

Supplementary Material

Weak associations between pubertal development and psychiatric and behavioral problems.

Emily Smith-Woolley, MSc^{1*a}, Kaili Rimfeld, MSc^{1*}, Robert Plomin, PhD¹.

**Joint first authors*

Abstract:

Pubertal development has been associated with adverse outcomes throughout adolescence and adulthood. However, much of the previous literature has categorized outcome variables and pubertal timing measures for ease of mean difference or odds-ratio interpretation. We use a UK-representative sample of over 5 000 individuals drawn from the Twins Early Development Study to extend this literature by adopting an individual differences approach and emphasizing effect sizes. We investigate a variety of psychiatric and behavioral measures collected longitudinally at ages 11, 14 and 16, for multiple raters and for males and females separately. In addition, we use two measures of pubertal development: totals on the Pubertal Development Scale at each age, as well as age of menarche for girls. We found that pubertal development, however assessed, was linearly associated with a range of psychiatric and behavioral outcomes; however, the effect sizes of these associations were modest for both males and females with most correlations between $-.10$ and $.10$. Our systematic analysis of associations between pubertal development and psychiatric and behavioral problems is the most comprehensive to date. The results showing linearity of the effects of pubertal development support an individual differences approach, treating both pubertal development and associated outcomes as continuous rather than categorical variables. We conclude that pubertal development explains little variance in psychiatric and behavioral outcomes (less than 1% on average). The small effect sizes indicate that the associations are weak and should not warrant major concern at least in non-clinical populations.

A description of the 38 outcomes measures used in the paper:

Age 11 measures

Self (child) report

1. Strengths and Difficulties Questionnaire: Behavior Problems – 20 items (Goodman, 1997; Goodman & Scott, 1999; Goodman, 2001): A dimensional and developmental measure of child mental health for children aged 3-16 years. Children are required to answer statements on a 3 point Likert scale (Not true; Quite true; Very true). It taps into 4 domains:

- *SDQ anxiety scale*: Derived from 5 items. Item example: 'I worry a lot'
- *Conduct problems*: Derived from 5 items. Item example: 'I get very angry and often lose my temper'.
- *Hyperactivity/inattention*: Derived from 5 items. Item example: 'I am easily distracted, I find it difficult to concentrate'.
- *Peer relationship problems*: Derived from 5 items. Item example: 'I am usually on my own. I generally play alone or keep to myself'

A total score is created by taking a mean of all 20 items, requiring at least 10 to be non-missing.

2. Moods and Feelings Questionnaire (MFQ) Short version – 11 items (Angold, Costello, Messer, Pickles, Winder & Silver, 1995): A brief questionnaire based on DSM-III-R criteria for depression. It is measured on a 3-point Likert scale (Not true; Quite true; Very true) and includes a series of descriptive phrases regarding how the subject has been feeling or acting recently. For example: 'I felt I was no good anymore'; 'I felt lonely'; 'I hated myself'

A total score is derived by taking the mean of the 11 MFQ items plus 2 SDQ items in the booklets ('I am often unhappy, downhearted or tearful' and 'I am restless, I cannot stay still for long'). The total requires at least 11 of the 13 items to be non-missing.

3. Multidimensional Peer Victimization Scale – 16 items (Mynard & Joseph, 2000): Designed to assess four types of peer victimization by asking children to comment on the frequency by answering statements on a 3 point Likert scale (Not at all; Once; More than once).

- *Social manipulation*: Derived from 4 items. Item example: 'How often during this school year has another pupil tried to get me into trouble with my friends'.
- *Attacks on property*: Derived from 5 items. Item example: 'How often during this school year has another pupil taken something of mine without permission'.
- *Verbal victimization*: Derived from 4 items. Item example: 'How often during this school year has another pupil called me names'.
- *Physical victimization*: Derived from 4 items. Item example: 'How often during this school year has another pupil punched you'.

A total score is derived by taking the mean of all 16 items, requiring at least 8 to be non-missing.

Parent report

4. Strengths and Difficulties Questionnaire: Behavior Problems – 20 items (Goodman, 1997; Goodman & Scott, 1999; Goodman, 2001): A dimensional and developmental

measure of child mental health for children aged 3-16 years. Parents are required to answer statements on a 3 point Likert scale (Not true; Quite true; Very true). It taps into 4 domains:

- *SDQ anxiety scale*: Derived from 5 items. Item example: 'S/he worries a lot'
- *Conduct problems*: Derived from 5 items. Item example: 'S/he gets very angry and often loses her/his temper'.
- *Hyperactivity/inattention*: Derived from 5 items. Item example: 'S/he is easily distracted, s/he finds it difficult to concentrate'.
- *Peer relationship problems*: Derived from 5 items. Item example: 'S/he is usually on her/his own. S/he generally play alone or keeps to her/himself'

A total score is created by taking a mean of all 20 items, requiring at least 10 to be non-missing.

5. Moods and Feelings Questionnaire (MFQ) Short version – 11 items (Angold et al, 1995): A brief questionnaire based on DSM-III-R criteria for depression. It is measured on a 3-point Likert scale (Not true; Quite true; Very true) and includes a series of descriptive phrases regarding how the subject has been feeling or acting recently. For example: 'S/he felt s/he was no good anymore'; 'S/he felt lonely'; 'S/he hated her/himself'

A total score is derived by taking the mean of the 11 MFQ items plus 2 SDQ items in the booklets ('S/he is often unhappy, downhearted or tearful' and 'S/he feels restless, S/he cannot stay still for long'). The total requires at least 11 of the 13 items to be non-missing.

6. Antisocial Process Screening Device – 20 items (Frick & Hare, 2001): An assessment to detect antisocial processes in youth. Parents rate children on a 3-point scale (Not true, Somewhat true and Certainly true). This questionnaire taps into 3 domains:

- *APSD Narcissism*: Derived from 7 items. Item example: 'Seems to think s/he is better than other people'.
- *APSD Callous-unemotional*: Derived from 6 items. Item example: 'Does not show feelings or emotions'
- *APSD Impulsivity*: Derived from 5 items. Item example: 'Acts without thinking of the consequences'.

A total score is derived by taking the mean of all 20 items, requiring at least 10 of the items to be non-missing.

7. Childhood Asperger Syndrome Test – 30 items (Scott, Baron-Cohen, Bolton & Brayne, 2002; Williams, Scott, Stott, Allison, Bolton, Baron-Cohen & Brayne, 2005): A tool to screen for Asperger Syndrome and related social and communication in a non-clinical setting.

Parents are required to rate each statement by indicating 'yes' or 'no'. It taps into 3 domains:

- *Social scales*: Derived from 11 items. Item example: 'Does s/he join in playing games with other children easily?'
- *Non-social scale*: Derived from 7 items. Item example: 'Does s/he mostly have the same interests as his/her peers?'
- *Communication scale*: Derived from 12 items. Item example: 'Is s/he good at turn-taking in conversation?'

A total score is created by taking the mean of all 30 items, requiring at least 15 of them to be non-missing.

8. Conners ADHD – 18 items (Conners, 2003):

A clinical tool for obtaining parental reports of childhood behavior problems, specifically inattentive and hyperactive-impulsive symptoms. Parents are asked to rate their twins'

behaviour on a 4 point Likert scale ('Not true at all' to 'Very much true'). It taps into 2 domains:

- *Inattention scale*: Derived from 9 items. Item example: 'Does not follow through on instructions and fails to finish work, schoolwork or chores'
- *Hyperactivity-Impulsivity*: Derived from 9 items. Item example: 'Interrupts conversations'

A total score is derived by taking the mean of the 18 items, requiring at least 9 items to be non-missing.

Age 14 measures

Self (child) report

9. Multidimensional Peer Victimization Scale – 16 items (Mynard & Joseph, 2000):

See age 12 self-report Victimization scale.

10. Conners ADHD – 18 items (Conners, 2003)

A clinical tool for obtaining reports of childhood behavior problems, specifically inattentive and hyperactive-impulsive symptoms. Individuals are asked to rate their behaviour on a 4 point Likert scale ('Not true at all' to 'Very much true'). It taps into 2 domains:

- *Inattention scale*: Derived from 9 items. Item example: 'I make careless mistakes or have trouble paying close attention to details'
- *Hyperactivity-Impulsivity*: Derived from 9 items. Item example: 'I am restless or overactive'

A total score is derived by taking the mean of the 18 items, requiring at least 9 items to be non-missing.

Parent report

11. Multidimensional Peer Victimization Scale – 16 items (Mynard & Joseph, 2000):

See age 12 parent-report Victimization scale

12. The Adolescent Autism Spectrum Quotient – 38 items (Baron-Cohen et al., 2001)

This questionnaire was adapted from the adult version, measuring the degree to which an individual shows autistic traits. Parents were required to rate their children on 38 statements on a 4-point Likert scale (Definitely disagree to Definitely agree). For example: 'If s/he tries to imagine something, s/he finds it very easy to create a picture in her/his mind'; 'S/he is fascinated by dates'; 'S/he likes to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.)'

A total score is derived by taking the mean of the 38 items, requiring at least 19 items to be non-missing.

13. Antisocial Process Screening Device – 20 items (Frick & Hare, 2001):

See age 12 parent-report antisocial personality disorder scale.

14. Conners ADHD – 18 items (Conners, 2003)

See age 12 parent-report Conners scale

Age 16 measures

Self (child) report

15. Strengths and Difficulties Questionnaire: Behavior Problems – 20 items (Goodman, 1997; Goodman & Scott, 1999; Goodman, 2001):

See self-report SDQ scale at age 12.

16. Childhood Anxiety Sensitivity Index – 18 items (Silverman, Fleisig, Rabian, & Peterson, 1991): This is a child-reported questionnaire measuring anxiety sensitivity (i.e., the belief that anxiety symptoms have negative consequences). Responses are rated on a 3 point Likert scale ('Not true' to 'Very true'). For example: 'I don't want other people to know when I feel afraid'; 'I get scared when I feel nervous'. A total score is derived by taking the mean of the 18 items, requiring at least 9 items to be non-missing.

17. Moods and Feelings Questionnaire (MFQ) Short version – 11 items (Angold et al, 1995): See age 12 self-report MFQ measure

18. Strengths and Weaknesses of ADHD Symptoms and Normal Behaviour Scale – 18 items (Swanson, Schuck, Porter, Carlson, Hartman, Sergeant et al, 2005): This behavior rating scale is based on DSM-5 criteria for ADHD diagnosis measuring inattentive, hyperactive, and impulsive behaviors. Children are asked to compare themselves to other people of their age on a 7-point scale from far below average to far above average:

- *Inattention scale*: Derived from 9 items. Item example: 'I sustain attention on tasks or leisure activities'.
- *Hyperactivity scale*: Derived from 9 items. Item example: 'I sit still (control movement of hands/ feet)'.

A total score is derived by taking the mean of the 18 items, requiring at least 9 items to be non-missing.

19. Inventory of Callous-Unemotional Traits (ICUT) total scale – 24 items (Frick, 2004; Kimonis, Frick, Skeem, Marsee, Cruise, Munoz et al, 2008). This questionnaire assessing callous and unemotional traits has three subscales (callous scale, unemotional scale and an uncaring scale). Individuals were required to rate 24 statements about their thoughts and behaviors over the past 6 months on a 4 point Likert scale ('Not at all true' to 'Definitely true').

- *Callous scale*: Derived from 11 items. Item example: 'I do not feel remorseful when I do something wrong.'
- *Unemotional scale*: Derived from 5 items. Item example: 'I do not show my emotions to others'.
- *Uncaring scale*: Derived from 8 items. Item example: 'I seem very cold and uncaring to others'.

A total score is derived by taking the mean of the 24 items, requiring at least 12 items to be non-missing.

20. Abbreviated Autism Quotient total scale – 13 items (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001): This questionnaire measures the degree to which an individual shows autistic traits. For each of the 13 items, individuals are asked whether they agree or disagree ('Definitely disagree' to 'Definitely agree') to 13 statements. These are part of two scales:

- *Social scale*: Derived from 8 items. Item example: 'I find it hard to make new friends'.

- *Attention to Detail scale*: Derived from 5 items. Item example: 'I usually notice car number plates or similar strings of information'.

A total score is derived by taking the mean of the 13 items, requiring at least 7 items to be non-missing.

21. Paranoid Checklist – 15 items (Fenigstein, & Venable, 1992): This is a questionnaire to investigate paranoid thoughts. Individuals are asked to rate how often they have had paranoid thoughts ('Not at all' to 'Daily'), for example: 'I might be being observed or followed'; 'People would harm me if given an opportunity'; 'I am under threat from others'. A total score is derived by taking the mean of the 15 items, requiring at least 8 items to be non-missing.

22. Cardiff Anomalous Perceptual Scale – 9 items (Bell, Halligan & Ellis, 2006): A measure of perceptual anomalies, this questionnaire has 9 items which individuals are asked to rate the frequency of on a 6 point scale from 'Not at all' to 'Daily'. For example: 'Hear noises or sounds when there is nothing about to explain them?'; 'See things that other people cannot?'; 'Hear voices commenting on what you're thinking or doing?'. A total score is derived by taking the mean of the 9 items, requiring at least 5 items to be non-missing.

23. 'Myself' subscale of Cognition Checklist for Mania: Revised for Grandiosity and Peters Delusion Inventory – 8 items (Beck, Colis, Steer, Madrak, & Goldberg, 2006; Peters, Joseph, Day, & Garety, 2004). Individuals are asked to answer 8 statements on their thoughts and feelings over the last month on a 4 point scale from 'Not at all' to 'Completely'. For example: 'I have many great ideas'; 'I have special abilities that others do not'; 'I am much more unique than anyone else'. A total score is derived by taking the mean of the 8 items, requiring at least 4 items to be non-missing.

24. Cognitive Disorganization – 11 items (Mason, Linney & Claridge, 2005): This scale measuring poor attention and concentration requires individuals to answer 11 items by answering either 'Yes' or 'No'. For example: 'Do you frequently have difficulty in starting to do things?'; 'Do you find it difficult to keep interested in the same thing for a long time?'; 'Is it hard for you to make decisions?'. A total score is derived by taking the mean of the 11 items, requiring at least 6 items to be non-missing.

25. Temporal Experience of Pleasure Scale (TEPS) for Hedonia – 10 items (Gard, Germans Gard, Kring, and John, 2006). Participants were asked to rate the statements according to how much they apply to them on a 6-point scale ('Very false for me' to 'Very true for me'), based on their thoughts and feelings over the last month. For example: 'When something exciting is coming up in my life, I really look forward to it'; 'I don't look forward to things like eating out at restaurants'. A total score is derived by taking the mean of the 10 items, requiring at least 5 items to be non-missing.

26. Introvertive Anhedonia scale – 7 items (Mason et al, 2005). This scale measures the inability to experience pleasure in what other people may find enjoyable. Individuals were asked to answer either 'Yes' or 'No' to a series of statements, for example: 'Are you too independent to get involved with other people?'; 'Are there very few things that you have ever enjoyed doing?'. A total score is derived by taking the mean of the 7 items, requiring at least 4 items to be non-missing.

27. Eating Problems scale – 4 items (Stice, Fisher & Martinez, 2004). Four items were chosen from this 22 item questionnaire, which taps into perceptions about body-image. The questions were: 1) Have you felt fat? 2) Have you had a definite fear that you might gain weight or become fat? 3) Has your weight influenced how you think about yourself as a person? 4) Has your shape influenced how you think about yourself as a person?

Participants were asked to rate their feelings on a 7 point scale from 'Not at all' to 'Extremely'. A total score is derived by taking the mean of the 4 items, requiring at least 2 items to be non-missing.

28. Multidimensional Peer Victimization Scale – 6 items (Mynard & Joseph, 2000): Shortened 6 item version of the victimization questionnaire at ages 12 and 14. A total score is derived by taking the mean of the 6 items, requiring at least 3 items to be non-missing.

29. Sleep quality – 11 items (Buysse, Reynolds, Monk, Berman, Kupfer, 1989; Bastien, Vallieres, & Morin, 2001; Morin, 1993): This measure was created by taking 7 items from the Pittsburgh Sleep Quality Index and 4 items from the Insomnia Severity Index. Items included: 'How many hours of actual sleep do you get at night'; 'During the past month, how often have you had trouble staying awake while eating meals or engaging in social activity'; 'How worried/distressed are you about your current sleep problem'. Items were recoded so that they all had values from 0-3 and items were summed together to create a total.

30. Substance use – 38 items (Adapted from ALSPAC 'Life of a 16+ Teenager' questionnaire): This measure asks questions about adolescence alcohol, tobacco, cannabis and 'other drug' use. The total score was created by summing together items scores.

31. Self-Report Delinquency Scale – 13 items (Elliott, Huizinga & Ageton, 1985) This questionnaire measures the frequency, severity and the age of delinquent events by asking individuals to first state whether they had participated in a delinquent event ('Yes' or 'No'), the age that they had participated (Less than age 5...16+) and how many times they had participated (once...20+). Items ranged from 'stealing something' to 'using force to get money from a teacher or another adult at school'. To create the total score, the sum of age and frequency were derived for each delinquency item and then the sum of these items were taken.

Parent report

32. Strengths and Difficulties Questionnaire: Behavior Problems – 20 items (Goodman, 1997; 2001):
See parent-report SDQ scale at age 12.

33. Anxiety-Related Behaviours Questionnaire (ARBQ) – 19 items (Eley, Bolton, O'Connor, Perrin, Smith & Plomin, 2003): This is a parent-reported, 19-item questionnaire on anxiety-related behaviours in children. Items are rated on a 3 point scale ('Never', 'Sometimes', 'Often'), representing five dimensions: General Distress or negative mood (for example: 'Often seems worked up, on edge or tense'; Separation Anxiety (for example: 'Is often extremely upset or distressed when parent leaves, wound up or stressed'; Fears (for example: 'Is afraid of animals or insects (like dogs, spiders, or snakes)'; Obsessive-Compulsive Behaviours (for example: 'Tends to check things are done exactly right'), and Shyness/Inhibition (for example: 'Tends to be shy and timid'). A total score is derived by taking the mean of the 19 items, requiring at least 10 items to be non-missing.

34. Moods and Feelings Questionnaire (MFQ) Short version – 11 items (Angold et al, 1995):
See Moods and Feelings Parent-report questionnaire at age 12.

35. Conners ADHD – 18 items (Conners, 2003)
See age 12 parent-report Conners scale

36. Scale for the Assessment of Negative Symptoms – 10 items (Andreasen, 1984): This is a measure of negative symptoms in schizophrenia comprising 10 items of: affective

flattening (2 items), alogia (2 items), avolition-apathy (2 items), anhedonia-associality (2 items) and attention (2 items). Parents are asked to rate how strongly they agree ('Not at all' to 'Definitely agree') on a 4-point Likert scale. Items include: 'Often fails to smile or laugh at things others would find funny'; 'Often sits around for a long time doing nothing'; 'Is often inattentive and appears distracted'. A total score is derived by taking the mean of the 10 items, requiring at least 5 items to be non-missing.

37. Inventory of Callous-Unemotional Traits (ICUT) total scale – 24 items (Frick, 2004; Kimonis, et al, 2008). This questionnaire assessing callous and unemotional traits has three subscales (callous scale, unemotional scale and an uncaring scale). Parents were required to rate 24 statements about their children's behavior over the past 6 months on a 4 point scale ('Not at all true' to 'Definitely true').

- *Callous scale*: Derived from 11 items. Item example: 'My child shows no remorse when he/she has done something wrong'.
- *Unemotional scale*: Derived from 5 items. Item example: 'My child does not show emotions'.
- *Uncaring scale*: Derived from 8 items. Item example: 'My child seems very cold and uncaring'.

A total score is derived by taking the mean of the 24 items, requiring at least 12 items to be non-missing.

38. Autism Spectrum Quotient (AQ) – 28 items (Baron-Cohen et al, 2001): This questionnaire measures the degree to which an individual shows autistic traits. Parents were required to rate their children on 28 statements on a 4-point Likert scale ('Definitely disagree' to 'Definitely agree'). This questionnaire taps into 4 domains:

- *Social scale*: Derived from 9 items. Item example: 'Finds it hard to make new friends'.
- *Attention Switching scale*: Derived from 8 items. Item example: 'Does not get upset if his/her daily routine is disturbed' (reversed).
- *Imagination scale*: Derived from 6 items. Item example: 'If trying to imagine something, finds it easy to create a picture in his/her mind' (reversed).
- *Attention to Detail scale*: Derived from 5 items. Item example: 'usually notices car number plates or similar strings of information'.

A total score is derived by taking the mean of the 28 items, requiring at least 14 items to be non-missing.