JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Koerselman, Kristian; Pekkarinen, Tuomas

Title: Cognitive consequences of the timing of puberty

Year: 2018

Version: Accepted version (Final draft)

Copyright: © 2018 Elsevier B.V.

Rights: CC BY-NC-ND 4.0

Rights url: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the original version:

Koerselman, K., & Pekkarinen, T. (2018). Cognitive consequences of the timing of puberty. Labour Economics, 54. https://doi.org/10.1016/j.labeco.2018.05.001

Accepted Manuscript

Cognitive consequences of the timing of puberty

Kristian Koerselman, Tuomas Pekkarinen

PII: S0927-5371(18)30041-1
DOI: 10.1016/j.labeco.2018.05.001
Reference: LABECO 1648

To appear in: *Labour Economics*

Received date: 12 September 2017
Revised date: 2 May 2018
Accepted date: 8 May 2018

Please cite this article as: Kristian Koerselman, Tuomas Pekkarinen, Cognitive consequences of the timing of puberty, *Labour Economics* (2018), doi: 10.1016/j.labeco.2018.05.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Highlights

- Late pubertal development is correlated with background variables.
- Late developers exhibit lower cognitive skill growth between ages 7 and 16.
- Behavior and motivation are largely unaffected conditional on background variables.
- Boys' effects persist in attainment and adult earnings.
- Girls' attainment and earnings estimates have same sign but are less precise.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Cognitive consequences of the timing of puberty

Kristian Koerselman[a,b,*], Tuomas Pekkarinen[c,d,e]

[a] *Finnish Institute for Educational Research, P.O. Box 35, FI-40014 University of Jyväskylä, Finland*
[b] *Jyväskylä University School of Business and Economics, P.O. Box 35, FI-40014 University of Jyväskylä, Finland*
[c] *VATT Institute for Economic Research, Arkadiankatu 7, P.O. Box 1279, FI-00101 Helsinki, Finland*
[d] *Aalto University School of Business, P.O. Box 21210, FI-00076 Aalto, Finland*
[e] *IZA Institute of Labor Economics, Schaumburg-Lippe-Straße 5–9, D-53113 Bonn, Germany*

Abstract

In this paper, we use British cohort data to study the degree to which children at different stages of pubertal development at age 11 and 16 differ in their cognitive and behavioral outcomes at age 16 as well as in their completed educational attainment and adult earnings. Controlling for age 7 cognitive skills, region of birth, father's socioeconomic status and parents' education, we show that boys' late pubertal development is associated with lower levels of cognitive skills at 16, lower final educational attainment, and lower earnings in adulthood. For girls we find a similar negative relationship between late puberty and cognitive skills, but only imprecise relationships between the timing of puberty and adult outcomes. We fail to find evidence of strong associations with either motivation or problem behavior in adolescence for either gender, suggesting a more direct link between pubertal and cognitive development.
JEL: *I20, J16*

Keywords: puberty, gender, adolescence, skills, attainment, earnings

1. Introduction and background

In recent years, the long-term effects of early life experience have been the subject of intensive economic research (Currie and Almond, 2011). While much of this literature has focused on the effects of shocks and interventions during early childhood, the cognitive and psychological development that takes place in puberty has received less attention. Yet puberty is a period when developmental differences across individuals widen and become more salient, while at

*Corresponding author
Email addresses: `kristian.w.koerselman@jyu.fi` (Kristian Koerselman), `tuomas.pekkarinen@vatt.fi` (Tuomas Pekkarinen)

ACCEPTED MANUSCRIPT

the same time puberty coincides with important educational decisions for many adolescents.

In this paper, we study the relationship between the timing of pubertal development on the one hand, and cognitive test scores during adolescence as well as final educational attainment and labor market earnings in adulthood on the other. Our goal is to examine whether individuals who are still going through the physical and cognitive changes associated with puberty during the time when they have to make decisions about post-compulsory education are at a disadvantage compared to individuals who have already reached maturity at this time. For this purpose, we use data from the British National Child Development Study. The NCDS is unique in that it combines information on pubertal development with rich behavioral, educational and labor market outcomes at various ages.

The correlation between educational outcomes and the timing of puberty has been studied in several earlier papers. There are well-known differences in the timing of puberty, both within and between gender.[1] The literature has mostly concentrated on within-gender differences. For boys, it is almost universally found that early developers do better in school. For example Dubas et al. (1991) find that late developing boys have lower grades, at least in middle school, and Koivusilta and Rimpelä (2004) find that late developing boys are less likely to choose the academic track in high school, and consequentially have lower levels of educational attainment in adulthood. Evidence on girls is less conclusive with some studies, such as Dreber et al. (2011), finding evidence of a negative relationship while others, such as Koivusilta and Rimpelä (2004) find no relationship. There is less evidence on the correlation of the timing of puberty and labor market outcomes, although Persico et al. (2004) do show that the early teenage growth spurt, a variable closely linked to puberty, is positively correlated with male adult earnings.

There are various reasons to suspect that the correlations reported in the existing literature may be misleading. First of all, since the measurement of puberty is intrusive when performed by a professional, many existing studies rely on self-assessments. The available evidence however suggests that self-assessments of pubertal development are unreliable (Rasmussen et al., 2015). Furthermore, it is typically unclear whether children who develop unusually eary or late disproportionately refrain from opting in to these kinds of studies.

Second, as shown in Susman and Dorn (2009), the timing of puberty is correlated with childhood circumstances such as malnutrition, obesity, stress, and other problems in early life. These factors are likely to have direct effects on educational outcomes, and can be correlated with other background factors that affect education negatively. Though it is debatable what exactly should be held constant when discussing the effects of pubertal development, it seems clear that differences in childhood circumstances should be adjusted for. This

[1] At age 15–16, most girls have emerged from puberty, while a sizable portion of boys are typically still undergoing important physical changes (Tanner, 1962; Steinberg, 2014).

ACCEPTED MANUSCRIPT

is an issue which is often overlooked in existing studies.

Simple correlations of educational attainment and timing of puberty also leaves open questions about the mechanisms through which puberty affects educational outcomes. The previous literature has shown that puberty is correlated with the development of both cognitive and non-cognitive skills. For example a meta-study by Newcombe and Dubas (1987) shows that pubertal development and IQ are positively correlated among both boys and girls. On the other hand, puberty has been shown to be correlated with behavioral changes that may in turn affect educational outcomes negatively (Peper and Dahl, 2013). Disentangling these mechanisms is important to inform decisions on the appropriate interventions to help adolescents.

In the present study, we try to address these issues by using data from the NCDS. The NCDS has sampled members of the British 1958 birth cohort, with regular follow-ups that are still being carried out today. A unique feature of these data is that they contain information on the timing of pubertal development as asserted by a medical professional at relatively low levels of non-response. Furthermore, the pubertal development information in the NCDS can be linked to measures of early-age ability and rich information on family background. This information allows us to control for most typical confounding factors. In addition, the NCDS contains information on test scores and questions that can be used to measure both cognitive outcomes and behavioral changes during adolescence. These variables can inform us about the mechanisms through which pubertal development potentially affects educational and labor market outcomes.

We contribute to the literature by showing that boys who are still at early stages of puberty at age 16 score lower in cognitive tests at age 16 than their more mature peers even after controlling for age 7 test scores and various background variables. For girls we find a similar, albeit weaker relationship between pubertal development at age 11 and age 16 test scores. At the same time, we fail to find evidence of large behavioral differences along the pubertal development axis for either boys or girls once family background and age 7 skills are controlled for. When it comes to adult outcomes, we find that for boys late pubertal development is also reflected in lower final educational attainment and lower earnings. For girls the relationship between the timing of puberty and adult outcomes is too imprecise to warrant strong conclusions.

These findings imply that late pubertal development may have long-lasting consequences through its association with cognitive outcomes during adolescence. In particular, late developing boys may be at a disadvantage compared to their more mature classmates when faced with important decisions on post-compulsory education. As argued in some earlier studies such as Pekkarinen (2008), education systems that concentrate such decisions around age 16 when pubertal development varies strongly among boys, may exacerbate and lock in the consequences of boys' differential development.

The remainder of the paper proceeds as follows. We start by presenting the data and the empirical strategy used in our analysis. We present our results in the third section. The fourth section concludes.

ACCEPTED MANUSCRIPT

2. Data

The main challenge in studying the relationship between pubertal development and adolescent or adult outcomes is to find data sets that combine reliable measures of both. In this paper, we use the British National Child Development Study (University of London, 2014; 2008a; 2008b). The NCDS aims to follow all those born in Great Britain during the week starting on the 3rd of March 1958. As such it can be seen as approximately representative of the 1958 birth cohort. The NCDS sample has been followed to the present day.

To the best of our knowledge, the NCDS is the only available data set to contain medical records of the timing of pubertal development, rich background information, test scores at several ages, and information on final educational attainment and labor market outcomes. It has for example been used previously by authors such as Persico et al. (2004) and Case and Paxson (2008) to study the relationship between adolescent height and adult outcomes.

2.1. Puberty measures

NCDS sample members were subjected to medical examinations including measurement of pubertal development, first in the 1969 sweep when sample members were 11 years old, and again in 1974 sweep at age 16. The age 11 examination recorded information on the development of pubic hair for both boys and girls, and on the development of boys' genitalia and on girls' breast development. The development along these dimensions was classified into five categories, from least to most developed.

At age 16, the pubertal development of the respondents was assessed in three dimensions. For both boys and girls, development of pubic and axillary hair was categorized into four categories (1 absent; 2 sparse; 3 intermediate; 4 adult). In addition, the age 16 examination gathered information on girls' breast development and boys' development of facial hair, both categorized into three levels (1 absent; 2 sparse; 3 adult). All of these measurements roughly follow Tanner's stages of pubertal development (Tanner, 1962), which is still the standard scale used in the assessment of pubertal development today.

To reduce the dimensionality of the puberty measures, we create nine-level scalar summary measures of pubertal development at ages 11 and 16 by adding the age-specific measures of pubertal development together, and rescaling the resulting sums to the 0–1 interval. On this measure, a value of 0 thus indicates that the respondent's pubertal development was assessed to be 'absent' on all of the different dimensions whereas a value of 1 means that the respondent's development was 'adult' on all of the same measures. For the sake of exposition, we furthermore subdivide the scale into 'low', 'mid' and 'high' categories when individuals have one of the lowest three, middle three, or highest three possible levels of the pubertal development respectively.

Table 1 shows various summary statistics of the puberty measures in the NCDS data by gender at ages 11 (Panel A) and 16 (Panel B), while Figure 1 shows histograms of the pubertal development index by gender and age. As becomes clear both from the figure and from the table, 11 is a relatively early

ACCEPTED MANUSCRIPT

age to measure pubertal development for individuals born in 1958, with most individuals having one of the lowest three values on the pubertal development index. Though the (rounded) mean of the pubertal development index is 0.2 for both boys and girls at age 11, the histogram reveals that individuals with middle to high levels of pubertal development at 11 are overwhelmingly girls. The variation in the age 11 measure is thus larger between girls than between boys.

By contrast, at age 16 boys differ much more from each other in their measured pubertal development than girls do. Whereas girls are clearly concentrated at the higher end of the distribution with a mode of 1 indicating full maturity, boys are more evenly spread across the distribution. Indeed, at the lower end of the distribution there are almost only boys.

Table 1: Age 11 and age 16 measures of pubertal development

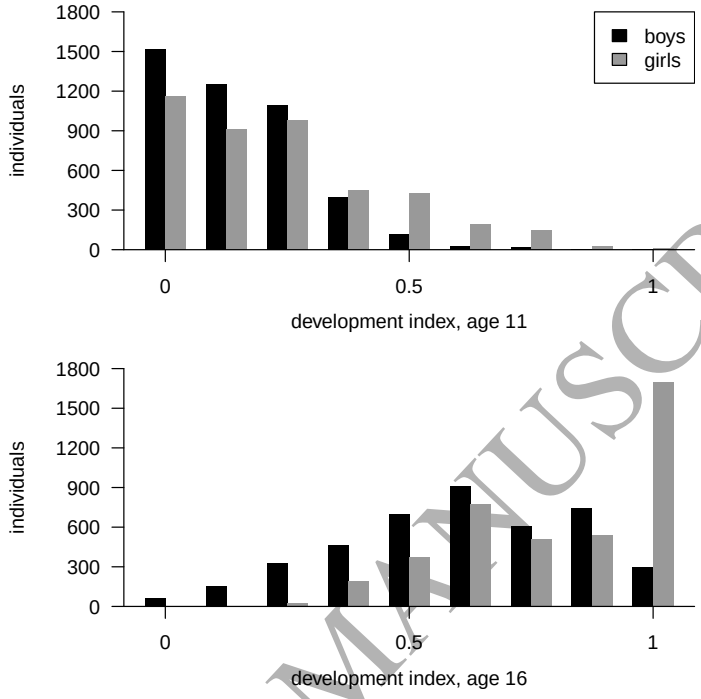
Panel A: age 11				relative frequency				
	mean	min	max	"1"	"2"	"3"	"4"	"5"
Boys								
Genitalia	1.8	1	5	0.38	0.45	0.15	0.01	0.00
Pubic hair	1.4	1	5	0.64	0.32	0.03	0.01	0.00
Sum	3.2	2	10					
Index	0.2	0	1					
Girls								
Breast development	2.0	1	5	0.36	0.36	0.20	0.06	0.01
Pubic hair	1.9	1	5	0.41	0.37	0.14	0.07	0.01
Sum	3.9	2	10					
Index	0.2	0	1					

Panel B: age 16				relative frequency			
	mean	min	max	"1"	"2"	"3"	"4"
Boys							
Pubic hair	3.4	1	4	0.02	0.09	0.39	0.49
Axillary hair	2.8	1	4	0.12	0.26	0.33	0.29
Facial hair	1.7	1	3	0.37	0.54	0.09	
Sum	7.9	3	11				
Index	0.6	0	1				
Girls							
Pubic hair	3.5	1	4	0.00	0.06	0.33	0.58
Axillary hair	3.3	1	4	0.02	0.18	0.30	0.49
Breast development	2.6	1	3	0.00	0.37	0.63	
Sum	9.4	3	11				
Index	0.8	0	1				

Notes: The sum of the individual pubertal development measures ranges from 2 to 10 at age 11, and from 3 to 11 at age 16. The index rescales this linearly to 0–1 in both cases. At age 11, pubertal development is assessed linearly from 1 to 5. At age 16 the scale for the variables that take four values is 1: Absent, 2: Sparse, 3: Intermediate, 4: Adult, and for the variables that take three values 1: Absent, 2: Sparse, 3: Adult.

ACCEPTED MANUSCRIPT

Figure 1: The distribution of the age 11 and age 16 pubertal development indices.



Notes: The figure shows histograms of the age 11 and age 16 pubertal development indices derived from the raw puberty measures in Table 1. Black bars indicate frequencies for boys, grey bars for girls.

Table 2 shows the transition matrix between the lowest, middle, and the highest thirds of the pubertal development index at ages 11 and 16. The table illustrates that measurement is relatively consistent between ages and within individuals, as witnessed by the low number of individuals in the lower triangular part of the matrix, but also that the speed of pubertal development differs between individuals. 31% of boys for example go from low development at age 11 to high development at age 16.

The pubertal development information in the NCDS corresponds reasonably well with stylized facts about pubertal development. In what follows we will mostly use the age 16 development index as our measure of pubertal development. This choice is motivated by the fact that we are interested in age 16 cognitive outcomes and therefore want to measure the stage of pubertal development at that particular age. However, since the precision of our estimators suffers from the low levels of variation in girls' pubertal development at age 16, we complement our analysis by showing results based on age 11 pubertal development as well.

ACCEPTED MANUSCRIPT

Table 2: Transition table

Boys

	age 16		
	low	mid	high
Age 11: low	0.12	0.44	0.31
Age 11: mid	0.01	0.04	0.07
Age 11: high	0.00	0.00	0.00

Girls

	age 16		
	low	mid	high
Age 11: low	0.01	0.26	0.44
Age 11: mid	0.00	0.06	0.19
Age 11: high	0.00	0.01	0.04

Notes: Both the age 11 and the age 16 the pubertal development indices take nine different values, which we further categorize into 1–3: low; 4–6: mid; and 7–9: high. The table shows the joint distribution of age 11 and age 16 pubertal development over these categories separately by gender.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

2.2. Background variables

Parental background and childhood circumstance correlate with the timing of puberty, but may have independent effects on educational attainment. For this reason, we use the rich information available in the NCDS to keep such factors constant in our comparisons. We control for father's and mother's educational attainment by ISCED level, as well as for levels of father's social status based on his occupation when the sample member was born. In the NCDS, the social class of the father is based on the 1958 interviews of the subjects' mothers who were asked about their husbands' occupations, and separated into six levels. We add as a separate category that the father is missing or that the information is otherwise unavailable. We similarly add separate categories for missing parental education information. Finally, we add dummy variables for eleven regions of residence at age 16.

In addition to the family background variables, the NCDS contains rich information on age 7 cognitive skills, which allows us to control for otherwise unobserved differences between sample individuals. These include the results of word recognition and word comprehension tests, a copying designs test to assess perceptuo-motor abilities, a draw-a-man test to assess general mental and perceptual abilities, and an arithmetic test.

Table 3 reports by gender the mean values of all background variables except region of birth. As should be expected, parental background variables do not differ by gender. There are on the other hand clear gender differences in mean test scores at age 7. Whereas boys score better in the arithmetic and copying designs tests, girls score better in the draw-a-man and reading tests.

2.3. Adolescent outcomes

The NCDS follow-up at age 16 provides us with measures of both cognitive and behavioral outcomes. Sample members were administered mathematics and reading tests. We combine the resulting test scores into a measure of cognitive ability at age 16. By design, the raw reading scores are strongly negatively skewed, and to avoid overweighting students with poor reading skills as well as to aid in the interpretation of regression coefficients, we map each set of scores onto a standard normal distribution.

Puberty is also often seen as being associated with changes in school motivation or in problem behavior. The NCDS questionnaire asks students to agree or disagree with the following statements: 'I feel school is a waste of time', 'I am quiet in the classroom and get on with my work', 'I think homework is a bore', 'I find it difficult to keep my mind on my work', 'I never take work seriously', 'I don't like school', 'I think there is no point in planning for the future; you should take things as they come' and 'I'm always ready to help the teacher'. In addition to these self-assessments, teachers were asked whether they think the student member is hardworking. We use these questionnaire items to construct a measures of motivation, effort and self-discipline. We further create a measure of problem behavior by combining survey questions about whether the sample member has been in contact with the police, has ever been to court, whether

ACCEPTED MANUSCRIPT

Table 3: Sample means of main covariates and outcomes by gender.

	girls	boys
Age 7 scores		
Arithmetic	-0.04	0.04
Copying	-0.03	0.03
Draw-a-man	0.05	-0.05
Reading	0.12	-0.11
Father's social status		
Professional	0.05	0.05
Managerial technical	0.17	0.16
Skilled nonmanual	0.08	0.09
Skilled manual	0.36	0.37
Semi-skilled	0.15	0.14
Unskilled	0.04	0.04
No information	0.15	0.15
Mother's education		
ISCED 5	0.03	0.03
ISCED 3	0.22	0.22
ISCED 2	0.59	0.59
ISCED 1	0.01	0.01
No information	0.14	0.16
Father's education		
ISCED 5	0.04	0.04
ISCED 3	0.19	0.19
ISCED 2	0.57	0.56
ISCED 1	0.01	0.01
No information	0.19	0.20
Age 16 outcomes		
Score 16	-0.07	0.07
Motivation	0.11	-0.11
Problems	0.20	0.31
Attainment		
Years of education	13.77	14.14
Income		
Log earnings	8.96	9.85

Notes: Background variables and age 16 problem behavior are expressed in sample proportions. Age 7 test scores, age 16 test scores, and motivation are are expressed in sample standard deviations. Years of education and log earnings are in absolute units.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

the parents have arguments about their child drinking, as well as two measures of truancy.[2]

We need to combine each group of outcome variables into one. Unlike the pubertal development index, neither cognitive skills nor motivation has a natural, easily interpreted scale. We therefore extract the first principal component from either group of variables using Principal Component Analysis or PCA. Principal components are essentially weighted sums of the original variables. Variables receive more weight if they correlate more strongly with the other variables in the group, the intuition being that they are less noisy measures of the underlying trait. For the cognitive skill scores, only two measures are available, and so the first principal component reduces to the unweighted mean of the two normalized scores. For the motivation-related variables, the weights are available in Appendix A. We will express estimates of effects on cognitive skill scores and motivation in sample standard deviations throughout.

We follow a different approach for problem behavior. For three quarters of the sample no problems are indicated at all while for the remaining quarter the child will usually be reported to display more than one problem at once, even if a greater number of problem indications do not necessarily indicate graver problems. This makes a summed measure hard to interpret. We therefore measure whether the child displays problem behavior at all with an indicator variable equaling one when the child has been reported to have problems on any of these measures.

2.4. Adult outcomes

As the NCDS follows individuals into adulthood, we can use later sweeps of the survey to construct measures of adult outcomes. We use the educational attainment measure from the seventh NCDS sweep at age 46. The attainment information in the seventh sweep is cumulative, so that the information would for example also be available for an individual missing from the seventh sweep, but not from the sixth. The exact degrees available to individuals as well as their associated grading scales differ from region to region, and we convert all attainment information to years of education to maintain comparability across regions.

As a measure of earnings, we take mean log earnings in the most recent job measured at ages 33 and 42. Different respondents report amounts corresponding to different payment periods, and so we annualize all amounts before logarithmizing. For individuals who report take home pay but not earnings, we impute earnings from take home pay within the sweep. We use the mean of the two logarithmized yearly earnings observations in our analysis.

[2]It is not clear that any of the variables we use have should be interpreted in the same way for boys as for girls. This is the case particularly strongly for the pubertal development variables as well as for problem behavior (cf. Lundberg, 2017). For this reason, we model boys and girls separately throughout our analysis.

ACCEPTED MANUSCRIPT

Table 3 also shows the mean values of our outcome variables by gender. In the case of some of the variables, gender differences differ from what we would observe for adolescents and young adults today. In particular, boys of the 1958 birth cohort scored higher in the age 16 tests than girls, and went on to have higher levels of completed educational attainment. Other gender differences are less surprising by today's standards: girls had higher levels of academic motivation and were less likely to get into trouble during adolescence. At ages 33 and 42 they also had lower earnings than men.

2.5. Missing observations and imputation

Like most cohort studies that follow individuals from birth to adulthood and subjects respondents to extensive interviews, the NCDS suffers from considerable non-response and attrition problems. Non-response on specific variables can be solved to a degree. There are several groups of variables in our analysis that we think of as measures of a single latent trait. These are the age 7 cognitive skill scores, pubertal development measures at a specific age, measures of motivation and self-discipline, measures of problem behavior, the two age 16 test scores, and earnings at the two different ages. Many individuals have missing values on some but not all of the measures within a trait, and completely discarding such individuals would result in a loss of valuable information.

In practice, we deal with missingness by first dropping all the individuals who have missing values on all measures of any single trait needed in the analysis. For about a third of the remaining individuals we have information on at least one measure for each trait, but missing values on some measures within traits. In the main sample, the information for this one third is for example about 90% complete on average. For these individuals, we impute the missing values from the other measures within each group by using the regularized iterative PCA imputation algorithm under the assumption that missingness on individual measures within the trait is random.(cf. Josse and Husson, 2012, section 3) Intuitively, this procedure is similar to filling in missing values with the mean of the (appropriately scaled) available variables for the same individual.

Our measure of age 16 cognitive skills consists of only two subscales. In this case, the missing values on one of the tests are replaced by the test score in the non-missing test. For example, an individual for whom the standardized score in age 16 mathematics is lacking will thus get it imputed by his or her standardized age 16 reading score. For the age 7 scores, we also use age 7 teacher ratings of oral ability, awareness of the world, reading skills, creativity, and quickness to learn to help impute test scores.

We find imputation of missing subscores much preferable to completely dropping the affected individuals from the sample. We nevertheless include variations of our main regression specifications in Appendix C where we exclude any individuals with imputed information. The estimates thus obtained turn out to be substantially similar to those based on the main samples.

Table 4 illustrates how missingness affects our sample sizes. The number of individuals observed in the NCDS at birth is 17416. Even after imputation, the number of individuals drops to less than half when we require key age 0, 7 and

ACCEPTED MANUSCRIPT

age 16 as well as pubertal development information to be available, though this is mostly due to missing observation on age 7 and age 16 test scores.

Table 4: Sample sizes

	individuals
At birth	17416
Age 7 and age 16 scores known	10618
Pubertal development additionally known (main sample)	8378
Attainment sample	6268
Earnings sample	6885

Notes: The table shows the size of the NCDS sample at birth, including 812 stillbirths and other early deaths. In total, about a tenth of the sample is removed from the target population through emigration or death by age 16. At that age, we have acceptable levels of information covariates and outcomes for about ten thousand sample members, for 8378 of which pubertal development at 16 is also known. Attrition is higher when we require attainment or earnings to be known in addition to pubertal development and covariates.

The large drop in usable sample size raises the obvious concern that the selection to our analysis sample is endogenous with respect to pubertal development. We explore this in Table 5 where we report the correlation of father's socioeconomic status, age 7 test scores, and age 11 pubertal development index with the probability of being observed in our age 16 analysis sample separately for boys and girls. As is clear from Table 5, attrition out of our main sample is negatively correlated with family background and age 7 cognitive skills. However, this does not necessarily imply that attrition causes biased estimates of the relationship with pubertal development. In Appendix B we compare the raw association between pubertal development and age 16 test scores between the main estimation sample and a specially constructed sample which also includes individuals with missing information on various other variables used in the main analysis. The results remain substantially unchanged between the two samples. More worrying for the interpretation of our estimates would be if individuals are missing from the data for reasons related to pubertal development. Reassuringly however, the last column of Table 5 however suggests that differences in pubertal development do not predict missingness in our sample.

Table 5: Correlates of missingness.

	SES	scr7	dev11
Missing from main sample, boys	-0.05	-0.11	0.00
Missing from main sample, girls	-0.06	-0.11	0.00

Notes: The table shows the correlation between pairwise complete observations between the individual not being included in the main regression sample, and a linearized index of father's social status, age 7 test scores, and age 11 pubertal development respectively.

ACCEPTED MANUSCRIPT

3. Results

To get a first glimpse of the relationship between pubertal development, covariates, and outcomes, we report their raw correlations in Table 6. Age 16 test scores, final educational attainment, and adult earnings correlate more strongly with age 16 pubertal development among boys, and more strongly with age 11 pubertal development among girls. Using age 16 pubertal development, behavioral outcomes such as motivation or problem behavior hardly correlate with pubertal development at all. For girls, age 11 development does correlate with age 16 motivation and problems, but the small correlations overall suggest that the characterization of puberty as a period of hormone-induced storm and stress is an oversimplification (cf. Flannery et al., 1994; Arnett, 1999; Steinberg and Morris, 2001; Steinberg, 2014).

The correlation table also shows us how pubertal development is correlated with father's social status and own age 7 scores. This suggests that pre-pubertal differences between individuals may be a cause of the correlations between pubertal development and outcome variables.

The background and outcome variables are all strongly correlated with each other with the expected signs. This suggests that they all contain useful amounts of signal, and that the lack of correlation between age 16 pubertal development and motivation or problem behavior is unlikely to be due to measurement error dominating the variation in the latter two variables.

In Figure 2 we examine the relationship between pubertal development and age 16 test scores more closely by plotting the mean test scores in standard deviation units for each value of our age 16 pubertal development measure. Circle areas in this figure are proportional to the number of individuals at each level of pubertal development. As can be seen from Figure 2, there is a clear positive relationship between physical maturity and age 16 test scores for both genders. For boys, the relationship is concave. For girls, the exact shape of the relationship is harder to determine because most girls are concentrated at the high end of the development index.

ACCEPTED MANUSCRIPT

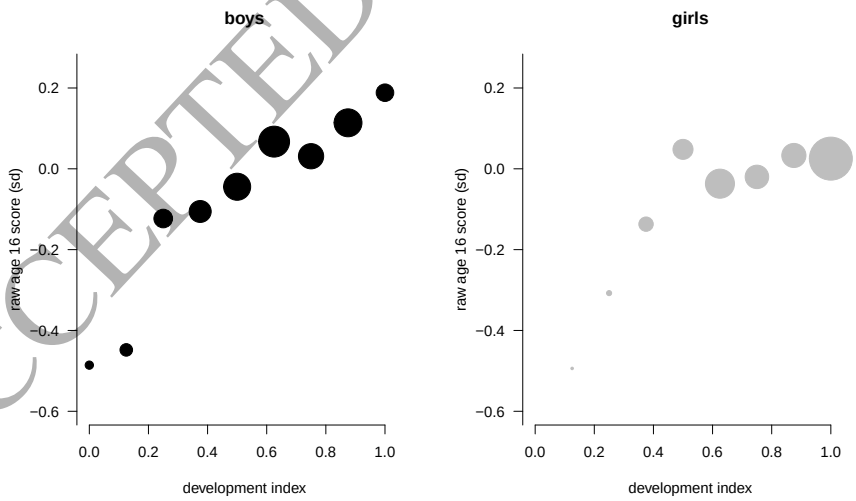
ACCEPTED MANUSCRIPT

Table 6: Correlations between pubertal development, the main background variables, and the main outcome variables.

	dev11	dev16	SES	scr7	motv	prob	scr16	att	ear
Boys									
Development index 11	1.00	0.23	0.03	0.03	-0.01	0.01	0.03	0.02	0.03
Development index 16	0.23	1.00	0.03	0.08	0.02	-0.01	0.12	0.05	0.08
Linearized SES	0.03	0.03	1.00	0.33	0.22	-0.21	0.39	0.28	0.24
Age 7 score	0.03	0.08	0.33	1.00	0.27	-0.22	0.72	0.41	0.35
Motivation	-0.01	0.02	0.22	0.27	1.00	-0.41	0.40	0.34	0.23
Problems	0.01	-0.01	-0.21	-0.22	-0.41	1.00	-0.33	-0.22	-0.15
Age 16 score	0.03	0.12	0.39	0.72	0.40	-0.33	1.00	0.54	0.39
Attainment	0.02	0.05	0.28	0.41	0.34	-0.22	0.54	1.00	0.29
Earnings	0.03	0.08	0.24	0.35	0.23	-0.15	0.39	0.29	1.00
Girls									
Development index 11	1.00	0.22	0.02	0.06	0.08	-0.07	0.09	0.04	0.05
Development index 16	0.22	1.00	0.02	0.03	0.03	0.02	0.03	0.02	0.02
Linearized SES	0.02	0.02	1.00	0.33	0.21	-0.19	0.41	0.31	0.18
Age 7 score	0.06	0.03	0.33	1.00	0.26	-0.21	0.72	0.41	0.28
Motivation	0.08	0.03	0.21	0.26	1.00	-0.41	0.39	0.30	0.21
Problems	-0.07	0.02	-0.19	-0.21	-0.41	1.00	-0.31	-0.19	-0.08
Age 16 score	0.09	0.03	0.41	0.72	0.39	-0.31	1.00	0.51	0.33
Attainment	0.04	0.02	0.31	0.41	0.30	-0.19	0.51	1.00	0.36
Earnings	0.05	0.02	0.18	0.28	0.21	-0.08	0.33	0.36	1.00

Notes: The table shows the correlation between pairwise complete observations of age 11 pubertal development, age 16 pubertal development, a linearized index of father's social status, the first principal component of age 7 test scores, measures of motivation and self-discipline as well as of problem behavior, the first principal component of age 16 test scores, years of education, and mean log earnings.

Figure 2: Raw pubertal development gradients in age 16 test scores separately for boys and girls. The outcome is expressed in sample standard deviations. Circle areas are proportional to the number of individuals at each level.



ACCEPTED MANUSCRIPT

3.1. Age 16 outcomes

Table 7 reports the results from regressions of various age 16 outcomes on pubertal development by gender. As Figure 2 suggests that the relationship between test scores and pubertal development is concave, we include both the pubertal development index and its square as explanatory variables, and also report F-statistics for their joint significance in the regression table. To facilitate the interpretation of the estimates, we use the regression coefficients to calculate the implied test score differentials between fully mature and fully prepubescent individuals (D_i: *1 vs 0*). We calculate the associated standard error from the relevant elements of the variance-covariance matrix of the estimator. We further run separate sets of regressions with the linear and quadratic pubertal development terms replaced with dummy variables for having either one of the lowest three out of nine levels of pubertal development (D_i: *low vs mid*), or having one of the highest three out of nine levels (D_i: *high vs mid*).

As can be seen from Column (1) of Table 7 the raw relationship between pubertal development and test scores is significantly positive both for boys and for girls. For boys the estimate implies that moving from fully prepubescent (i.e. a development index equal to zero) to fully mature (i.e. development index equal to one) is associated with an increase in test scores of 0.60 standard deviations whereas the similar figure for girls is 0.46. In other words, the test score gap between fully developed individuals and individuals who are still at the onset of puberty is estimated to be approximately half a standard deviation. The estimated differences are clearly larger at lower end of the development index scale than at the higher end, where most individuals are found. The difference between the middle and the lowest third of the development scale is 0.25 for boys and 0.40 for girls whereas the difference between the highest and the middle third is only 0.11 for boys and 0.05 for girls. Of note are also the much larger standard errors for the *low vs mid* comparisons for girls, illustrating the low numbers of girls with low levels of pubertal development at 16.

In Column (2) of Table 7 we add the scores from the four age 7 cognitive skill tests as controls to our regressions. This has the effect of reducing the estimated joint coefficients of pubertal development and its square by approximately 40%. Moving from fully prepubescent to fully mature now implies an increase in test scores of 0.38 standard deviations for boys and of 0.24 standard deviations for girls. When we also include controls for father's social status, the educational attainment of both parents, and child's region of birth in Column (3), the estimated effect of pubertal development is further reduced to 0.35 for boys and reduced to 0.11 for girls. These results illustrate how the raw correlation between the timing of pubertal development and other outcomes reported in the existing literature are in part reflecting a selection into the timing of puberty rather than a more direct effect. At the same time, the age 16 score difference by pubertal development conditional on all of these variables is still precisely estimated and large for boys. For girls, the conditional estimate is smaller and still positive, but at much lower levels of precision.

As was discussed at the outset, one channel through which pubertal development is often perceived to affect educational outcomes is through its effects

ACCEPTED MANUSCRIPT

Table 7: Regression estimates of the pubertal development gradient in age 16 outcomes.

	(1)	(2)	(3)	(4)	(5)	(6)
outcome	score 16	score 16	score 16	motivation	problems	score 16
Intercept	yes	yes	yes	yes	yes	yes
Age 7 scores	no	yes	yes	yes	yes	yes
Other controls	no	no	yes	yes	yes	yes
Motivation and problems	no	no	no	no	no	yes
Boys						
Age 16 development (D_i)	1.13	0.68	0.48	-0.06	0.07	0.51
	(0.27)	(0.20)	(0.19)	(0.26)	(0.12)	(0.18)
Age 16 development squared (D_i^2)	-0.53	-0.30	-0.13	0.07	-0.05	-0.16
	(0.23)	(0.17)	(0.16)	(0.22)	(0.10)	(0.15)
D_i: 1 vs 0	0.60	0.38	0.35	0.01	0.03	0.35
	(0.07)	(0.05)	(0.05)	(0.07)	(0.03)	(0.05)
D_i: low vs mid	-0.25	-0.12	-0.11	0.02	-0.02	-0.12
	(0.05)	(0.04)	(0.03)	(0.05)	(0.02)	(0.03)
D_i: high vs mid	0.11	0.09	0.10	0.03	-0.01	0.09
	(0.03)	(0.02)	(0.02)	(0.03)	(0.01)	(0.02)
$F(D_i = 0 \wedge D_i^2 = 0)$	34.80	26.87	26.39	0.12	0.34	30.38
$P(> F)$	0.00	0.00	0.00	0.89	0.71	0.00
N	4261	4261	4261	4261	4261	4261
R^2	0.02	0.48	0.53	0.09	0.06	0.59
adj. R^2	0.02	0.48	0.53	0.08	0.06	0.59
Girls						
Age 16 development (D_i)	1.07	0.65	0.31	0.69	-0.15	0.15
	(0.55)	(0.41)	(0.39)	(0.55)	(0.23)	(0.37)
Age 16 development squared (D_i^2)	-0.61	-0.41	-0.20	-0.36	0.13	-0.11
	(0.37)	(0.28)	(0.26)	(0.37)	(0.15)	(0.25)
D_i: 1 vs 0	0.46	0.24	0.11	0.33	-0.02	0.04
	(0.20)	(0.14)	(0.14)	(0.20)	(0.08)	(0.13)
D_i: low vs mid	-0.40	-0.11	-0.05	-0.25	0.07	0.01
	(0.17)	(0.13)	(0.12)	(0.17)	(0.07)	(0.11)
D_i: high vs mid	0.05	0.02	0.02	0.04	0.01	0.01
	(0.03)	(0.02)	(0.02)	(0.03)	(0.01)	(0.02)
$F(D_i = 0 \wedge D_i^2 = 0)$	4.14	1.56	0.33	2.84	1.09	0.10
$P(> F)$	0.02	0.21	0.72	0.06	0.34	0.90
N	4117	4117	4117	4117	4117	4117
R^2	0.00	0.44	0.51	0.08	0.05	0.56
adj. R^2	0.00	0.44	0.51	0.08	0.04	0.56

Notes: Outcomes are age 16 cognitive skills, age 16 motivation, and age 16 problem behavior. Standard errors have been added in parentheses. D_i: *1 vs 0* indicates the estimated regression-adjusted difference between fully mature and fully immature individuals $E(y_i|D_i = 1, ..) - E(y_i|D_i = 0, ..)$ based on the quadratic specification, with its standard error calculated from the associated variance-covariance matrix. D_i: *low vs mid* and D_i: *high vs mid* indicate the regression-adjusted differences $E(y_i|D_i \leq \frac{1}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ and $E(y_i|D_i > \frac{2}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ from a separate model with one indicator variable for the lowest three out of nine levels of the pubertal development index, and one indicator variable for the highest three. The remaining regression statistics are derived from the quadratic model.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

on behavior and other non-cognitive outcomes. In Columns (4) and (5) of Table 7, we explore this channel by regressing our age 16 measures of motivation and problem behavior on pubertal development. Both regressions include the same controls as the regression of cognitive skills in Column (3). For boys, our estimates are close to zero for both outcomes, and not significantly different from zero. For girls the estimate for motivation is positive and large, but so imprecisely estimated that we cannot rule out even a negative relationship. In the case of problem behavior, the correlation with pubertal development is close to zero also for girls. These results confirm the pattern already observed in the raw correlations and do not give support for the hypothesis that the timing of pubertal development would be associated with large behavioral or motivational problems around age 16 for boys, at least conditional on pre-pubertal differences.

Behavioral and motivational problems are also not the channel through which pubertal development affects cognitive skill scores. When we change the specification of Column (3) to also include motivation and problem behavior as controls in Column (6), the estimates are virtually unchanged for boys and the implied difference between the fully mature and prepubescent boys is again 0.35. Given that both our motivation and problem measures are reasonably strongly correlated with cognitive skills in and of themselves, it is unlikely that they would simply be too noisy to pick up an effect. We thus interpret our estimates in Columns (4) through (6) as suggestive evidence for the observed test score gradient in pubertal development being due to a relationship with the timing of cognitive rather than non-cognitive development. For girls, the estimates remain too noisy to draw strong conclusions, and the estimate of the relationship with cognitive skill scores remains statistically indistinguishable from zero both before and after controlling for motivation and problems.

3.2. Adult outcomes

As the results in Table 7 suggest that pubertal development may have cognitive consequences during adolescence, it is natural to ask whether these effects are also visible in adult outcomes such as final educational attainment and earnings. In Table 8 we report the results of specifications similar to Table 7 Columns (1) and (3), but with final educational attainment expressed in years of education as the outcome variable. The estimated relationship between pubertal development and educational attainment are of similar magnitude for boys as for girls, at about three quarters of a year of education without controls in Column (1), and at a difference of about half a year once parental background and age 7 test scores are taken into account in Column (2). Though the estimates are similar across gender, the estimates for girls are so imprecise that they become almost uninterpretable. This can also be seen from the apparently inconsistent results between the quadratic and dummy variable specifications for girls.

In Columns (3) and (4), we repeat the same analysis with mean log earnings at ages 33 and 42 as the outcome. The raw relationship between pubertal development and earnings is positive for both boys and girls but precisely estimated only for the former. The estimated unconditional relationship implies that as

ACCEPTED MANUSCRIPT

Table 8: Regression estimates of the pubertal development gradient in educational attainment and earnings.

specification	(1)	(2)	(3)	(4)
outcome	attainment	attainment	log earnings	log earnings
Intercept	yes	yes	yes	yes
Age 7 scores	no	yes	no	yes
Other controls	no	yes	no	yes
Boys				
Age 16 development (D_i)	2.26	1.02	0.38	0.28
	(0.91)	(0.82)	(0.18)	(0.17)
Age 16 development squared (D_i^2)	-1.49	-0.58	-0.16	-0.12
	(0.78)	(0.70)	(0.15)	(0.14)
D_i: 1 vs 0	0.76	0.45	0.21	0.16
	(0.24)	(0.21)	(0.05)	(0.04)
D_i: low vs mid	-0.29	-0.12	-0.08	-0.05
	(0.16)	(0.15)	(0.03)	(0.03)
D_i: high vs mid	0.05	0.08	0.05	0.04
	(0.11)	(0.10)	(0.02)	(0.02)
$F(D_i = 0 \wedge D_i^2 = 0)$	5.30	2.10	10.84	6.70
$P(> F)$	0.01	0.12	0.00	0.00
N	3043	3043	3335	3335
R^2	0.00	0.20	0.01	0.15
adj. R^2	0.00	0.19	0.01	0.14
Girls				
Age 16 development (D_i)	1.36	1.18	0.29	-0.08
	(1.80)	(1.63)	(0.51)	(0.49)
Age 16 development squared (D_i^2)	-0.65	-0.68	-0.15	0.09
	(1.21)	(1.09)	(0.35)	(0.33)
D_i: 1 vs 0	0.71	0.50	0.14	0.01
	(0.62)	(0.57)	(0.18)	(0.17)
D_i: low vs mid	-0.20	0.01	-0.01	0.03
	(0.55)	(0.50)	(0.16)	(0.15)
D_i: high vs mid	0.14	0.08	0.04	0.03
	(0.10)	(0.09)	(0.03)	(0.03)
$F(D_i = 0 \wedge D_i^2 = 0)$	1.64	0.52	0.70	0.39
$P(> F)$	0.19	0.59	0.50	0.68
N	3225	3225	3550	3550
R^2	0.00	0.20	0.00	0.10
adj. R^2	0.00	0.19	0.00	0.09

Notes: the outcome is age 46 educational attainment measured in years of education in Columns (1) and (2) and mean log earnings at ages 33 and 42 in Columns (3) and (4).. Standard errors have been added in parentheses. D_i: *1 vs 0* indicates the estimated regression-adjusted difference between fully mature and fully immature individuals $E(y_i|D_i = 1, ..) - E(y_i|D_i = 0, ..)$ based on the quadratic specification, with its standard error calculated from the associated variance-covariance matrix. D_i: *low vs mid* and D_i: *high vs mid* indicate the regression-adjusted differences $E(y_i|D_i \leq \frac{1}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ and $E(y_i|D_i > \frac{2}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ from a separate model with one indicator variable for the lowest three out of nine levels of the pubertal development index, and one indicator variable for the highest three. The remaining regression statistics are derived from the quadratic model.

adults, very early maturing boys earn approximately 0.21 log points, or 23 percent more than very late maturing boys. Controlling for age 7 test scores and family background reduces this difference to a still relatively precisely estimated 17 percent. For girls, standard errors are again too large to interpret the results.

ACCEPTED MANUSCRIPT

3.3. Age 11 pubertal development

As shown in Section 2, at age 16 there is much less variation in measured pubertal development among girls than among boys, and this in turn contributes to the lack of precision in the estimates for girls. Even if our primary interest is pubertal development at age 16, we repeat part of our analysis using the age 11 pubertal development measure, where there is relatively more variation among girls.

The first three columns of Table 9 mirror the first three columns of 7, but with the age 16 development index now replaced with the age 11 development index. As could be expected, the estimates for boys now have low precision, and we will mostly ignore them. The quadratic specification suggests an inverted U-shaped pattern for girls, both conditionally and unconditionally, with both very early and very late maturing girls performing badly, even if the conditional D_i: *high vs mid* estimates in Columns (2) and (3) suggest that early maturing girls may not do worse at all.

Figure 3 shows the regression-adjusted test scores associated with the estimates of Column (3) in Table 9. As can be seen from the figure, the relationship does indeed look inverted U-shaped, but the vast majority of individuals are located on the upward sloping part of the curve. Indeed, the reason that the *high vs mid* point estimates in the regression tables are positive for both genders is exactly because so few individuals have been assessed to have one of the top two levels of age 11 pubertal development.

Differences in girls' motivation in Column (4) shows a similar pattern as in test scores, with estimates not clearly statistically distinguishable from zero, but suggestive of a small positive relationship for most girls. Estimates of the relationship with problem behavior are also small in Column (5). As a consequence, when we again adjust the test score estimate for motivation and problem behavior in Column (6), the estimates are virtually unchanged from Column (3).

Turning to the attainment and earnings estimates in Table 10, we can see that precision is again too low to draw any conclusions for boys, while estimates for girls are somewhat imprecise but surprisingly similar to the age 16 pubertal development estimates for boys.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 9: Regression estimates of the age 11 pubertal development gradient in age 16 outcomes.

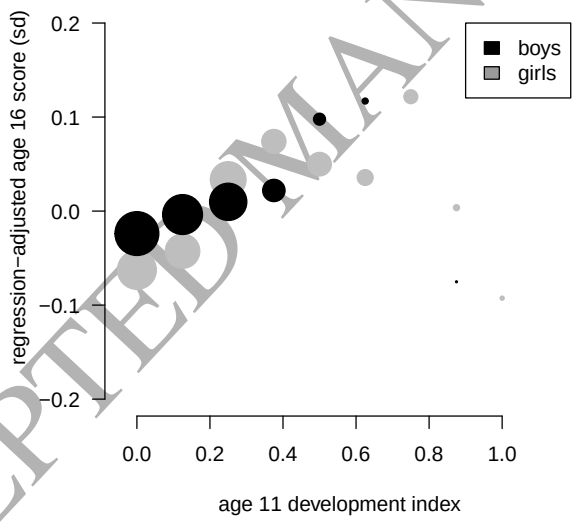
	(1)	(2)	(3)	(4)	(5)	(6)
outcome	score 16	score 16	score 16	motivation	problems	score 16
Intercept	yes	yes	yes	yes	yes	yes
Age 7 scores	no	yes	yes	yes	yes	yes
Other controls	no	no	yes	yes	yes	yes
Motivation and problems	no	no	no	no	no	yes
Boys						
Age 11 development (D_i)	0.52	0.15	0.02	-0.10	0.18	0.08
	(0.23)	(0.17)	(0.16)	(0.22)	(0.10)	(0.15)
Age 11 development squared (D_i^2)	-0.48	0.21	0.39	0.13	-0.30	0.29
	(0.46)	(0.34)	(0.32)	(0.43)	(0.20)	(0.30)
D_i: 1 vs 0	0.04	0.36	0.41	0.02	-0.12	0.37
	(0.28)	(0.21)	(0.19)	(0.25)	(0.12)	(0.18)
D_i: low vs mid	-0.06	-0.05	-0.05	0.01	-0.03	-0.06
	(0.05)	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)
D_i: high vs mid	0.09	0.23	0.26	-0.13	-0.05	0.27
	(0.22)	(0.16)	(0.15)	(0.21)	(0.10)	(0.14)
$F(D_i = 0 \wedge D_i^2 = 0)$	4.60	5.20	4.17	0.14	1.60	5.34
$P(> F)$	0.01	0.01	0.02	0.87	0.20	0.00
N	4424	4424	4424	4424	4424	4424
R^2	0.00	0.47	0.53	0.09	0.07	0.59
adj. R^2	0.00	0.47	0.52	0.08	0.07	0.59
Girls						
Age 11 development (D_i)	0.78	0.51	0.47	0.79	-0.09	0.31
	(0.18)	(0.14)	(0.13)	(0.19)	(0.08)	(0.12)
Age 11 development squared (D_i^2)	-0.56	-0.40	-0.39	-0.80	0.04	-0.24
	(0.26)	(0.19)	(0.18)	(0.26)	(0.11)	(0.17)
D_i: 1 vs 0	0.22	0.12	0.08	-0.01	-0.05	0.07
	(0.11)	(0.08)	(0.08)	(0.11)	(0.05)	(0.07)
D_i: low vs mid	-0.19	-0.10	-0.08	-0.06	0.01	-0.07
	(0.03)	(0.03)	(0.02)	(0.03)	(0.01)	(0.02)
D_i: high vs mid	-0.03	0.04	0.03	0.05	-0.05	0.01
	(0.07)	(0.06)	(0.05)	(0.08)	(0.03)	(0.05)
$F(D_i = 0 \wedge D_i^2 = 0)$	21.00	14.53	12.52	12.12	2.65	6.79
$P(> F)$	0.00	0.00	0.00	0.00	0.07	0.00
N	4315	4315	4315	4315	4315	4315
R^2	0.01	0.44	0.52	0.09	0.06	0.57
adj. R^2	0.01	0.44	0.51	0.08	0.06	0.56

Notes: Outcomes are age 16 cognitive skills, age 16 motivation, and age 16 problem behavior. Standard errors have been added in parentheses. D_i: *1 vs 0* indicates the estimated regression-adjusted difference between fully mature and fully immature individuals $E(y_i|D_i = 1, ..) - E(y_i|D_i = 0, ..)$ based on the quadratic specification, with its standard error calculated from the associated variance-covariance matrix. D_i: *low vs mid* and D_i: *high vs mid* indicate the regression-adjusted differences $E(y_i|D_i \leq \frac{1}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ and $E(y_i|D_i > \frac{2}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ from a separate model with one indicator variable for the lowest three out of nine levels of the pubertal development index, and one indicator variable for the highest three. The remaining regression statistics are derived from the quadratic model.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Figure 3: Regression-adjusted pubertal development gradients in age 16 test scores for boys (black) and girls (grey) using pubertal development at 11. The outcome is the age 16 test score expressed in sample standard deviations. Circle areas are proportional to the number of individuals in each bin. Black circles represent boys, grey circles girls. The figure shows that with the exception of a small number of very early developing individuals, age 16 test scores rise in pubertal development conditional on covariates.



ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 10: Regression estimates of the age 11 pubertal development gradient in educational attainment and earnings.

specification	(1)	(2)	(3)	(4)
outcome	attainment	attainment	log earnings	log earnings
Intercept	yes	yes	yes	yes
Age 7 scores	no	yes	no	yes
Other controls	no	yes	no	yes
Boys				
Age 11 development (D_i)	0.65	-0.23	0.36	0.22
	(0.77)	(0.70)	(0.15)	(0.14)
Age 11 development squared (D_i^2)	-0.51	0.80	-0.53	-0.37
	(1.59)	(1.45)	(0.30)	(0.28)
D_i: 1 vs 0	0.13	0.57	-0.17	-0.15
	(0.94)	(0.89)	(0.19)	(0.17)
D_i: low vs mid	0.11	0.10	-0.04	-0.02
	(0.15)	(0.14)	(0.03)	(0.03)
D_i: high vs mid	1.17	1.10	-0.28	-0.25
	(0.84)	(0.77)	(0.16)	(0.15)
$F(D_i = 0 \wedge D_i^2 = 0)$	0.83	0.23	3.38	1.26
$P(> F)$	0.43	0.80	0.03	0.28
N	3450	3450	3817	3817
R^2	0.00	0.19	0.00	0.15
adj. R^2	0.00	0.18	0.00	0.14
Girls				
Age 11 development (D_i)	0.78	0.04	0.23	0.14
	(0.57)	(0.51)	(0.16)	(0.15)
Age 11 development squared (D_i^2)	-0.38	0.31	-0.12	-0.05
	(0.80)	(0.72)	(0.22)	(0.21)
D_i: 1 vs 0	0.40	0.35	0.11	0.08
	(0.34)	(0.31)	(0.10)	(0.09)
D_i: low vs mid	-0.27	-0.12	-0.06	-0.03
	(0.10)	(0.09)	(0.03)	(0.03)
D_i: high vs mid	-0.02	0.05	0.04	0.04
	(0.22)	(0.20)	(0.06)	(0.06)
$F(D_i = 0 \wedge D_i^2 = 0)$	3.39	0.94	3.54	1.68
$P(> F)$	0.03	0.39	0.03	0.19
N	3653	3653	4082	4082
R^2	0.00	0.20	0.00	0.10
adj. R^2	0.00	0.19	0.00	0.09

Notes: the outcome is age 46 educational attainment measured in years of education in Columns (1) and (2) and mean log earnings at ages 33 and 42 in Columns (3) and (4).. Standard errors have been added in parentheses. D_i: *1 vs 0* indicates the estimated regression-adjusted difference between fully mature and fully immature individuals $E(y_i|D_i = 1, ..) - E(y_i|D_i = 0, ..)$ based on the quadratic specification, with its standard error calculated from the associated variance-covariance matrix. D_i: *low vs mid* and D_i: *high vs mid* indicate the regression-adjusted differences $E(y_i|D_i \leq \frac{1}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ and $E(y_i|D_i > \frac{2}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ from a separate model with one indicator variable for the lowest three out of nine levels of the pubertal development index, and one indicator variable for the highest three. The remaining regression statistics are derived from the quadratic model.

3.4. Robustness

In the results presented so far, we retain individuals in the sample with partially missing information by imputing missing values within traits. This naturally raises the question about the sensitivity of our results to this imputation procedure. If we simply refrain from using any individuals with any missing information our sample sizes are reduced by 7% to 54% depending on the outcome variable used. In appendix Table C.13 we replicate the regressions of age 16 test scores as well as adult educational attainment and earnings on pubertal development at age 16 with samples where individuals with any missing values are dropped. As can be seen from Table C.13, the results are not driven by imputed observations. Even if we drop the observations for which we have imputed

ACCEPTED MANUSCRIPT

some of the trait measures, we still find that boys' late pubertal development is associated with lower test scores at age 16, lower years of education, and lower earnings.

The age 16 cognitive skill score is a combination of math and English subscores. Both because of gender differences in the two subjects and because of differences in how and where individuals learn these specific skills, it is of interest to separate the overall test score into these two separate outcomes. Appendix Table D.14 shows the result of this exercise. As can be seen from the table, point estimates are larger for English than for math, but the results are otherwise similar between the subscores.

4. Conclusions

The results presented in this paper suggest that late pubertal development of boys can have long-lasting consequences. Boys who were more or less prepubescent at age 16 score lower in cognitive tests at that age than should otherwise be expected. Their lower cognitive performance in adolescence also seems to be reflected in adult outcomes. Late developing boys have lower levels of educational attainment and earnings in adulthood than those who were more mature by age 16 even after controlling for factors such as age 7 cognitive skills and family background. At the same time, we also show that part of the raw relationship between pubertal development and various outcome measures can be explained by selection into the timing of puberty, for example by parental background.

Results for girls are imprecise when using age 16 pubertal development because large proportions of girls have already gone through pubertal development at that age. In part this reflect biological fact rather than a measurement problem. Using age 11 measures of pubertal development however reveals a similar upward-sloping relationship between pubertal development and age 16 cognitive skills for the vast majority of girls as we find for boys. The precision of our estimates is unfortunately too low to draw strong conclusions about girls' adult outcomes, but estimates are consistent with similar patterns existing for girls as for boys.

In the earlier literature it has been argued that the fact that compulsory education ends at age 16 may exacerbate the consequences of developmental differences at that age (e.g. Pekkarinen, 2008). Our results are consistent with this argument, but our data unfortunately do not allow us to study this question more explicitly, nor does it allow us to determine whether for example an educational system that would stream individuals based on their development would weaken the association between the timing of puberty and educational outcomes.

Gender differences in the timing of pubertal development are often cited as a potential factor behind the much discussed gender differences in educational attainment. Regretably, the NCDS data are not ideal for the study of factors behind the type of gender differences that we observe in education today. In the British 1958 birth cohort, male educational attainment is still *higher* than

ACCEPTED MANUSCRIPT

female attainment. Even though girls were more motivated and disciplined than boys in this cohort, and displayed less problem behavior, some of the causes of the current female dominance in educational attainment were clearly either not present at the time when members of this cohort made their educational choices, or were at the time more than compensated for by other factors. It is also worth remembering that while we find relatively large differences in outcomes when comparing fully mature to fully immature individuals, the differences in pubertal development are much smaller between typical individuals, even across gender. This limits the scope for how much of the variation in educational outcomes differences in pubertal development can hope to explain.

5. Acknowledgments

Funding: This work was supported by the Strategic Research Council at the Academy of Finland [grant number: 293445]. The authors appreciate the useful help and insightful comments from Sharon Simonton, Jonas Vlachos, Roope Uusitalo, Petri Böckerman, Mika Kortelainen, participants in various workshops and conferences, as well as from the editor and the two anonymous referees.

References

References

Arnett, J. J. (1999). Adolescent storm and stress, reconsidered. *American psychologist*, *54*, 317.

Case, A., and Paxson, C. (2008). Stature and status: Height, ability, and labor market outcomes. *Journal of Political Economy*, *116*.

Currie, J., and Almond, D. (2011). Human capital development before age five. In *Handbook of labor economics* (pp. 1315–1486). Elsevier volume 4.

Dreber, A., Von Essen, E., and Ranehill, E. (2011). Age at pubertal onset and educational outcomes. *Available at SSRN 1934188*, .

Dubas, J., Graber, J., and Petersen, A. (1991). The effects of pubertal development on achievement during adolescence. *American Journal of Education*, *99*, 444–460.

Flannery, D. J., Torquati, J. C., and Lindemeier, L. (1994). The method and meaning of emotional expression and experience during adolescence. *Journal of adolescent Research*, *9*, 8–27.

Josse, J., and Husson, F. (2012). Handling missing values in exploratory multivariate data analysis methods. *Journal de la Société Française de Statistique*, *153*, 79–99.

ACCEPTED MANUSCRIPT

Koivusilta, L., and Rimpelä, A. (2004). Pubertal timing and educational careers: a longitudinal study. *Annals of human biology*, *31*, 446–465.

Lundberg, S. J. (2017). Father absence and the educational gender gap. *IZA Discussion Paper*, *10814*.

Newcombe, N., and Dubas, J. (1987). Individual differences in cognitive ability: are they related to timing of puberty. In R. Lerner, and T. Foch (Eds.), *Biological-psychological interactions in early adolescence: a life-span perspective*. Hillsdale: Erlbaum.

Pekkarinen, T. (2008). Gender differences in educational attainment: Evidence on the role of tracking from a finnish quasi-experiment. *The Scandinavian Journal of Economics*, *110*, 807–825.

Peper, J. S., and Dahl, R. E. (2013). The teenage brain surging hormones—brain-behavior interactions during puberty. *Current directions in psychological science*, *22*, 134–139.

Persico, N., Postlewaite, A., and Silverman, D. (2004). The effect of adolescent experience on labor market outcomes: The case of height. *Journal of Political Economy*, *112*, 1019–1053.

Rasmussen, A. R., Wohlfahrt-Veje, C., de Renzy-Martin, K. T., Hagen, C. P., Tinggaard, J., Mouritsen, A., Mieritz, M. G., and Main, K. M. (2015). Validity of self-assessment of pubertal maturation. *Pediatrics*, *135*, 86–93.

Steinberg, L. (2014). *Adolescence*. McGraw-Hill.

Steinberg, L., and Morris, A. (2001). Adolescent development. *Journal of Cognitive Education and Psychology*, *2*, 55–87.

Susman, E. J., and Dorn, L. D. (2009). Puberty: Its role in development. *Handbook of adolescent psychology*, .

Tanner, J. (1962). *Growth at adolescence.*. Springfield, Ill.

University of London. Institute of Education. Centre for Longitudinal Studies (2008a). [dataset] National Child Development Study, Sweep 5, 1991, 2nd Edition. City University. Social Statistics Research Unit, [original data producer(s)]. UK Data Service. SN: 5567, http://doi.org/10.5255/UKDA-SN-5567-1.

University of London. Institute of Education. Centre for Longitudinal Studies (2008b). [dataset] National Child Development Study, Sweep 7, 2004–2005, 3rd Edition. UK Data Service. SN: 5579, http://doi.org/10.5255/UKDA-SN-5579-1.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

University of London. Institute of Education. Centre for Longitudinal Studies (2014). [dataset] National Child Development Study, Childhood Data, Sweeps 0–3, 1958–1974, 3rd Edition. National Birthday Trust Fund, National Children's Bureau, [original data producer(s)]. UK Data Service. SN: 5565, http://doi.org/10.5255/UKDA-SN-5565-2.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Appendix A. Principal components

Table A.11: Motivation factor loadings

variable	weight
I feel school is a waste of time	0.37
I am quiet in the classroom and get on with my work	0.30
I think homework is a bore	0.36
I find it difficult to keep my mind on my work	0.35
I never take work seriously	0.39
I don't like school	0.41
I think there is no point in planning for the future	0.21
I'm always ready to help the teacher	0.20
Teacher assessment: hardworking	0.34

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Appendix B. Selective attrition

Table B.12: Selective attrition.

specification	(1)	(2)
outcome	score 16	score 16
Intercept	yes	yes
Age 7 scores	no	no
Other controls	no	no
Boys		
Age 16 development (D_i)	1.13	1.17
	(0.27)	(0.26)
Age 16 development squared (D_i^2)	-0.53	-0.58
	(0.23)	(0.22)
D_i: 1 vs 0	0.60	0.59
	(0.07)	(0.07)
D_i: low vs mid	-0.25	-0.24
	(0.05)	(0.05)
D_i: high vs mid	0.11	0.10
	(0.03)	(0.03)
$F(D_i = 0 \wedge D_i^2 = 0)$	34.80	37.02
$P(> F)$	0.00	0.00
N	4261	4826
R^2	0.02	0.02
adj. R^2	0.02	0.01
Girls		
Age 16 development (D_i)	1.07	1.11
	(0.55)	(0.53)
Age 16 development squared (D_i^2)	-0.61	-0.67
	(0.37)	(0.35)
D_i: 1 vs 0	0.46	0.44
	(0.20)	(0.18)
D_i: low vs mid	-0.40	-0.37
	(0.17)	(0.16)
D_i: high vs mid	0.05	0.03
	(0.03)	(0.03)
$F(D_i = 0 \wedge D_i^2 = 0)$	4.14	3.43
$P(> F)$	0.02	0.03
N	4117	4595
R^2	0.00	0.00
adj. R^2	0.00	0.00

Notes: the outcome is the standardized age 16 test score. Column (1) shows the unadjusted relationship between pubertal development and test scores within the estimation sample from Table 7. In Column (2), observations which are missing on other variables but non-missing on pubertal development and age 16 test scores have also been added to the sample. The relationship between the two variables is almost unchanged between the columns. D_i: *1 vs 0* indicates the estimated regression-adjusted difference between fully mature and fully immature individuals $E(y_i|D_i = 1, ..) - E(y_i|D_i = 0, ..)$ based on the quadratic specification, with its standard error calculated from the associated variance-covariance matrix. D_i: *low vs mid* and D_i: *high vs mid* indicate the regression-adjusted differences $E(y_i|D_i \leq \frac{1}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ and $E(y_i|D_i > \frac{2}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ from a separate model with one indicator variable for the lowest three out of nine levels of the pubertal development index, and one indicator variable for the highest three. The remaining regression statistics are derived from the quadratic model.

ACCEPTED MANUSCRIPT

Appendix C. Imputed observations removed

Table C.13: Regression estimates of the pubertal development gradient, unimputed information only.

	(1)	(2)	(3)	(4)	(5)
outcome	score 16	score 16	score 16	attainment	log earnings
Intercept	yes	yes	yes	yes	yes
Age 7 scores	no	yes	yes	yes	yes
Other controls	no	no	yes	yes	yes
Boys					
Age 16 development (D_i)	1.37	0.88	0.71	1.05	0.36
	(0.32)	(0.24)	(0.23)	(0.84)	(0.21)
Age 16 development squared (D_i^2)	-0.70	-0.43	-0.27	-0.59	-0.18
	(0.28)	(0.20)	(0.19)	(0.72)	(0.18)
D_i: 1 vs 0	0.67	0.45	0.44	0.46	0.18
	(0.09)	(0.06)	(0.06)	(0.22)	(0.05)
D_i: low vs mid	-0.29	-0.17	-0.16	-0.17	-0.06
	(0.06)	(0.04)	(0.04)	(0.15)	(0.04)
D_i: high vs mid	0.12	0.10	0.11	0.08	0.04
	(0.04)	(0.03)	(0.03)	(0.10)	(0.03)
$F(D_i = 0 \wedge D_i^2 = 0)$	30.37	25.37	27.71	2.15	5.33
$P(> F)$	0.00	0.00	0.00	0.12	0.00
N	2917	2917	2917	2832	1742
R^2	0.02	0.47	0.53	0.20	0.17
adj. R^2	0.02	0.47	0.52	0.20	0.15
Girls					
Age 16 development (D_i)	0.91	0.54	0.14	0.82	0.02
	(0.66)	(0.50)	(0.47)	(1.68)	(0.69)
Age 16 development squared (D_i^2)	-0.47	-0.30	-0.06	-0.49	-0.02
	(0.45)	(0.34)	(0.32)	(1.13)	(0.46)
D_i: 1 vs 0	0.44	0.24	0.08	0.33	0.00
	(0.23)	(0.18)	(0.16)	(0.56)	(0.25)
D_i: low vs mid	-0.49	-0.19	-0.11	-0.01	0.15
	(0.23)	(0.18)	(0.17)	(0.52)	(0.24)
D_i: high vs mid	0.06	0.03	0.02	0.04	0.03
	(0.04)	(0.03)	(0.03)	(0.10)	(0.04)
$F(D_i = 0 \wedge D_i^2 = 0)$	3.65	1.39	0.36	0.19	0.01
$P(> F)$	0.03	0.25	0.70	0.83	0.99
N	2758	2758	2758	2978	1633
R^2	0.00	0.42	0.50	0.20	0.11
adj. R^2	0.00	0.42	0.50	0.19	0.09

Notes: outcomes are age 16 cognitive skills, years of education, and mean log earnings. Standard errors have been added in parentheses. D_i: *1 vs 0* indicates the estimated regression-adjusted difference between fully mature and fully immature individuals $E(y_i|D_i = 1, ..) - E(y_i|D_i = 0, ..)$ based on the quadratic specification, with its standard error calculated from the associated variance-covariance matrix. D_i: *low vs mid* and D_i: *high vs mid* indicate the regression-adjusted differences $E(y_i|D_i \leq \frac{1}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ and $E(y_i|D_i > \frac{2}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ from a separate model with one indicator variable for the lowest three out of nine levels of the pubertal development index, and one indicator variable for the highest three. The remaining regression statistics are derived from the quadratic model.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Appendix D. Separating math and English skills

Table D.14: Pubertal development and Math and English scores.

specification	(1)	(2)
outcome	math 16	English 16
Intercept	yes	yes
Age 7 scores	yes	yes
Other controls	yes	yes
Boys		
Age 16 development (D_i)	0.35	0.52
	(0.20)	(0.20)
Age 16 development squared (D_i^2)	-0.12	-0.13
	(0.17)	(0.17)
D_i: 1 vs 0	0.23	0.40
	(0.05)	(0.05)
D_i: low vs mid	-0.07	-0.13
	(0.04)	(0.04)
D_i: high vs mid	0.06	0.12
	(0.03)	(0.02)
$F(D_i = 0 \wedge D_i^2 = 0)$	10.10	31.28
$P(> F)$	0.00	0.00
N	4261	4261
R^2	0.44	0.47
adj. R^2	0.44	0.46
Girls		
Age 16 development (D_i)	0.35	0.21
	(0.43)	(0.39)
Age 16 development squared (D_i^2)	-0.28	-0.08
	(0.29)	(0.26)
D_i: 1 vs 0	0.06	0.13
	(0.15)	(0.14)
D_i: low vs mid	-0.03	-0.06
	(0.13)	(0.12)
D_i: high vs mid	-0.01	0.04
	(0.02)	(0.02)
$F(D_i = 0 \wedge D_i^2 = 0)$	1.22	1.62
$P(> F)$	0.30	0.20
N	4117	4117
R^2	0.40	0.47
adj. R^2	0.39	0.47

The table shows estimates from a specification similar to the main specification in Table 7, Column (3), but separately using the mathematics and English subscales as the outcome variable. D_i: *1 vs 0* indicates the estimated regression-adjusted difference between fully mature and fully immature individuals $E(y_i|D_i = 1, ..) - E(y_i|D_i = 0, ..)$ based on the quadratic specification, with its standard error calculated from the associated variance-covariance matrix. D_i: *low vs mid* and D_i: *high vs mid* indicate the regression-adjusted differences $E(y_i|D_i \leq \frac{1}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ and $E(y_i|D_i > \frac{2}{3}, ..) - E(y_i|\frac{1}{3} < D_i \leq \frac{2}{3}, ..)$ from a separate model with one indicator variable for the lowest three out of nine levels of the pubertal development index, and one indicator variable for the highest three. The remaining regression statistics are derived from the quadratic model.

ACCEPTED MANUSCRIPT