, measuring internalizing and externalizing behaviour. In addition the subscales provide data on social competence, learning and engagement in education, sleep, unintentional injury, drug and alcohol use, conduct disorders, disruptive behaviour
disorders, depression and anxiety, and self harm. The YSR also identifies the presence of physical illness and disability.

The YSR is supplemented by the following questionnaires.

a) Selected measures of self competency taken from the Raine cohort (45), including Cowen’s Self Efficacy (46) and Adolescent Self Perception Profile (47).

b) 12 selected items from Motivation and Engagement; Enjoyment of School, Academic Buoyancy and Class Participation scales (48, 49). These results can be linked to the Australian National Assessment Program Literacy and Numeracy scores (50).

c) A questionnaire is used to document sleep/wake patterns, and sleep disorders in the adolescent’s natural sleeping environment (51). In addition, a convenience subset of adolescents will use the Actiwatch home monitor system annually.

d) Direct questions on unintentional injury type in the preceding 12 months provided by RI, selected questions from The Australian School Students Drug & Alcohol Survey (52), branched age appropriate sexuality questions on romantic relationships, sexual feelings, history of STI and pregnancy (23) and the 40 item validated Australian self-reported delinquency scale by Mak (53) in those adolescents who score highly on the YSR subscale for externalizing problems.

e) The Short Moods and Feeling Questionnaire (13 item) in all subjects (54). The short form has been validated as a self-reported unidimensional measure of symptom severity of childhood depression in community samples.
f) The 16 item Deliberate Self Harm inventory (55) for those who respond with a 1 or 2 for YSR item 18 or 91 (relating to self-harm or suicide).

g) Selected questions from Health Behaviors in School-aged Children for physical inactivity, nutrition and tobacco (56).

h) One SMS question which is a linear analogue self-assessment scale question on mood providing information on mood fluctuation (57).

i) Tanner Stage of puberty by self report using standardized line drawings which allow determination of accuracy of this method of pubertal staging, which is the main method used in epidemiological and non-clinical studies.

1.3.6.2. Questionnaires – parent/guardian

a) Demographic data for family and adolescent.

b) Child Behavior Checklist (CBCL) validated questionnaire for parents of children aged 6–18 years to corroborate the adolescent’s YSR (58).

c) The validated Macmaster’s Family Assessment Device (59) to obtain information on family and local environmental factors, which are potential confounders.

After the final 3-year questionnaire all adolescents will be asked whether they believed that the speed of their puberty was faster, the same as or slower than their age peers and whether the onset of their puberty was earlier, the same as or later compared to their age peers. Their parent/guardian will be asked the same questions in relation to their child.
1.3.6.3. Physical examination – adolescents only

Weight and body composition is measured in light clothing using Tanita TBF-300 Pro Body Composition Analyzer (60). Height, and foot length as an indicator of growth stage (61) are measured on a portable stadiometer (to 0.1 cm) and with a metal ruler (to 0.5 cm) respectively, and waist (62) with a tape (to 0.1 cm) using standard techniques. Body Mass Index, as a measure of overweight (kg/m²) and waist circumference are expressed as individual z-scores based upon age and sex related reference charts (63). Blood pressure (BP) and pulse rate are measured using an automated BP monitor, under standardised conditions.

1.3.7. Laboratory – adolescents only

Fasting bloods are collected using local anaesthetic cream and separated serum and spot urine samples frozen at −80°C, as well as buffy coat in one sample. Total blood draw is approximately 30 ml. Menstrual cycle stage will be recorded in post-menarchal females, with the perimenstrual week avoided.

1.3.7.1. Blood

Testosterone and oestradiol will be measured using AP5000 LC Tandem MS in the Andrology laboratory, ANZAC Research Institute (64) (DHEA, DHT and oestrone will also be measured). There will be adequate blood collected to measure other potential biological variables of interest, which will form the basis of future studies but which are not integral to this study. These variables include LH, FSH, SHBG and IGF-1, inhibin, anti-Mullerian hormone, ACTH, cortisol, growth hormone, prolactin, oxytocin, TSH, thyroxine, triiodothyronine and Vitamin D (all of which are hormones relevant to puberty) (4), full blood count, ferritin, glucose, liver function,
urea, electrolytes, full lipid profile, creatinine, and insulin, fatty acid profile, leptin, adiponectin, resistin, TNF alpha, CRP and Interleukin 6 (as indicators of cardiovascular risk and also to derive eGFR) (65). There is the potential to be able to study genes relevant to puberty and sex steroid action; KiSS-1, KAL1, FGFR1, GnRHR, GRPR-54, TAC3, neurokininB, androgen and oestrogen receptor, steroidogenic enzymes such as CYP, StAR, P450sc, P450 (17alpha) and 17beta-HSD (66), and genes relevant to body composition and insulin resistance, such as leptin gene and product, MC4R, FTO, IL-6, and UCP-2 (67).

1.3.7.2. *Urine*

A first morning, fasting urine sample will be collected at home. Urinary testosterone and oestradiol will be measured, also using AP5000 LC Tandem MS. Urinary hormones, although less well standardized than blood, are essential to define hormone trajectories which can never be accomplished by a single annual sample. Sperm (indicating spermarche) will be measured in male urine samples (68). There will be adequate urine to measure creatinine and microalbumin, but these measures are not integral to the current study.

1.3.8. **Geocoding**

Residential address will be geocoded (69) to obtain additional information such as neighbourhood socioeconomic status, access to services and amenities and other environmental exposures, which may be study confounders for which adjustment will be required.
1.3.9. Statistical analysis

The main independent variables are tempo of puberty and age of onset at puberty. Both will be measured by testosterone and oestradiol, in blood and urine. Data analysis will have three main components.

1. The first component is testing the reliability and validity of measures under focus. These analyses will centre on: (a) descriptive, reliability and item functioning; (b) factor analysis to test factor structure and validity of measures; (c) ANOVAs for preliminary tests of main and interaction effects of background socio-demographics; and (d) tests of (in)variance in factor structure to ensure congruence in measurement properties across sub-groups to justify pooling sample data for subsequent modeling throughout the project (70).

2. The second component is a correlational one and assesses the impact of tempo and onset of puberty over time. Specifically, for example, by assessing in the one analytic model (e.g. in linear regression models for continuous variables such as YSR scores for sensation seeking and in logistic regression models for dichotomous outcomes such as depression or sexual debut) the effects of Time 1 onset and tempo of puberty on Time 2 outcomes after controlling for Time 1 outcomes, it is possible to get a sense of the relative salience of the onset, and tempo of puberty in one time period over factors in a following time period (71). Also in these longitudinal analyses, key socio-demographic (and other) factors such as gender, school grade and socioeconomic factors can be included to get a further sense of puberty by controlling for other potential influences. The three monthly urine samples for the biological markers of puberty can be used to regress time on time measures, with
tempo being defined by the residuals. Thus the larger and the more positive a residual over time, the more rapid the tempo of change.

3. The third component is based on mean-level differences (e.g. repeated measures) and explores for significant upward and downward shifts in means on target outcomes as a function of main effects (e.g. early and late onset; rapid and slow tempo) and interactions (e.g. early onset/rapid tempo; late onset/rapid tempo; early onset/slow tempo and late onset/slow tempo) of puberty effects that also control for key socio-demographic (and other) factors as covariates (72). Taken together, these analyses (and adaptations of them) allow for integrative tests of the substantive and methodological issues at hand.

Analyses will be conducted using SAS, Stata and WinBUGS (for repeated measures analyses).

1.4. Discussion

This study aims to extend previous puberty research which has sought to consider the relationship between the biological events of puberty and the complex, longer state of adolescence and its associated health issues. As a longitudinal study it commences at a younger age in order to capture the transition into puberty, as defined by the first elevations of testosterone and oestradiol in blood and urine. Thus puberty is defined using biological markers, which are of fundamental importance to the project. Previous studies have relied on self report of pubertal stage, which is a lagging and insensitive measure, that correlates only weakly with measured puberty hormones (5). This will be the first study to focus on the biological tempo of puberty and its
impact on adolescent health and wellbeing, with frequent biological sampling and the introduction of new, ultrasensitive state-of-the-art, mass spectrometric steroid assays. Previous studies have used immunoassay with unextracted serum, which has proven analytically inaccurate and unreliable (although faster) and which further places in doubt previously described relationships between the clinical features of puberty and its hormonal determinants.

The utility of this new knowledge, by the translation of hormone data into real life, is to better direct interventions at a time when developmental trajectories remain plastic and capable of alteration and without the need for biological sample collection. It can be done because ‘puberty phenotypes’ such as early or late onset (out of step with peers) or rapid tempo (rapid physical change) and which deviate most from average age of onset/average tempo (similar to the majority) are already recognizable using simple observation by clinicians, parents, educators and adolescents. What is missing is an understanding of the direction and magnitude of influence of intrinsic tempo and timing of puberty on adolescent behaviour, wellbeing and health and how hormonal changes interact with external factors. The indirect, non-biological evidence is that the magnitude of intrinsic hormonal influence is significant.

There is no suggestion that altering the onset or tempo of puberty is a realistic intervention (despite its feasibility using modern steroidal or gonadotrophin releasing hormone analogue therapy (73, 74)). It is however of great interest to determine whether the dynamic hormonal changes of puberty trigger behavioural or physical problems. If puberty hormones do have an effect, it would be possible to target established, evidence-based directed interventions to at-risk adolescents, presumably those whose puberty patterns are dissimilar to the majority. These could include
evidence-based interventions in sleep hygiene, unintentional injury, mental health, and cardiovascular risk. In schools, where year levels are defined by age, rather than pubertal stage, information will be provided to better assist school personnel to identify and deal with the wide developmental range in any given year group.

Adolescents, their parents and families will benefit from understanding how hormones influence adolescent behaviours. The community will benefit if knowledge gained from this study is used to improve adolescents’ health and wellbeing.

1.5. References

Appendix A: ARCHER Study Methodology


47. Harter S. The Self-Perception Profile for Adolescents. 1988, University of Denver, Denver, USA


Appendix A: ARCHER Study Methodology

68. Laron Z. Age at first ejaculation (spermarche)–the overlooked milestone in male development. Pediatr Endocrinol Rev. 2010;7(3):256-257.
APPENDIX B – ARCHER STUDY COHORT DESCRIPTION

1.1. Abstract

Aims:

To provide an overview of preliminary findings from the longitudinal ARCHER (Adolescent Rural Cohort study of Hormones, Health, Education, Environments and Relationships) study of rural adolescent health.

Methods:

The ARCHER study is a three-year multidisciplinary longitudinal rural study of adolescents (recruited at 9 to 14 years) from two regional centres in the Central West of NSW. The ARCHER study includes an extensive yearly survey of adolescents and their parent(s)/guardian using questionnaires, anthropometry, blood and urine collection. Measures include universal aspects of adolescent health and wellbeing, such as education, health risk behaviours, mental and physical health. Data analysis was performed to explore trends over time and by age.

Results:

342 young people were recruited to ARCHER with 82% retention across four Waves of data collection over 3 years. At baseline, the median age was 11 years, 45% were female and 11% were Aboriginal or Torres Strait Islander adolescents. Serum testosterone and oestradiol levels confirmed that the majority of the adolescents were in early puberty. The young people came from diverse socioeconomic backgrounds, although participating families were generally more affluent and better educated than the broader regional population. Across the four Waves, between 10 and 18%
experienced significant depressive symptoms and rates increased with age.

According to parents/guardians, only approximately 1 in 10 young people met national guidelines for physical activity, and between 16 to 22% lived in a household with at least one smoker. Prevalence of overweight and obesity ranged from 26 to 28% overall (females 29 to 32%; males 23 to 25%). Associations between gender, mental and physical health and stage of puberty were present.

Relevance:

Adolescence is a critical epoch laying durable foundations for lifelong health and wellbeing. Understanding and promoting rural adolescent health and wellbeing will contribute to addressing known rural health inequities.

Conclusions:

We have successfully recruited and maintained a cohort to answer novel research questions. Data generated will further our understanding of puberty and its effects as well as providing insight into the specific determinants of health for young people growing up in rural Australia.
1.2. Introduction

Adolescence is a critical phase of life that involves complex biological, physical, cognitive and social changes for the young person. These changes combine and interact with environmental factors such as socioeconomic status, education and familial/peer relationships to potentially alter trajectories of future health and wellbeing (Sawyer et al., 2012; Viner et al., 2012). Puberty is the biological event defining early adolescence. Many assumptions and assertions are made about the authentic, underlying biological effects of the major puberty hormones – testosterone in males and oestradiol in females - on adolescent health, wellbeing and behaviour. To date, the definitive longitudinal studies with appropriately frequent measures of puberty hormones have yet to be reported. Most studies rely on physical staging of puberty, a late effect of puberty hormone action (Dorn et al., 2006). The effect of arguably the most dramatic hormone changes in life and their interactions with other determinants of health and changes in adolescence remain poorly understood. Additionally, the frequency of many characteristic behaviours and recognised morbidities of adolescents increase after the physical events of puberty are complete. These observations suggest that the individual patterns of puberty hormones – both timing and tempo – may be influential. This is the hypothesis of the ARCHER study (Adolescent Rural Cohort study of Hormones, Health, Education, Environment and Relationships) which aims to identify the genuine biological influence of puberty hormones on adolescent health, wellbeing and behaviour (Steinbeck et al., 2012).

Although about one-third of 10 to 13 year olds live in regional or remote areas (32.8%; ABS, 2011), the health determinants of Australian rural and regional...
adolescents are poorly documented and under-researched. Known inequalities of rural populations include lower incomes, poorer levels of education and employment and access to services. The further from metropolitan areas young people live the higher the burden of disease and mortality rates, accidents are more common, diet is poorer, there are greater levels of alcohol abuse and poorer access to health care (AIHW, 2011). These inequalities in social determinants of health and the specific health needs of young people in non-metropolitan areas are not well understood.

This preliminary analysis of the ARCHER study focusses on several important aspects of adolescent health from the myriad studied, namely sleep, depression and overweight and obesity (O&O). It is well recognised that sleep patterns change during adolescence. Adolescents tend to have decreased total sleep duration and increased differences between weeknight and weekend sleep (Owens, 2014). The known consequences of their poor sleep are wide-ranging and include poorer academic achievement, increased risk-taking behaviours, obesity and poorer mental health (Chaput et al., 2016; Shochat et al., 2014). This time in life is also characterised by substantial cognitive, social-cognitive and social-organisational changes that have significant implications for academic wellbeing, including lower academic motivation and engagement (Martin, 2009; Martin et al., 2015). Depression is the leading cause of disability amongst young people worldwide (WHO, 2014). During adolescence, the prevalence of depression increases, which is commonly attributed to puberty, although empirical evidence is almost entirely lacking (Balzer et al., 2015; Duke et al., 2014). The impacts of depression in young people include poor functioning at school or work, poor relationships with family and friends, risk-taking behaviour and suicide (Lawrence et al., 2015). The prevalence of O&O
increased substantially in young Australians during the latter part of the 20th century, similar to global trends (Hardy et al., 2011; Ng et al., 2014). Although this increase appears to have plateaued in Australia in the last decade (ABS, 2013a), adolescent O&O remains an important public health issue. Australia has one of the highest rates of youth obesity among developed countries (Ng et al., 2014). Reduced physical activity is also a well-known risk factor for O&O. The Australian National Guidelines for Physical Activity and Sedentary Behaviour recommend moderate to vigorous activity for at least 60 minutes every day of the week (Australian Department of Health, 2014).

The known contribution of adolescent health and health-related behaviours to adult health and wellbeing, and the dearth of research particularly into rural youth, emphasise the importance of this unique ARCHER cohort. This paper provides an overview of this rural youth cohort including: i) hormones and pubertal status; (ii) physical health, health risk behaviours, sleep and environment; (iii) mental health; and (iv) overweight and obesity; and examines predictors of mental health and overweight and obesity in this rural sample.

1.3. Methods

The detailed methodology of the multidisciplinary ARCHER study has already been published (Steinbeck et al., 2012). Briefly, young people in school years 5 to 7, and their parents/guardians, were recruited from two regional centres, Dubbo and Orange, in the state of New South Wales, Australia and surrounding rural areas (see Figure 1). Recruitment was primarily through schools, with information sent home to parents/guardians. Other recruitment methods included flyers on community
noticeboards, through sporting clubs and through local digital and print media. The lower year 5 school limit was selected to ensure that participating adolescents were capable of self-report yet were captured before puberty was too advanced. Annual surveys of both adolescents and parents/guardians included demographics (individual and household), socioeconomic factors, sleep patterns, health risk behaviours, mental health and physical activity/lifestyle factors. Young people were asked to complete the Short Mood and Feelings Questionnaire (SMFQ; Angold et al., 1995) and high levels of depressive symptoms were defined by scores of 11 or higher. They self-reported their Tanner stage and answered a question on perceived stage of pubertal development: “Compared to other kids your age and sex, what is your stage of puberty?” responding either ‘behind’, ‘the same’ or ‘in front’.
Figure 1. Australian remoteness areas (with detail for the Central West of NSW). Source: ABS Australian Standard Geographic Standard (ASGS) Digital Boundaries; ABS, 2013b.
In addition, data on anthropometry and puberty hormones in serum were collected annually from adolescents (only Wave 1 serum hormones are available at this stage). Serum testosterone and oestradiol were measured by liquid chromatography, mass spectrometry (PMID 19747904, 26090565). Weight was measured using Tanita TBF-300 Pro Body Composition Analyzer. Height was measured on a portable stadiometer (to 0.1 cm) using standard techniques. Body Mass Index (kg/m²) was converted to grades of thin, normal weight, overweight and obese using the International Obesity TaskForce cut-offs for age and sex (Cole and Lobstein, 2012). The study population all completed the baseline questionnaires and measurements between June 2011 and August 2013. Participation rates are shown in Table 1. While the overall retention rate was 82% (n=281/342), between adjacent Waves the retention ranged from 92% to 95%. Participants’ baseline characteristics were compared with data for local government areas (LGAs) containing over 5% of ARCHER households. These included Cabonne, Dubbo, Orange, and Wellington. From the 2011 Census, there were 5,754 adolescents aged 10 to 13 years resident in these four LGAs, 24,690 parents (aged 30 to 49 years) of adolescents aged 10 to 13 years and 35,258 dwellings.
Table 1. Young ARCHER participants with complete survey data over four annual Waves of data collection

<table>
<thead>
<tr>
<th>Wave</th>
<th>Enrolled</th>
<th>Survey data available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 1</td>
<td>342</td>
<td>339</td>
</tr>
<tr>
<td></td>
<td>did not continue/moved out of area</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>skipped Wave 2 (rejoined Wave 3)</td>
<td>1</td>
</tr>
<tr>
<td>Wave 2</td>
<td>314 (91.8%)</td>
<td>313</td>
</tr>
<tr>
<td></td>
<td>did not continue/moved out of area*</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>skipped Wave 3 (rejoined Wave 4)</td>
<td>1</td>
</tr>
<tr>
<td>Wave 3</td>
<td>296 (94.3%)</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td>did not continue/moved out of area</td>
<td>18</td>
</tr>
<tr>
<td>Wave 4</td>
<td>281 (94.9%)</td>
<td>281</td>
</tr>
</tbody>
</table>

* n=2 joined study at Wave 4
N.B. % represents proportion retained from previous Wave.
1.3.1. Data analysis

Measures of central tendency and dispersion are given as medians and interquartile ranges (IQR) due to skewed distributions for age, age at menarche and total sleep times. Associations between categorical data were estimated using the chi-squared test. Paired comparisons across the four Waves for typical hours slept on weekdays and on weekends were made using the non-parametric Friedman Test. Sleep compensation was defined as sleeping at least one hour longer on weekends than on school nights, and was based on self-reported total sleep time (TST). The relationship between age and sleep compensation was explored using a Spearman’s correlation ($r_s$). Separate generalised estimating equation (GEE) models were used to predict depressive symptoms and O&O across the four Waves of data. Variables included in each of the regression models were age, socioeconomic status (highest household occupation), number of days of moderate to vigorous physical activity (MVPA, self-reported), self-reported TST during school nights and during the weekend, and either Tanner stage or perceived comparative stage of pubertal development (same, behind or in front). All analyses were carried out using IBM SPSS Statistics v24. Alpha was set at 0.05 for all comparisons.

1.3.2. Ethics

Approval for the ARCHER study was granted by The University of Sydney Human Research Ethics Committee (HREC 13094 approved 2010). Both parental (or guardian) consent and child assent to participate were obtained.
1.4. **Results**

Median age of the 342 young participants at baseline was 11 years, 45% were female and 11% were Aboriginal or Torres Strait Islander adolescents. The sample was representative of the region in terms of gender (proportion in LGA 49% female), but slightly underrepresented youth of Aboriginal background (proportion in LGA 17%). The young people came from diverse socioeconomic backgrounds. Approximately a quarter (26%) of households included an adult with a managerial occupation, whereas for 17% the highest household occupation was sales, machinery operation, labouring or home duties. In comparison to the LGAs, participating families were generally more affluent (more likely to be employed: ARCHER 87% vs LGA 81%; and to own a house: ARCHER 80% vs LGA 68%) and better educated (ARCHER 49% with tertiary education vs LGA 20%). Only one young person had left school at Wave 4.

1.4.1. **Puberty and hormones**

The majority of the adolescents were in early puberty according to testosterone and oestradiol levels in plasma at baseline (Wave 1) (see Figures 2a-b). Fifty six percent of females and 49% of males rated themselves as pre-pubertal or early-pubertal (Tanner stage 1 to 2; see Table 2). Just under one quarter of females (23.2%) and 10.1% of males rated their stage of puberty as ‘behind’ compared to others their age and sex (see Table 2).
Figure 2a. Serum testosterone (nmol/L) at baseline for males (n=166).

Figure 2b. Serum oestradiol (pmol/L) at baseline for females (n=133).
Table 2. Pubertal stage at baseline, by age group and sex.

<table>
<thead>
<tr>
<th></th>
<th>9 to 10 years (n=88)</th>
<th>11 to 12 years (n=202)</th>
<th>13 to 14 years (n=49)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tanner Stage 1 or 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>77.3%</td>
<td>51.8%</td>
<td>8.2%</td>
</tr>
<tr>
<td>Males</td>
<td>71.4%</td>
<td>49.5%</td>
<td>9.7%</td>
</tr>
<tr>
<td><strong>Perceived comparative stage of puberty</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behind</td>
<td>23.9%</td>
<td>15.8%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Same</td>
<td>60.2%</td>
<td>65.8%</td>
<td>77.6%</td>
</tr>
<tr>
<td>In Front</td>
<td>15.9%</td>
<td>18.3%</td>
<td>20.4%</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behind</td>
<td>33.3%</td>
<td>22.3%</td>
<td>5.6%</td>
</tr>
<tr>
<td>Same</td>
<td>51.3%</td>
<td>60.6%</td>
<td>72.2%</td>
</tr>
<tr>
<td>In Front</td>
<td>15.4%</td>
<td>17.0%</td>
<td>22.2%</td>
</tr>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behind</td>
<td>16.3%</td>
<td>10.2%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Same</td>
<td>67.3%</td>
<td>70.4%</td>
<td>80.6%</td>
</tr>
<tr>
<td>In Front</td>
<td>16.3%</td>
<td>19.4%</td>
<td>19.4%</td>
</tr>
</tbody>
</table>
1.4.2. Physical health, health risk behaviours and sleep

According to parent/guardian reports, only approximately 1 in 10 young people met national guidelines for physical activity (i.e. at least 60 minutes every day of moderate to vigorous physical activity) (8.5 to 11.2% across the study Waves). In contrast, between 13.5 and 23.7% of young people reported at least 60 minutes of MVPA each day. Across the four time points, between 16.3 and 21.5% lived in a household with at least one smoker. The proportions of young people reporting smoking ‘even part of a cigarette’ (but more than ‘just a few puffs’) increased from 1.5% (n=5/339) at Wave 1 to 7.5% (21/281) at Wave 4. The proportions reporting having had even part of an alcohol drink, beyond a few sips, was 2.1% (7/339) at Wave 1, and rose to a quarter of the sample (24.9%, 70/281) by Wave 4. Other drug use was minimal, with no reported amphetamine use at all, and marijuana use between 0.6% and 3.2% across the Waves. Approximately one in five (19.3%) young people reported having an illness, disability or handicap at least once during the study (point prevalence between 6.8% and 11.8%). Of these 66 young people, 42% (n=28) reported having asthma.

Total sleep hours decreased across the four Waves: from a median for weekdays of 10 hours down to 8.8 hours (p < 0.001) and on weekends from 9.3 hours down to 9 hours (p < 0.01). Across the Waves, sleep compensation (sleeping at least one hour longer on weekends than on school nights) was reported by between 26.7% and 40.9% of adolescents. The difference in weekend and school night total sleep hours was significantly positively correlated with age (rs = 0.15, p < 0.001), indicating
increased compensation with age. There was also increased weekend sleep compensation by Tanner stage for both females and males (see Figure 3).
Figure 3. Weekend sleep compensation by sex and Tanner Stage. Data are regardless of Wave; N=1,195.
1.4.3. Mental health

Just under a third (30.7%) of young people had experienced significant depressive symptoms over the four Waves, with point prevalence ranging from 10.5% (Wave 2) to 18.1% (Wave 4). Depression increased in prevalence with age: 12.0% of 9 to 11 year olds, 13.9% of 12 to 14 year olds and 17.2% of 15 to 17 year olds. For females, depressive symptoms increased in prevalence with advancing Tanner stage (12% at stage 1 and 27% at stage 5; p < 0.01), whereas for males there was no clear pattern (see Figure 4a). Perceptions of comparative pubertal development were associated with depressive symptoms for males (p < 0.001). Among young males who perceived themselves as ‘behind’ the pubertal stage of their peers, 22% had depressive symptoms compared to 14% of those who thought they were ‘in front’ and 6% of those ‘the same’ as others (see Figure 4b). The association of perceived pubertal stage was not significant in females.
Figure 4a. Significant depressive symptoms by sex and Tanner stage. Data are regardless of Wave; N=1,216.

Figure 4b. Significant depressive symptoms by sex and perceived comparative stage of puberty. Data are regardless of Wave; N=1,225.
1.4.4. **Overweight and obesity**

Prevalence of O&O ranged from 26.4 to 28.1% across the Waves (see Table 3). While prevalence of O&O was always higher in females (females: 29-32% vs males: 23-25%), it was not significantly higher than males during any Wave. O&O was significantly associated with increasing Tanner stage in females (p < 0.05), but not males (see Figure 5a). For females who thought their pubertal stage was behind others their age and sex, 16% were O&O, compared to 31% of those ‘the same’ and 52% of those ‘in front’ (p < 0.001). For males, the rates of O&O for those ‘behind’ were 22%, ‘the same’ 23% and ‘in front’ 31%, not a significant pattern (see Figure 5b).

Age at menarche was available for 84% of females, and the median age was 13 (IQR 1.7). One-fifth (26/129) reached menarche before 12 years, 27.1% between 12 and < 13 years and 52.7% from 13 years onwards. Considering data from all Waves, there was a significant relationship between age at menarche and O&O, 45.9% (n=45/98) of those having an earlier (before 12 years) menarche were O&O, compared with 29.7% (35/118) in the average age menarche group and 26.1% (63/241) in the later (13 years onwards) menarche group (p < 0.01).
<table>
<thead>
<tr>
<th>Grade</th>
<th>Wave 1 (n=341)</th>
<th>Wave 2 (n=313)</th>
<th>Wave 3 (n=293)</th>
<th>Wave 4 (n=274)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thin</td>
<td>6.5%</td>
<td>7.3%</td>
<td>5.8%</td>
<td>8.4%</td>
</tr>
<tr>
<td>Normal</td>
<td>67.2%</td>
<td>66.5%</td>
<td>66.2%</td>
<td>63.5%</td>
</tr>
<tr>
<td>Overweight</td>
<td>17.3%</td>
<td>18.2%</td>
<td>18.4%</td>
<td>17.9%</td>
</tr>
<tr>
<td>Obese</td>
<td>9.1%</td>
<td>8.0%</td>
<td>9.6%</td>
<td>10.2%</td>
</tr>
</tbody>
</table>
Figure 5a. Overweight and obesity by sex and Tanner Stage. Data are regardless of Wave; N=1,203.

Figure 5b. Overweight and obesity by sex and perceived comparative pubertal stage. Data are regardless of Wave; N=1,212.
1.4.5. Predictors of mental health

Models were developed to predict mental health (SMFQ significant depressive symptoms) over the four Waves. Depressive symptoms in females was associated with greater age, lower SES and fewer hours sleep on weekdays. Depression in males was associated with perceived pubertal stage, with significantly lower odds of depressive symptoms if perceived stage was 'the same', compared to perceiving self to be ‘behind’ (OR 0.25, 95% CI 0.10-0.58) (see Table 4).
Table 4. Odds of significant depressive symptoms (SMFQ) across the four Waves.

<table>
<thead>
<tr>
<th>SMFQ caseness</th>
<th>Females</th>
<th></th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio</td>
<td></td>
<td>Odds ratio</td>
</tr>
<tr>
<td></td>
<td>(95% confidence interval)</td>
<td></td>
<td>(95% confidence interval)</td>
</tr>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Model 2</td>
</tr>
<tr>
<td>Age</td>
<td>1.24 (0.98-1.57)</td>
<td>1.23 (1.04-1.46)†</td>
<td>1.02 (0.74-1.41)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.04 (0.79-1.37)</td>
<td></td>
</tr>
<tr>
<td>SES</td>
<td>1.19 (1.03-1.37)†</td>
<td>1.19 (1.04-1.37)†</td>
<td>1.12 (0.97-1.29)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.11 (0.96-1.29)</td>
<td></td>
</tr>
<tr>
<td>MVPA (SR)</td>
<td>1.02 (0.90-1.17)</td>
<td>1.01 (0.89-1.16)</td>
<td>0.94 (0.80-1.10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.96 (0.81-1.13)</td>
<td></td>
</tr>
<tr>
<td>School night TST</td>
<td>0.75 (0.62-0.91)‡</td>
<td>0.75 (0.62-0.91)‡</td>
<td>1.17 (0.85-1.60)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.14 (0.84-1.55)</td>
<td></td>
</tr>
<tr>
<td>Weekend TST</td>
<td>0.85 (0.72-1.01)</td>
<td>0.86 (0.73-1.02)</td>
<td>0.84 (0.69-1.03)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.86 (0.70-1.07)</td>
<td></td>
</tr>
<tr>
<td>Tanner Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Stage 2</td>
<td>1.10 (0.23-5.15)</td>
<td></td>
<td>0.51 (0.17-1.51)</td>
</tr>
<tr>
<td>Stage 3</td>
<td>0.81 (0.16-4.18)</td>
<td></td>
<td>0.37 (0.10-1.46)</td>
</tr>
<tr>
<td>Stage 4</td>
<td>1.11 (0.20-6.09)</td>
<td></td>
<td>0.68 (0.21-2.25)</td>
</tr>
<tr>
<td>Stage 5</td>
<td>0.81 (0.12-5.26)</td>
<td></td>
<td>0.59 (0.15-2.31)</td>
</tr>
<tr>
<td>Comparative Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behind</td>
<td>0.63 (0.32-1.24)</td>
<td></td>
<td>0.25 (0.10-0.58)‡</td>
</tr>
<tr>
<td>Same</td>
<td>0.99 (0.46-2.12)</td>
<td></td>
<td>0.62 (0.25-1.57)</td>
</tr>
</tbody>
</table>

Model 1 includes Tanner stage as a predictor; Model 2 includes Comparative stage of pubertal development as a predictor.
MVPA (SR), moderate to vigorous physical activity, self-reported
SES, Socioeconomic Status, higher ratings indicate poorer SES
SMFQ, Short Mood and Feelings Questionnaire
TST, total sleep time.
† p < 0.05; ‡ p < 0.01
1.4.6. **Predictors of O&O**

Tanner stage and perceived comparative pubertal stage were significant predictors for female O&O, but not males (see Table 5), with increased odds of O&O with higher Tanner stage or perceiving self as ‘same’ or ‘in front’ compared to being ‘behind’ peers in pubertal development. Weekend total sleep time was a significant predictor for males but not females, with O&O males having shorter weekend TST (median 8 hours) than normal weight males (median 9 hours). Self-reported MVPA was a significant predictor for both sexes, with O&O associated with fewer days of sustained physical exercise.
Table 5. Odds of Overweight and Obesity across the four Waves.

<table>
<thead>
<tr>
<th></th>
<th>Females Odds ratio (95% confidence interval)</th>
<th>Males Odds ratio (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Age</td>
<td>0.73 (0.56-0.95)†</td>
<td>0.89 (0.74-1.07)</td>
</tr>
<tr>
<td>SES</td>
<td>1.10 (0.93-1.32)</td>
<td>1.12 (0.94-1.34)</td>
</tr>
<tr>
<td>MVPA (SR)</td>
<td>0.82 (0.73-0.93)‡</td>
<td>0.83 (0.74-0.93)‡</td>
</tr>
<tr>
<td>School night TST</td>
<td>0.89 (0.73-1.09)</td>
<td>0.86 (0.71-1.05)</td>
</tr>
<tr>
<td>Weekend TST</td>
<td>0.92 (0.77-1.11)</td>
<td>0.95 (0.80-1.14)</td>
</tr>
<tr>
<td>Tanner Stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Stage 2</td>
<td>1.45 (0.55-3.83)</td>
<td>0.84 (0.30-2.39)</td>
</tr>
<tr>
<td>Stage 3</td>
<td>2.68 (0.86-8.35)</td>
<td>1.16 (0.34-3.90)</td>
</tr>
<tr>
<td>Stage 4</td>
<td>2.67 (0.71-10.1)†</td>
<td>1.23 (0.30-5.00)</td>
</tr>
<tr>
<td>Stage 5</td>
<td>5.95 (1.30-27.3)†</td>
<td></td>
</tr>
<tr>
<td>Comparative stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behind</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same</td>
<td>2.37 (1.25-4.49)‡</td>
<td>1.31 (0.71-2.40)</td>
</tr>
<tr>
<td>In Front</td>
<td>5.48 (2.43-12.4)‡</td>
<td>2.00 (0.94-4.28)</td>
</tr>
</tbody>
</table>

Model 1 includes Tanner stage as a predictor; Model 2 includes Comparative stage of pubertal development as a predictor.

MVPA (SR), moderate to vigorous physical activity, self-reported

SES, Socioeconomic Status, higher ratings indicate poorer SES

SMFQ, Short Mood and Feelings Questionnaire

TST, total sleep time.

† p < 0.05; ‡ p < 0.01; P < 0.001
1.5. Discussion

Special features of the ARCHER study are its rural location and large sample size for a study which involves intense biological sampling and family data collection. Few other studies nationally or internationally have achieved this sample size or focussed on rural young people. Three local birth cohorts, Dunedin and Christchurch from New Zealand (Marie et al., 2008; Poulton et al., 2006), and Raine from Western Australia (McKnight et al., 2012) have less frequent review and/or limited biological data collection. The Copenhagen Puberty study includes 4,000 children, of which only 209 provided consent for six-monthly blood samples (Aksglaede et al., 2009).

The strengths of the ARCHER study are the detailed description of biological puberty, the data collection from both adolescents and their parents and the high retention in the study. An additional strength of the ARCHER study is that the data are obtained from young people growing up in regional and rural NSW, Australia. This location allows the investigators to consider additional aspects of rural health. The local community, including young people, have been involved since the conception of the study through ARCHER Community Consultative Committees and focus groups (Cooper Robbins et al., 2012). Their input has guided the study questions, methods, recruitment and retention, an additional strength of this cohort study. The ARCHER cohort is not a representative population sample. As with all recruitment into adolescent longitudinal studies with an intensive biological sampling methodology (Jeffery et al., 2012; Sorensen et al., 2010) the makeup of the sample is largely dictated by those who volunteer, and may therefore have some selectivity which may create participation bias.
The present analysis demonstrates that there is sufficient power and variability in the ARCHER study to draw conclusions about the biological effects of puberty hormones while also providing a sample that is both diverse and has the capacity to report on a range of rural and regional adolescent determinants of health, health risk factors and outcomes. The collection of the detailed annual questionnaires and anthropometry with the biological data in the ARCHER study will allow for the first time, the future investigation of the interactions of puberty hormones with the social, environmental and cultural factors that impact on and determine adolescent physical and mental health. To provide a clear description of the individual tempo of puberty the ARCHER study will use mass spectrometry-based urinary steroid assays to monitor hormone change in participant’s urine at three monthly intervals.

This cohort exhibits expected adolescent behaviours and outcomes including changing sleep patterns, increases in health risk behaviours (such as smoking and alcohol consumption), mental health symptoms and O&O. A substantial proportion of the cohort were captured in early puberty, as indicated by serum hormones and self-report. This is important in order to study the longitudinal relationship between puberty and health. We demonstrated known relationships between pubertal stage and obesity in females. Interestingly, recent reports indicate an association between higher androgens and obesity in adolescent females (Cree-Green et al., 2015). However, models in this paper only used self-reported pubertal staging and perceived comparative development. Once longitudinal serum hormone data are available, the ARCHER study will be able to investigate relationships between the timing and tempo of puberty and health outcomes.
The predictive models in this study demonstrate the complex inter-relationships between physical and social determinants of health, with household socioeconomic status an important factor for female mental health and sleep having a relationship with both mental health and O&O. Specifically, increased school night sleep was associated with lower odds of significant depressive symptoms in females and increased weekend sleep with lower odds of O&O in males. There is little known about sleep in rural adolescents, which may be affected by specific rural determinants or cultural values and attitudes towards health risk behaviours (Reichenberger et al., 2016). To the authors’ knowledge this is the first study in rural Australian adolescents.

Approximately 30% of the young participants in the ARCHER study experienced significant depressive symptoms at some stage of the study. While data indicate that rates of high psychological distress in secondary students in NSW are no higher in rural-regional than in metropolitan Local Health Districts (LHDs) (13.2% versus 13.4%), hospitalisations for intentional self-harm among 15 to 24 year olds are significantly higher (326.9 to 422.5 in regional and remote areas versus 280.9 per 100,000 population in major cities) (HealthStats NSW), making prevention of depression in young people an important rural public health issue. We confirmed the known relationship, for females, with pubertal development and depressive symptoms. Interestingly it was only perceived comparative pubertal stage that was an important predictor for males.

Another important public health issue for rural youth is O&O. Rates of O&O among secondary students aged 12 to 17 years in NSW in 2014 were higher in rural and
regional LHDs (23.0%) than in metropolitan LHDs (19.7%) (NSW HealthStats). Little is known about any specific factors that are associated with overweight and obesity in young people living in regional and rural Australia. Rural and regional areas are diverse and differ from urban areas not only from an environmental perspective but also socioeconomically and culturally. Lower population densities, fewer transport options, and reduced access to health services challenge health promotion activities that might be effective in urban areas. More information is needed about O&O in rural and regional young people to better inform interventions (Pucius, 2008). As puberty is a risk factor for obesity (Jasik and Lustig, 2008), a study that captures early adolescence has a major scientific advantage. In the current preliminary analysis, we confirmed the relationship between physical activity and O&O, and found that pubertal factors were relevant for females and that sleep, particularly on the weekend, was a significant factor for males. This cohort provides a valuable data source to determine further, unidentified factors associated with O&O in rural adolescents.

1.5.1. Conclusion

We have successfully recruited and maintained an adolescent cohort to answer novel research questions. The ARCHER study provides a unique opportunity to investigate young people as they transition through adolescence. It is already contributing to our knowledge of rural adolescent health, and will allow the future investigation of the interactions of puberty hormones with the social, environmental and cultural factors that impact on and determine adolescent physical and mental health. Data generated will further our understanding of puberty and its effects as well as providing insight
Appendix B: ARCHER Study Cohort Description

into the specific determinants of health for young people growing up in non-
metropolitan NSW and importantly will allow for the development of appropriate,
targeted interventions in at risk groups.

1.6. References

Akselgaard L, Sørensen K, Petersen JH, Skakkebaek NE, Juul A: Recent decline in age at breast

Angold A, Costello EJ, Messer SC, Pickles A, Winder F, Silver D: Development of a short
questionnaire for use in epidemiological studies of depression in children and adolescents. Int


Australian Bureau of Statistics. ‘Census of Population and Housing’, TableBuilder. Findings based on
use of ABS TableBuilder data, 2011.

Australian Department of Health. Australia’s Physical Activity and Sedentary Behaviour Guidelines
act-guidelines#/apa1317] (accessed 3 February 2016)

no. PHE 140 Canberra: AIHW, 2011.


Chaput J-P, Dutil C. Lack of sleep as a contributor to obesity in adolescents: impacts on eating and

Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight

people”: Adolescents’ Views on Engaging Young People in Longitudinal Research. J Res

Cree-Green M, Newcomer BR, Coe G, Newnes L, Baumgartner A, Brown MS, Pyle L, Reusch JE,
Nadeau KJ. Peripheral insulin resistance in obese girls with hyperandrogenism is related to
oxidative phosphorylation and elevated serum free fatty acids. Am J Physiol. 2015; 308(9):
E726-E733.

Dorn LD, Dahl RE, Rojhan Woodward H, Biro F: Defining the boundaries of early adolescence: a
user’s guide to assessing pubertal status and pubertal timing in research with adolescents.
App Dev Sci. 2006; 10:30-56.

Duke S, Balzer B, Steinbeck K. Testosterone and its effects on human male adolescent mood and

Hardy L, Diaz PE, King L, Cosgrove C, Bauman A. NSW Schools Physical Activity and Nutrition

1135:265-279.

Jeffery AN, Metcalfe BS, Hosking J, Streeter AJ, Voss LD, Wilkin TJ: Age before stage: insulin
resistance rises before the onset of puberty: a 9-year longitudinal study (EarlyBird 26).

Lawrence D, Johnson S, Hafekost J, Boterhoven De Haan K, Sawyer M, Ainley J, Zubrick SR. The
Mental Health of Children and Adolescents. Report on the second Australian Child and
Appendix B: ARCHER Study Cohort Description


APPENDIX C - REQUIREMENT FOR SPECIFIC GRAVITY AND CREATININE ADJUSTMENTS FOR URINARY STEROIDS AND LUTEINIZING HORMONE CONCENTRATIONS IN ADOLESCENTS


Supplementary Material available online:
1.1. **Abstract**

1.1.1. **Objectives**

Urinary hormone concentrations are often adjusted to correct for hydration status. We aimed to determine whether first morning void urine hormones in growing adolescents require adjustments and, if so, whether urinary creatinine or specific gravity are better adjustments.

1.1.2. **Design and methods**

The study population was adolescents aged 10.1 to 14.3 years initially who provided fasting morning blood samples at 0 and 12 months \( n = 343 \) and first morning urine every three months \( n = 644 \). Unadjusted, creatinine and specific gravity-adjusted hormonal concentrations were compared by Deming regression and Bland–Altman analysis and grouped according to self-rated Tanner stage or chronological age. \( F \)-ratios for self-rated Tanner stages and age groups were used to compare unadjusted and adjusted hormonal changes in growing young adolescents. Correlations of paired serum and urinary hormonal concentration of unadjusted and creatinine and specific gravity-adjusted were also compared.

1.1.3. **Results**

Fasting first morning void hormone concentrations correlated well and were unbiased between unadjusted or adjusted by either creatinine or specific gravity. Urine creatinine concentration increases with Tanner stages, age and male gender whereas urine specific gravity was not influenced by Tanner stage, age or gender. Adjustment by creatinine or specific gravity of urinary luteinizing hormone,
estradiol, testosterone, dihydrotestosterone and dehydroepiandrosterone concentrations did not improve correlation with paired serum concentrations.

1.1.4. Conclusions

Urine steroid and luteinizing hormone concentrations in first morning void samples of adolescents are not significantly influenced by hydration status and may not require adjustments; however, if desired, both creatinine and specific gravity adjustments are equally suitable.

1.2. Introduction

Measurements of urinary gonadotropins and steroids in children and adolescents emerged as methods to estimate pubertal development and gonadal function early in the immunoassay era (1-3). Urine sampling provides an integrated measurement especially for hormones such as luteinizing hormone (LH) (4) secreted in pulsatile manner or diurnally like sex steroids in early puberty and is more acceptable to children and adolescents than venipuncture. The relatively high hormone concentrations in urine compared with blood or saliva, together with the ability to concentrate urine, is advantageous for assays with low sensitivity or analytes at low concentrations. However, an inherent problem of using urine is the wide and unregulated variation reflecting the individual’s fluid status.

Urine dilution or concentration creates corresponding changes in urine solute concentrations so that adjustment of urine concentration may be required to avoid misinterpreting hormone excretion due to variation in hydration (5). Osmolality, specific gravity (SG) and creatinine measurements are used to adjust hydration.
Although measurement of osmolality by freezing point depression is considered the reference method, it is laborious, time consuming and expensive and so is usually replaced by SG and creatinine measurements particularly for large-scale field studies. Urinary SG is measured using a refractometer to compare light refraction of a urine sample against pure water standard or by reagent strips which measure the ionic strength of urine by colour changes. Urine SG of sample is normalized to a population reference value. While SG measurement has been largely superseded by urine creatinine adjustment in clinical laboratories, SG adjustment for urine dilution remains standard in antidoping laboratories and is used in some toxicology studies. Creatinine adjustment is based on the assumption that (a) this end-product formed endogenously from muscle creatine is released into the bloodstream and excreted in urine at a constant rate depending only on total muscle mass and (b) endogenous hormones and creatinine undergo renal excretion at the same rate. Yet, creatinine excretion rate may be influenced by the growing muscle mass during puberty leading to potential systemic errors in using creatinine adjustments.

Some but not other studies suggest creatinine or SG adjustment for measurement of urinary substances although such adjustments may be either unnecessary or even introduce additional measurement errors. Furthermore, none has focused on situations where creatinine is changing systematically due to somatic growth. Thus, the present study aimed to determine whether the first morning void hormonal assessments carried out in growing young adolescents at various stages of pubertal progression require adjustments and, if so, to determine whether creatinine or SG adjustment was better.
1.3. **Materials and Methods**

1.3.1. **Samples**

Adolescents aged 10.1 to 14.3 years initially were recruited from local secondary schools in the state of New South Wales (NSW), Australia. Ethical approval was obtained from the Human Research Ethics Committee, University of Sydney (HREC 13094). Fasting morning blood samples were collected at 0 and 12 months \((n = 343)\), and first morning urine collected three monthly after 12 h fasting at home by the adolescents between 7:00 and 8:30 \((n = 644)\). Postmenarcheal girls provided samples in the mid-follicular phase (days 7–10) with the assumption of 28 to 32-day cycle. Serum and urine samples were stored at \(-80^\circ\text{C}\) until analysis. The adolescents also provided a self-rating of puberty using line drawings based on the Tanner stages at 0 and 12 months.

1.3.2. **Assays**

Urine SG was measured by immersing a reagent strip (ChoiceLine 10, Roche Diagnostics) in freshly voided urine samples. Dipstick colour changes were compared visually with the colour chart to estimate the SG. Urine samples were subsequently stored at \(-80^\circ\text{C}\) and subjected to three freeze–thaw cycles for LH, creatinine and steroid analysis. The urine samples were first thawed and assayed for LH measurements with the Immulite 1000 LH (Siemens) as described previously (18). The within-assay coefficients of variation were <10%. The samples underwent second freeze-thaw cycle for the creatinine measurements. Urine creatinine concentrations were determined by the colorimetric alkaline-picrate (Jaffé) method (CREJ2, Roche Diagnostics, Cat. No. 04810716 190) on a Cobas C501 analyzer.
Calibrators (Roche Diagnostics, Cat. No. 10759350 190) were used for this automated system to generate a linear curve ranging between 375 and 55,000 µmol/L and the limit of detection of 375 µmol/L.

The final urine thaw was for urine steroid measurement. Urinary and serum estradiol (E2), testosterone (T), dihydrotestosterone (DHT) and dehydroepiandrosterone (DHEA) were measured by liquid chromatography tandem mass spectrometry (LC–MS/MS) as modified from a previously described method for serum (19) and adapted for urine specimens following deconjugation, with details described in the Supplementary Materials.

The LH and steroid concentrations were adjusted to standard SG of 1.020 according to the formula \[ \text{hormone concentrationsample} \times \frac{(1.020 - 1)}{\text{SGsample} - 1} \] (20) and to standard CR measurement of the present study adolescent population (12.40 mmol/L \( [n = 644; \ F \ 331] \)) using the formula \[ \text{hormone concentration}_{\text{sample}} \times \frac{\text{Creatinine}_{\text{population}}}{\text{Creatinine}_{\text{sample}}} \] where Creatinine_{population} was defined as the mean of the urinary creatinine of the whole sample.

### 1.3.3. Data Analysis

The steroid and LH concentrations unadjusted and adjusted for SG or creatinine were compared by Deming (orthogonal) regression and deviance (Bland–Altman) analysis using MedCalc software. Based on using the same analyte with different adjustments, the variance ratio in the Deming regression was assumed to be unity. Non-independence according to variations in the number of samples provided by each individual was ignored in this analysis. Descriptive statistics including mean,
standard deviation (SD) and standard error of mean (SEM) were performed by SPSS version 21. The $F$-ratios of age and Tanner stage group comparison for each hormone were calculated by one-way ANOVA separately for each gender.

1.4. Results

In first morning urine void samples ($n = 644$), the mean (SD, range) creatinine concentration was 12.4 (4.5, 1.4–31.5) mmol/L with an overall gender difference being higher in men ($P < 0.05$). The SG was 1.020 (0.0054, 1.005–1.030) without significant gender difference ($P = 0.054$). Urine creatinine concentrations were progressively increased according to chronological age and to Tanner stage (Figure 1) for both genders. Pooling genders, there were significant differences in urine creatinine concentrations by age and Tanner stage ($P < 0.05$, two-way ANOVA) but not for urine SG according to age ($P = 0.29$) or Tanner stages ($P = 0.22$) (data not shown).

Urinary LH, E2, T, DHT and DHEA concentrations, adjusted for either SG or creatinine, are compared according to Deming regression line, and the deviance plots are shown in Figure 2. For each urinary hormone concentration, there was a good correlation between the SG and creatinine-adjusted concentrations ($R^2: 0.69–0.85$) free from proportional bias between adjustment methods.

Similarly, Deming regression and Bland–Altman comparison between the unadjusted and adjusted hormone concentrations with either SG or creatinine (Table 1) also demonstrate lack of bias whether adjusted or not by either creatinine or SG.
The mean, SD and F-ratios of the unadjusted, SG-adjusted and creatinine-adjusted hormones according to Tanner stages and age groups in women and men are shown in Supplementary Tables 3 and 4 and Supplementary Tables 5 and 6, respectively. The results show consistent estimates and progression according to age and Tanner stage of unadjusted, SG-adjusted and creatinine-adjusted urine LH and steroid concentrations. The mean, SD and F-ratios of serum hormones according to Tanner stage and age groups are shown in Supplementary Tables 7 and 8.

The correlation coefficient of paired urinary and serum hormone concentrations is given in Table 2. The unadjusted and adjusted (creatinine and SG) urinary LH, E2, T, DHT and DHEA concentration showed similar correlation against serum. The samples were also grouped into three creatinine and SG percentile ranges (up to 25th percentile, between 25th to 75th percentile and above 75th percentile), and regression analysis was performed between unadjusted/adjusted urine hormone concentrations against serum hormone concentrations (data not shown). There were no improvements in the correlation coefficient values within the groups. Dividing the same percentiles according to gender also did not improve the correlations between the urine unadjusted/adjusted hormones against serum hormones concentrations (data not shown).
Figure 1. Plot of urinary creatinine measurements of adolescents groups according to age (left panel) and Tanner stage (right panel). Data represent the creatinine mean and SEM. For the age plot, samples were from 3-monthly intervals (n = 644) whereas for the Tanner stage plot, samples were from 0 and 12 months (n = 359).
**Figure 2.** Comparison of urinary LH, E₂, T, DHT and DHEA concentrations adjusted by SG and creatinine. Comparison was made according to Deming regression analysis (left panels) and Bland–Altman plots (right panels). For the Deming plots, the slope is shown as a solid line and line of identity in fine dotted line. Insets are the regression formula and 95% confidence limits on the intercept and slope. The Bland–Altman plots represent the differences between creatinine and SG-adjusted hormone concentrations against the averages of the hormone concentrations adjusted with the two correction method. The solid line and the dashed lines represent the observed average and the 95% limit of confidence (±1.96 SD), respectively.
Table 1. Comparison of unadjusted against SG- and creatinine-adjusted urinary hormone measurements.

<table>
<thead>
<tr>
<th></th>
<th>UA vs SG</th>
<th></th>
<th></th>
<th>UA vs CR</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Deming Regression</td>
<td>Bland-Altman</td>
<td>Deming Regression</td>
<td>Bland-Altman</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slope (95% CI)</td>
<td>Intercept (95% CI)</td>
<td>r² (95% CI)</td>
<td>Mean</td>
<td>95% CI</td>
<td>Slope (95% CI)</td>
</tr>
<tr>
<td>LH (IU/L)</td>
<td>0.94 (0.84 to 1.03)</td>
<td>0.03 (−0.67 to 0.73)</td>
<td>0.91 (0.90 to 0.92)</td>
<td>0.6</td>
<td>8.3, −7.2</td>
<td>1.03 (0.94 to 1.11)</td>
</tr>
<tr>
<td>E2 (nmol/L)</td>
<td>0.92 (0.75 to 1.09)</td>
<td>0.02 (−0.19 to 0.22)</td>
<td>0.92 (0.91 to 0.94)</td>
<td>0.4</td>
<td>7.4, −6.6</td>
<td>1.08 (0.93 to 1.22)</td>
</tr>
<tr>
<td>T (nmol/L)</td>
<td>0.98 (0.89 to 1.06)</td>
<td>−0.17 (−0.75 to 0.42)</td>
<td>0.95 (0.94 to 0.96)</td>
<td>01.4</td>
<td>40, −37</td>
<td>1.20 (1.08 to 1.31)</td>
</tr>
<tr>
<td>DHT (nmol/L)</td>
<td>1.12 (0.74 to 1.50)</td>
<td>−0.63 (−1.94 to 0.68)</td>
<td>0.95 (0.94 to 0.95)</td>
<td>0.7</td>
<td>12, −11</td>
<td>1.53 (0.71 to 2.35)</td>
</tr>
<tr>
<td>DHEA (nmol/L)</td>
<td>0.90 (0.81 to 0.99)</td>
<td>0.65 (−0.92 to 2.22)</td>
<td>0.87 (0.85 to 0.89)</td>
<td>5.2</td>
<td>62, −52</td>
<td>1.07 (0.94 to 1.20)</td>
</tr>
</tbody>
</table>

The slope, intercept and 95% confidence interval were determined by Deming regression. The mean and 95% CI (±1.96 SD) were derived from the Bland–Altman plots.

UA: unadjusted; SG: specific gravity adjusted; CR: creatinine adjusted; $R^2$: correlation of determination; CI: confidence interval; E2: serum estradiol; T: testosterone; DHT: dihydrotestosterone; DHEA: dehydroepiandrosterone; LH: luteinizing hormone. For the Deming regression, the variance ratio was assumed to be unity.

*All the P values were <0.0001.
Table 2. Pearson’s correlation coefficient and confidence intervals (in parentheses) of paired urinary and serum LH, E2, T, DHT and DHEA (n = 343).

<table>
<thead>
<tr>
<th>Urine vs. serum</th>
<th>LH</th>
<th>E₂</th>
<th>T</th>
<th>DHT</th>
<th>DHEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>UA</td>
<td>0.56 (0.48–0.63)</td>
<td>0.73 (0.67–0.77)</td>
<td>0.79 (0.74–0.82)</td>
<td>0.44 (0.35–0.52)</td>
<td>0.63 (0.56–0.69)</td>
</tr>
<tr>
<td>SG</td>
<td>0.57 (0.50–0.64)</td>
<td>0.72 (0.66 to 0.77)</td>
<td>0.79 (0.74–0.83)</td>
<td>0.42 (0.33–0.50)</td>
<td>0.60 (0.52–0.66)</td>
</tr>
<tr>
<td>CR</td>
<td>0.56 (0.48–0.63)</td>
<td>0.79 (0.74–0.82)</td>
<td>0.80 (0.76–0.84)</td>
<td>0.35 (0.25–0.44)</td>
<td>0.66 (0.59–0.71)</td>
</tr>
</tbody>
</table>

UA: unadjusted; SG: specific gravity adjusted; CR: creatinine adjusted; E₂: serum estradiol; T: testosterone; DHT: dihydrotestosterone; DHEA: dehydroepiandrosterone; LH: luteinizing hormone.
1.5. Discussion

Urinary measurement of reproductive hormones is a convenient means to evaluate pubertal status and gonadal function for field population studies. In clinical settings, adjustment based on the assumption of stable urine creatinine excretion is commonly used to adjust for variations in hydration although other techniques such as regression normalization or log transformation are proposed (10, 21). As an end metabolite of muscle creatine, urine creatinine is determined by total muscle mass in addition to other factors such as age, gender, diet (meat consumption), physical activity and body mass index, some of which exert their effects via changes in muscle mass (6, 17, 22, 23). Hence, one aim of the present study was to determine for the first time whether creatinine adjustment was valid or required for longitudinal studies of growing adolescents.

Our findings confirm that the first morning urine creatinine concentration increases with age and Tanner stages and was higher among men. However, adjustment for urine creatinine was no better or worse than adjustment for SG or even no adjustment. This may reflect the fact that we studied first morning void urine samples which control hydration, whereas similar interpretation may not apply to urine sampled at random when hydration state may vary more. Our findings are consistent with previous studies showing prominent intra- and interindividual variability in creatinine excretion of second morning and 24 h urine samples in adults due to variable fluid intake (17). Significantly higher creatinine concentrations in morning versus afternoon (24), in evening spot samples (6) and creatinine loss due to multiple freeze–thaw cycles have also been reported (25, 26) all of which introduce systematic errors in use of urine creatinine for dilution adjustments. Thus, although
studies have suggested alternative adjustment based on SG in adult humans and primates (27, 28), none has focused on the need for SG adjustments in first morning voids of growing adolescents.

SG is readily measured by reagent strip for field studies without needing a laboratory. Previous studies demonstrate good agreement between SG measurements by reagent strip versus refractometer (29, 30) or osmolality (30-32). SG measurement by reagent strip is widely used in clinical applications (32, 33). Although refractometer urine SG may be influenced by disease states leading to high serum protein or glycosuria (7, 34), reagent strip SG is not affected by glucose, only minimally by urea and albumin, but may be affected by the rare instances of alkaline urine (31). Urine SG reading may also be influenced by diet, environment and the renal reabsorption capacity (35). Among adolescents, we find that urine SG measured with reagent strips is systematically not influenced by age or gender consistent with previous reports (22, 36).

Limits of acceptable creatinine and SG measurements vary between studies. Generally, urine is considered too dilute when the SG and creatinine concentrations are lower than 1.010 and 0.5 g/L (4.4 mmol/L), respectively, and too concentrated where SG and creatinine concentrations higher than 1.030 (or 1.035) and 3 g/L (26.5 mmol/L), respectively (17, 35). However, due to the standardized method of collection and hydration (first morning void), the present study did not discard any samples as too dilute or too concentrated.

The present study demonstrated that the fasting first morning void urine hormone concentrations adjusted by creatinine correlated well with those adjusted by SG in
this adolescent population. This is consistent with previous reports that used randomly collected or timed urine collection from children and adults showing good correlations when creatinine and SG adjustments were compared directly (5, 12, 21, 23, 35, 37) or with adjustment according to both (5, 27, 28) including a reduced variation using these adjustments in some studies (10, 12). However, the present study shows that neither of the adjustment methods for first morning void urine sample of adolescents were significantly improved compared with unadjusted hormone concentrations. These observations are consistent with previous reports for creatinine adjustment of urine steroid measurements in adult women (38, 39).

In studies where the urinary hormone concentrations were correlated with paired circulating serum concentrations, the urinary unadjusted concentration or concentration expressed by volume of urine correlates better than the adjustment based on analyte to creatinine ratios (11, 13, 40), although some studies have shown improved correlation with creatinine adjustments (41, 42). The present study demonstrated that the urinary hormone concentrations adjusted with creatinine, and SG did not improve the correlation with paired serum concentrations. These samples were also grouped into three creatinine and SG percentile ranges (25th, 25–75th and 75th) to replicate non-fasting conditions with wider variation in hydration status. However, no improvement was observed in terms of correlation between the unadjusted or adjusted urine hormone and paired serum concentrations. These findings further support that the adjustments may not be necessary for first morning void urine samples.

In conclusion, the present study shows that adjustment of urinary steroid and LH concentration for hydration state may not be required for first morning void
specimens of even growing adolescents. If adjustments are required, then either creatinine or SG is equally suitable and provides comparable results. Reagent strip SG measurements are simple and sufficiently reliable, economical and time-saving for large numbers of urine sampling in long-term field studies.

1.6. References


APPENDIX D: SUPPLEMENTARY MATERIAL

FOR CHAPTER 3

Supplement A: Search Strategy

MEDLINE

1. Adolescent/
2. schools/ or students/
3. exp Puberty/
4. (adol* or teen* or juvenile* or youth* or student*).tw.
5. pubert*.tw.
6. 1 or 2 or 3 or 4 or 5
7. Adolescent Psychology/ or Adolescent Psychiatry/
8. behav*.tw.
9. adolescent behaviour/ or Behaviour/ or behavioural symptoms/ or affective symptoms/ or aggression/ or agonistic behaviour/ or bullying/ or depression/ or self-injurious behaviour/ or self mutilation/ or stress, psychological/ or drinking behaviour/ or alcohol drinking/ or drug-seeking behaviour/ or impulsive behaviour/ or risk reduction behaviour/ or risk-taking/ or social behaviour/ or Risk-Taking/ or Accidents/ or accidents, home/ or accidents, traffic/ or Dangerous Behaviour/ or impulsive behaviour/ or compulsive behaviour/ or behaviour, addictive/
10. (Social* adj3 (behav* or conform* or adjustment or dominan*)).tw.
11. (Behav* adj3 (competitive or cooperative)).tw.
12. (Risk adj3 (taking or behav*)).tw.
13. (impulsiv* or dangerous* or dangerous behav* or hazardous or delinqu* or antisocial behav* or conduct disorder* or oppositional defian*).tw.
14. exp Substance-Related Disorders/
15. (drug taking or drug abuse or smok* or tobacco or alcohol or addiction* or substance abuse or drug dependenc*).tw.
16. (accident* or crash* or traffic accident*).tw.
17. exp Self Concept/
18. Body Image/ or personal autonomy/
19. (Self adj3 (concept* or image* or esteem or perception*)).tw.
20. exp aggression/ or bullying/
21. Violence/ or Juvenile Delinquency/ or Student Dropouts/ or Social Behaviour Disorders/
22. (aggress* or violen* or bully* or bullies).tw.
23. mental disorders diagnosed in childhood/ or "attention deficit and disruptive behaviour disorders"/ or child behaviour disorders/ or "feeding and eating disorders of childhood"/ or Antisocial Personality Disorder/ or Conduct Disorder/
24. social behaviour/ or aggression/ or competitive behaviour/ or cooperative behaviour/ or helping behaviour/ or shyness/ or social dominance/ or social identification/ or social isolation/ or social stigma/
25. exp Self-injurious behaviour/
26. exp Suicide/
27. emotions/ or affect/ or irritable mood/ or anger/ or rage/ or anxiety/ or apathy/ or boredom/ or happiness/ or hate/ or hostility/
28. (emotion* or mood* or bored* or hostil* or apath* or frustrat*).tw.
29. exp mood disorders/
30. mental disorders/ or Depression/ or anxiety disorders/ or eating disorders/ or mood disorders/ or sleep disorders/ or substance-related disorders/
31. (depress* or suicid* or parasuicid* or self harm* or self injur* or self destruct* or self mutilat*).tw.
32. (anxiet* or nervous* or anxious).tw.
33. motivation/ or achievement/ or "conflict (psychology)"/ or drive/ or goals/ or "power (psychology)"/
34. (motivat* or ambition*).tw.
35. exp Sleep Disorders/
36. sleep/ or sleep disorders/ or dyssomnias/ or sleep deprivation/ or sleep disorders, circadian rhythm/ or sleep disorders, intrinsic/
37. (sleep* or insomnia or sleep disorder or late waking).tw.
38. or/7-37
39. exp Estrogens/
40. exp Oestradiol/
41. (estrogen or oestradiol or ?oestradiol or ?estrogen*).mp.
42. 39 or 40 or 41
43. 6 and 38 and 42
65. limit 64 to humans
Supplement B: Table of Excluded Studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attie et al.</td>
<td>1989</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Avgoustinaki et al.</td>
<td>2012</td>
<td>Subjects outside of age range; no methodology for hormone measurement provided</td>
</tr>
<tr>
<td>Baker et al.</td>
<td>2007</td>
<td>Oestradiol was combined with other markers of pubertal development; no analysis</td>
</tr>
<tr>
<td>Balada et al.</td>
<td>1993</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Barrac et al.</td>
<td>2010</td>
<td>Subjects were a small number of elite runners and thus not a representative community sample</td>
</tr>
<tr>
<td>Benjet et al.</td>
<td>2001</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Benjet et al.</td>
<td>2002</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Blyth et al.</td>
<td>1985</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Boettinger et al.</td>
<td>2010</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Brambilla et al.</td>
<td>2001</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Brooker et al.</td>
<td>2012</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Bruinsma et al.</td>
<td>2006</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Chandrashekar et al.</td>
<td>2001</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Colzato et al.</td>
<td>2012</td>
<td>Outside age range; assessed cognitive, not behavioural outcomes</td>
</tr>
<tr>
<td>Cotrufo et al.</td>
<td>2000</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Culbert et al.</td>
<td>2011</td>
<td>Review article</td>
</tr>
<tr>
<td>Daitzman et al.</td>
<td>1980</td>
<td>Male only</td>
</tr>
<tr>
<td>de Water et al.</td>
<td>2013</td>
<td>Females on oral contraceptive pill included (continuous contraception e.g. Mirena excluded)</td>
</tr>
<tr>
<td>DeBruine et al.</td>
<td>2005</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Deng et al.</td>
<td>2011</td>
<td>Outside age range; looking at menstrual cycle changes</td>
</tr>
<tr>
<td>DeRose et al.</td>
<td>2011</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Dick et al.</td>
<td>2001</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Dorgan et al.</td>
<td>2003</td>
<td>Oestradiol was a dependent variable</td>
</tr>
<tr>
<td>Drapela et al.</td>
<td>2006</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Dubas et al.</td>
<td>1991</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Duncan et al.</td>
<td>1985</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Durante et al.</td>
<td>2009</td>
<td>Assertive mating is not a relevant behaviour</td>
</tr>
<tr>
<td>Edelstein et al.</td>
<td>2012</td>
<td>Oestradiol was a dependent variable</td>
</tr>
<tr>
<td>Edelstein et al.</td>
<td>2010</td>
<td>One third of females on oral contraceptive pill; behaviour is sexual</td>
</tr>
<tr>
<td>Edwards et al.</td>
<td>2011</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Ehrhardt et al.</td>
<td>1981</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Finkelstein et al.</td>
<td>1997</td>
<td>Exogenous oestradiol used</td>
</tr>
<tr>
<td>Fujisawa et al.</td>
<td>2012</td>
<td>Outcome of interest, loneliness, was not considered a behaviour or mood</td>
</tr>
<tr>
<td>Ge et al.</td>
<td>2002</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Ge et al.</td>
<td>2001</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Gearing et al.</td>
<td>2001</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Goddings et al.</td>
<td>2012</td>
<td>Neuroimaging study without validated behavioural or affective measure</td>
</tr>
<tr>
<td>Hayward et al.</td>
<td>2002</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Inoff-Germain et al.</td>
<td>1988</td>
<td>Subjective assessment of behaviour</td>
</tr>
<tr>
<td>Joinson et al.</td>
<td>2012</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Klump et al.</td>
<td>2006</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Klump et al.</td>
<td>2008</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Lazarro et al.</td>
<td>1996</td>
<td>Eating disorder cohort with no controls (oestradiol is affected by malnutrition)</td>
</tr>
<tr>
<td>Llewellyn et al.</td>
<td>2012</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Mendle et al.</td>
<td>2012</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Pahlen et al.</td>
<td>2005</td>
<td>Review</td>
</tr>
<tr>
<td>Paikoff et al.</td>
<td>1991</td>
<td>Oestradiol and behaviour not contemporaneous</td>
</tr>
<tr>
<td>Pajer et al.</td>
<td>2006</td>
<td>Some subjects on oral contraceptive pill</td>
</tr>
<tr>
<td>Reynolds et al.</td>
<td>2011</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Reynolds et al.</td>
<td>2012</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Riecher-Rössler et al.</td>
<td>1994</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Schellerman</td>
<td>2011</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Offermans et al.</td>
<td>2005</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Study</td>
<td>Year</td>
<td>Issue</td>
</tr>
<tr>
<td>---------------------</td>
<td>------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Schiller et al.</td>
<td>2012</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Schwartz et al.</td>
<td>2012</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Smiarowska et al.</td>
<td>2002</td>
<td>No control group</td>
</tr>
<tr>
<td>Soldin et al.</td>
<td>2011</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Susman et al.</td>
<td>1996</td>
<td>Oestradiol was a dependent variable</td>
</tr>
<tr>
<td>Swart et al.</td>
<td>1996</td>
<td>No oestradiol measurement</td>
</tr>
<tr>
<td>Vermeersch et al.</td>
<td>2008</td>
<td>Male only</td>
</tr>
</tbody>
</table>
APPENDIX E: SUPPLEMENTARY MATERIAL

FOR CHAPTER 4

Supplement A: Search Strategy for MEDLINE

1. exp adolescent/
2. exp schools/
3. exp students/
4. exp puberty/
5. (adol* or teen* or juvenile* or youth*).tw.
6. pubert*.tw.
7. (student* or school-student* or high-school student*).tw.
8. (school* or high-school*).tw.
9. or/1-8
10. exp "behavior and behavior mechanisms"/ or exp "psychological phenomena and processes"/ or exp "behavioral disciplines and activities"/
11. exp Substance-Related Disorders/
12. exp Sleep Disorders/
13. exp Accidents, Traffic/ or exp Accidents/ or exp Accidents, Home/
14. behav*.tw.
15. (self concept* or self image* or self esteem or self perception* or body image or personal autonomy).tw.
16. (risk taking or risk behav*).tw.
17. (aggress* or violen* or bully* or bullies or cyber bull*).tw.
18. depress*.tw.
19. (suicid* or parasuicid* or self harm* or self injur* or self destruct* or self mutilation).tw.
20. (anxiet* or nervous*).tw.
21. (dangerous* or dangerous behav* or hazardous).tw.
22. (impulsiv* or impulsive behav*).tw.
23. (emotion* or moodiness or boredom or hostilit* or apath* or frustrat*).tw.
24. (drug taking or drug abuse or smok* or tobacco or alcohol or addiction* or substance abuse or drug dependenc*).tw.
25. (motivat* or ambition*).tw.
26. (self concept* or self discrepane*).tw.
27. (delinqu* or antisocial behav* or conduct disorder* or oppositional defian*).tw.
28. (social behav* or social conformity or social adjustment or social dominance or competitive behav* or cooperative behav*).tw.
29. (sleep* or insomnia or sleep disorder or late waking).tw.
30. (accident* or crash* or traffic accident*).tw.
31. or/10-30
32. exp Testosterone/
33. exp androgens/
34. testosterone.tw.
35. androgen*.tw.
36. or/32-35
37. 9 and 31 and 36
38. limit 37 to humans
### Supplement B: Characteristics of excluded studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angold, A, et al.</td>
<td>1999</td>
<td>Time of testosterone collection not documented</td>
</tr>
<tr>
<td>Banks, T, et al.</td>
<td>1996</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Bauman, KE, et al.</td>
<td>1989</td>
<td>Time of testosterone collection not documented</td>
</tr>
<tr>
<td>Biro, FM, et al.</td>
<td>2006</td>
<td>Review article</td>
</tr>
<tr>
<td>Book, AS, et al.</td>
<td>2001</td>
<td>Meta-analysis; all age groups</td>
</tr>
<tr>
<td>Brambilla, F, et al.</td>
<td>2001</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Bricker, J</td>
<td>2011</td>
<td>Focus on sexual behavior</td>
</tr>
<tr>
<td>Bricker, JB, et al.</td>
<td>2006</td>
<td>Focus on sexual behavior</td>
</tr>
<tr>
<td>Brooks, JH, et al.</td>
<td>1996</td>
<td>Young offenders: no normal controls</td>
</tr>
<tr>
<td>Buchanan, CM</td>
<td>1989</td>
<td>No specific moods reported</td>
</tr>
<tr>
<td>Cafri, G, et al.</td>
<td>2006</td>
<td>No measurement of testosterone</td>
</tr>
<tr>
<td>Campbell, A, et al.</td>
<td>1997</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Christiansen, K</td>
<td>2001</td>
<td>Review</td>
</tr>
<tr>
<td>Cota-Robles, S, et al.</td>
<td>2002</td>
<td>No measurement of testosterone</td>
</tr>
<tr>
<td>Dabbs, JM, et al.</td>
<td>1991</td>
<td>Young offenders: no normal controls</td>
</tr>
<tr>
<td>Daitz, RJ, et al.</td>
<td>1978</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Dmitrieva, TN, et al.</td>
<td>2001</td>
<td>No data for relationship in control group</td>
</tr>
<tr>
<td>Dorn, LD, et al.</td>
<td>2003</td>
<td>No measures of behavior</td>
</tr>
<tr>
<td>Dorn, LD, et al.</td>
<td>1999</td>
<td>No measurement of testosterone</td>
</tr>
<tr>
<td>Dorn, LD, et al.</td>
<td>2009</td>
<td>No data for relationship in control group; outside age range</td>
</tr>
<tr>
<td>Drigotas, SM, et al.</td>
<td>1993</td>
<td>Testosterone level and behavior not contemporaneous</td>
</tr>
<tr>
<td>Felson, RB, et al.</td>
<td>2002</td>
<td>No measurement of testosterone</td>
</tr>
<tr>
<td>Field, T</td>
<td>2002</td>
<td>Review</td>
</tr>
<tr>
<td>Forbes, EE, et al.</td>
<td>2010</td>
<td>No direct data between testosterone and behavior, (testosterone and fMRI)</td>
</tr>
<tr>
<td>Fujisawa, TX, et al.</td>
<td>2011</td>
<td>No measurement of testosterone</td>
</tr>
<tr>
<td>Gallup, AC, et al.</td>
<td>2010</td>
<td>No measurement of testosterone</td>
</tr>
<tr>
<td>Gallup, AC, et al.</td>
<td>2007</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Granger, DA, et al.</td>
<td>1994</td>
<td>Letter which did not present original data</td>
</tr>
<tr>
<td>Halpern, CT, et al.</td>
<td>1992</td>
<td>Study of testosterone</td>
</tr>
<tr>
<td>Halpern, CT, et al.</td>
<td>1994</td>
<td>Focus on sexual behavior</td>
</tr>
<tr>
<td>Halpern, CT, et al.</td>
<td>1993</td>
<td>Focus on sexual behavior</td>
</tr>
<tr>
<td>Halpern, CT, et al.</td>
<td>1997</td>
<td>Focus on sexual behavior</td>
</tr>
<tr>
<td>Halpern, CT, et al.</td>
<td>1998</td>
<td>Focus on sexual behavior</td>
</tr>
<tr>
<td>Hassler, M</td>
<td>1991</td>
<td>Cognitive rather than behavioral outcomes</td>
</tr>
<tr>
<td>Hassler, M</td>
<td>1991</td>
<td>Cognitive rather than behavioral outcomes</td>
</tr>
<tr>
<td>Hassler, M</td>
<td>1992</td>
<td>Cognitive rather than behavioral outcomes</td>
</tr>
<tr>
<td>Hassler, M</td>
<td>1992</td>
<td>Cognitive rather than behavioral outcomes</td>
</tr>
<tr>
<td>Kaiser, H, et al.</td>
<td>2006</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Kerschbaum, HH, et al.</td>
<td>2006</td>
<td>Time of testosterone collection not documented; outside age range</td>
</tr>
<tr>
<td>Kreuz, LE, et al.</td>
<td>1972</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Kulik, HE, et al.</td>
<td>2000</td>
<td>Induction of puberty with exogenous androgen therapy</td>
</tr>
<tr>
<td>Laznišatová, J</td>
<td>2004</td>
<td>Cognitive rather than behavioral outcomes</td>
</tr>
<tr>
<td>Laznišatová, J</td>
<td>2001</td>
<td>Cognitive rather than behavioral outcomes</td>
</tr>
<tr>
<td>Li, C, et al.</td>
<td>1993</td>
<td>Outside age range</td>
</tr>
<tr>
<td>Maras, A, et al.</td>
<td>2003</td>
<td>Not a community sample</td>
</tr>
<tr>
<td>Maras, A, et al.</td>
<td>2003</td>
<td>Not a community sample</td>
</tr>
<tr>
<td>Martin, CA, et al.</td>
<td>2006</td>
<td>ADHD, no normal controls</td>
</tr>
<tr>
<td>Martin, CA, et al.</td>
<td>1999</td>
<td>No data on testosterone alone</td>
</tr>
<tr>
<td>Mattsson, A, et al.</td>
<td>1980</td>
<td>Young offenders only: no normal controls</td>
</tr>
<tr>
<td>McDermott, K, et al.</td>
<td>1994</td>
<td>Review article</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Year</td>
<td>Title/Notes</td>
</tr>
<tr>
<td>---------------------</td>
<td>------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Nerozzi, D</td>
<td>1984</td>
<td>Psychologically disturbed adolescents: no normal controls</td>
</tr>
<tr>
<td>Newman, ML</td>
<td>2004</td>
<td>PhD dissertation, outside age range</td>
</tr>
<tr>
<td>Nottelmann, ED, et al.</td>
<td>1990</td>
<td>Review</td>
</tr>
<tr>
<td>Ostatnikova, D, et al.</td>
<td>2000</td>
<td>Cognitive rather than behavioral outcomes</td>
</tr>
<tr>
<td>Paikoff, RL, et al.</td>
<td>1991</td>
<td>Testosterone level and behavior not contemporaneous,</td>
</tr>
<tr>
<td>Popma, A, et al.</td>
<td>2007</td>
<td>Delinquent boys: no normal control group</td>
</tr>
<tr>
<td>Post, JM</td>
<td>2010</td>
<td>Comment</td>
</tr>
<tr>
<td>Reynolds, MD, et al.</td>
<td>2007</td>
<td>Testosterone level and behavior not contemporaneous</td>
</tr>
<tr>
<td>Rowe, R, et al.</td>
<td>2004</td>
<td>Time of testosterone collection not documented</td>
</tr>
<tr>
<td>Scerbo, AS, et al.</td>
<td>1994</td>
<td>Disruptive children only: no normal controls</td>
</tr>
<tr>
<td>Schaal, B, et al.</td>
<td>1995</td>
<td>Letter</td>
</tr>
<tr>
<td>Sinkkonen, J, et al.</td>
<td>1998</td>
<td>Time of testosterone collection not documented</td>
</tr>
<tr>
<td>Smith, EA, et al.</td>
<td>1985</td>
<td>Focus on sexual behavior</td>
</tr>
<tr>
<td>Strong, RK, et al.</td>
<td>2000</td>
<td>Male and female data pooled for correlation between testosterone and behavior</td>
</tr>
<tr>
<td>Susman, EJ, et al.</td>
<td>1996</td>
<td>Effect of aggression and impulse control on testosterone</td>
</tr>
<tr>
<td>Susman, EJ, et al.</td>
<td>1988</td>
<td>Review</td>
</tr>
<tr>
<td>Udry, J</td>
<td>1991</td>
<td>Time of testosterone collection not documented</td>
</tr>
<tr>
<td>Udry, JR, et al.</td>
<td>1985</td>
<td>Focus on sexual behavior</td>
</tr>
<tr>
<td>Updegraff KA et al.</td>
<td>2006</td>
<td>Focus on romantic relationships</td>
</tr>
<tr>
<td>van Bokhoven, I, et al.</td>
<td>2006</td>
<td>Behaviors averaged across ages</td>
</tr>
<tr>
<td>van Gozzen, SH, et al.</td>
<td>2000</td>
<td>No measurement of testosterone</td>
</tr>
<tr>
<td>Vermeersch, H, et al.</td>
<td>2010</td>
<td>Data of testosterone and behavior previously reported in Vermeersch 2008</td>
</tr>
</tbody>
</table>