Assessing Pubertal Timing and Dynamics in a Large Longitudinal Sample of Adolescent Males

by

Ammar Khairullah

A thesis submitted in conformity with the requirements for the degree of Masters of Science
Institute of Medical Science
University of Toronto

© Copyright by Ammar Khairullah 2013
Assessing Pubertal Timing and Dynamics in a Large Longitudinal Sample of Adolescent Males

Ammar Khairullah

Masters of Science
Institute of Medical Science
University of Toronto
2013

Abstract

Pubertal timing and dynamics play an important role in physical and psychological development. This thesis describes three height-based indices that can be used to assess pubertal timing, and a measure of average exposure to testosterone derived from testosterone trajectories to assess pubertal dynamics. These parameters are derived in a sample of 500 male adolescents from the Avon Longitudinal Study of Parents and Children (ALSPAC). The indices of pubertal timing are shown to correlate with one another in ALSPAC. In a separate cross-sectional study, we show the association between one index and both testosterone levels and self-reported pubertal stage. The measure of average exposure to testosterone characterizes the timing and the magnitude of the rise in participants’ testosterone levels during adolescence. Future investigations may employ the methods outlined to account for timing and dynamics of puberty in relation to adolescent development.
Acknowledgments

I would like to take this opportunity to thank Dr. Tomáš Paus for the opportunity to work on an exciting project that posed many educational challenges, and for mentoring me as I navigated through this degree. Thank you to Dr. Mallar Chakravarty and Dr. Mark Palmert for the insightful questions and discussions during committee meetings.

Thank you to the members of the Paus lab: Rosanne Aleong, Marc Berman, Erin Dickie, Eileen Ding, Leon French, Courtney Gray, Klara Mareckova, Melissa Pangelinan, Marzia Pesaresi, Nick Qiu, Deborah Schwartz, Angelita Wong. The unmatched camaraderie and support I found from our lab makes for a great training environment.

I wish to acknowledge the Baycrest Centre for Geriatric Care and the University of Toronto, and the people within who helped me with so many things over the course of my studies. Thank you to our collaborators at the University of Bristol and the Pennsylvania State University for the important part you played in this work. I would like to thank the government of Ontario, the Wiseman family, and the Institute of Medical Science for financial support.

Last but not least, a big thank-you to my family: Dad, Mom and Khadija. I can always count on getting support, good advice, and words of motivation from all of you. And Janet, thanks for patiently listening and always knowing the right things to say when I needed to hear them.
Contributions

Dr. Tomáš Paus originated the ideas for this research, and holds the NIH grant that supports this project. Staff employed by the Avon Longitudinal Study of Parents and Children as well as the Saguenay Youth Study collected the data presented in this thesis. Dr. Laura Klein, Dr. Courtney Whetzel and their staff at the Pennsylvania State University conducted ELISAs for hormone quantification. Dr. Suzanne Ingle at the University of Bristol provided us with time-corrected testosterone values through applying multilevel modeling techniques. Dr. Melissa Pangelinan facilitated my learning of MATLAB. Drs. Erin Dickie and Shafagh Fallah provided statistical advice, as did Dr. Margaret May (University of Bristol). Dr. Rosanne Aleong facilitated communication between the different sites working on this project. Drs. Tomáš Paus, Margaret May, Suzanne Ingle, Kate Tilling, Laura Howe, Laura Klein, Courtney Whetzel, Elizabeth Susman, Zdenka Pausova, Louis Richer, Gabriel Leonard, Michel Perron, and Suzanne Veillette provided feedback on the manuscripts.
# Table of Contents

Title Page .........................................................................................................................i

Abstract ..........................................................................................................................ii

Acknowledgments ....................................................................................................... iii

Contributions ............................................................................................................... iv

Table of Contents .......................................................................................................... v

List of Tables ................................................................................................................. x

List of Figures .............................................................................................................. xi

Chapter 1 – Introduction .............................................................................................. 1

1.1 Puberty ......................................................................................................................... 1

1.2 Pubertal Timing ........................................................................................................... 2

1.3 The Hypothalamic-Pituitary-Gonadal Axis ................................................................. 3

1.4 Initiation of Puberty ..................................................................................................... 4

1.5 Testosterone Production ............................................................................................ 6

1.6 Testosterone in the Blood Stream .............................................................................. 6

1.7 Mechanisms of Testosterone Action .......................................................................... 7

1.8 Testosterone Levels During Male Puberty ................................................................. 8

1.9 The Effect of Testosterone on Development ............................................................ 8

1.10 Changes Associated with Pubertal Status ............................................................... 11

1.11 Changes Associated with Pubertal Timing ............................................................... 12

1.12 Patterns of Pubertal Changes in Males .................................................................. 13

1.13 The Evolution of Pubertal Status Assessment ......................................................... 14
Chapter 2 - Height-based Indices of Pubertal Timing in Male Adolescents

2.1 Abstract

2.2 Introduction

2.3 Methods

2.3.1 The Avon Longitudinal Study of Parents And Children

2.3.2 Age at Peak Height Velocity (APHV)

2.3.3 Height Difference in Standard Deviations (HDSDS)

2.3.4 Percent Achieved of Adult Stature (PAAS)

2.3.5 Interpretation of the Indices of Pubertal Timing

2.3.6 Using Quintiles to Investigate Concordance of Height-Based Pubertal Timing Indices

2.3.7 The Saguenay Youth Study

2.3.8 Anthropometry and PAAS Scores

2.3.9 Pubertal Development Scale

2.3.10 Serum Testosterone

2.3.11 Statistical Analyses

2.4 Results

2.4.1 Correlation between three height-based pubertal timing indices in ALSPAC
2.4.2 Concordance of Indices using Quintiles ........................................................................ 40
2.4.3 Association of the Percent Achieved of Adult Stature Index with Measures of Pubertal Status ................................................................................................................. 40
2.4.4 Comparing Performance of the Pubertal Timing Indices in the Two Studies ............ 43

2.5 Discussion ................................................................................................................. 47
2.5.1 Association of the PAAS Index with Measures of Pubertal Status ......................... 48
2.5.2 Age at Peak Height Velocity .................................................................................. 49
2.5.3 Height Difference in Standard Deviations .............................................................. 50
2.5.4 Percent Achieved of Adult Stature ........................................................................ 51
2.5.5 Comparison of PAAS to Pubertal Questionnaires ................................................. 52
2.5.6 Other Considerations ............................................................................................ 53

2.6 Acknowledgements ................................................................................................... 55

2.7 Supplementary Section: ............................................................................................ 56

Chapter 3 - Testosterone trajectories and reference ranges in a large longitudinal sample of male adolescents ................................................................................................................. 57

3.1 Abstract ...................................................................................................................... 58

3.2 Introduction ............................................................................................................... 59

3.3 Methods ...................................................................................................................... 60
3.3.1 Meta-analysis of Published Data ........................................................................... 60
3.3.2 The Avon Longitudinal Study of Parents and Children ........................................ 61
3.3.3 Quantification of Testosterone and Sex Hormone Binding Globulin ...................... 62
3.3.4 Testosterone’s Circadian Rhythm ......................................................................... 63
3.3.5 Measure of Average Exposure to Testosterone Throughout Adolescence .......... 64
3.3.6 Growth spurt and Characteristics of Testosterone Trajectories ............................. 65
3.3.7 Free and Bioavailable Testosterone ...................................................................... 65
3.3.8 Statistics ............................................................................................................... 66
5.5 Preliminary Findings in the ALSPAC sample .................................................. 100

5.6 Testosterone, White Matter, and Psychopathology ........................................... 101

References ............................................................................................................. 103
List of Tables

| Table 2-1 – Descriptive Statistics, Distribution Characteristics, and Intercorrelations of Pubertal Timing Indices in the Avon Longitudinal Study of Parents and Children | 39 |
| Table 2-2 – Range of Quintiles for each Pubertal Timing Index | 41 |
| Table 2-3 - Descriptive Statistics and Distribution Characteristics of the Saguenay Youth Study Sample | 44 |
| Table 3-1 - Meta-Analysis of Total Testosterone Values | 68 |
| Table 3-2 - Meta-Analysis including ALSPAC data | 69 |
| Table 3-3 - Hormone Data from the Avon Longitudinal Study of Parents and Children | 70 |
List of Figures

Figure 2-1 – Concordance between Indices of Pubertal Timing 42

Figure 2-2 – Association between PAAS and Serum Testosterone 45

Figure 2-3 – Association between PAAS and PDS score 46

Figure 3-1 – Testosterone exposure and trajectory characteristics 71

Figure 3-2 – Testosterone exposure quintiles and testosterone trajectories 72

Figure 3-3 - APHV quintiles and testosterone trajectories 73
Chapter 1 – Introduction

The motivation for the work presented in this thesis stems from an ongoing project to characterize the role of testosterone in maturation of white matter in the male brain during adolescence. The timing and dynamics of testosterone are to be examined in the context of pubertal timing and testosterone trajectories throughout adolescence, respectively. A valid and logical methodology for studying the aforementioned features of male puberty is essential before testing hypotheses relating testosterone to properties of white matter. This thesis comprises of two manuscripts submitted for publication to peer-reviewed journals. The first (Chapter two) describes height-based methods of ascertaining timing of pubertal onset, and the second (Chapter three) describes a cumulative measure of testosterone obtained from trajectories over adolescence. The primary dataset is a sample of 513 males from the Avon Longitudinal Study of Parents and Children. This general introduction provides the relevant background and literature review to set the stage for the two manuscripts, which are followed by a general discussion and future directions.

1.1 Puberty

Puberty is an important milestone of life that initiates the transition from childhood into adulthood. Upon the completion of this developmental phase, the individual becomes sexually mature and capable of reproduction (Karlberg, 2002). Puberty dictates the development of primary sexual characteristics; in males this involves growth
and development of the penis (Marshall & Tanner, 1970). Numerous secondary sexual characteristics also develop, including growth of pubic, facial and body hair, and deepening of the voice. Other physical changes also occur during this period such as increases in stature, weight, muscle mass and pulmonary volume (Karlberg, 2002). On average, males show the first signs of pubertal changes at approximately 11 years of age and complete their development by approximately 16 years (Herman-Giddens et al., 2012; Marshall & Tanner, 1970). Studying the phase of puberty is important as there are numerous important changes taking place in various domains of the individual, including physical, emotional, cognitive and social development (Steinberg et al., 2006).

1.2 Pubertal Timing

The timing of pubertal onset is variable among individuals within a population; it is a critical consideration when studying puberty and development. First, it is important to distinguish clearly pubertal status from pubertal timing. The former refers to the degree of physical maturation that an individual has attained based on one or more measures. These measures may include hormone levels, genital development, pubic hair growth, and voice breaking in males (Shirtcliff, Dahl, & Pollak, 2009). Closely related but distinct is pubertal timing, which refers to pubertal status in relation to peers of the same sex and age. Cognition, differences in body composition, the emergence of psychopathology, and other outcomes are related to variation in pubertal timing (Kaltiala-Heino, Kosunen, & Rimpelä, 2003; Kindblom et al., 2006; Sanders & Soares, 1986). The effects of pubertal status and pubertal timing on development will be
revisited later in greater detail. The next section describes the physiology underlying male puberty and discusses factors associated with the initiation of puberty.

1.3 The Hypothalamic-Pituitary-Gonadal Axis

The physiological processes of puberty are intimately connected with the hormones of the Hypothalamic-Pituitary-Gonadal (HPG) axis (Hiort, 2002; Reiter, 1987). As the name implies, the main components of this axis are the hypothalamus, the pituitary gland, and the gonads, each of which is an important player in the negative feedback system that regulates the levels of sex hormones in the body. Neurons found in the preoptic area of the hypothalamus secrete Gonadotropin-releasing hormone (GnRH), a trophic peptide hormone, into the hypophyseal portal system of blood vessels (Romeo, 2003). These vessels link the hypothalamus to the anterior pituitary gland, where GnRH effects the production and secretion of the gonadotropins (Hiort, 2002), namely Luteinizing hormone (LH) and Follicle-stimulating hormone (FSH).

In males, LH is responsible for stimulating Leydig cells to produce testosterone from cholesterol via five enzymatic steps (Hiort, 2002) through a process known as steroidogenesis. On the other hand, FSH initiates the production of spermatozoa through a process known as spermatogenesis. The increased peripheral levels of testosterone feed back on the hypothalamus and pituitary gland to control the levels of sex hormones produced by the gonads (Reiter, 1987).
1.4 Initiation of Puberty

In males, the HPG axis is extremely active during the end of prenatal life and during the initial period of neonatal life, maintaining high levels of gonadal hormones similar to those observed in adulthood (Reiter, 1987). It then lies dormant throughout childhood until it reawakens with the initiation of puberty, allowing for the elevation of sex hormone concentrations in the body (Palmert & Boepple, 2001; E. J. Susman & Rogol, 2004). Pulsatile GnRH release occurs at all ages but the first hormonal sign of puberty within the HPG axis is an increase in levels of LH and FSH secretion, which occurs before external signs of pubertal development are detectable (E. J. Susman & Rogol, 2004). Although the precise molecular machinery that reactivates the HPG axis to initiate puberty is not well known, the protein Kisspeptin-54 is thought to play an important role in this process. Peripheral infusions of the Kisspeptin-54 peptide in normal human males stimulate function of the HPG axis and upregulate levels of LH and FSH (Dhillo et al., 2005). Activity of the pineal gland has also been associated with pubertal regulation through secretion of melatonin, the hormone that controls the circadian rhythm of the body (Aleandri, Spina, & Morini, 1996). Administration of melatonin delayed puberty in male Djungarian hamsters (Buchanan & Yellon, 1991) and levels of this hormone have been noted to fall abruptly with advancing development in boys (Silman, Leone, Hooper, & Preece, 1979). It has been suggested that hormones secreted during adrenarche, the increased androgenic steroid secretion from the adrenal glands, may also play in pubertal initiation. Adrenarche typically precedes the onset of puberty, occurring between the ages of 6 – 8 years (Hiort, 2002). Since gonadal puberty can take place without adrenarche and vice-versa, the precise
relationship, if any, between these two processes is not clearly delineated (Cutler & Loriaux, 1980).

The action of these molecules within the brain and periphery that might have a role in initiation of puberty may be linked to genetic factors, or exposure to environmental factors such as nutrition, light, stressors and endocrine disruptors (Parent et al., 2003). Genes and the environment are contributing to the variance observed in the onset of pubertal timing. In humans, there is a 4-to-5 year variation in the onset of puberty across individuals despite relatively similar life conditions (Parent et al., 2003).

With pubertal onset, the HPG-axis elevates and sustains levels of gonadal hormones in the circulatory system. These hormones are now able to exert their physiological effects on the body, resulting in the multitude of physical changes that are observed during this period of development. Testosterone, released from Leydig cells in the testes, is the main androgen in male development. Since testosterone levels correlate with pubertal stage (Granger et al., 2003), males that have an earlier onset of puberty will experience elevated levels of testosterone earlier than on-time and late maturers. This phenomenon may, in part, account for differences related to pubertal timing. Given the importance of testosterone in the male pubertal processes, the following sections will provide an overview of testosterone production, its various states in the blood stream, and its mechanisms of action at the target tissue.
1.5 Testosterone Production

As described earlier, reawakening of the HPG axis results in stimulation of the pituitary gland, which then releases gonadotropins, one of which is luteinizing hormone. This hormone acts on Leydig cells of the testes via the LH receptor, a member of the G-protein coupled receptor family. Activated LH receptors stimulate adenylate cyclase through GTP-binding proteins, whereby there is an increase in cyclic Adenosine Mono Phosphate (cAMP) production (Hiort, 2002; B. R. Zirkin & Chen, 2000). The cAMP then stimulates steroid production through mechanisms that may involve regulating protein synthesis at the ribosomal level, or by modulating gene expression after travelling into the nucleus. An influx of cholesterol through the mitochondrial membrane is achieved by StAR, an active transporter. This is followed by cleavage of cholesterol’s side chain in the mitochondria by the P450cc enzyme which yields pregnenolone (Hiort, 2002). A series of three more intermediate enzymatic steps produces testosterone, which is then secreted from the cell by the endoplasmic reticulum (B. Zirkin, Dykman, Kromann, Cochran, & Ewing, 1982; B. R. Zirkin & Chen, 2000).

1.6 Testosterone in the Blood Stream

Not all testosterone molecules in the bloodstream are physiologically active. There are three states in which testosterone exists in the blood: bound to Sex Hormone Binding Globulin (SHBG), bound to albumin, or unbound (Södergård, Bäckström, Shanbhag, & Carstensen, 1982; Vermeulen, Verdonck, & Kaufman, 1999). The unbound or free fraction of testosterone is biologically active, whereas either of the
bound fractions of testosterone needs to dissociate from their binding protein before they can bind to an androgen receptor and exert an effect. The binding between SHBG and testosterone molecules is of a high affinity, preventing unbinding of testosterone (Södergård et al., 1982). In comparison, the binding between albumin and testosterone is weaker, allowing some testosterone molecules to dissociate and subsequently exert an effect on target tissues (see Södergård 1982 for the association constants). Given concentrations of total testosterone and SHBG in the blood, level of bioavailable testosterone can be calculated; derived from the unbound and albumin-bound fractions of testosterone (Södergård et al., 1982; Vermeulen et al., 1999).

1.7 Mechanisms of Testosterone Action

Unbound testosterone diffuses into target cells in order to exert a physiological effect. Testosterone bound to SHBG may also be taken up into the cell through receptor-mediated pathways (Hiort, 2002). Once inside the cell, testosterone may be converted to estradiol by the aromatase enzyme, or converted into dihydrotestosterone (DHT) through the reducing action of the 5-alpha reductase enzyme (Krause, 2006; Stoffel-Wagner, 2001). If the testosterone molecule is not aromatized or reduced, it can elicit a response from the cell via the Androgen Receptor (AR). The AR is a nuclear receptor present in the cytoplasm of the cell that forms a homodimer with another AR upon activation after the binding of testosterone (Hiort, 2002; Poletti, 2004). The newly formed homodimer moves from the cytoplasm into the cell nucleus. The main function of the AR is that of a DNA-binding transcription factor; as such the activated complex finds and binds to specific hormone response elements on the DNA (Poletti, 2004).
Specific target gene transcription is increased or down-regulated depending on the complex interactions between the receptor complex and the basal transcription machinery (Poletti, 2004). Thus, testosterone exerts its biological effect through this modulation of gene transcription via the Androgen Receptor. The next section pertains to the levels of testosterone during male puberty, and the effects associated with testosterone, as well as pubertal status and timing.

1.8 Testosterone Levels During Male Puberty

Levels of total testosterone in pre-pubertal males tend to be lower than 0.5 nmol/L, whereas the levels can rise to 18 nmol/L and higher in post-pubertal males (Andersson et al., 1997; Crofton et al., 2002). I have conducted a meta-analysis of normative ranges of total testosterone by age. The results are presented in Chapter three, illustrating the rapidly rising levels of testosterone throughout male adolescence.

1.9 The Effect of Testosterone on Development

Testosterone plays a role in the multitude of physical changes noted during puberty: genital changes, larynx growth and deepening of the voice, increases in bone and muscle mass, growth spurt, increase in mass of erythrocytes, thickening of skin, and hair growth on the trunk (Krause, 2006). A longitudinal study of young male soccer players showed that developing physical strength was related to changes in serum testosterone (Hansen, Bangsbo, Twisk, & Klausen, 1999). In addition to physical
changes, testosterone has been implicated in a variety of behavioural and cognitive processes, and is associated with psychopathology.

In 12 year old males, testosterone reactivity, measured through changes in testosterone levels in a 20-minute period, predicted family problems one year later (Marceau, Dorn, & Susman, 2012). A study of males aged 12 – 15 years by Rowe and colleagues found that testosterone was associated with behaviours of nonaggressive conduct disorder in boys with deviant peers, and leadership behaviours in boys with non-deviant peers (Rowe, Maughan, Worthman, Costello, & Angold, 2004). Pubertal levels of testosterone were found to be negatively associated with performance on a mental rotation task of spatial ability during early adulthood in males (Vuoksimaa, Kaprio, Eriksson, & Rose, 2012).

When investigating the effect of sex steroids on the pubertal brain, Peper and colleagues found a positive relationship between testosterone and absolute volume of grey matter in the brain in boys between 10 -15 years (Peper et al., 2009). Perrin and colleagues showed that testosterone plays a role in the increase of volume of white matter and decrease in white matter Magnetization Transfer Ratio (implying a thicker axon caliber) in male adolescents between the ages of 12 -18 years (Perrin et al., 2008).

Furthermore, extreme levels of testosterone are associated with different psychopathologies and prodromal symptoms. Male patients, with negative symptoms of chronic schizophrenia, aged approximately 35 years, and taking antipsychotic and anticholinergic medication, had lower levels of plasma testosterone than healthy controls (Akhondzadeh et al., 2006). Levels of day-time and night-time testosterone, as
well as 24-hour mean secretion of testosterone were found to be lower in depressed male patients in comparison with healthy controls (Schweiger et al., 1999). A study by Sankar and colleagues found that lower testosterone and an increased number of CAG repeats in the Androgen Receptor gene are predictors of negative affect in males with moderate to severe depression (Sankar & Hampson, 2012). Male adolescents with prodromal symptoms of psychosis had lower testosterone levels when compared with non-clinical controls (Van Rijn et al., 2011). Taken together, these results point to a relationship between lowered levels of testosterone and psychopathology. On the other hand, in cases with artificially increased levels of testosterone such as in users of anabolic androgentic steroids (AAS), we also see evidence of psychiatric disorders. In a study of athletes with a history of AAS use, Pope and Katz found that a striking 23% of users had reported major mood syndromes; mania, hypomania, or major depression (Pope Jr & Katz, 1994).

In addition, testosterone has been linked to a number of externalizing and internalizing symptoms. Male adolescents at high risk for psychopathology with elevated scores of externalizing behaviors were found to have higher levels of plasma testosterone (Maras et al., 2003). A positive relationship between testosterone levels and aggression scores was identified in a separate study that examined children with disruptive behavior (Scerbo & Kolko, 1994). Bokhoven and colleagues found that males who developed a criminal record as adults had higher testosterone levels at age 16 years compared with males who did not obtain a record (Van Bokhoven et al., 2006). Testosterone’s association with internalizing symptoms includes an increase in plasma concentrations following a psychological stress session in anxious males, but not in controls (Gerra et al., 2000). Furthermore, socially anxious males experienced a drop in
testosterone levels after a loss to an experiment confederate in a rigged competition, whereas testosterone levels did not decrease in controls (Maner, Miller, Schmidt, & Eckel, 2008).

I will now review changes found to be associated with pubertal status and timing. The findings correspond with results presented in relation to testosterone, which may be one of the driving forces behind associations identified with the status and timing of puberty.

1.10 Changes Associated with Pubertal Status

The literature implicates both pubertal status and timing as playing a role in various changes throughout the individual’s development. Ge and colleagues showed that pubertal status in males is related to both internalizing and externalizing symptoms as assessed from both child self-report and caregiver reports (Ge, Brody, Conger, & Simons, 2006). Boys that reported higher levels of negative affect tended to be further along in the status of their genital development (E. Susman, Dorn, & Chrousos, 1991). In a study of 102 boys, progression in pubertal maturation was associated with decreased cohesion with their mothers and less frequent communication with their fathers (Steinberg, 1987). An investigation from Flannery and colleagues showed that pubertal status in boys was related to sexual experience and delinquency (Flannery, Rowe, & Gulley, 1993).
1.11 Changes Associated with Pubertal Timing

In addition to pubertal status, variation in pubertal timing should also be taken into account when studying adolescent development, as it has been associated with differences in various domains including cognition, body composition, and emergence of psychopathology. Late maturers within college students fared better compared with early maturers on mental rotation tasks that test spatial abilities (Sanders & Soares, 1986). Furthermore, earlier onset of puberty predicts a greater body mass index and increased central fat mass, measured using Computed Tomography, at 19 years of age in men (Kindblom et al., 2006). Early maturers were found to have more intense conflict with parents, but not an increase in frequency of conflict, than on-time maturers (Steinberg, 1987).

Adolescence is a period of vulnerability when certain psychiatric disorders tend to emerge (Paus, Keshavan, & Giedd, 2008). Since pubertal onset coincides with adolescence, atypical timing of puberty may play a role in the development of psychopathology. A study done in Finnish youth found that very early or late puberty in males was associated with self-reported depression scores obtained using the 13-item Beck Depression Inventory (Kaltiala-Heino et al., 2003). Graber and colleagues reported a higher lifetime prevalence of disruptive disorder behaviour and higher rates of current substance use in male participants from a study of senior high school students in Oregon (Graber, Seeley, Brooks-Gunn, & Lewinsohn, 2004). Schizotypy was assessed using the Schizotypal Personality Questionnaire in a group of medical students, and examined in relation to pubertal timing. For male participants, the Unreality subscale was associated with both extremes of pubertal timing, and
participants who scored highly on the Withdrawn subscale tended to be late maturers (Gruzelier & Kaiser, 1996). These examples highlight the importance of considering timing of pubertal onset as well as pubertal status in relation to various spheres of development. In order to carefully investigate these two variables, a proper understanding of how they are assessed and measured is critical.

1.12 Patterns of Pubertal Changes in Males

In the 1970s, Tanner and colleagues undertook one of the first detailed longitudinal studies of puberty in order to investigate the patterns of physical changes in sex characteristics over time in a sample of 228 boys in Britain. Marhsall and Tanner (1970) characterized stages of pubertal development using changes seen in male genitalia and pubic hair growth into five different stages ranging from Stage 1 (pre-pubertal) to Stage 5 (post-pubertal). These are now commonly referred to in the literature as Tanner Stages. Pictures and criteria for each stage are provided in their report to serve as a standard when assessing pubertal development in this manner.

Table 1 and Figure 3 from this paper notes the average age when each stage is achieved, and clearly illustrates that there is considerable variation across individuals in their age at any particular stage of pubertal development. Additionally, individuals take a different amount of time to complete the whole process of pubertal development. Some participants traversed Genital Stage 2 through 5 in less time than it took others to progress from Genital Stage 2 to 3 (Marshall & Tanner, 1970). This work by Marshall and Tanner served to provide an important framework, which is now commonly used as
a basis for assessment of pubertal status in the study of adolescents across the globe. The next section discusses methods to assess pubertal status based on the Tanner Stages, as well as other approaches used in the literature.

1.13 The Evolution of Pubertal Status Assessment

The gold standard for assessing pubertal status has been a clinician conducting a physical exam on a participant, classifying them into Tanner’s stages based on development of physical characteristics (Coleman & Coleman, 2002; Dorn, Dahl, Woodward, & Biro, 2006; Shirtcliff et al., 2009). Due to potential participant discomfort with the intimate nature of this assessment, it is challenging to integrate physical exams into non-clinical settings. Discomfort is especially noted in male participants, since palpation of the testes is an important component of the male physical exam. In a study where physical examinations were conducted by a nurse practitioner, 17% of 82 male participants refused to undergo the assessment (Shirtcliff et al., 2009). An alternative to palpation of the testes is to take photographs of the genitalia, but this is also difficult to implement in a research setting (Petersen, Tobin-Richards, & Boxer, 1983).

In order to circumvent the problems posed by physical exams, Petersen and colleagues developed a self-report scale, the Pubertal Development Scale (PDS), on which participants rate their development on the following criteria: pubertal growth spurt, body hair, skin changes, facial hair and voice change for boys, and breast development for girls (Petersen, Crockett, Richards, & Boxer, 1988). An alternative self-report measure uses photographs of the different Tanner stages of genital development and
pubic hair growth and allows the participant to indicate which of the images they currently most closely resemble (Marshall & Tanner, 1970).

As the levels of gonadal hormones advance pubertal maturation, measuring these sex steroids is an alternative strategy to the physical or self-report assessments when studying pubertal development. Testosterone levels are correlated with pubertal stages in boys (Granger et al., 2003), yet there is overlap in hormone levels across pubertal stage, and therefore an imperfect match of hormones with the physical methods described above (Andersson et al., 1997; Crofton et al., 2002; Shirtcliff et al., 2009). This may be due to the circadian rhythm of the hormones and time of measurement, or perhaps different individuals need to attain different thresholds of hormone levels to advance pubertal development.

1.14 Concordance between Measures of Pubertal Status

Dorn and Biro reviewed studies examining concordance between the combinations of different methods used in pubertal stage assessment (Dorn & Biro, 2011). A study on the concordance of self-report with physical examinations found that Tanner Stages assessed through the PDS and physical exam matched 26% of the time in a population of South-African males aged between 10-18 years (Norris & Richter, 2008). Physical exam matched self-report measures of genitalia and pubic hair Tanner Stages using line drawings 49% and 66% of the time, respectively, in a sample of 172 Chinese boys between the ages of 8-18 years (Chan et al., 2008), and 75% and 83%, respectively, in 24 male Canadian athletes between the ages of 12 and 17 years (Leone
Percent agreement between the PDS and line drawings, the two self-report methods of assessment, was 39% in a sample of 1,366 males aged 9-16 years (Bond et al., 2006).

Shirtcliff and colleagues conducted a study comparing hormones, physical examination, the Pubertal Development Scale and a picture-based self-report measure in a sample including 82 boys ranging from 9 to 14 years of age (Shirtcliff et al., 2009). The stage obtained from PDS gonadal and pubic hair questions matched the stage assessed through physical exam 54% and 60% of the time, respectively. The same comparison between the picture-based assessment and physical exam found concordance 41% and 54% of the time. The percentage of participants that reported the same genitalia stage and pubic hair stage on the two self-report measures were 37% and 47% respectively. The authors used structural equation modeling to investigate the strength of the association between the physical and self-report assessments with testosterone levels. The physical exam and the PDS scores were found to capture levels of testosterone well, but the picture-based assessment did not (Shirtcliff et al., 2009).

1.15 Assessing Pubertal Timing

There are numerous methods available to assess pubertal timing reported in the literature. With cross-sectional data, the participant’s status on important checkpoints of puberty can be compared to the rest of the study sample or to population norms to ascertain relative pubertal timing. The ratio obtained from comparing bone age to
chronological age can be used to infer pubertal timing (Dorn et al., 2006). The availability of appropriate longitudinal data corresponding to the period where pubertal development started can be used to designate the age of pubertal onset for a participant as the earliest age at which Tanner Stage 2 was measured (Coleman & Coleman, 2002; Marshall & Tanner, 1970). An important consideration for the use of longitudinal Tanner staging assessments to investigate pubertal development is that it is difficult to identify the exact moment at which an individual reaches a stage. One can only conclude that a change occurred at some point in time between two consecutive differing assessments in an individual.

Another possibility in pubertal timing assessment outside of conducting physical examinations or measuring hormones involves the use of height measurements. Height has many advantages as a variable, including simplicity, objectivity, and non-invasiveness of measurement. Furthermore, it is a common variable that may already be found in many datasets, allowing height-based measures of pubertal timing to be calculated retrospectively. Indices of pubertal timing that can be derived from measurements of height are: Age at Peak Height Velocity (Marshall & Tanner, 1970), Height Difference in Standard Deviations (Wehkalampi et al., 2008), and Percent Achieved of Adult Stature. The derivation of these indices and their relationship with questionnaire-based and hormonal measures of puberty are explored further in Chapter two.

This thesis pertains to the methods of assessing pubertal timing and testosterone trajectories in males, to be used in an investigation into the role of pubertal timing and dynamics in the maturation of the male adolescent brain. The motivation and framework
of the overarching project, the approach of Population Neuroscience, and the cohort studies that are used in our work are described in the subsequent sections.

1.16 Sexual Maturation, Hormones, Brain, Psychopathology

As described in earlier sections, extreme levels of testosterone and pubertal timing have been associated with psychopathology such as depression, schizophrenia, and prodromal psychiatric symptoms. Paus (2010) reviews the literature linking abnormal properties of white matter in the brain to psychopathology (Paus, 2010a). A logical question arises as to whether testosterone is implicated in psychopathology because it plays a role in the maturation of white matter. There is emerging evidence that shows the role of testosterone in maturation of the white matter in the male adolescent brain (Perrin et al., 2008). This cross-sectional investigation, studying adolescents between the ages of 12-18, higher levels of testosterone were found to be associated with an increased relative volume of white matter in males. We are interested in dissecting further the role of testosterone in maturation of the male adolescent brain through the use of a longitudinal design. The goal is to characterize the role of testosterone dynamics and the closely linked concept of pubertal timing throughout adolescence on properties of white matter in the brain in early adulthood. A better understanding of this phenomenon will enable us to understand if and how extreme levels and timing of testosterone in males affects mental health vis-à-vis the white matter of the adolescent brain. An inherent challenge with a research question such as this is related to the collection and analysis of longitudinal data in a large sample of male adolescents. The discipline of Population Neuroscience is well suited to
be implemented in such an investigation, and is discussed further in the following sections.

1.17 Population Neuroscience

Investigations employing a twin-study design have revealed that between 50% and 90% of variance observed in brain phenotypes across the population is explained by genetic factors, depending on the specific trait being assessed (Peper, Brouwer, Boomsma, Kahn, & Hulshoff Pol, 2007). In addition to genes, environmental factors contribute to the inter-individual differences observed in the structural and functional phenotypes of the brain. Novel strategies beyond lesion studies and small neuroimaging studies are required in order to disentangle the complex interplay between genetic factors and environmental exposures that participate in shaping the brain.

The field of Population Neuroscience lies at the intersection of the disciplines of genetics, epidemiology and cognitive neuroscience. By employing methodological tools from genetics and epidemiology to characterize the ‘exposures’ in the form of genes and environmental factors respectively, in tandem with using brain-imaging techniques such as Magnetic Resonance Imaging (MRI) to study outcome phenotypes, we can begin to ask questions about how our brains are affected by our genes, environmental exposures, and the interactions between them (Paus, 2013).

Studies that characterize the normal development of children while utilizing large sample sizes and comprehensive genotypes and phenotypes are positioned well to answer questions about the factors shaping development. A large cohort empowers
these investigations to find small and important effects that may not be picked up in studies with small sample sizes.

Numerous large-scale studies that fulfill these requirements are ongoing or have been completed in various corners of the globe. In studies of these magnitudes, it is impossible to set up every research question and hypothesis that will be pursued prior to acquiring data; therefore the extent of data collection is of great breadth out of necessity. A lot of thought also goes into designing protocols that will yield data flexible enough to be manipulated into different measures and utilized in a variety of manners for different investigations. It is the responsibility of the investigator to obtain judiciously the most salient data from within the vast datasets and use creative methods to derive meaning from them.

1.18 Saguenay Youth Study

The Saguenay Youth Study (SYS) was designed to investigate the genetic and environmental factors that play a role in the development of the brain and bodies of adolescents. It is based in the Saguenay Lac-Saint-Jean (SLSJ) region of Quebec, Canada, where it started recruitment in the early 2000s and completed data collection on just over 1000 adolescents between the ages of 12-18 in 2012. The design of the SYS is described in detail elsewhere (Pausova et al., 2007). Briefly, the study was designed as a cross-sectional investigation that utilized numerous phenotyping platforms over the course of several study visits to characterize participants' mental, cardiovascular and metabolic health. Samples of DNA were also collected from
participants (and their biological parents). Use of a family based design allowed adolescents and their siblings to be recruited along with their parents. Additionally, half the sample was exposed to maternal cigarette smoking while the non-exposed participants from the other half were matched to them by maternal education. The sample is composed entirely of white Caucasians, has an average IQ of 105, and is of lower socio-economic status (SES) than the general population of the region, given the inclusion criteria stipulating that half of the adolescent’s mother smoked during pregnancy and the matching procedure based on maternal education.

Participants underwent a Magnetic Resonance Imaging (MRI) protocol that included sequences to obtain T1-weighted, T2-weighted, Proton-Density Weighted, Magnetization Transfer, and abdominal images (Pausova et al., 2007). Assessments of environmental exposures cover all stages of life; the prenatal period (e.g. smoking, alcohol), infancy (e.g. breastfeeding), childhood (e.g. food availability, maternal care, stressful life events), and adolescence (e.g. diet, sleep). Assessments with a detailed cardiovascular and metabolic protocol were conducted in addition to measures of behaviour and cognitive abilities.

1.19 The Avon Longitudinal Study of Parents and Children

A large-scale trans-generational prospective observational study known as the Avon Longitudinal Study of Parents and Children (ALSPAC) had been initiated in the early 1990s within the Bristol area of the United Kingdom and is currently still ongoing. Within the time period between 1990 and 1992, a number of pregnant women were
recruited into the study, resulting in a total of 14541 pregnancies being included in the sample. With the birth of their children, and additional recruitment efforts adding other children into the cohort, ALSPAC consists of a total of 20390 participants. Boyd and colleagues present a detailed profile of the ALSPAC cohort (Boyd et al., 2012).

This large cohort has been followed frequently over the past 23 years via questionnaires sent to homes, and invitations to attend clinics at which additional data was collected. Over the first 18 years of the lives of the participants, a total of 68 datapoints exist from 35 questionnaires sent to the participants, 24 questionnaires sent to their mothers or main caregivers, and initiations to nine clinic sessions. As such, ALSPAC contains a wealth of phenotypes, as well as genetic data and biological samples. Some examples of phenotypes derived from questionnaires include data on development and puberty, mental health, cognitive ability, substance use, and assessments from teachers. The clinic sessions collected measures pertaining to anthropometry, bioimpedance, diet, cognition, behaviour, social relationships, mental health and substance use. DNA samples were collected from 11343 participants, and 8365 of them have genome-wide data available. The ALSPAC biobank holds various biological samples collected from a young age through the clinics sessions, including but not limited to blood, urine, hair, toenails, teeth, saliva and placenta. Furthermore, linkage of participants’ health and administrative data has been done which helps enrich the ALSPAC data resource.
1.20 Cross-Sectional and Longitudinal Observational Studies

Difference in study design is one of the key considerations between data obtained from the SYS and ALSPAC studies. The SYS is cross-sectional in design and measured participants once between the ages of 12 and 18 years. The longitudinal design of the ALSPAC study allowed for each participant to be studied on numerous occasions from birth until early adulthood. A proper understanding of the limitations of each design should be taken into consideration before beginning an investigation.

Cross-sectional observational studies are excellent tools for exploring and identifying associations between different variables. The study design is relatively easier to implement than a longitudinal design, as participants are studied at one time-point. Since these studies are an observational snapshot of their participant, it is difficult to make inferences about cause-and-effect relationships when variables are found to be associated with each other (Gordis, 2008).

Longitudinal studies are more powered than cross-sectional investigations to disentangle cause-and-effect relationships by virtue of the temporal separation in the exposures and outcomes (Paus, 2010b). Interesting associations identified in a cross-sectional study may be pursued in the context of a longitudinal dataset in order to understand the temporal aspects of the association. A challenge faced by longitudinal studies is maintaining high levels of participation, as individuals from the study cohort may choose to drop out while the study is ongoing.
1.21 Methodologies to Assess Pubertal Timing and Testosterone Trajectories

In a longitudinal dataset such as ALSPAC, there is a need for methodologies to synthesize data into meaningful variables that can be used in analyses to test hypotheses. Methods to assess pubertal timing and trajectories of testosterone are of particular relevance to our project investigating the role of testosterone in maturation of white matter in the male brain during adolescence. Chapter two of this thesis describes the derivation of three different height-based indices of pubertal timing, and Chapter Two focuses on describing a cumulative measure of testosterone exposure from the testosterone trajectories throughout male adolescence.
Chapter 2 - Height-based Indices of Pubertal Timing in Male Adolescents

Ammar Khairullah, Margaret T May, Kate Tilling, Laura D Howe, Gabriel Leonard, Michel Perron, Louis Richer, Suzanne Veillette, Zdenka Pausova, Tomáš Paus

In Press: International Journal of Developmental Science
2.1 Abstract

It is important to account for timing of puberty when studying the adolescent brain and cognition. The typical methods of assessing pubertal status may not be feasible in some studies, especially for males. Using data from a sample of 478 males from a longitudinal study, we describe the calculations of three independent height-based markers of pubertal timing: Age at Peak Height Velocity (APHV), Height Difference in Standard Deviations (HDSDS), and Percent Achieved of Adult Stature (PAAS). These markers correlate well with each other. In a separate cross-sectional study, we show that the PAAS marker correlates well with testosterone levels and self-reported pubertal-stage scores. We conclude by discussing key considerations for investigators when drawing upon these methods of assessing pubertal timing.
2.2 Introduction

Puberty is a period of gonadal maturation and, in turn, the development of secondary sex characteristics (Sizonenko, 1972). Pubertal onset begins a secretory cascade of gonadal hormones that target numerous organs, including the brain. The age at onset of puberty is not constant across the population. The variation in pubertal timing has been studied in relation to different phenomena occurring during adolescence, including cognition and the emergence and frequency of psychopathology. For example, among college students, late maturers tended to do better than early maturers on mental rotation, a test of spatial ability (Sanders & Soares, 1986). Another study found that early puberty in females, and both very early and late puberty in males, was associated with depression (Kaltiala-Heino et al., 2003). These findings demonstrate that timing of pubertal development is an important variable that should be taken into consideration when studying the adolescent brain.

Pubertal onset and timing of puberty can be studied by assessing pubertal stages, and an individual’s progression through them. The assessment of pubertal stages is traditionally achieved through a physical examination conducted by a trained clinician (Marshall & Tanner, 1970), through which an individual is classified into one of the five stages of pubertal development often referred to as ‘Tanner stages’. This mode of assessment is considered to be the gold standard (Dorn et al., 2006). In large population-based studies, however, measuring pubertal status in this fashion is not feasible, mainly due to the potential participant discomfort associated with the intimate nature of the assessments. This is particularly relevant for male participants as palpation of the testes is an important part of the examination (Dorn et al., 2006). In a
study where physical examinations were conducted by a nurse practitioner, 17% of 82 male participants refused to undergo the assessment (Shirtcliff et al., 2009). In lieu of physical examinations, an often-used strategy is the administration of questionnaires answered by the participant or his/her parent.

There are two common ways in which pubertal staging is achieved through such questionnaires. The first involves the use of picture-based questions employing drawings or photographs of children at different Tanner stages. Alternatively, participants are assessed through the Pubertal Development Scale (PDS), a series of questions (without pictures) that ask about their physical growth (height), body hair, skin changes, breast changes, and male-specific questions regarding facial hair and voice changes (Dorn et al., 2006; Petersen et al., 1988). Shirtcliff and colleagues (2009) report fair concordance, as defined by Landis and Koch (1977), between genital stage assessed through physical exam with the picture-based questionnaire ($\kappa = .36$) and PDS ($\kappa = .36$). On both types of questionnaires, adolescents overestimated pubertal maturation when they were at lower stages of development compared with their peers, and underestimated development when they were at higher stages relative to their peers (Shirtcliff et al., 2009). The literature suggests that self-reported measures should be used only as crude estimates of pubertal development (Wu, Schreiber, Klementowicz, Biro, & Wright, 2001).

Since growth hormone (GH) plays a key role in regulating height and is closely linked with sex steroids, particularly testosterone, the adolescent growth spurt occurs along with pubertal maturation (Rose et al., 1991; Veldhuis, Roemmich, Richmond, & Bowers, 2006). Therefore, in lieu of observing progression through pubertal stages,
increases in height can be used to assess pubertal timing in an objective manner. Height measurements have the advantage of being simple, non-invasive, and inexpensive to incorporate into a study design. Furthermore, height can already be found in many datasets, allowing for height-based measures of pubertal timing to be derived retrospectively. Using data from two population-based studies, here we show the utility of height measurements for estimating pubertal timing in males through three different height-based indices; we explore how these measures compare with each other and with self-reported pubertal stage. Furthermore, we estimate the associations of a height-based measure of pubertal status with questionnaire-based pubertal stage assessment and levels of total testosterone in a separate sample.

The first index, Age at Peak Height Velocity (APHV), identifies the age at which a characteristic feature of puberty, the growth spurt, takes place (Tanner & Whitehouse, 1976). The second index, Height Difference in Standard Deviations (HDSDS), allows us to ascertain pubertal timing relative to the rest of the study sample based on the shift of their height standard-scores from height at 14 years to final height, as described by Wehkalampi et al. (2008). The mechanism by which this index operates is discussed in Section 3.3. The third index incorporates the Khamis-Roche method of predicting adult stature to derive Percent Achieved of Adult Stature (PAAS), which informs us of relative pubertal timing based on how close participants are to their final height predicted by their parental stature and anthropometric measurements (Khamis & Roche, 1994).
2.3 Methods

Data from two population-based cohorts, the Avon Longitudinal Study of Parents and Children (ALSPAC) and the Saguenay Youth Study (SYS), were used for our analyses. Detailed descriptions of the ALSPAC and SYS datasets have been reported previously (Boyd et al., 2012; Pausova et al., 2007). Ethical approval for ALSPAC was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. The SYS was approved by the local ethics committee. Consent and assent forms were completed by parents/guardians and participants, respectively.

2.3.1 The Avon Longitudinal Study of Parents And Children

ALSPAC is a birth cohort study that recruited 14,541 pregnant women resident in Avon, UK with an expected delivery date between April 1991 and December 1992. Since then the children and their parents have been studied extensively, with longitudinal data available from numerous self-administered questionnaires sent to their homes, linkages to medical records, and through examinations carried out during study clinics.

The children were invited to participate in nine clinic examinations over the course of the study, roughly at the following years of age: 7, 8, 9, 10, 11, 12, 13, 15, and 17. Clinic staff measured height to the last complete mm using a Harpenden Stadiometer, and weight to the nearest 50g using a Tanita Body Fat Analyser (TBF 305) at each visit. Maternal and paternal heights are available from questionnaires completed by the parents. Outside of the clinics, nine pubertal questionnaires were sent over time by mail to participants’ homes inviting them to answer questions about growth of
genitalia and pubic hair with the help of pictures corresponding to different Tanner stages. Our sample of 478 males is composed of a subset of the ALSPAC cohort that had been recruited for further study using magnetic resonance imaging (MRI). In the next section, we describe the process of deriving three independent height-based indices of pubertal timing from the ALSPAC height data.

2.3.2 Age at Peak Height Velocity (APHV)

Closely spaced longitudinal measurements of height are required for the calculation of APHV. We have developed a script using MATLAB (Mathworks, Natick, MA USA) to automate the process of APHV estimation. First, the longitudinal measurements of height and the respective age in months are used to plot the participant’s height over time. The cubic spline-interpolation function from the MATLAB Basic Fitting toolbox (Mathworks, Natick, MA USA) is used to generate a curve through the data-points (Supplementary Figure 1) as reported previously (Ramsay, Altman, & Bock, 1994; Sherar et al., 2011). Plotting the derivative of this height curve produces the participant’s growth curve, and indicates height velocity at each month (Supplementary Figure 2). The maximum height velocity and corresponding age in months are identified, providing an estimate of the PHV (cm/month) and the APHV (months) respectively. The spacing between ages of observation (in months) varies across participants, with a mean of 15.25, SD of 6.27, and range
Supplementary Figures

Supplementary Figure 1
ALSPAC Participant Height Curve

Supplementary Figure 2
ALSPAC Participant Growth Curve
between 2 and 49 (months). A biologically driven restriction is placed on the script to ensure that the APHV is recalculated if it falls earlier than an age of 10 years, which is under three standard deviations from established norms (Marshall & Tanner, 1970).

### 2.3.3 Height Difference in Standard Deviations (HDSDS)

Unlike APHV, calculation of the HDSDS requires only two height measurements: height at the average age at Peak Height Velocity for the population and final height, 14 and 17.5 years respectively (Wehkalampi et al., 2008). We interpolated height at exactly 14 years and 17.5 years for all participants using the fitted spline described in Section 2.3.2. This step is not necessary if participants have height measurements at ages 14 and 17.5 years, but this was not the case in our sample. Participants missing data from the closest clinic to 14 and 17.5 years were excluded from the analysis. Internally derived standardized scores of height at 14 years (z14y) were calculated for each participant, and the same was done for height at 17.5 years (z17.5y). The following formula is then used: HDSDS = z14y – z17.5y

### 2.3.4 Percent Achieved of Adult Stature (PAAS)

We used the Khamis-Roche method to predict final stature for each participant based on their age, height, weight and parental stature (Khamis & Roche, 1994). To use this as an index of pubertal timing, we calculated the percentage of adult stature achieved by age 14. The average height of the two parents (mid-parental stature) and both height and weight at age 14 (acquired by using a fitted spline as described in Section 2.3.2) were obtained for all participants; height was measured in inches and
weight in pounds. These values, along with age and sex based coefficients provided by
Khamis and Roche (i.e, specifically for males at age 14) were used in the following
equation:

\[ \text{Predicted adult stature} = -6.4299 + 0.59151 \times \text{height} - 0.09776 \times \text{weight} + 0.58762 \times \text{mid-parental stature} \]

To calculate PAAS:

\[ \text{PAAS} = \left( \frac{\text{Height at 14}}{\text{Predicted adult stature}} \right) \times 100\% . \]

2.3.5 Interpretation of the Indices of Pubertal Timing

We calculated standardized-scores of APHV for all participants and multiplied
these values by -1. Furthermore, the percentages calculated for all the participants
through PAAS were converted into standardized scores. This way, across each index,
participants have a score that is on a similar scale and centered around zero, where a
score of zero denotes an exactly average maturer. A positive score on an index
indicates the participant is a relatively early maturer, whereas a negative score shows
that a participant is a late maturer when compared with the rest of the sample.

2.3.6 Using Quintiles to Investigate Concordance of Height-Based
Pubertal Timing Indices

To investigate concordance across the three indices beyond correlation
coefficients, we ranked and divided values obtained for each index into quintiles. The
quintiles were assigned scores from one through five: the first quintile (score of one)
contains the lowest scores and indicates the latest maturers while the last quintile
(score of five) contains the highest scores and indicates the earliest maturers. These quintiles should not be confused with the five Tanner Stages of pubertal status.

We calculated the percentage of participants with the exact quintile score on the HDSDS or PAAS indices when compared with the APHV index. Furthermore, we calculated the percentage of participants that were classified into either the same quintile or a higher or lower adjacent quintile in the HDSDS or PAAS indices given their APHV quintile score. Additionally, unweighted kappa and kappa with linear weighting were calculated between APHV and the other two indices.

2.3.7 The Saguenay Youth Study

The SYS is a cross-sectional study of 1,024 adolescents of French Canadian origin living in the Saguenay-Lac-Saint-Jean region of Quebec, Canada. Recruitment of the adolescent participants (12 to 18 years of age) began in high schools, followed by a structured telephone interview with the mother conducted by a research nurse. All participants were assessed using an extensive phenotyping protocol, including MRI of the brain and body, to characterize a broad range of phenotypes relevant for their mental, cardiovascular, and metabolic health (see Pausova et al., 2007 for details). Importantly for our analyses, anthropometric measurements, scores on Pubertal Development Scale, and plasma levels of sex hormones were also obtained. Data from the 496 male participants were used in our analyses.
2.3.8 Anthropometry and PAAS Scores

As the SYS data are cross-sectional, only the PAAS can be derived. Height, weight, and mid-parental stature along with the appropriate age and sex-based coefficients, available for each half year from 12 years to 17.5 years, for males were used to estimate adult stature for each participant using the Khamis-Roche method as described above. PAAS scores were calculated and transformed into standardized z-scores.

2.3.9 Pubertal Development Scale

All participants filled out the Pubertal Development Scale (PDS) (Petersen et al., 1988), an eight-item self-reported measure of physical development based on the Tanner stages. The male participants answer questions about their growth in stature, pubic hair, and voice changes which are used for classification into one of five categories of pubertal status: (1) prepubertal, (2) beginning pubertal, (3) midpubertal, (4) advanced pubertal, and (5) postpubertal.

2.3.10 Serum Testosterone

Fasting blood samples were taken in the morning (between 8:00 and 9:00 A.M.) and analyzed via radioimmunoassay (Testosterone RIA DSL-4000; Diagnostic Systems Laboratory) to measure serum levels of total testosterone (ng/ml) and sex-hormone-binding globulin.
2.3.11 Statistical Analyses

Correlations were used to examine the relationships between different indices of height-based pubertal timing obtained in ALSPAC. Fixed effects ANOVA models were used to examine the relationship between data from the ALSPAC pubertal questionnaire data and the three pubertal timing indices. In the Saguenay Youth Study, linear regression was used to relate PAAS to total testosterone. Fixed effects ANOVA models were used to examine PAAS scores and testosterone levels across the five Pubertal Development Scale groups and the responses to the pubic-hair question on the PDS questionnaire. The pubic-hair question was selected from the PDS as an analogous question was present in the ALSPAC questionnaires, and thus can serve as a means of comparison between the two types of questionnaires vis-à-vis the height-based indices. All statistical analyses were done using JMP 9 for Macintosh (SAS Institute, Inc., Cary, NC).

2.4 Results

2.4.1 Correlation between three height-based pubertal timing indices in ALSPAC

Our subset from the Avon Longitudinal Study of Parents And Children consisted of 478 male participants. Of these, a total of 18 were excluded: 8 for missing four or more height measurements, 1 for having an anomalous height measurement, and 9 for very high Age at Peak Height Velocity estimates that are imprecise due to spline interpolation at the far extreme of the height measurements. The remaining sample of 460 participants had APHV estimated. Comparison of means with two-tailed t-tests
between excluded participants and the remaining sample show the 18 excluded participants as having shorter stature in the four clinics between 12 and 17 years of age ($r^2 = .02, .01, .04, .01; p<.05$). This effect is present because the excluded group contains the nine late maturers with imprecise APHV estimates. As growth spurt before the age of 10 years is unlikely (Marshall & Tanner, 1970), we recalculated APHV with this age as a lower limit for the 53 participants whose estimates lay below 10 years. The mean recalculated APHV for these participants was 161.42 months, which matches closely the mean APHV for the full sample of 160.26 months. For the calculation of Height Difference in Standard Deviations score, 34 participants who were missing important data from the clinics adjacent to ages 14 and 17.5 had their height interpolations removed as described in Section 2.3.3, leaving 426 participants that could have HDSDS scores calculated. Similarly, after excluding participants with missing data required for the Khamis-Roche calculations (mainly due to missing maternal and/or paternal height), 354 participants had Percent Achieved of Adult Stature scores calculated.

Table 2-1 presents statistics describing the distribution of scores from each index of pubertal timing and pairwise correlations between the three indices, run after standardizing both APHV and PAAS, and multiplying APHV by -1 as described in Section 2.3.5. All three indices are strongly correlated; the strongest correlation is found between the APHV and HDSDS scores ($n=426; r=.76, p<.0001, 95\% \text{ Confidence Interval [CI]} .72 - .80$). The correlation between APHV and PAAS is also good ($n=354; r=.59, p<.0001, 95\% \text{ CI } .52 - .66$) as is the correlation between HDSDS and PAAS ($n=325; r=.63, p<.0001, 95\% \text{ CI } .56 - .69$).
**Table 2-1 - Descriptive Statistics, Distribution Characteristics, and Intercorrelations of Pubertal Timing Indices in the Avon Longitudinal Study of Parents and Children**

<table>
<thead>
<tr>
<th>Index</th>
<th>n</th>
<th>Mean</th>
<th>(SD)</th>
<th>Range</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Correlation with APHV</th>
<th>Correlation with HDSDS</th>
<th>Correlation with PAAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>APHV</td>
<td>460</td>
<td>160.26</td>
<td>13.04</td>
<td>120,189</td>
<td>-0.32</td>
<td>0.24</td>
<td>1</td>
<td>0.76</td>
<td>0.59</td>
</tr>
<tr>
<td>HDSDS</td>
<td>426</td>
<td>-0.015</td>
<td>0.82</td>
<td>(-2.18,1.85)</td>
<td>-0.25</td>
<td>-0.58</td>
<td>0.76</td>
<td>1</td>
<td>0.63</td>
</tr>
<tr>
<td>PAAS at 14y (%)</td>
<td>354</td>
<td>93.02</td>
<td>2.18</td>
<td>(83.6,99.9)</td>
<td>-0.3</td>
<td>0.61</td>
<td>0.59</td>
<td>0.63</td>
<td>1</td>
</tr>
</tbody>
</table>
2.4.2 Concordance of Indices using Quintiles

Table 2-2 presents the ranges contained within the quintiles described in Section 2.3.6 on all three indices. As can be seen in Figure 2-1, correct classification into the same or neighboring APHV quintile occurred approximately 90% across each quintile for HDSDS, and approximately 80% for PAAS. Kappa was calculated for each comparison: for APHV and HDSDS, the unweighted ($\kappa = .41; p<.0001$) and linear-weighted kappa ($\kappa = .64; p<.0001$) indicate moderate and substantial agreement respectively (Landis & Koch, 1977). For APHV and PAAS, the unweighted ($\kappa = .21; p<.0001$) and linear-weighted kappa ($\kappa = .43; p<.0001$) indicate fair to moderate agreement.

2.4.3 Association of the Percent Achieved of Adult Stature Index with Measures of Pubertal Status

Data from the Saguenay Youth Study dataset were used for these analyses. Of the 496 participants, a total of 70 were excluded: 38 participants were over the age of 17.5 and could not be included because the Khamis-Roche coefficients were not estimated beyond this age, 30 participants were missing at least one parental height, 1 participant did not have height and weight recorded,
Table 2-2 - Range of Quintiles for each Pubertal Timing Index

<table>
<thead>
<tr>
<th>Index</th>
<th>Quintile 1</th>
<th>Quintile 2</th>
<th>Quintile 3</th>
<th>Quintile 4</th>
<th>Quintile 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>APHV</td>
<td>(-1.92, -.89)</td>
<td>(-.89, -.07)</td>
<td>(.07, .13)</td>
<td>(.13, .75)</td>
<td>(.75, 2.80)</td>
</tr>
<tr>
<td>(z-score)</td>
<td>(189, 174)</td>
<td>(174, 162)</td>
<td>(162, 159)</td>
<td>(159, 150)</td>
<td>(150, 120)</td>
</tr>
<tr>
<td>APHV</td>
<td>(-2.18, -.79)</td>
<td>(.76, -.25)</td>
<td>(.24, .23)</td>
<td>(.23, .70)</td>
<td>(.71, 1.85)</td>
</tr>
<tr>
<td>(months)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDSDS</td>
<td>(-4.26, -.92)</td>
<td>(.92, -.27)</td>
<td>(.26, .32)</td>
<td>(.34, .78)</td>
<td>(.79, 3.12)</td>
</tr>
<tr>
<td>PAAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(z-score)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAAS</td>
<td>(83.6, 91.0)</td>
<td>(91.0, 92.4)</td>
<td>(92.4, 93.7)</td>
<td>(93.7, 94.7)</td>
<td>(94.7, 99.9)</td>
</tr>
<tr>
<td>(percent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2-1

+: % Participants within the same or adjacent quintile on HDSDS and APHV

X: % Participants within the same or adjacent quintile on PAAS and APHV

O: % Participants within the same HDSDS and APHV quintiles

◊: % Participants within the same PAAS and APHV quintiles
and 1 participant had an anomalous maternal height. Furthermore, two participants did not have PDS scores recorded and thus were not included in the analyses requiring PDS data. Only 216 participants had testosterone data; the remaining samples are yet to be analyzed. Table 2-3 presents the sample mean and standard deviations of important variables used in calculation of PAAS and further analyses.

Figure 2-2 shows that standardized scores from the PAAS index correlate strongly with testosterone levels in male adolescents \( n=216; r=.73; p<.0001; 95\% \text{ CI } .66 - .79 \), thus explaining over half of the variance \( n=216; r^2=.53; p<.0001 \). Figure 2-3 depicts the ANOVA comparing PAAS scores across groups based on PDS scores \( n=424; r^2=.45; p<.0001 \). We see a similar association between pubic-hair stage and PAAS \( n=424; r^2=.44; p<.0001 \). The ANOVA comparing testosterone levels across PDS score groups \( n=214; r^2=.44; p<.0001 \) shows a strong relationship, as does testosterone across pubic-hair stages \( n=214; r^2=.38; p<.0001 \).

**2.4.4 Comparing Performance of the Pubertal Timing Indices in the Two Studies**

We used ANOVA models to assess performance of the three indices of pubertal timing in ALSPAC in relation to the Genitalia and Pubic Hair stages ascertained with the fifth mailed pubertal questionnaire, completed at approximately 13.5 years of age (range: 13.1-13.8 years). Pubic Hair stage consistently outperformed Genitalia stage in explaining variance, often by more than 10\%: APHV by Pubic Hair stage \( n=338, r^2=.38; p<.0001 \); APHV by
Table 2-3 - Descriptive Statistics and Distribution Characteristics of the Saguenay Youth Study Sample

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean</th>
<th>(SD)</th>
<th>Range</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (months)</td>
<td>426</td>
<td>175.96</td>
<td>18.64</td>
<td>(144, 212)</td>
<td>0.14</td>
<td>-1.07</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>426</td>
<td>166.24</td>
<td>10.65</td>
<td>(140.5, 188.5)</td>
<td>-0.35</td>
<td>-0.56</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>426</td>
<td>60.41</td>
<td>16.56</td>
<td>(28.8, 153.5)</td>
<td>1.07</td>
<td>2.77</td>
</tr>
<tr>
<td>Maternal height (cm)</td>
<td>426</td>
<td>162.88</td>
<td>6.27</td>
<td>(147.3, 177.8)</td>
<td>0.09</td>
<td>-0.5</td>
</tr>
<tr>
<td>Paternal height (cm)</td>
<td>426</td>
<td>175.18</td>
<td>6.48</td>
<td>(151.1, 203.2)</td>
<td>0.04</td>
<td>0.92</td>
</tr>
<tr>
<td>PAAS (%)</td>
<td>426</td>
<td>94.18</td>
<td>5.00</td>
<td>(81.9, 101.0)</td>
<td>-0.52</td>
<td>-0.84</td>
</tr>
<tr>
<td>PDS</td>
<td>424</td>
<td>3.27</td>
<td>0.81</td>
<td>(1, 5)</td>
<td>-0.42</td>
<td>-0.11</td>
</tr>
<tr>
<td>Testosterone (ng/ml)</td>
<td>216</td>
<td>5.06</td>
<td>2.60</td>
<td>(.1, 11.3)</td>
<td>-0.18</td>
<td>-0.61</td>
</tr>
</tbody>
</table>
Figure 2-2 – Linear regression shows that 53% of variance in testosterone is explained by Percent Achieved of Adult Stature.
Figure 2-3 – Analysis of variance shows that 45% of variance in Percent Achieved of Adult Stature is explained by Pubertal Development Scale groups.
Genitalia stage \((n=312, r^2=.22; p<.0001);\) HDSDS by Pubic Hair stage \((n=314, r^2=.43; p<.0001);\) HDSDS by Genitalia stage \((n=291, r^2=.27; p<.0001);\) PAAS by Pubic Hair stage \((n=268, r^2=.31; p<.0001);\) PAAS by Genitalia stage \((n=247, r^2=.18; p<.0001).\)

To obtain an analogous comparison between questionnaire based stages and pubertal timing index in SYS within a similar narrow respondent age-range as with the ALSPAC questionnaire, we restricted our analysis shown in Figure 2-3 to participants between the ages of 13 and 14. This ANOVA between PAAS and PDS scores using the narrowed SYS age group \((n=118; r^2=.22; p<.0001)\) showed that the PDS scores in SYS performed slightly better than the ALSPAC questionnaire Genitalia Stage but not as well as staging based on Pubic Hair, which were shown earlier to explain 18% and 31% of variance respectively. The ANOVA between pubic-hair stage in SYS and PAAS using the narrowed age group \((n=118; r^2=.25; p<.0001)\) explains slightly less variance than the analogous question in ALSPAC which explains 31% of variance.

2.5 Discussion

As shown with the ALSPAC data, the three height-based indices of pubertal timing correlate well with each other, indicating that they capture similar information about the timing of puberty from height data. When we categorized index scores into quintiles, we found moderate to substantial congruence with linear-weighted kappa between the three different methods. For our purposes of assessing relative pubertal timing between participants, exact concordance between the indices is not essential; we are able to distinguish early from late matures if the score from one index falls into an equal or adjacent quintile in another index. For instance, we are able to infer that a
participant is a relatively early maturer if he receives scores of 4, 4, and 5 on the three indices, but not if he scores – for example - 1, 3, and 5.

We compared quintile scores from Height Difference in Standard Deviations and Percent Achieved of Adult Stature against the Age at Peak Height Velocity quintile scores, which were used as the standard. APHV scores are likely the most reliable given that its derivation takes into account all the longitudinal height data available from the participant. The HDSDS takes into account just two height data-points from the individual, and PAAS uses only one. The average percentage of participants' HDSDS and PAAS scores correctly classified within one quintile on APHV is 91.75% and 79.59% respectively. Thus we are confident that a participant’s pubertal timing as indicated by the HDSDS or PAAS indices will likely match what is obtained from APHV.

2.5.1 Association of the PAAS Index with Measures of Pubertal Status

It would be interesting to compare our three indices of pubertal timing with pubertal stage assessed through clinical examinations in a future study containing the appropriate data to see how well these height-based measures perform when compared to the gold standard. This would also answer the question as to which of the three indices most closely matches physical features of puberty (genitalia and pubic hair growth) that are assessed through the clinical examination.

Until we are able to answer these questions, it is reassuring to know that the height-based indices correlate with each other within ALSPAC, but it is essential that they are associated with hormone trends and methods of pubertal stage assessment. The SYS data illustrates that PAAS is indicative of pubertal timing by demonstrating that
it explains over half of the variance in testosterone levels and 45% of variance in Pubertal Development Scale (PDS) scores of male participants. By extension, the APHV and HDSDS indices likely also reflect changes in testosterone and PDS levels given their strong correlation with the PAAS; this will be verified in the future. It is important to note that the strength of the association between the various indices of pubertal timing and physical features of pubertal development would vary with age, showing an inverted U-shaped relationship. Pubertal timing indices will not distinguish physical development at pre-pubertal and post-pubertal ages when there is little to no variance in physical development. They will match physical development closest between the ages of 13 to 15 years, when there is most variance in the physical characteristics of the male population. We have described the process used to calculate each index, and will now discuss the background, interpretations, and key considerations in the use of the three pubertal timing indices.

2.5.2 Age at Peak Height Velocity

In 1976, Tanner published norms for APHV in males (Tanner & Whitehouse, 1976). Since then, many studies have used a number of different methods to calculate APHV, ranging from simple mid-year estimations (e.g., Lindgren, 1976) to spline interpolants (e.g., Sherar et al., 2011). Our script is automated and efficient, allowing for robust and consistent estimates in large datasets. In order to obtain a good estimate of APHV, it is important to obtain numerous closely spaced measurements of height between the ages of 10 and 17 years. At least five height measurements in this time
period spaced no more than two years apart should be used. Naturally, additional measurements more closely spaced together will yield more precise estimates of APHV.

A noteworthy observation about our APHV estimates is that the mean lies at approximately 13.3 years. This contrasts with the reported average estimate of 14 years from Tanner (Tanner & Whitehouse, 1976). Secular trends towards earlier female maturation have been studied (Euling et al., 2008) and are generally accepted. An analogous pattern towards earlier pubertal onset in males has been recently identified (Herman-Giddens et al., 2012). Our APHV estimates are congruent with claims of a secular trend towards earlier maturation in males.

2.5.3 Height Difference in Standard Deviations

Calculation of HDSDS only requires two height data-points, height at age 14 and final height (Wehkalampi et al., 2008). Since adult stature in males is not achieved until past age 20 and these data are not yet available, we used height at 17.5 years at which time most males have slowed their growth to less than 1cm annually (Roche & Davila, 1972). This index informs us of pubertal timing in a participant by describing the shift of their rank in the population based on height from an age when only half of the population has experienced a growth spurt compared with an age when growth has ceased. The logic is as follows: if an individual is an early maturer, he would have experienced a growth spurt already at 14 years, the average age for peak height velocity in the population (Marshall & Tanner, 1970). When comparing his rank in the population based on height at age 14 to final height, it should be apparent that his rank has decreased since the rest of the population caught up to him in growth. A late
maturer, on the other hand, will have a lower rank at age 14 but a relatively higher rank in final height due to the fact that he will experience a growth spurt in the interim. This change in relative position within the population is calculated using a difference in height standard scores; a positive HDSDS score indicates early maturation and a negative score indicates late maturation. Although the average APHV in our sample is slightly lower than that of 14 years reported elsewhere (Marshall & Tanner, 1970), the strong correlation we find between APHV and HDSDS indicates that the use of height at 14 years does not adversely affect this index.

2.5.4 Percent Achieved of Adult Stature

The Khamis-Roche method allows the prediction of an individual's adult stature by using age and sex dependant coefficients along with the individual's current height, weight, and mid-parental stature (Khamis & Roche, 1994). Through dividing current height by the Khamis-Roche prediction, we calculate the individual's Percent Achieved of Adult Stature. Participants who mature early would have attained a higher PAAS at age 14 when compared with their late maturing counterparts. This method will be most useful between the ages of 13 to 15 where the variation of pubertal stages across the population is at a maximum.

It is important to note that the original paper from Khamis and Roche contained incorrectly recorded coefficients; therefore the revised paper containing the corrected coefficients should be used (Erratum in: Pediatrics. 1995; 95:457). Furthermore, the Khamis-Roche coefficients were derived using a Caucasian sample of males and females from southwestern Ohio, thus careful consideration is required from the
investigator before extrapolating these equations to other populations. A crucial assumption inherent in the Khamis-Roche method used for derivation of the PAAS is that height is a highly heritable trait. The heritability of height in males ranges by country from .87 to .93 in the large twin samples from the GenomEUtwin study (Silventoinen et al., 2003). As environmental exposures do still play a role, it is important to exercise caution when applying the Khamis-Roche coefficients across a sample where some children and parents may have grown up in different environments. For example, immigration may lead to children being reared in different nutritional environments from their parents, and as a consequence parental height might cease to be a predictor of final height. As such, the Khamis-Roche method and consequently the PAAS method are best suited for populations where both parents and children were exposed to similar non-genetic effects, as is seen in the Saguenay Youth Study sample.

2.5.5 Comparison of PAAS to Pubertal Questionnaires

The questionnaires used to assess pubertal timing in ALSPAC and SYS are notably different in the use of pictures. The ALSPAC questionnaires contained four questions that asked about voice changes, armpit hair, pubic hair, and genital growth; the latter two questions were accompanied by drawings of the five Tanner stages to aid participants in answering accurately. Conversely, the PDS questionnaire used by the SYS asked questions about growth in height, presence of body and facial hair, acne, and voice chances. There were no pictures, nor any questions pertaining directly to genitalia. Answers from this set of questions are combined into a composite score that allows estimation of Tanner stage.
In analyses with questionnaire-derived measures of pubertal stage and the pubertal timing indices shown in Section 2.4.4, the Pubic Hair question from ALSPAC performed the best. The PDS score from SYS also performed well, with the genitalia staging from ALSPAC coming in third. One issue that the ALSPAC questionnaires suffered from was missing data. The prevalence of missing data ranged from 15% to 36% over the various pubertal questionnaires sent out to participants. Participants may potentially be uncomfortable with answering questions about genitalia as evidenced by patterns of missing data in the five questionnaires sent after age 13. The genitalia questions are left unanswered 5%-7% more often than the pubic hair questions. The PDS questionnaire presents a viable alternative for use in studies, as it does not contain a question asking about genitalia. Missing data did not play a role in our subset of the SYS data, where less than half a percent of participants failed to complete the PDS questionnaire.

2.5.6 Other Considerations

It may not be the case that height data from exactly age 14 or 17.5 is available for all participants. Each ALSPAC clinic, despite having a planned target age, measured participants in a wide age range. For example, the minimum, average, and maximum ages at attendance at the 9-year clinic were 8.75, 9.9, and 11.67 years respectively. Since ALSPAC contained longitudinal data, we were able to use spline interpolation to estimate height at the exact ages of 14 and 17.5 years for subsequent HDSDS and PAAS calculations.
Although we use population-based cohorts to demonstrate the derivation and correspondence between these indices, they can be used in smaller samples. This is because the scores of each index are standardized, and thus provide information on a participant’s relative timing of maturation compared to the rest of the sample. The data provided in Tables 2-1 and 2-2 may prove useful as a reference. We hope that other investigators will find our descriptions helpful in order to account for pubertal timing when studying the adolescent brain.
2.6 Acknowledgements

The authors would like to thank the following for their assistance: Drs. Mallar Chakravarty, Erin Dickie, Kate Northstone, Mark Palmert, and Melissa Pangelinan. AK received student financial support from the Ontario Graduate Scholarship and the Max and Ruth Wiseman Graduate Student Fellowship. TP is the Tanenbaum Chair in Population Neuroscience at the Rotman Research Institute, University of Toronto.

The Saguenay Youth Study project is funded by the Canadian Institutes of Health Research (TP, ZP), Heart and Stroke Foundation of Quebec (ZP), and the Canadian Foundation for Innovation (ZP). We thank all families who took part in the Saguenay Youth Study and the following individuals for their contributions in acquiring data: Manon Bernard (database architect, The Hospital for Sick Children), Jacynthe Tremblay and her team of research nurses (Saguenay Hospital), Helene Simard and her team of research assistants (Cégep de Jonquière) and Rosanne Aleong (program manager, Rotman Research Institute).

The UK Medical Research Council and the Wellcome Trust (Grant Ref: 092731) and the University of Bristol provide core support for ALSPAC. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. LDH is funded by a UK Medical Research Council Population Health Scientist fellowship (G1002375). This publication is the work of the authors and Dr. Margaret May will serve as guarantor for the contents of this paper. This research was specifically funded by NIH, 5R01MH085772.
2.7 Supplementary Section:

Cubic Spline (Hundley, 2009): Given \( n \) data points, \((x_1, y_1), \ldots, (x_n, y_n)\), a cubic spline is a piecewise-defined function of the form:

\[
S_1(x) = y_1 + b_1(x - x_1) + c_1(x - x_1)^2 + d_1(x - x_1)^3 \quad \text{for } x \in [x_1, x_2]
\]

\[
S_2(x) = y_2 + b_2(x - x_2) + c_2(x - x_2)^2 + d_2(x - x_2)^3 \quad \text{for } x \in [x_2, x_3]
\]

\[\vdots\]

\[
S_{n-1}(x) = y_{n-1} + b_{n-1}(x - x_{n-1}) + c_{n-1}(x - x_{n-1})^2 + d_{n-1}(x - x_{n-1})^3 \quad \text{for } x \in [x_{n-1}, x_n]
\]
Chapter 3 - Testosterone trajectories and reference ranges in a large longitudinal sample of male adolescents

Ammar Khairullah, Laura C. Klein, Suzanne M. Ingle, Margaret T. May,
Courtney A. Whetzel, Elizabeth J. Susman, Tomáš Paus

Submitted: Journal of Adolescent Health
3.1 Abstract

Purpose – Pubertal dynamics plays an important role in physical and psychological development of children and adolescents. We aim to provide reference ranges of plasma testosterone in a large longitudinal sample, and to describe a measure of testosterone trajectories.

Methods – Using data reported in the literature, we conducted a meta-analysis of testosterone levels in males between the ages of 6 and 19 years. This reference range was supplemented by the longitudinal measurements of plasma testosterone obtained from 513 males in the Avon Longitudinal Study of Parents and Children. We use integration of a cubic spline fitted to each participant’s testosterone trajectory to calculate a measure of average exposure to testosterone over adolescence.

Results - ALSPAC testosterone data are congruent with other reports in the literature. The measure of average exposure to testosterone is associated with different features of testosterone trajectories and the growth spurt during puberty.

Conclusions - The average exposure to testosterone is a useful measure for future investigations using testosterone trajectories to examine pubertal dynamics.
3.2 Introduction

Puberty (activation of the hypothalamic-pituitary-gonadal axis) and adolescence (maturation of adult social and cognitive behaviours) are intertwined (Sisk & Foster, 2004). Variations in pubertal trajectories have been associated with those in physical and behavioral development. For example, an earlier onset of puberty predicts a higher body mass index and central fat-mass in 19-year old men (Kindblom et al., 2006). Early pubertal timing and faster pubertal tempo appear to be associated with depressive symptoms in male adolescents (Mendle, Harden, Brooks-Gunn, & Graber, 2010).

The activation of the hypothalamic-pituitary-gonadal (HPG) axis stimulates the male testes and, in turn, increases production of testosterone (Grumbach, 2002; Hiort, 2002). In males, testosterone is a sex steroid that plays a critical role in numerous pubertal processes, including genital growth (Huang et al., 2012), change in body composition (Hansen, Bangsbo, Twisk, & Klausen, 1999), and maturation of the brain (Peper et al., 2009; Perrin et al., 2008). Given testosterone’s crucial role in puberty, its dynamics may be driving associations previously ascribed to pubertal timing and tempo (Mendle et al., 2010), thus underscoring the importance of studying testosterone trajectories in relation to healthy development.

A number of human studies have investigated changes in testosterone levels throughout the lifespan of males. We reviewed the literature for reports that provided normal ranges of total testosterone during male adolescence, pooled together data from the five studies identified (Elmlinger, Kühnel, Wormstall, & Döller, 2005; Hero, Wickman, Hanhijärvi, Siimes, & Dunkel, 2005; Lee, Jaffe, & Midgley, 1974; Schnakenburg, Bidlingmaier, & Knorr, 1980; Starka, Duskova, & Hill, 2008), and report the weighted
values. Longitudinal measures of testosterone are a pre-requisite for investigating testosterone trajectories in adolescent males. To this end, we quantified plasma levels of total testosterone in 513 typically developing males from the Avon Longitudinal Study of Parents and Children (ALSPAC), where multiple blood samples were collected during participants’ childhood and adolescence. We also include ranges for sex hormone binding globulin (SHBG) and the calculated free and bioavailable fractions of testosterone, as these are of interest when studying the effect of testosterone on the brain and body. Finally, we describe a measure of average exposure to testosterone throughout adolescence derived from participants’ testosterone trajectories, and examine the trajectories in context of the adolescent growth spurt, as indexed by Age at Peak Height Velocity (APHV), and several characteristics of the testosterone trajectories.

3.3 Methods

3.3.1 Meta-analysis of Published Data

In this meta-analysis, we included studies that reported either a mean (and standard deviation) or a median (with range) for values of total testosterone measured from plasma or serum samples in males between the ages of 6 and 20 years. To identify studies for this meta-analysis, we conducted a PubMed search in January 2012, using combinations of the following search terms: ‘testosterone’, ‘androgen’, ‘normal’, ‘reference range’, ‘adolescence’, ‘childhood’, ‘serum’, and ‘plasma’. We excluded studies for the following reasons: presenting data in a graph without an accompanying
table, presenting data in subgroups of large age-ranges (greater than 2-year intervals), or reporting reference ranges for salivary testosterone or free testosterone without total testosterone. Furthermore, within the selected studies, we included only those age subgroups containing at least 10 participants.

First, we standardized the units for total testosterone into nanomoles per liter using a conversion factor where necessary (1 ng/mL = 3.467 nmol/L). For studies reporting median and range, we used methods described by Hozo and colleagues to estimate the mean and standard deviation (SD) (Hozo, Djulbegovic, & Hozo, 2005). When pooling data across studies for each age subgroup within the meta-analysis, we used equations from Borenstein and colleagues to compute the weighted mean and corresponding standard deviation (Borenstein, Hedges, Higgins, & Rothstein, 2009).

3.3.2 The Avon Longitudinal Study of Parents and Children

We used blood samples obtained from participants in the Avon Longitudinal Study of Parents and Children (ALSPAC), a birth cohort that recruited 14,541 pregnant women resident in Avon, United Kingdom with an expected delivery date between April 1991 and December 1992. Since then, the children and their parents have been studied extensively with longitudinal data available from self-administered questionnaires sent to their homes, linkages to medical records, and through clinical examinations carried out during study “clinics” (Boyd et al., 2012). Ethical approval for ALSPAC was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.
Blood samples were taken at select clinics over the course of the study, roughly at the following years of age: 7, 9, 11, 13, 15, and 17. We obtained and analyzed sex steroids from the blood samples of a subset of 513 male participants who had been recruited for further study using magnetic resonance imaging (MRI). We did not utilize blood samples from the 7-year clinic, as most individuals are likely to have very low or undetectable levels of sex steroids at this age. All participants had a maximum of five and a minimum of three blood samples across these five clinics (5 samples: n=232; 4 samples: n=213; 3 samples: n=68).

3.3.3 Quantification of Testosterone and Sex Hormone Binding Globulin

The blood samples, in the form of lithium heparin plasma, were sent from ALSPAC to the Biomarker Core Laboratory at The Pennsylvania State University, where enzyme-linked immunosorbent assays were conducted - using commercially available kits - to determine plasma concentrations of testosterone (EIA-1559; DRG International; Springfield, New Jersey, USA) and SHBG (IB79131; Immuno-Biological Laboratories, Inc.; Minnesota, MN, USA). For testosterone and SHBG, the sample test volumes were 25 ul and 10 ul, respectively. The assays had lower limits of sensitivity of 0.083 ng/mL (testosterone) and 0.77 nmol/L (SHBG), with average inter- and intra-assay coefficients of variation less than 10%. The assay sensitivity is calculated by subtracting two standard deviations from the mean of 20 identical runs of the zero standard.
For quality-control purposes, ten percent of participant samples were tested in duplicate, balanced across plates while the rest of the samples were run in singlet to increase efficiency. All samples from a participant were run on the same plate. Of the samples run in duplicate, test values that varied by more than 5% were subject to repeat testing. The first-well values from all samples are used in data analyses for consistency. For testosterone samples that yielded undetectable hormone values or values below the lower limit of sensitivity for the assay, we used the lower bound value of the detectable range for the assay: 0.083 ng/mL.

3.3.4 Testosterone’s Circadian Rhythm

Since testosterone demonstrates a circadian rhythm (Plymate, Tenover, & Bremner, 1989) and the time of venipuncture was not standardized across study participants due to logistical reasons, we used multilevel modeling to predict total testosterone measurements at a standard time of day (12PM). The effect of time of venipuncture on testosterone levels differed across age groups; therefore separate models were fitted for each age (9, 11, 13, 15 and 17 years). In each of these models age was included as a continuous variable and time of sample was included as a categorical variable split into half-hour increments. The models fitted for ages 15 and 17 also included age squared, as there was evidence of a quadratic effect of age on testosterone. We assessed model fit by comparing the observed with the predicted values. There were clear differences between observed and predicted values at age 13 indicating that models for age 13 did not fit well. We hypothesized that this was because age 13 is the time when pubertal changes are most likely to happen. We therefore fit the
age 13 model separately for those who had and had not reached age at peak height velocity. Age was not explicitly included in these models and again, time of sample was modeled as a categorical variable using half-hour increments.

3.3.5 Measure of Average Exposure to Testosterone Throughout Adolescence

We calculated a measure of average exposure to testosterone throughout adolescence for participants based on the area under the curve of their testosterone trajectory. The longitudinal measurements of total testosterone and the respective age in months at venipuncture were used to plot the participant’s total testosterone values during his childhood and adolescence. The shape preserving spline-interpolation function from the MATLAB Basic Fitting toolbox (Mathworks, Natick, MA USA) was used to generate a curve through the data-points. The area under the curve (integral) was calculated, with respect to ground, for the section of the spline between the first and last blood sample. This is similar to the method applied by Pruessner and colleagues in context of cortisol sampled at multiple points within an hour (Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). To accommodate variations between participants’ age ranges from the first to the last clinic, we standardized this variable by dividing the integral value for each participant by their exact age range (in months) between the first and last clinic. The resulting value is an estimate of the arithmetic mean of monthly measurements, if testosterone were sampled each month throughout the participant’s adolescence.
3.3.6 Growth spurt and Characteristics of Testosterone Trajectories

For each participant, we calculated Age at Peak Height Velocity (APHV) using up to nine height measurements collected at each ALSPAC clinic, from 7 through 17 years of age, as described previously (Khairullah et al., 2013). Briefly, the height data for each participant are fitted with a cubic spline from the Basic Fitting Toolbox in MATLAB (Mathworks, Natick, MA USA). Taking the derivative of the height curve gives us the growth curve, from which we ascertain the Peak Height Velocity and the corresponding age (in months) when this occurs.

We estimated the analogous measures for testosterone data, Peak Testosterone Change and Age at Peak Testosterone Change, by applying the same spline method. This calculation was restricted to participants who had blood collected at all five clinics (n=232), given that fewer than five data-points would likely lead to imprecise estimates.

3.3.7 Free and Bioavailable Testosterone

To calculate levels of free testosterone and to derive values of bio-available testosterone, we used the Södergård equation in conjunction with the SHBG concentrations and adjusted values of total testosterone (Södergård, Bäckström, Shanbhag, & Carstensen, 1982).
3.3.8 Statistics

We used Stata version 12.1 (StataCorp LP, College Station, TX) for multilevel modeling to adjust values of Total Testosterone. Other statistical analyses used JMP 9 for Macintosh (SAS Institute, Inc., Cary, NC): simple linear regression was used to relate the measure of average testosterone exposure to APHV, Age at Peak Testosterone Change, Peak Testosterone Change, and Total Testosterone at 17 years. In order to visualize participants’ testosterone trajectories and to investigate whether they varied in form between those with higher or lower average testosterone, we divided participants into quintiles based on the measure of average exposure to testosterone and graphed trajectories separately for each quintile. The five graphs were compared by visual inspection.

3.4 Results

3.4.1 Meta-analysis

Table 3-1 reports the pooled data from the meta-analysis. On average, there is little testosterone (<1 nmol/L) found in males before the age of 10 years, at which point the levels begin to rise. Between 10 and 15 years of age, the plasma testosterone levels increase nearly seven-fold. The rate of increase slows as the males reach adult levels (~15 nmol/L) between age 16 and 17 years.
3.4.2 ALSPAC: Testosterone levels

The models we fitted to remove the effect of circadian rhythm explained 1.7%, 9.5%, 5.8%, and 6.3% of the variance in total testosterone at the 9, 11, 15 and 17-year clinics, respectively. Models for the 13-year clinic were fitted separately for participants who had not yet reached APHV and those that had passed APHV; the variance explained was 17.3% and 22.1%, respectively. Note that the relatively high proportion of variance in the unadjusted levels of testosterone during the 13-year old clinic is due to the combination of rising testosterone levels and morning-afternoon sampling schedule in this clinic (see Discussion for details). Time of sample explained less than 0.1% of the variance in the adjusted testosterone values at each clinic.

Table 3-2 reports the meta-analysis after incorporating the adjusted total testosterone values from ALSPAC. Table 3-3 provides descriptive statistics for the distributions of age, time at venipuncture, and concentrations of unadjusted and adjusted total testosterone, SHBG, free testosterone and bioavailable testosterone. These values are similar to those reported in previous studies (Table 3-1).
TABLE 3-1 – Meta-Analysis of Total Testosterone Values (nmol/L)

<table>
<thead>
<tr>
<th>Age</th>
<th>Study</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th># Studies</th>
<th>n</th>
<th>Weighted Mean</th>
<th>Weighted SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 - 7</td>
<td>Schnakenburg 1980</td>
<td>18</td>
<td>0.74</td>
<td>0.15</td>
<td>105</td>
<td>1.6</td>
<td>2.8</td>
<td>2</td>
<td>123</td>
<td>0.74</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 - 9</td>
<td>Elminger 2005</td>
<td>32</td>
<td>0.54</td>
<td>0.20</td>
<td>25</td>
<td>0.92</td>
<td>0.27</td>
<td>3</td>
<td>217</td>
<td>0.68</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 - 11</td>
<td>Starka 2008</td>
<td>45</td>
<td>2.07</td>
<td>1.22</td>
<td>23</td>
<td>2.25</td>
<td>1.12</td>
<td>34</td>
<td>2.43</td>
<td>0.76</td>
<td>60</td>
<td>1.04</td>
<td>1.39</td>
<td>5</td>
<td>403</td>
<td>2.14</td>
<td>0.51</td>
</tr>
<tr>
<td>12 - 13</td>
<td>Lee 1974</td>
<td>74</td>
<td>8.18</td>
<td>4.91</td>
<td>40</td>
<td>10.61</td>
<td>3.18</td>
<td>102</td>
<td>9.96</td>
<td>2.08</td>
<td>118</td>
<td>5.17</td>
<td>2.09</td>
<td>5</td>
<td>781</td>
<td>8.03</td>
<td>1.27</td>
</tr>
<tr>
<td>14 - 15</td>
<td>Hero 2005</td>
<td>87</td>
<td>10.06</td>
<td>5.07</td>
<td>48</td>
<td>14.54</td>
<td>4.95</td>
<td>616</td>
<td>9.7</td>
<td>7.4</td>
<td>83</td>
<td>14.60</td>
<td>1.93</td>
<td>5</td>
<td>890</td>
<td>13.46</td>
<td>1.57</td>
</tr>
<tr>
<td>18 - 20</td>
<td></td>
<td>56</td>
<td>18.10</td>
<td>3.99</td>
<td>23</td>
<td>15.35</td>
<td>4.90</td>
<td>302</td>
<td>14.5</td>
<td>8.2</td>
<td>3</td>
<td>381</td>
<td>16.70</td>
<td>2.89</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3-2 – Meta-Analysis including ALSPAC Data

<table>
<thead>
<tr>
<th>Age</th>
<th># Studies</th>
<th>n</th>
<th>Weighted Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 - 7</td>
<td>2</td>
<td>123</td>
<td>0.74 (0.15)</td>
</tr>
<tr>
<td>8 - 9</td>
<td>4</td>
<td>556</td>
<td>0.78 (0.08)</td>
</tr>
<tr>
<td>10 - 11</td>
<td>6</td>
<td>944</td>
<td>2.00 (0.46)</td>
</tr>
<tr>
<td>12 - 13</td>
<td>6</td>
<td>1164</td>
<td>8.04 (1.23)</td>
</tr>
<tr>
<td>14 - 15</td>
<td>6</td>
<td>1406</td>
<td>13.58 (1.43)</td>
</tr>
<tr>
<td>16 - 17</td>
<td>6</td>
<td>900</td>
<td>16.72 (1.68)</td>
</tr>
<tr>
<td>18 - 19</td>
<td>4</td>
<td>442</td>
<td>16.84 (1.97)</td>
</tr>
</tbody>
</table>
Table 3-3 – Hormone Data from the Avon Longitudinal Study of Parents and Children

<table>
<thead>
<tr>
<th>Age</th>
<th>n</th>
<th>Time of Venipuncture</th>
<th>Total Testosterone (nmol/L)</th>
<th>Time Corrected Total Testosterone (nmol/L)</th>
<th>SHBG (nmol/L)</th>
<th>Free Testosterone (pmol/L)</th>
<th>Bio-available Testosterone (nmol/L)</th>
<th>% Reached APHV</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.82(.30); 9.42-11.58</td>
<td>441</td>
<td>13.78(2.05); 10.83-16.87</td>
<td>.72(.58); .29-4.39</td>
<td>.82(.09); .73-1.39</td>
<td>92.28(43.64); 4.33-262.29</td>
<td>8.38(3.33); 2.72-30.92</td>
<td>.17(.07); .06-6.4</td>
<td>1</td>
</tr>
<tr>
<td>11.69(.21); 11.33-13.08</td>
<td>482</td>
<td>13.74(2.14); 10.75-19.38</td>
<td>1.47(1.83); .29-15.24</td>
<td>1.55(1.23); .08-10.71</td>
<td>74.90(37.45); 4.88-267.52</td>
<td>20.36(23.00); 0.97-193.54</td>
<td>.42(.47); .02-3.98</td>
<td>7</td>
</tr>
<tr>
<td>13.81(.17); 13.08-14.67</td>
<td>414</td>
<td>13.81(2.13); 9.75-19.05</td>
<td>6.87(5.10); .29-35.53</td>
<td>9.05(4.66); 3.50-28.30</td>
<td>45.41(26.03); 3.53-164.24</td>
<td>186.86(125.34); 20.60-582.22</td>
<td>3.84(2.57); .42-11.96</td>
<td>69</td>
</tr>
<tr>
<td>15.38(.25); 14.5-17.33</td>
<td>464</td>
<td>10.31(1.97); 8.00-14.00</td>
<td>15.18(5.48); .85-45.62</td>
<td>14.81(2.72); 7.40-29.82</td>
<td>30.81(13.32); 4.01-89.09</td>
<td>356.94(96.07); 81.86-638.33</td>
<td>7.33(1.97); 1.68-13.11</td>
<td>95</td>
</tr>
<tr>
<td>17.70(.33); 16.58-19.17</td>
<td>413</td>
<td>10.66(1.99); 8.33-15.00</td>
<td>17.05(5.41); 1.63-42.48</td>
<td>16.50(2.65); 8.68-28.59</td>
<td>26.25(12.37); 3.35-140.09</td>
<td>438.81(96.12); 56.72-780.88</td>
<td>9.01(1.97); 1.17-16.04</td>
<td>100</td>
</tr>
</tbody>
</table>

Values are reported as Mean(SD); Range. Time of venipuncture is calculated as hour plus minutes/60.

The data is taken from the following ALSPAC Focus Clinics: Focus at 9, Focus at 11, Teen Focus 2, Teen Focus 3, Teen Focus 4.
Figure 3-1 - Average testosterone exposure is inversely related to (A) timing of growth spurt and (B) timing of the largest testosterone increase, but is positively related to (C) magnitude of the largest testosterone increase and (D) total testosterone at 17 years.
Figure 3-2 - Higher versus lower average exposure to testosterone is associated with earlier rising testosterone trajectories that remain relatively high, and earlier onset and greater magnitude of peak change in testosterone. Trajectories are plotted with adjusted testosterone values (see Methods).
Figure 3-3 - Earlier versus late growth spurt is associated with earlier rising testosterone trajectories, and earlier onset of peak change in testosterone. Trajectories are plotted with adjusted testosterone values (see Methods).
3.4.3 ALSPAC: Growth Spurt and Characteristics of Testosterone Trajectories

Eight participants with very high APHV estimates (>200 months) were excluded, as described previously (Khairullah et al., 2013). The mean age of growth spurt, defined as APHV, is 159.83 months (SD: 13.33 months), with a range of 120 to 189 months.

The distribution of Age at Peak Testosterone Change in our sample is bimodal (see Discussion), with peaks at 153 and 174 months (Mean: 162.87, SD: 10.42). On the other hand, values of the Peak Testosterone Change are normally distributed (Mean: 0.63 [(nmol/L)/month], SD: 0.21).

3.4.4 Measure of Average Exposure to Testosterone

We excluded the participants missing two blood samples (n=68) as their calculated values on this measure were lower compared with participants missing one or none. Values from participants missing one sample were not different from participants with all five blood samples. The values within the subgroup of participants missing one sample, however, differed based on which clinic’s sample was missing (n=213; $r^2=0.08; p=0.002$). Participants missing a sample from the second or third clinics had higher estimates, whereas those missing a sample from the fourth or fifth clinic had lower estimates. We used residuals from this model to estimate corrected values (mean + residual) in this subgroup. A total of 445 participants had a measure of average exposure to testosterone calculated (Mean: 8.63 nmol/L; SD: 2.15 nmol/L; Range: 3.34 – 19.50 nmol/L).
The measure of average exposure to testosterone is associated with: (A) Age at Peak Height Velocity \((n=437; r^2=0.23; b=-3.01; t=-11.28; p<.0001)\); (B) Age at Peak Testosterone Change \((n=232; r^2=0.39; b=-3.21; t=-12.03; p<.0001)\); (C) Peak Testosterone Change \((n=232; r^2=0.06; b=0.03; t=3.68; p=0.0003)\); (D) Testosterone at the 17 year clinic \((n=409; r^2=0.28; b=0.68; t=12.73; p<.0001)\). These associations are shown in Figure 3-1.

The means (and SDs) of the quintiles of average exposure to testosterone were:

1st quintile: 5.92 (0.91); 2nd: 7.43 (0.29); 3rd: 8.43 (0.30); 4th: 9.59 (0.39); and 5th: 11.72 (1.56) nmol/L.

Figure 3-2 shows participants’ testosterone trajectories and testosterone-change trajectories (calculated as Testosterone at Present Clinic – Testosterone at Previous Clinic) for each of the average-testosterone quintiles. There are clear differences in the shape of these trajectories across the average-testosterone quintiles: participants in the fifth quintile demonstrate a rise in testosterone relatively earlier and of greater magnitude whereas those within the first quintile tend to experience a later and lower increase in testosterone. A notable feature of the testosterone-change trajectories is the peak, indicating when a participant’s testosterone increased the most. Almost all participants from Quintile 5 experience an increase in testosterone-change before age 14 years, whereas the participants in Quintile 1 experience their increase almost exclusively after age 14 years.

Figure 3-3 shows testosterone trajectories and testosterone-change trajectories by APHV quintiles. We find early rising and steep trajectories for early maturers (APHV Quintile 1), and conversely, late rising and gentler trajectories for late maturers (APHV
Quintile 5). Similarly to what was seen in Figure 3-2, participants in APHV Quintile 1 experience an increase before age 14 years, whereas those in APHV Quintile 5 experience their increase after age 14 years. Note that the change trajectories in Figure 3-2 and 3-3 show only the 232 participants with all five blood samples.
3.5 Discussion

The quantification of the testosterone in male ALSPAC participants yields values that match with ranges reported in other studies. Comparing Table 3-1 and Table 3-3, we see similar trends in the mean values of total testosterone between 9 and 19 years of age. Of the five studies included in the meta-analysis, the ALSPAC data resemble the most (Elmlinger et al., 2005; Lee et al., 1974) values reported in the two studies by Elmlinger and Lee. Other studies report slightly lower values in their samples between the ages of 12 to 15 years. Our synthesized results combining the literature with ALSPAC serve as useful reference ranges for future investigations that pertain to testosterone in typically developing male adolescents (Table 3-2).

We also describe a simple measure for characterizing the average exposure to testosterone over adolescence based on longitudinal measures. We use an integral to derive this measure, as simply taking the arithmetic mean of the participant’s assayed testosterone measurements may produce a biased estimate, for example if the samples were clustered towards the beginning or end of adolescence. Using the integral of a fitted spline helps minimize this bias.

This measure is associated with timing of the pubertal growth spurt and several characteristics of testosterone trajectories: the age and magnitude of the rise in testosterone, and testosterone levels at the end of the trajectory. Participants who have a higher measure of average exposure to testosterone tend to achieve an early growth spurt (known to be associated with earlier activation of the HPG axis (Stanhope, Pringle, & Brook, 1988)). They also show an earlier and larger rise in testosterone levels, and have higher levels of testosterone at 17 years. Since testosterone levels
remain high once they accelerate, early and large increases in testosterone result in a
greater exposure to testosterone over adolescence. Note that the distribution of Age at
Peak Testosterone Change is bimodal because our blood samples are taken at two-
year intervals. We predict that this distribution would be closer to normal if the frequency
of sampling was greater.

Studying testosterone trajectories in a large sample is made challenging by
circadian rhythms that cause fluctuations in hormone levels over the time course of a
day (Plymate et al., 1989; Sharma et al., 1989). As such, the time of blood draw is an
important consideration in these studies. In blood collected every 45 minutes over a
course of 25.5 hours (in five healthy males), total testosterone peaked at approximately
0500h and 1000h, and troughed at approximately 1400h and 2100h (Cooke, McIntosh,
& McIntosh, 1993). The ALSPAC blood samples were obtained throughout the day and
required statistical adjustment to minimize effects of the circadian rhythm on the
hormone values. This was particularly the case in data from the third clinic (Teen Focus
2; ~13 years) where a greater effect of the circadian rhythm was observed; during this
clinic, time of sampling ranged from 0900h to 1900h. The two preceding clinics also
collected blood samples throughout the day, but the data were not strongly affected by
the circadian rhythm given the low testosterone values present in participants at these
young ages. The data from clinics at ages 15 and 17 years are not affected as much,
likely because nearly all the blood samples were taken before the 1400h nadir. In the
first three clinics, where many participants were measured after this nadir, the time
adjustment increased their values and therefore increased the mean total testosterone
for these clinics. In comparison, the last two clinics have a proportionally smaller
adjustment that decreased the mean testosterone slightly. Ideally, all participants in the
study would have been sampled at the same time of day, preferably in the morning hours, but this was not possible for logistical reasons.

Testosterone, particularly its bioavailable fraction, is an important androgen that exerts an effect on multiple biological targets in the body. In addition to the multitude of physical changes initiated during puberty, bioavailable testosterone has been implicated in development and maturation of the adolescent brain (Peper et al., 2009; Perrin et al., 2008), as well as in psychopathology such as depression and risk of psychosis (Sankar & Hampson, 2012; van Rijn et al., 2011). Investigations into the effects of hormones have usually been conducted in a cross-sectional manner with Venipuncture to collect blood for hormone quantification carried out in temporal proximity to measurement of the outcome of interest (e.g., body composition or behavior). Inquiry into the effect of testosterone exposure over a long time period has rarely been conducted.

We plan to embark on such studies using data in ALSPAC, where the database includes longitudinally collected blood samples and measures of psychopathology throughout adolescence, in addition to a single time-point of Magnetic Resonance Imaging in late adolescence to characterize the morphology of the brain. In such a design, longitudinal measures can be collapsed into a measure of average exposure and then related to an outcome measured at the end of the study as has been done elsewhere, for example, in the context of air pollution and cardiopulmonary mortality (Pope, Burnett, & Thun, 2002). Our measure of testosterone exposure over adolescence showed associations with salient characteristics of testosterone trajectories. In this way, we can study the effect of testosterone dynamics on shaping the adolescent brain, and then examine the relationship between testosterone and risk
of psychopathology. Other investigations aiming to examine longitudinal effects of hormones may find utility in the measure we describe, shown to be associated with characteristics of the hormone trajectory.
3.6 Acknowledgements

The authors would like to thank the following for their assistance: Mary McDiarmid, Drs. Rosanne Aleong, Mallar Chakravarty, Shafagh Fallah, Kate Northstone, Mark Palmert, Zdenka Pausova, and Sue Ring. Furthermore we would like to thank Dr. Luboslav Starka for providing means and standard deviations for data reported in Starka et al., 2008. AK received student financial support from the Ontario Graduate Scholarship and the Max and Ruth Wiseman Graduate Student Fellowship. TP is the Tanenbaum Chair in Population Neuroscience at the Rotman Research Institute, University of Toronto. This research was funded by NIH, 5R01MH085772 (to TP). Dr. Tomáš Paus had full access to all the data in the study and takes responsibility for the integrity of the data and accuracy of the data analysis. The authors declare no conflicts of interest.

The UK Medical Research Council and the Wellcome Trust (Grant Ref: 092731) and the University of Bristol provide core support for ALSPAC. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. This publication is the work of the authors and Dr. Margaret May will serve as guarantor for the contents of this paper.
Chapter 4 - General Discussion

The main contributions of this thesis are the development of methodologies to establish height-based indices of pubertal timing and a measure of testosterone trajectories. Longitudinal height measurements during adolescence are used to extract the Age at Peak Height Velocity (APHV), or Height Difference in Standard Deviations (HDSDS). If longitudinal data is not available, a single height and weight measurement, preferably taken at or close to 14 years of age along with parental stature, can be used to derive the Percent Achieved of Adult Stature (PAAS) as shown. Data from the Saguenay Youth Study shows strong associations between the PAAS index and both PDS questionnaire and testosterone levels; PAAS explains 45% and 53% of the variance in these measures respectively. The advantages and disadvantages of each of these three indices will be explored in a later section.

As the APHV, HDSDS, and PAAS indices are highly correlated with one another, we can hypothesize that the same strong association that PAAS shares with questionnaire-based assessments and hormones will also exist with the other indices of pubertal timing. This is addressed in part in Chapter three, where we see that APHV explains 21% of variance in the measure of average testosterone exposure. This method incorporates the integral of each participant’s testosterone curve over adolescence, which is then standardized based on the interval of the participant’s age between the earliest and last clinic. The measure derived in this way provides a quantification of participants' testosterone trajectories based on the average exposure to testosterone throughout their adolescence.
The following sections of the general discussion pertain to the limitations of the work, the challenges faced along the way, and other considerations relevant to implementation of these methods. A description of future directions follows this discussion, where we will delineate how the methods described in this thesis will be applied to an investigation of the role of testosterone timing and dynamics in maturation of white matter in brain of male adolescents.

4.1 Picture Based Assessment of Pubertal Stage

The initial obstacle that became apparent from the ALSPAC dataset was the presence of biologically implausible responses in the pubertal questionnaires filled out either by the participant and/or his parent. Specifically, the question asking participants to identify their Tanner Genital Stage with pictures posed a problem. The responses it yielded deviated greatly from the norms established by Marshall and Tanner. For example, the mean age for achieving Tanner Genital Stage 2 reported by Marshall and Tanner is 11.64 years, with a standard deviation of 1.07 years (Marshall & Tanner, 1970). Only a small number of participants should reach this stage by 9.5 years (two standard deviations under the mean), but in our sample, 68% and 73% of respondents indicated that they were at Tanner Genital Stage 2 or greater in the 8 and 9 year questionnaires respectively.

Other studies have also found that this particular picture-based assessment of pubertal status yields data that do not correspond well with different measures of pubertal status (Shirtcliff et al., 2009). The work by Shirtcliff and colleagues
demonstrates that the picture-based assessment shows the poorest concordance with testosterone levels in boys, in comparison with both a physical examination and the Pubertal Developmental Scale. Since pubertal status assessed through the PDS captured changes in testosterone levels well (Shirtcliff et al., 2009), the presence of pictures in the other questionnaire appear to interfere with accuracy of self-report in males. Since an alternative questionnaire method of ascertaining pubertal development was not available in the ALSPAC database, the invalid data from these pubertal questionnaires provided the impetus to develop methodologies for assessing pubertal timing.

4.2 Height Measurements

The growth spurt is an important feature of puberty (Karlberg, 2002; Marshall & Tanner, 1970). Since ALSPAC had numerous longitudinal measurements of height for most participants, the prospect of pursuing derivation of height-based indices of pubertal timing was formed. We used measurements of height recorded from participants by ALSPAC staff using a Harpenden Stadiometer at the ALSPAC Focus Clinics.

Determining the best strategy for managing missing measurements posed a challenge when using the longitudinal height data. Approximately 70% of participants in the analyzed sample had a full complement of nine height measurements, but 4% were missing 3 or more data-points. We have dealt with the missing data through the use of spline interpolation, as described in the next section.
4.3 Age at Peak Height Velocity

We fit a cubic spline to each participant’s height measurements. The spline interpolation creates a curve connecting the participant’s measurements, even if some of the measurements are missing. Taking this interpolated spline as the participant’s height curve, differentiation yields a growth curve, from which APHV is easily extracted. The spline method gives us an estimate for the exact month of the growth spurt.

Naturally, we might expect that APHV estimations obtained from a participant with one or more missing data-points may be less precise compared with an estimate made from a full set of height measurements. Nevertheless, we find that the distributions of APHV for participants with one or two missing data-points are almost identical, in terms of mean and SD, to the distribution for participants with all measurements. This could be a fortuitous consequence of the fact that the clinic closest to age 14, the reported average APHV in the population, and the adjacent clinics on either side had low incidences of missing data compared with other clinics. The percentage of missing data in these three clinics with crucial height measurements ranged from 1.7-3.3%. In comparison, other clinics, particularly at the younger ages, had up to 10.3% of data missing.

Important benefits of the APHV measure are that it has norms established in other samples, and the fact that it is not study dependant. Given the same set of height measurements, the estimate produced by the calculation for each participant will remain constant. This is in contrast to the HDSDS index of pubertal timing, where the same set of height measurements may yield different HDSDS scores for a participant depending on the sample he is being compared to. Thus, the APHV estimate is most useful if
comparisons across studies are desirable. The difficulty is that many studies would not be able to calculate APHV, as a longitudinal design with numerous height measurements over adolescence is required. A weakness in using APHV estimates to assess pubertal timing lies in fact that APHV does not coincide with pubertal onset. The duration of the time period between these two events may vary across participants.

4.4 Height Difference in Standard Deviations

The Height Difference in Standard Deviations measure does not require numerous longitudinal data-points. As described in the methods section of Chapter two, the two crucial data-points are height at age 14 and 17.5 years (Wehkalampi et al., 2008). Availability of numerous longitudinal data-points allowed us to use spline interpolation to estimate height at exactly these ages. This allowed us to eliminate variability in the calculated HDSDS due to measurement of participants at different ages. HDSDS calculated using the heights recorded as is from the Teen Focus 2 and Teen Focus 4 clinics, which are the closest to age 14 and 17.5 years, correlated strongly with the HDSDS calculated using the heights from spline interpolation. A downside to the HDSDS index is its inability to account for variations in the magnitude of the growth spurt. It is possible for a participant with an early and small growth spurt to have the same HDSDS score as a participant with a slightly later and larger growth spurt. This is not a critical concern given the strong correlation between APHV and HDSDS, but nevertheless is a noteworthy feature of the measure.
An important consideration when calculating HDSDS is the average Age at Peak Height Velocity and the age at final stature for the sample being studied. The methods outlined in Chapter two are predicated on the fact that the average APHV in males is approximately 14 years (Tanner & Whitehouse, 1976), and that males have reached very close to their final height by 17.5 years of age (Roche & Davila, 1972). Evidence of secular trends in male puberty (Herman-Giddens et al., 2012) may require revision of the ages employed in the calculation for the index to retain maximum utility; the height at 14 years would need to be replaced by a height measurement from the age at PHV in the population. If this method is to be applied in female sample, data from the corresponding ages need to be substituted; these are 11.5 and 17.5 years for the average APHV and age at final height respectively (Wehkalampi et al., 2008).

4.5 Percent Achieved of Adult Stature

The Percent Achieved of Adult Stature index can be applied to cross-sectional datasets but is contingent on the availability of the participant’s height, weight, and parental stature. As with HDSDS, spline interpolated height at 14 was used in PAAS calculations. An identical interpolation method was used for weight at 14, which is another component of the Khamis Roche prediction equation. The PAAS calculated using interpolated height and weight showed a strong correlation with the same index calculated using the exact height and weight measured at the Teen Focus 2 clinic.

A challenge in applying the equations from Khamis-Roche to derive PAAS lies in extrapolation of the coefficients for prediction of final height provided by the original
paper to the sample of interest. The coefficients formulated by Khamis and Roche originate from a cohort of Caucasian males and females from southwestern Ohio (Khamis & Roche, 1994). Careful consideration is required from the investigator before extrapolating these coefficients to other populations. It would be ideal if coefficients derived in a similar fashion as the original study were available from the population being studied for application to the study sample. Since this was not the case for either ALSPAC or the Saguenay Youth Study, the Khamis and Roche coefficients were extrapolated to both these samples. If this were an inappropriate step to take, we would see this reflected in a poor association between the HDSDS and PAAS indices in ALSPAC. This is because the PAAS is analogous to the HDSDS index, but uses a predicted final height instead of an actual final height. The correlation between these two indices is strong, supporting the use of the original coefficients in the ALSPAC sample.

4.6 Biological Samples

The height-based indices informed us of pubertal timing, and testosterone trajectories from the participants give us information on the dynamics of puberty. The ALSPAC bio-bank (Boyd et al., 2012) provided biological samples for hormone quantification purposes. Of interest to our study, blood samples had been collected in participants at Focus Clinics at a maximum of six time-points; ages 7, 9, 11, 13, 15, and 17. We chose to omit samples from age 7 when planning hormone quantification, as males are not expected to have started puberty or have experienced a spike in
testosterone levels by this age (Andersson et al., 1997; Crofton et al., 2002; Elmlinger et al., 2005; Starka et al., 2008).

4.7 Storage Media and ELISA Kits

We identified a potential problem related to a mismatch between the storage medium and requirements of the ELISA assay kit, namely its validation using sodium heparin as an anticoagulant. Blood samples from most participants were collected with tubes containing lithium heparin. Additionally, a small minority of the samples was unavailable as lithium plasma, and for these participants we were to instead obtain samples stored using Ethylenediaminetetraacetic acid (EDTA). It was unclear whether we could proceed with the lithium heparin and EDTA plasma samples for the assays validated using sodium heparin.

In order to test the validity of the ELISA kits with alternate anticoagulants, we designed a pilot study and collected blood samples from 10 healthy volunteers (5 females), and partitioned each individual’s sample into four different blood collection media: sodium heparin plasma, lithium heparin plasma, EDTA plasma, and serum. The four samples were run on the ELISA kits for testosterone and SHBG, and correlations across the values obtained from different media were run. The results obtained from each collection media were very strongly related, with \( r > .99 \) for each testosterone comparison and \( r > 0.96 \) for each SHBG comparison. Proceeding onwards with our lithium heparin plasma and EDTA samples being assayed on the ELISA kits available was appropriate in light of the results from the pilot study.
4.8 Establishing Assay Protocol

Once the preparatory groundwork for assaying of the ALSPAC samples was complete, an important decision regarding assay protocol remained to be resolved. This pertained to the number of repeats of each assay that needed to be completed. The need for reliable data had to be balanced with the cost of the assays. The decision would make a significant impact on expenses, on the order of several thousands of dollars, given the large volume of samples being assayed (>2000) and the expensive assays being used. By consulting with our collaborators at the Pennsylvania State University, a cost-effective strategy was developed. Assays would be run in singlet for 90% of the samples and in duplicate for the remaining 10%. The reliability of the data would be checked through coefficients of variation (CV) from the assays run in duplicate. As the CVs were lower than 10% (2% for testosterone, 5% for SHBG), there was no requirement to rerun all samples in duplicate.

Some of the participant’s testosterone assays were found to have undetectable testosterone levels. This was particularly frequent in the early clinics at age 9 and 11 years. In order to differentiate participants that had a blood sample and were assayed from those missing a blood sample, undetectable concentrations were replaced with the average minimum detectable concentration of the assay as stated by the kit manual. Furthermore, assays for some samples returned values below the minimum threshold defined by the kit. This is possible because the minimum detectable concentration of the assay is calculated as the average minimum concentration detected over a number of kits, and each ELISA plate is an experiment by itself with a slightly different standard
curve, allowing for different minima. The values from these samples were also replaced with the average minimum detectable concentration.

4.9 Circadian Rhythm of Testosterone

The diurnal variation in male testosterone levels has been studied in great detail (Cooke et al., 1993; Plymate et al., 1989). As far as possible, it is best to collect blood from the study sample at a standardized time of day to circumvent the confounding effect of the circadian rhythm. Given the scale of the ALSPAC study, this would have been very challenging logistically – both for the researchers and the participants (and their families). As a result the participants were sampled at various times throughout the day. Variation in testosterone levels as a function of the circadian rhythm is clearly seen when plotting total testosterone levels against time of sample. The effect is particularly striking in the data from the Teen Focus 2 clinic at 13 years, where many participants have attained high levels of testosterone and there is a wide spread in the time of blood sampling.

Removing the effect of time from the testosterone levels is crucial. Any results obtained with uncorrected data might be confounded by the circadian rhythm. The best strategy to correct for time would be to build a model to predict testosterone level at a standard time of day for all participants. Application of multilevel modeling methods provided good estimates of time corrected levels of testosterone, predicted as if measured at 1200h, for four of the five clinics (all except for Teen Focus 2). For the Teen Focus 2 clinic, we fit separate models for participants that had passed their APHV
and those that had yet to reach APHV to improve the time corrected predictions. Without this step, the modeling process could not differentiate between participants that had low levels of testosterone due to time of sampling or later pubertal timing. Thus, it attempted to adjust all values on the same scale leading to inflated predictions for participants who are late maturers.

4.10 Testosterone Trajectories

Having corrected the testosterone data for the effect of time, building and analyzing each participant’s testosterone trajectories was the next step. Simple measures such as sum of the testosterone levels at different time-points, or the slope of a linear regression line through the participant’s data served as a starting point but were too crude to use as independent variables. For example, a participant missing the last blood sample would not have a comparable sum measure to a participant with a full complement of samples. The slope measure is not ideal because we are fitting a linear model to data that we know by nature is not linear (Elmlinger et al., 2005; Hero et al., 2005; Lee et al., 1974; Schnakenburg et al., 1980; Starka et al., 2008).

We attempted to use a latent class analysis as an exploratory approach to identify characteristics of trajectories that would enable classification of participants into different groups. Despite attempts to model the data with different parameters (exponent of term and number of classes), the output clustered a majority of participants' trajectories (>80%) into one class and distributed the remaining few
participants into other classes. The largely uneven distribution of participants into classes is problematic if we want to use class as an ordinal variable in our analyses.

Another approach was to quantify the average exposure to testosterone throughout adolescence. As described in Chapter three, a fitted spline model enabled calculation of area under each participant’s testosterone trajectory. The calculated integral was divided by the age interval of the participant between the first and last clinic in order to standardize the size of the integral and obtain a measure of average exposure to testosterone throughout adolescence. Splitting participants into five groups based on quintiles of this measure illustrated that characteristics of the trajectories, namely the timing and magnitude of the peak change in testosterone, were different across the quintile-based groups.

4.11 Fractions of Testosterone

Testosterone is present in three different states in the circulation: bound to Sex Hormone Binding Globulin (SHBG), bound to albumin, or unbound (Södergård et al., 1982; Vermeulen et al., 1999). The different combinations of these testosterone fractions that we consider is based on the research question being asked. Total testosterone comprises the sum of all three fractions described above. It informs us as to the functioning of the HPG axis and is relevant when examining testosterone levels vis-à-vis pubertal status and maturation.

If, however, we are interested in the physiological impact of testosterone, it is important to consider only the fractions of testosterone that are able to exert an effect
on target tissues in the body (Vermeulen et al., 1999). Testosterone bound to SHBG is prevented from dissociating due to a high binding affinity between the two molecules. Albumin-bound testosterone is able to dissociate more easily, and together with free testosterone comprises bioavailable testosterone (Södergård et al., 1982). This fraction of testosterone can move into target tissues and exert a physiological impact.

4.12 Sodergard Equation for Bioavailable Testosterone

The equations to calculate free and bioavailable testosterone fractions described by Sodergard and colleagues require the concentrations of total testosterone, albumin and SHBG. Vermeulen and colleagues reported unimportant changes in the equation’s output as a consequence of varying albumin concentrations up to 25%. Therefore, the use of an assumed albumin concentration of 43 g/L (6.2 x 10^4 mol/L) is justified when working with samples from patients without marked abnormalities in plasma protein composition (Vermeulen 1999). Sodergard and colleagues reported association constants for binding of testosterone to SHBG (5.97 x 10^8 M^{-1}) and albumin (4.06 x 10^4 M^{-1}) for use in their equations. With the molecule concentrations and association constants, we can calculate the concentration of SHBG-bound testosterone and albumin-bound testosterone, which then allows us to determine the fractions of free and bioavailable testosterone.
4.13 Effect of Other Androgen Metabolites

One potential weakness of this approach is that it does not take into account the role of other androgen metabolites in the bloodstream. While some of these molecules including Dihydrotestosterone (DHT), 5-androstenediol, and 3α-Androstanediol also bind to SHBG and albumin, Sodergard and colleagues showed that the androgen metabolites could be omitted from the equation without adversely affecting calculated concentrations. The difference between two calculated percentages of unbound concentrations of testosterone; one using the maximum metabolite concentrations found in the literature and the other using the minimum; was found to be less than .1%.

We calculated the concentration of bioavailable testosterone from measured total testosterone and SHBG concentration and the reported values of the relevant association constants in the literature. In subsequent investigations using the ALSPAC dataset, this measure can be used as a variable that is of more physiological relevance to the effect of testosterone on the body in comparison with total testosterone.
Chapter 5 - Future Directions

Having established methods to estimate pubertal timing and derive a measure of testosterone trajectories, we can proceed on to testing hypotheses in our ALSPAC sample. In the next sections, I will describe other data collected from participants that is of relevance to this overarching project. Following this, I will outline briefly the relevant findings from the literature, which we draw upon to make specific predictions to be tested in our sample. Our hypotheses pertain to relationships between testosterone, properties of white matter, and psychopathology.

5.1 Magnetic Resonance Imaging

A total of 507 male participants from ALSPAC participated in our study involving a detailed protocol of MRI acquisition. The specific MR sequences collected that are essential to our study include: T1-weighted images (T1W), magnetization transfer (MT) images, diffusion tensor imaging (DTI), and multi-component driven equilibrium single pulse observation of T1 and T2 (mcDESPOT). Processing of images from each of these sequences yields properties of white matter that we are interested in.

The T1W images allow us to derive volumes of grey and white matter in different brain compartments (e.g. hemisphere, lobe, or WM tract). Magnetization Transfer Ratio (MTR) is calculated using MT images, and serves as an indirect index of myelination (Schmierer et al., 2008; Schmierer, Scaravilli, Altmann, Barker, & Miller, 2004). Analysis of images from DTI outputs parameters of water diffusion: Mean Diffusivity (MD) and
Fractional Anisotropy (FA). The MD describes the magnitude of water diffusion in the voxel, whereas FA characterizes the degree of directionality of water diffusion (Paus, 2010a). Myelin Water Fraction (MWF) is obtained from mcDESPOT images. MWF is calculated as the proportion of signal from water within the lipid bilayers of the myelin sheath to the total water (present in extracellular and intracellular spaces), thereby functioning as second indirect index of myelin (Deoni, Rutt, Arun, Pierpaoli, & Jones, 2008; Gareau, Rutt, Karlik, & Mitchell, 2000; MacKay et al., 1994).

5.2 Salivary Testosterone

During the visit for MR scanning, participants provided a saliva sample collected via passive drool. It is important to use this method of collection when planning measurements of sex steroids (Granger et al., 2007). The samples will be assayed for salivary testosterone by our collaborators at the Pennsylvania State University. It will be interesting to see how salivary testosterone at time of scanning compares with total and bioavailable testosterone from the closest ALSPAC clinic (at age 17), and the average measure of testosterone throughout adolescence described in Chapter three.

5.3 Measures of Psychopathology

The ALSPAC dataset contains a wealth of measures, collected as early as 42 months of age, through which to assess psychopathology in participants (Boyd et al., 2012). Included in these measures are the Short Mood and Feelings Questionnaire
(SMFQ) to assess depressive symptoms (Angold & Stephen, 1995), the Clinical Interview Schedule – Revised (CIS-R) and psychosis-like symptoms (PLIKS) questionnaires (Horwood et al., 2008; Lewis, Pelosi, Araya, & Dunn, 2009).

5.4 Testosterone and the Brain

Our investigation in ALSPAC is motivated by the findings from Perrin et al (2008) in 408 participants (204 males) from the Saguenay Youth Study sample. The study identified a positive association between levels of bioavailable testosterone and relative volume of white matter in males, but not in females. This association was found to be stronger in males with the shorter Androgen Receptor gene, which codes for a more efficient AR, compared with males who had a longer AR gene. Furthermore, higher testosterone levels were also related to a decrease in MTR levels. The authors hypothesized that testosterone led to an increase in size of the axon caliber, but not the myelin sheath surrounding it, explaining the observed simultaneous increase in volume of white matter and decrease in MTR.

Another study using data from the males in the SYS sample showed that the apparent grey matter density of the putative corticospinal tract (CST), a white matter tract with very thick axons, increased with levels of bioavailable testosterone (Hervé et al., 2009). This finding is consistent with the hypothesis suggesting that testosterone is increasing the size of the axon caliber in particular, and therefore the g-ratio (diameter of axon caliber/total axon diameter) of axons in the male brain.
Rametti and colleagues studied the effect of at least six months of testosterone treatment on white matter FA in a group of 15 female to male (FtM) transsexuals (Rametti et al., 2012). Employing methods of tract based spatial statistics, they found a post-treatment increase of FA values in the right CST and the right superior longitudinal fasciculus (SLF) compared with pre-treatment measurements. Previous work by the same group showed that pre-treatment FtM transsexuals and control males had higher FA values than control females in the same WM tracts (Rametti et al., 2011).

Based on the findings reported above, we predict that earlier pubertal timing and higher testosterone levels during adolescence are associated with properties of white matter consistent with an increase in g-ratio. Participants with earlier Age at Peak Height Velocity and larger values for the average testosterone measure during adolescence will show increased global measures of relative volume of white matter, and both decreased MTR and MWF. Since MTR and MWF are indirect indices of myelin content (Laule et al., 2007), and the thickness of the myelin sheath is inversely proportional to g-ratio (Paus & Toro, 2009), it is logical that MTR and MWF would be inversely proportional to g-ratio as well. Fractional Anisotropy is thought to depend on microstructural features of fiber tracts, including alignment of individual axons, density of axons, and myelin content. As such, it is unclear whether and how FA relates to g-ratio. Drawing upon the findings from Rametti et al (2012), we can predict that participants with greater testosterone exposure over adolescence will have higher FA in the CST and SLF.
5.5 Preliminary Findings in the ALSPAC sample

Analyses using simple linear regression show that participants with a greater exposure to testosterone over adolescence have slightly lower relative volumes of white matter between the ages of 18 to 20 years. Building of more sophisticated models that control for possible confounders (age at scan, intelligence quotient, socioeconomic status) and – most importantly – incorporating measures of pubertal timing (i.e., APHV), are the next steps to pursue.

Some of the key differences between the SYS and ALSPAC studies that may explain these results are related to the age of the participants at MR scanning in the two studies (12-18 years in SYS, 18-20 years in ALSPAC) and timing of testosterone measurement in relation to the MR imaging. Participants had blood samples taken in close temporal proximity to their MR scanning session in the SYS, whereas the ALSPAC scanning session is separated from the nearest blood collection time-point by one year. Analyses with the salivary testosterone data from ALSPAC will allow us to make a direct comparison between results from the two studies, as the saliva was collected at the MR imaging visit. After establishing the relationship between testosterone and white matter, we can begin to bring in the measures of psychopathology and evaluate whether properties of white matter mediate relationships between testosterone and psychopathology.
5.6 Testosterone, White Matter, and Psychopathology

The measures of psychopathology available through the ALSPAC dataset allow us to test the idea that testosterone affects psychopathology through its role in shaping white matter.

For example, depressive disorder is associated with lower levels of testosterone (Schweiger et al., 1999). In men with shorter CAG repeat lengths in the AR gene, low levels of testosterone are related to depressive symptoms (Seidman, Araujo, Roose, & McKinlay, 2001). The use of anabolic steroids has been linked to depression, among other side effects (Pope Jr & Katz, 1994). Studies using MR imaging have found smaller white matter volumes in the frontal lobes of depressed adolescents (Steingard et al., 2002). Adults with major depressive disorder were found to have decreased FA in the left anterior limb of the internal capsule and the inferior parietal portion of the superior longitudinal fasciculus when compared with healthy controls (Zou et al., 2008).

We can assess depressive symptoms in the ALSPAC participants using data from the SMFQ and CIS-R questionnaires. A testable question based on the relevant literature is whether depressive symptoms are associated with lower testosterone exposure throughout adolescence. Additionally, we can use structural equation modeling to test if lower volumes of white matter in the frontal lobe and decreased FA in the SLF mediate this relationship.

Our findings will contribute to the understanding of etiology of psychopathology, and may potentially explain the manifestation of psychiatric illness through a pathway connected to the brain. A proper understanding of the set of causes for disease is
imperative to the development of preventative and therapeutic strategies to reduce the burden of disease on society.
References


adiposity in young adult males: the Gothenburg osteoporosis and obesity


Yousry, T. a, ... Miller, D. H. (2008). Quantitative magnetic resonance of
postmortem multiple sclerosis brain before and after fixation. *Magnetic Resonance
in Medicine, 59*, 268–277.
Schnakenburg, K., Bidlingmaier, F., & Knorr, D. (1980). 17-Hydroxyprogesterone,
androstenedione, and testosterone in normal children and in prepubertal patients


