Title: Multi-method assessment of pubertal timing and associations with internalizing psychopathology in early adolescent girls

Authors: Marjolein E. A. Barendse*[1], Michelle L. Byrne*[1,2], John C. Flournoy[3], Elizabeth A. McNeilly[1], Victoria Guazzelli Williamson[1], Ann-Marie Barrett[1], Samantha J. Chavez[1], Elizabeth A. Shirtcliff[4], Nicholas B. Allen[1,5], Jennifer H. Pfeifer[1]

* *Equal contributors*

1. Department of Psychology, University of Oregon, Eugene, OR, USA
2. Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Victoria, Australia
3. Department of Psychology, Harvard University, Cambridge, Massachusetts, USA
4. Human Development and Family Studies, Iowa State University, Ames, Iowa, USA
5. School of Psychological Sciences, University of Melbourne, Victoria, Australia

Corresponding author: name, address, telephone and fax numbers, and e-mail address

Acknowledgements: We would like to thank all participants and their families for their involvement in the study. We would also like to thank Danielle Cosme for her guidance and advice on specification curve analysis methods. This work was supported by the National Institute of Mental Health (R01MH107418; PI Pfeifer). Author MLB was supported by the National Institute Of Mental Health under Award Number K01MH111951. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Keywords: pubertal timing; depression; anxiety; adolescence; specification curve analysis

Abstract

Objective: Early pubertal timing has consistently been associated with internalizing psychopathology in adolescent girls. Here, we aimed to examine whether the association between timing and mental health outcomes varies by measurement of pubertal timing and internalizing psychopathology, differs between adrenarcheal and gonadarcheal processes, and is stronger concurrently or prospectively.

Methods: We assessed 174 female adolescents (age 10.0-13.0 at Time 1) twice, with an 18-month interval. Participants provided self-reported assessments of depression/anxiety symptoms and pubertal development, subjective pubertal timing, and date of menarche. Their parents/guardians also reported on the adolescent's pubertal development and subjective pubertal timing. We assessed salivary DHEA, testosterone and estradiol levels, and conducted clinical interviews to determine the presence of case level (DSM-IV and HiTOP) internalizing disorders. From these data, we computed 12 measures of pubertal timing at both time points, as well as 7 measures of internalizing psychopathology, and entered these in a Specification Curve Analysis.

Results and Conclusion: Overall, earlier pubertal timing was associated with increased internalizing psychopathology cross-sectionally and prospectively. However, results varied by measure of pubertal timing and psychopathology, with the strongest associations when pubertal timing was based on the Tanner Stage Line Drawings and when the outcome was case-level DSM-IV depression or HiTOP distress disorders. Timing based on hormone levels was not associated with internalizing psychopathology, suggesting that psychosocial mechanisms, captured by timing measures of visible physical characteristics, are more meaningful determinants of internalizing psychopathology than biological ones in early adolescent girls. Future research should precisely measure and test these psychosocial mechanisms.

Introduction

Adolescence is a sensitive period of life for neurobiological development and risk for psychopathology.[1,2] Girls are two to three times more likely to experience depression than boys from puberty onwards.[3] The substantial changes in social, physical, and hormonal development that occur during pubertal development can be related to mental health outcomes.[4] Pubertal timing, which is pubertal status or stage relative to same-age and same-sex peers, has repeatedly and independently been associated with risk for psychopathology.[5–7] In particular, many studies show that early timing (i.e., developing ahead of peers) is associated with increased risk for internalizing disorders like depression and anxiety[8], although some studies have found that this effect is small[9] or not statistically significant[3]. A recent meta-analysis[10] of 101 studies found that, overall, early timing is associated with more internalizing psychopathology, although this was moderated by the measurement method of pubertal timing.

Different methods may tap into two different groups of mechanisms proposed to drive the association between pubertal timing and mental health: psychosocial and biological mechanisms.[11] Biological processes include sensitivity of the brain to pubertal hormones, for example.[12] Meanwhile, psychosocial mechanisms might include negative self-perceptions of physical differences, or consequences of others overestimating an adolescent's social or cognitive maturity. Subjective timing (asking adolescents to rate their own pubertal timing) addresses psychosocial mechanisms more, whereas age at menarche or hormone levels relative to age both represent biological mechanisms, and physical maturation measures (e.g. the Pubertal Development Scale (PDS[13]) or the Tanner Staging Line Drawings (LD[14])) capture a combination of both since they are the direct result of hormonal changes but are also visible to the adolescent and people in their environment. In addition to distinctions between measures that capture biological vs. psychosocial mechanisms, another meaningful distinction in

measurement of pubertal timing lies in the different processes of puberty, adrenarche and gonadarche.[15] Compared to gonadarche, there is less research on adrenarche predicting psychopathology, even though it represents a period where adrenal hormones may be important mechanisms for brain development during the transition from late childhood to early adolescence.[16,17 18]

Differences in measurement or definition of internalizing psychopathology may also contribute to inconsistencies in associations with pubertal timing (for review, see[19]). A meta-analysis found a significant association between pubertal timing and both "distress" and "fear" psychopathology .[10] Importantly, however, they did not distinguish between symptomatic and diagnostic measures of psychopathology. Limiting outcomes to only case-level diagnoses may miss associations between pubertal timing and variation in subclinical symptoms, or have reduced power compared to continuous symptom-level variables with greater sample variance. On the other hand, focusing only on symptoms may obfuscate clinically-meaningful outcomes, and typically relies on self-report questionnaires, which can include subjective bias. It is also possible that discrete diagnostic categories alone may not fully capture the spectrum of mechanisms underlying developmental psychopathology. The categorical framework of the Diagnostic and Statistical Manual of Mental Disorders (DSM) may not fully capture the heterogeneity within disorders and the common co-occurrence between certain disorders. The Hierarchical Taxonomy of Psychopathology (HiTOP) is an example of a research-driven approach to classifying mental disorder, wherein the structure of psychopathology is conceptualized through higher order dimensions (e.g., internalizing) within which lower order subfactors (e.g., distress, fear) are embedded. Studies using both approaches have found associations between timing and internalizing disorders.[20–22]

Finally, the aforementioned meta-analysis[10] found that age of the sample did not moderate the association between early timing and psychopathology, but they only used cross-sectional data, and

cannot show if pubertal timing can predict later mental health outcomes. Of the handful of prospective longitudinal studies available, some show that various measures of pubertal timing [23–26] have been prospectively associated with internalizing psychopathology in later adolescence and sometimes through young adulthood, although not always.[27] There is also conflicting evidence whether early timing is related to internalizing psychopathology when controlling for history of psychopathology.[28,29] This has implications for identifying pubertal timing as a potential risk factor for the onset of mental health problems during adolescence, which could inform prevention and early intervention efforts.

Research Questions and Hypotheses

Previous research has established that early pubertal timing is a risk factor for internalizing psychopathology in adolescents. However, several substantive (i.e., mechanistic) and methodological questions remain. These are related to 1) the measurement of pubertal timing, 2) the relevance of adrenarcheal versus gonadarcheal processes, 3) the measurement of internalizing psychopathology, and 4) the existence of concurrent versus prospective associations between timing and psychopathology. The aim of the current longitudinal study was to determine the ways in which pubertal timing is cross-sectionally and prospectively associated with internalizing psychopathology in a sample of mostly White adolescent girls.[30] We focused on female adolescents because an important part of the analyses includes pubertal processes, which differ vastly between the sexes, and because girls become increasingly at risk for internalizing mental health problems during puberty.[3,31] To address the open questions discussed above, we applied specification curve analysis (SCA; also called multiverse analysis), a technique that allows researchers to examine and report all non-redundant, reasonable, and justifiable measurement and analytic specifications, and to identify the consequences of specification decisions.[32] The choices in the SCA included (detailed in Methods):

1. Different types of measurement methods of pubertal timing;
2. Within those types, measures of adrenarcheal vs. gonadarcheal processes;
3. Different types of measurement methods of internalizing psychopathology;
4. Cross-sectional and prospective associations between pubertal timing and internalizing psychopathology;
5. Inclusion of control variables (the covariates we considered were threat-related early life stress and pre-existing internalizing psychopathology);
6. If missing data was imputed or deleted listwise.

Based on the findings from the previous meta-analysis [10], we predicted that the largest effect sizes for the association between pubertal timing and internalizing problems would be for age of menarche and timing measured through self-reported Tanner scores. We did not make hypotheses about differences between timing of adrenarcheal and gonadarcheal processes, since no previous studies have compared these. We further expected to see associations with all forms of internalizing psychopathology. Finally, based on the literature to date, we expected both cross-sectional and prospective associations, but had no predictions about the relative strength of each compared to the other.

Methods

Participants

We recruited 174 female adolescents for this longitudinal study, primarily from schools. Inclusion criteria at enrollment included age 10.0-13.0 years; no developmental disability, psychotic disorder, or behavioral disorder; and no current use of psychotropic medication other than stimulants.

We used data from the first two time points (Time 1 and Time 2), which were 18 months apart (M age at Time 1 = 11.63, SD = 0.82; M age at Time 2 = 13.20, SD = 0.84). We administered all measures below at both time points. Full inclusion and exclusion criteria, racial and SES distribution of the sample and further details on the procedure can be found in the protocol paper.[30] We received ethics approval from the Institutional Review Board of the University of Oregon. Parents provided informed consent and adolescents assented to participate.

Measures of Pubertal Timing

Subjective Timing

We used the question in the Pubertal Development Scale[13] that asks about subjective impression of pubertal timing as a measure of subjective timing, both adolescent- and parent-reported: "Do you think your/your child's development is any earlier or later than most other girls your/her age?" This question was not used in the creation of the PDS score described below. This question is answered on a 5-point scale, ranging from "much earlier" to "much later".

Age at Menarche

We asked adolescents at every time point whether they had ever had their period and if yes, to report the date of menarche. Treatment of missing and inconsistent dates are in Supplemental Materials.

Residual-based timing variables

We additionally used the following measures of pubertal development: PDS, Tanner Stage LD, physical maturation composite scores, and hormone levels. These measures are described in detail below. We created timing variables from these by regressing the pubertal development variable linearly on age within each time point (i.e., two separate linear models, as a single linear model across age did not fit the data) and outputting the residuals.

Pubertal Development Scale (PDS)

Participants and parents completed the PDS. This questionnaire consists of five questions regarding the adolescent's secondary sexual characteristics. We converted answers on the self-reported and parent-reported PDS to Tanner stages[14] using validated conversion methods.[33]

Tanner Stage Line Drawings (LD)

The Tanner stage LD,[14] female version, consist of two sets of five drawings depicting breasts and pubic hair. For both sets, adolescents choose the image that most closely reflects their current stage of development. Scores range from 1 (prepubertal) to 5 (postpubertal).

Puberty, Gonadal and Adrenal Composites

We created an overall puberty composite by averaging the PDS and LD Tanner stages, as well as separate gonadal and adrenal composite scores. For these latter composites, we first calculated gonadal and adrenal scores on the PDS. The average of the adrenal PDS score and the lower body LD stage formed the adrenal composite, and the average of the gonadal PDS score and the upper body LD stage formed the gonadal composite.

Hormone Assessment

We asked participants to collect four saliva samples of 2mL at waking, with one week in between samples, and we assayed them for dehydroepiandrosterone (DHEA), testosterone, and estradiol.[30] For further details on the hormone protocol and analysis, refer to Supplemental Material.

Measures of internalizing psychopathology

Depressive symptoms

We measured depressive symptoms with the Center for Epidemiologic Studies Depression Scale for Children (CES-DC).[34,35] The CES-DC is a 20-item self-report measure of depression symptoms over

the past week with responses ranging from 0 (“Not at all”) to 3 (“A lot”), and a total maximum score of 60. The CES-DC has demonstrated excellent internal consistency and concurrent validity with the Children’s Depression Inventory[34] and DSM diagnoses, as well as good discriminant validity.[36]

Anxiety symptoms

Participants filled out the short form of the revised Screen for Child Anxiety Related Disorders (SCARED-R) as a measure of anxiety symptoms. The brief version of the SCARED-R screens for DSM-IV anxiety-related symptomatology through a 5-item multidimensional anxiety scale[37]. Answer options range from 0 (“Not True or Hardly Ever True”) to 2 (“Very True or Often True”). The measure has good internal consistency and concurrent validity.[37]

Diagnoses

Trained interviewers conducted clinical interviews at Time 1 and 2 with participants using the Schedule for Affective Disorders and Schizophrenia for School Aged Children (6–18 Years) Present and Lifetime Version Interview (K-SADS-PL).[38] Details on reliability are in Supplemental Material. Current and past diagnoses of major depressive disorder, dysthymia, adjustment disorder with depressed mood and depression-not otherwise specified based on the DSM-IV were combined in a binary ‘depressive disorder’ variable. We combined current and past diagnoses of generalized anxiety disorder (GAD), social anxiety disorder, separation anxiety disorder, panic disorder, agoraphobia, specific phobia, obsessive-compulsive disorder, post-traumatic stress disorder (PTSD), and anxiety disorder-not otherwise specified based on the DSM-IV in a binary ‘anxiety disorder’ variable. Further, we created an ‘internalizing disorder’ variable, counting everyone with either a depressive disorder, an anxiety disorder or both as having an internalizing disorder. Finally, diagnoses were also categorized using the Hierarchical Taxonomy of Psychopathology (HiTOP) method,[39] which produced additional ‘distress

disorder' (including depressive disorders, GAD and PTSD) and 'fear disorder' (the remaining anxiety disorders) variables.

Control variables

We considered two control variables: Time 1 internalizing psychopathology and early life stress. The Time 1 psychopathology measure always matched the outcome variable (e.g. if CES-DC at Time 2 was the outcome variable, CES-DC at Time 1 was considered as a control variable). As a measure of early life stress (ELS), participants filled out the Childhood Trauma Questionnaire[40] at Time 1. Previous research has demonstrated that the association between ELS and pubertal timing is limited to threat-related ELS.[41] Therefore, we excluded physical and emotional neglect from the total ELS score. To limit the ELS score to early life and before puberty, we only included items endorsed as having occurred before age 7.

Analyses

Data were analyzed in R v3.6.3. Scripts for analysis can be found on Github (DOI: 10.5281/zenodo.4269697).

Imputation

We imputed missing pubertal stage variables, subjective timing variables, psychopathology outcome variables, and control variables using multiple imputation (MI) with Amelia II in R,[42] since we considered these variables to be missing at random.[42] For a table of percentages of missing data by variable and further details on the imputation strategy, refer to Supplemental Material.

Specifications Considered

We considered 12 measures of pubertal timing and 7 measures of internalizing psychopathology, as described in the sections above. Additionally, we considered prospective and cross-sectional associations by including the Time 1 or Time 2 pubertal timing measure, respectively. The exception to this is age at menarche, for which we combined data across all time points. Further, we fit both models with multiply imputed data and complete-case analyses, as a sensitivity analysis due to the parent-reported PDS at Time 1 assumed to be missing not at random. Finally, we considered all possible combinations of the two control variables: no controls, Time 1 psychopathology only, ELS only, or both. This led to a total of 1288 specifications.

Specification Curve

We multiplied residual-based timing variables by -1 to align all pubertal timing variables in the same direction, i.e. higher values represent later timing. We fit linear regression models for continuous outcomes (depressive and anxiety symptoms) and logistic regression models for binary outcomes (diagnoses). All continuous variables were standardized before fitting the regression model. After running all specified models, we ranked them by their regression coefficient and plotted them in a specification curve (Figure 2). The bottom part of the specification curve visualizes how results differ depending on predictor, outcome and analytical decisions. For details on bootstrapping and inferential statistics, refer to Supplemental Material. In the results, a p-value <.05 would indicate that less than 5% of the null-hypothesis datasets had more specifications in the dominant direction, more significant specifications in the dominant direction or more extreme median point estimates, than the original dataset. Tables 2 and 3 show these three inferential statistics and their p-values, split by either the predictor or outcome variable. We adapted code for our analyses from code by Orben and colleagues[43] and code to plot the specification curve from the *specr* package in R3.6.3.

Results

Descriptives and Correlations Between Measures of Pubertal Timing

See Table 1 for descriptive information of the sample and the distribution of pubertal development and internalizing psychopathology at Time 1 and Time 2. There was substantial comorbidity of internalizing disorders: at Time 1, 21% of participants with an internalizing disorder had both an anxiety and a depressive disorder, at Time 2 this was 42%. The overlap between distress and fear disorders was 17% at Time 1 and 32% at Time 2. Figure 1 shows that correlations between the various measures of pubertal timing varied from weak to very strong, and were similar at both time points.

Table 1. Descriptive statistics and change over time in puberty and psychopathology measures

		Time 1 (S.D.)	Time 2 (S.D.)	Change (*p*)
Age		11.63 (0.82)	13.20 (0.84)	<.001
Self-report PDS stage		2.94 (0.98)	3.94 (0.97)	<.001
Parent-report PDS stage		2.74 (1.08)	3.96 (0.93)	<.001
LD Tanner stage		2.83 (0.91)	3.77 (0.74)	<.001
Puberty composite		2.90 (0.89)	3.85 (0.79)	<.001
Gonadal composite		2.95 (0.90)	3.84 (0.84)	<.001
Adrenal composite		2.87 (1.08)	3.86 (0.86)	<.001
Self-report subjective	Much earlier	3.0%	4.4%	.50

timing	Somewhat earlier	21.8%	17.6%	
	About the same	52.1%	57.9%	
	Somewhat later	19.4%	18.9%	
	Much later	3.6%	1.3%	
Parent-report subjective timing	Much earlier	3.9%	2.0%	.03
	Somewhat earlier	19.5%	17.6%	
	About the same	63.6%	64.7%	
	Somewhat later	13.0%	15.0%	
	Much later	0%	0.7%	
Age at menarche		12.38 (1.10)		NA
DHEA (pg/ml)		102.88 (116.23)	125.64 (91.67)	.007
Testosterone (pg/ml)		40.25 (20.95)	67.23 (24.41)	<.001
Estradiol (pg/ml)		0.91 (0.47)	0.97 (0.57)	.39
Depressive symptoms (CES-DC total)		13.19 (10.85)	15.07 (11.53)	.002
Anxiety symptoms (short SCARED-R mean)		0.36 (0.38)	0.38 (0.38)	.64
Internalizing disorder diagnosis		16.67%	30.67%	.002
Depressive disorder diagnosis		5.75%	19.02%	.001
Anxiety disorder diagnosis		14.37%	24.54%	.02
Distress disorder diagnosis		8.62%	22.70%	.001
Fear disorder diagnosis		11.49%	18.40%	.11

Note: descriptives are means with SD between brackets, or percentages per category. Raw hormone levels are presented, prior to log-transformation and correction for confounds. Change over time was tested with Wilcoxon's rank test for paired data (subjective timing variables), McNemar tests (diagnosis variables) or paired t-tests (other variables).

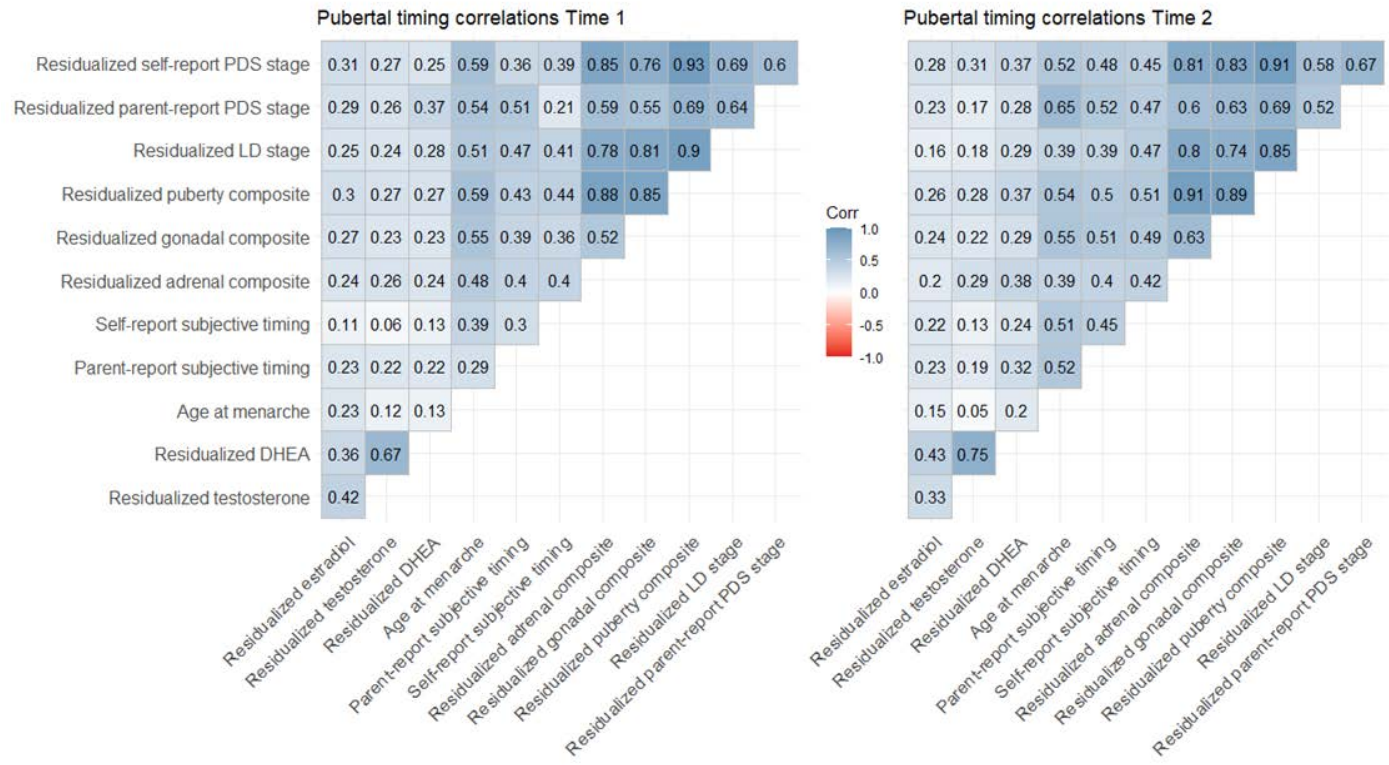

Figure 1. Correlations (Spearman's rho) between measures of pubertal timing at Time 1 and Time 2. Note that residual-based variables have been multiplied by -1 so that for all pubertal timing variables higher values indicate later timing.

Specification Curve and Overall Effects

Figure 2 shows that the associations were overwhelmingly in the negative direction. The strength and significance appeared to vary depending on how pubertal timing is defined. Comparing the observed associations to bootstrapped null models demonstrated that early pubertal timing was significantly associated with more internalizing psychopathology (median point estimate (i.e. regression coefficient) -

0.11, 95% confidence interval -0.13 to -0.10, $p < .001$; share of results in the negative direction 1069/1288, $p < .001$; share of significant results in the negative direction 121/1288, $p = .002$). As seen in Table 2, the strongest associations were found for residualized LD (Tanner stage) and residualized pubertal composite scores (i.e. LD and PDS combined). Table 3 demonstrates that pubertal timing has the strongest association with risk for depressive disorders.

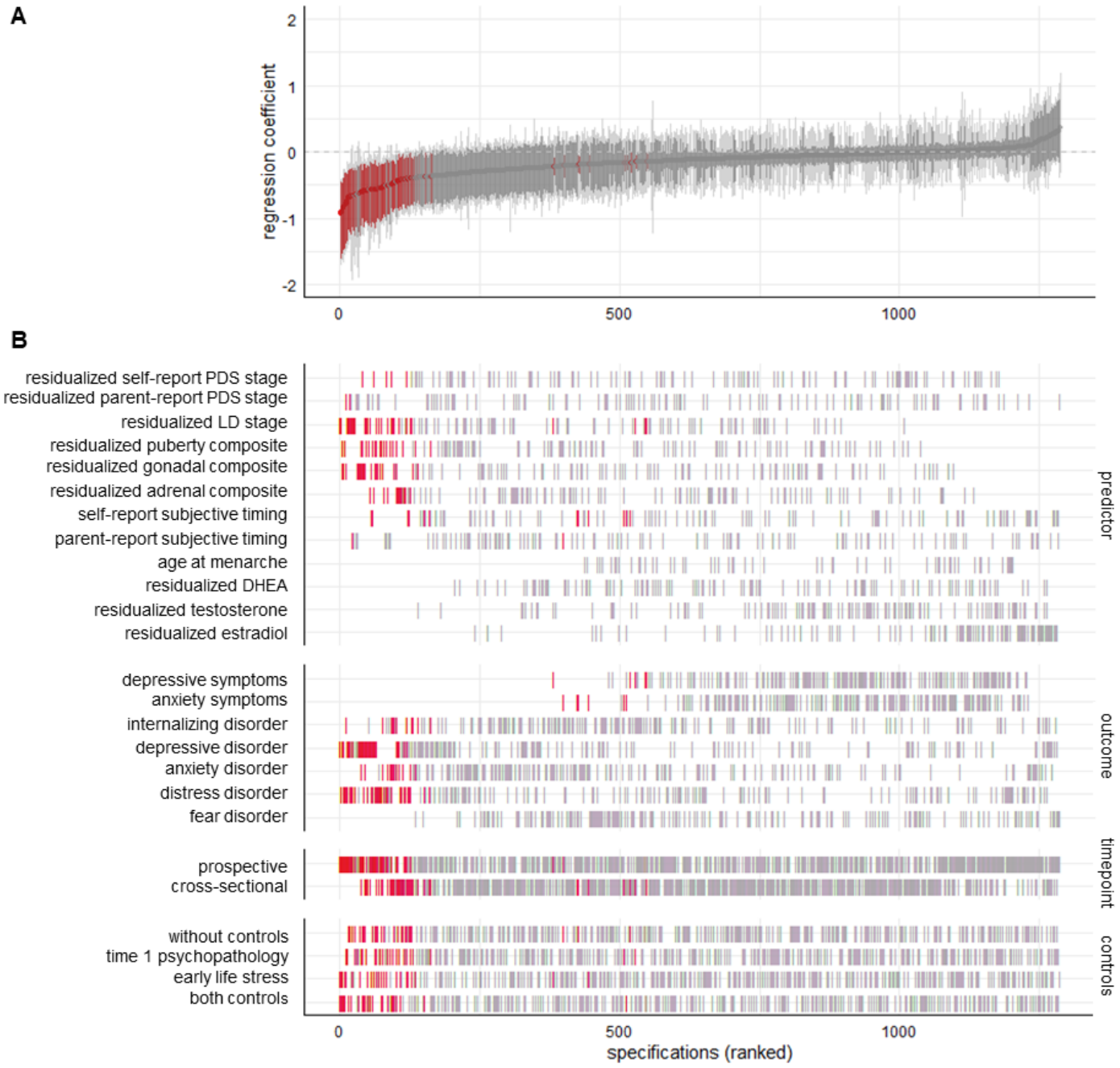

Figure 2. A) Specification curve for the association between pubertal timing and internalizing psychopathology. Each dot represents one specification, with red coloured dots representing significant models ($p<.05$). The red area around each dot is the bootstrapped 95% confidence interval. Specifications are ranked by their regression coefficient. All models were run with standardized data. B) Specifications sorted by predictor, outcome, time point of the predictor, and combination of control variables. Note that residual-based variables have been multiplied by -1 so that for all pubertal timing variables higher values indicate later timing.

Table 2. Inferential statistics for the association between different measures of pubertal timing and internalizing psychopathology.

	Prospective						**Cross-sectional**						**Combined**					
	Median point estimate [CI]		Share of results in negative direction		Share of sign. results in negative direction		Median point estimate [CI]		Share of results in negative direction		Share of sign. results in negative direction		Median point estimate [CI]		Share of results in negative direction		Share of sign. results in negative direction	
	observed	p	share	p	share	p	observed	p	share	p	share	p	observed	p	share	p	share	p
Residualized self-report PDS stage	-0.16 [-0.21; -0.08]	**<.001**	48/66	**<.001**	6	.232	-0.09 [-0.14; -0.06]	**.002**	52/66	**<.001**	0	1	-0.11 [-0.15; -0.07]	**<.001**	100/112	**<.001**	6	.914
Residualized parent-report PDS stage	-0.15 [-0.23; -0.09]	**.002**	48/66	**<.001**	2	.980	-0.12 [-0.15; -0.04]	**<.001**	50/66	**<.001**	0	1	-0.13 [-0.16; -0.08]	**<.001**	98/112	**<.001**	2	1
Residualized LD Tanner stage	-0.28 [-0.36; -0.20]	**<.001**	56/66	**<.001**	20	**<.001**	-0.29 [-0.34; -0.19]	**<.001**	56/66	**<.001**	18	**<.001**	-0.29 [-0.32; -0.21]	**<.001**	112/112	**<.001**	38	**<.001**
Residualized puberty composite	-0.23 [-0.30; -0.15]	**<.001**	56/66	**<.001**	17	**<.001**	-0.19 [-0.26; -0.13]	**<.001**	56/66	**<.001**	5	.448	-0.22 [-0.26; -0.16]	**<.001**	112/112	**<.001**	22	**<.001**
Residualized gonadal composite	-0.21 [-0.29; -0.18]	**<.001**	56/66	**<.001**	17	**<.001**	-0.17 [-0.21; -0.10]	**<.001**	53/66	**<.001**	7	.108	-0.18 [-0.23; -0.14]	**<.001**	109/112	**<.001**	24	**<.001**
Residualized adrenal composite	-0.18 [-0.23; -0.10]	**<.001**	50/66	**<.001**	9	**.014**	-0.19 [-0.22; -0.12]	**<.001**	56/66	**<.001**	6	.236	-0.18 [-0.21; -0.13]	**<.001**	106/112	**<.001**	15	**.008**
Self-report subjective timing	0.01 [-0.08; 0.03]	.384	30/66	.650	2	.994	-0.17 [-0.22; -0.13]	**<.001**	56/66	**<.001**	10	**.014**	-0.11 [-0.15; -0.07]	**<.001**	82/112	**<.001**	12	.162
Parent-report subjective timing	-0.17 [-0.21; -0.10]	**<.001**	45/66	**<.001**	2	.982	-0.18 [-0.20; -0.08]	**<.001**	51/66	**<.001**	0	1	-0.17 [-0.18; -0.10]	**<.001**	96/112	**<.001**	2	1
Age at menarche	NA		NA		NA		NA		NA		NA		-0.07 [-0.09; 0]	**.028**	38/56	**.012**	0	1
Residualized DHEA	-0.15 [-0.17; -.06]	**<.001**	48/66	**<.001**	0	1	-0.06 [-0.11; 0]	**.002**	44/66	**<.001**	0	1	-0.11 [-0.12; -0.04]	**<.001**	92/112	**<.001**	0	1
Residualized testosterone	-0.05 [-0.08; 0.03]	.060	36/66	.056	0	1	-0.03 [-0.08; 0.01]	.170	49/66	**<.001**	0	1	-0.04 [-0.07; 0.01]	**.030**	85/112	**<.001**	0	1
Residualized estradiol	0.08 [0.04; 0.14]	**.002**	53/66	**<.001**	0	1	-0.03 [-0.08; 0.02]	.396	36/66	**.032**	0	1	0.03 [0; 0.07]	.084	73/112	**.002**	0	1
All predictors combined	-0.11 [-0.13;-0.09]	**<.001**	510/616	**<.001**	75	**<.001**	-0.11 [-0.13;-0.10]	**<.001**	597/616	**<.001**	46	.29	-0.11 [-0.13;-0.10]	**<.001**	1069/1288	**<.001**	121	**.002**

Note: DHEA = Dehydroepiandrosterone; CI = confidence interval; LD = Tanner Stage Line Drawings; PDS = Pubertal Development Scale.

Note: "Prospective" statistics are from models of Time 1 pubertal timing and Time 2 outcomes, "cross-sectional" from models of Time 2 pubertal timing and outcomes, and "combined" from all of those models. Statistics are based on bootstrapped null models, see Methods for details.

Table 3. Inferential statistics for the association between pubertal timing and different measures of internalizing psychopathology.

	Median point estimate [confidence interval]		Share of results in negative direction		Share of sign. results in negative direction	
	observed	p	share	p	share	p
Depressive symptoms	-0.04 [-0.06; -0.03]	**<.001**	138/184	**<.001**	5	.580
Anxiety symptoms	-0.04 [-0.06; -0.03]	**<.001**	135/184	**<.001**	6	.280
Internalizing disorder	-0.17 [-0.21; -0.14]	**<.001**	165/184	**<.001**	12	.968
Depressive disorder	-0.29 [-0.34; -0.25]	**<.001**	162/184	**<.001**	43	**<.001**
Anxiety disorder	-0.20 [-0.24; -0.17]	**<.001**	165/184	**<.001**	13	.950
Distress disorder	-0.20 [-0.29; -0.19]	**<.001**	154/184	**<.001**	42	**<.001**
Fear disorder	-0.12 [-0.15; -0.07]	**<.001**	150/184	**<.001**	0	1

Note: Statistics are based on bootstrapped null models, see Methods for details.

Prospective vs. Cross-Sectional Associations

All pubertal timing measures except age at menarche were acquired at both Time 1 and Time 2, which allowed us to examine prospective (Time 1 predictor) and cross-sectional (Time 2 predictor) associations. Bootstrapping the pairwise difference between cross-sectional and prospective models showed that the median point estimate of the association between pubertal timing and internalizing psychopathology was equally strong prospectively as cross-sectionally (-0.12 prospectively and -0.11 cross-sectionally, bootstrapped $p = .82$; note that age at menarche was excluded from this comparison). Table 2 presents the inferential statistics for cross-sectional and prospective models separately. The bottom row shows the share of significant results in the negative direction for prospective models and for cross-sectional models. This share was significantly higher for prospective models (75/616 versus 46/616, bootstrapped $p = .002$).

Relevance of Imputing Missing Values and Including Control Variables

Bootstrapping the pairwise difference between models using imputed and complete-case data demonstrated that imputation did not change the point estimate of the association between pubertal timing and internalizing psychopathology (bootstrapped $p = .44$; models with imputed data had median point estimate = -0.12 and median SE = 0.19; for models with complete data it was -0.11 and 0.21, respectively). The median point estimate was also not dependent on the included control variables (no controls -0.12, Time 1 psychopathology -0.11, ELS -0.11, both -0.11; bootstrapped $p = .57$ for Time 1 psychopathology compared to no controls; bootstrapped $p = .34$ for ELS compared to no controls). The share of significant models in the negative direction was comparable across the different combinations of control variables: 31/322 for no controls, 32/322 for Time 1 psychopathology only, 27/322 for ELS only and 31/322 for both controls.

Discussion

This study applied a specification curve analysis to determine how pubertal timing is cross-sectionally and prospectively associated with internalizing psychopathology across different measurements of pubertal timing and internalizing psychopathology. The association was strongest when pubertal timing was Tanner Stage and the outcome was case-level DSM-IV depression or HiTOP distress disorders. Prospective associations were significantly more often significant than cross-sectional ones. The current results build on a previous meta-analysis,[10] with the main discrepancy that we did not find significant associations between age at menarche and internalizing psychopathology. Importantly, the current study is one of the few that has allowed comparative examination of these relationships within the same sample.

Age at Menarche

The above-mentioned meta-analysis [10] did not report effect sizes for age at menarche in relation to specific outcome categories (e.g. internalizing). Thus, it is possible that their reported effects of age at menarche are due to its association with non-internalizing mental health outcomes. Nevertheless, our results are still in contrast with several studies looking at depression and anxiety specifically that have found significant associations with age at menarche.[9,20,31,44]

The majority of our participants reached menarche within the course of our study, therefore allowing us to limit recall bias as much as possible. In contrast to the majority of previous studies, we examined age at menarche as a continuous variable instead of creating categories of 'early', 'normal', and 'late' timing. We did this to avoid choosing arbitrary cut offs for these categories. We conducted additional exploratory analyses to test a non-linear association between age at menarche and internalizing problems (see Supplemental Materials), but these did not change the pattern of results.

Age at menarche is a rough estimate of pubertal timing based on one milestone, the onset of menstruation. The process of puberty is not a singular event; onset of multiple processes can occur early or late compared to peers. Further, menarche is a late-occurring event typically occurring years after pubertal onset. So, another reason age at menarche may be different from the other metrics is that it mixes pubertal onset and pubertal duration. In contrast, a subjective measure of timing or an assessment of body changes can be measured at any (or multiple) points during the process of puberty. If you consider pubertal timing as where an adolescent is at any point in the process relative to peers, it could for example be "early" compared to peers at one stage, and "on time" compared to peers at another stage later. Therefore, measures of timing that do not rely on a single event may better capture sensitive periods of development (which may be prior to menarche), when timing matters most for the aetiology of mental health disorders.

Adrenarche vs. Gonadarche

Associations were similar for timing of adrenal and gonadal maturation. First, both the residualized adrenal composite and the residualized gonadal composite from self-report measures were significantly associated with internalizing psychopathology, even though these composites were only moderately correlated with each other (see Figure 1). Therefore, in girls aged approximately 10 to 14.5, both adrenarcheal and gonadarcheal processes may contribute to internalizing mental health problems.

Second, timing based on adrenal hormone (DHEA and testosterone) levels or gonadal hormone (estradiol) levels was not related to any measure of symptoms or disorder. Therefore, calculating pubertal timing from hormones may not be a useful method of associating timing with internalizing problems in early-to-mid adolescent girls. Hormone levels relative to age may represent aspects of pubertal timing that do not contribute to the mechanisms that are most relevant to the association between timing and internalizing psychopathology.

Measure of Internalizing Psychopathology

The results also varied by outcome measure of internalizing psychopathology, with the strongest associations for depressive disorders and "distress" disorders (i.e., depression, generalized anxiety and PTSD). These categorical variables were based on a clinical diagnostic interview, demonstrating that associations between early pubertal timing and depressive/distress psychopathology also exist when not solely based on the adolescent's perception. This adds to the results from the meta-analysis[10], where 80% of the studies included only measured symptoms as the outcome.

The median effect sizes of the association between pubertal timing and depressive disorders, as well as distress disorders, were much stronger than between pubertal timing and self-rated depressive

symptoms. Since case-level disorders were based on diagnostic interviews, these findings suggest the association is not simply a result of perceptual bias or self-report bias. They might even suggest that self-report bias obfuscates the association with pubertal timing, or alternatively, that pubertal timing might be most relevant in distinguishing more severe, case-level depression from moderate and low depressive symptoms. However, our results are still inconsistent with other studies that have found associations between pubertal timing and subclinical depressive symptoms.[9,21,22]

The bootstrapped inferential statistics point to no significant association between pubertal timing and anxiety symptoms or disorders (outside of those captured in the HiTOP distress category). This may be due, in part, to the heterogeneity of the anxiety disorder category in the DSM-IV. It is further possible that anxiety disorders and HiTOP fear disorders (phobias, SAD, panic, OCD) are less impacted by pubertal timing as they tend to develop earlier than depression.[45]

Cross-sectional vs. Prospective Associations

Interestingly, associations with psychopathology were more often significant prospectively than cross-sectionally. This could suggest that effects simply take time to emerge. Or, it could be that the timing of the *initial* steps in the pubertal process are particularly salient; therefore, the period of age 10-12 might be a sensitive window for capturing the aspects of pubertal timing that are relevant to internalizing mental health. Our prospective associations were measured over a time span of 18 months during early/mid adolescence, so we cannot draw any conclusions about associations with mental health at later ages.

Moreover, including the equivalent Time 1 psychopathology measure did not eliminate or weaken the results. Few previous studies have controlled for history of psychopathology when examining how early timing is related to internalizing psychopathology, and the two studies that did this

showed conflicting findings.[28,29] Our study explicitly tested the same associations with and without Time 1 psychopathology as a control variable and thus showed that associations between pubertal timing and internalizing psychopathology remained after controlling for this variable. Although our methods do not allow us to infer causality, these findings provide an indication that the likely direction of effect is from pubertal timing to internalizing psychopathology.

Testing Potential Mechanisms

The lack of effects of purely biological (hormonal) measures of timing, combined with the significant results for early timing based on self-reported bodily changes as well as (cross-sectionally) subjective timing, offers support for hypotheses linking early pubertal timing to internalizing psychopathology through social processes. Adolescents with early timing may be perceived as physically different from their peers by other people and/or themselves, and therefore others may treat them differently and/or they may feel negatively about themselves. Future research needs to test mediation models that include specific psychosocial measures, such as self-perception and treatment by others. As an example, in line with our proposed psychosocial mechanisms, the amount of sexual harassment experienced has been shown to mediate the link between early pubertal timing and depressive symptoms in girls.[46] Furthermore, these mechanisms may have an effect on the development of new social and romantic relationships that, in turn, influence mental health according to a recently proposed model[47] so changes in relationship functioning should also be measured.

Longitudinal measurements are especially important for determining mechanisms of the association between psychopathology and subjective timing. The measurement of subjective timing does not tell us if the early-maturing adolescent is simply noticing that they are physically different and not

being cognitively, affectively, or socially ready for the physical changes, or if the adolescent has a negative bias about themself already, due to their risk for depression that will emerge later.

Limitations and Future Directions

Our findings have to be considered in light of several limitations. Firstly, a minority of our participants were still pre-menarcheal and had to be set as missing on our age at menarche variable. However, as mentioned, we conducted additional post-hoc analyses of complete data that did not change the pattern of results. Nevertheless, analysis from longitudinal studies where all participants have completed menarche may uncover additional findings if there is substantial variance in later timing measured this way.

Also, our sample was 66% White, which is more diverse than the local population, but still restricted our ability to explicitly test whether the examined associations hold for all racial/ethnic groups. Since pubertal timing can vary by race/ethnicity,[48] this is an important consideration for future studies.

Further, we found that including threat-related early life stress as a control variable had no impact on the results. However, our sample had low levels of ELS. Future studies with more variability on this measure should continue to test this as a control variable. Finally, genetically-informed studies should be conducted to examine whether pubertal timing and psychopathology have overlapping genetic etiology, since both are partly heritable.[49,50]

Conclusions

The current study is one of the first to comprehensively examine the relationships between a wide range of measures of pubertal timing and internalizing psychopathology within the same sample.

Overall, this study of adolescent girls showed that self-reported and subjective measures of timing were associated with internalizing problems, but age-adjusted hormone levels were not. Furthermore, associations between timing and mental health were strongest for depressive and distress disorders. Future studies should examine mechanisms explaining the link between pubertal timing and internalizing psychopathology that can be targeted in prevention and intervention efforts. For these studies, we suggest that researchers carefully choose the method(s) of measurement for both pubertal timing and mental health. Ultimately, this research will assist clinicians in treating internalizing disorders in adolescent girls by highlighting biological and psychosocial risk factors.

References

1. Ladouceur CD, Peper JS, Crone EA, Dahl RE. White matter development in adolescence: The influence of puberty and implications for affective disorders. *Dev Cogn Neurosci*. 2012;2(1):36-54. doi:10.1016/j.dcn.2011.06.002
2. Paus T, Keshavan M, Giedd JN. Why do many psychiatric disorders emerge during adolescence? *Nat Rev Neurosci*. 2008;9(12):947-957. doi:10.1038/nrn2513
3. Angold A, Costello EJ, Worthman CM. Puberty and depression: the roles of age, pubertal status and pubertal timing. *Psychol Med*. 1998;28(1):51-61.
4. Patton GC, Viner R. Pubertal transitions in health. *Lancet*. 2007;369(9567):1130-1139. doi:10.1016/S0140-6736(07)60366-3
5. Angold A, Costello EJ. Puberty and depression. *Child Adolesc Psychiatr Clin N Am*. 2006;15(4):919-937, ix. doi:10.1016/j.chc.2006.05.013
6. Kaltiala-Heino R, Marttunen M, Rantanen P, Rimpel?? M. Early puberty is associated with mental health problems in middle adolescence. *Soc Sci Med*. 2003;57(6):1055-1064. doi:10.1016/S0277-9536(02)00480-X
7. Mendle J, Harden KP, Brooks-Gunn J, Graber JA. Development's tortoise and hare: pubertal timing, pubertal tempo, and depressive symptoms in boys and girls. *Dev Psychol*. 2010;46(5):1341-1353.
8. Graber JA. Pubertal timing and the development of psychopathology in adolescence and beyond. *Horm Behav*. 2013;64(2):262-269. doi:10.1016/j.yhbeh.2013.04.003
9. Stice E, Presnell K, Bearman SK. Relation of early menarche to depression, eating disorders, substance abuse, and comorbid psychopathology among adolescent girls. *Dev Psychol*. 2001;37(5):608-619. doi:10.1037//0012-1649.37.5.608
10. Ullsperger JM, Nikolas MA. A meta-analytic review of the association between pubertal timing and psychopathology in adolescence: Are there sex differences in risk? *Psychol Bull*. 2017;143(9):903-938. doi:10.1037/bul0000106
11. Rudolph KD. Puberty as a Developmental Context of Risk for Psychopathology. In: Lewis M, Rudolph KD, eds. *Handbook of Developmental Psychopathology*. Springer US; 2014:331-354. doi:10.1007/978-1-4614-9608-3_17
12. Ge X, Natsuaki MN. In Search of Explanations for Early Pubertal Timing Effects on Developmental Psychopathology. *Curr Dir Psychol Sci*. 2009;18(6):327-331. doi:10.1111/j.1467-8721.2009.01661.x
13. Petersen AC, Crockett L, Richards M, Boxer A. A self-report measure of pubertal status: Reliability, validity, and initial norms. *J Youth Adolesc*. 1988;17(2):117-133. doi:10.1007/BF01537962
14. Morris NM, Udry JR. Validation of a self-administered instrument to assess stage of adolescent development. *J Youth Adolesc*. 1980;9(3):271-280. doi:10.1007/BF02088471
15. Counts DR, Pescovitz OH, Barnes KM, et al. Dissociation of Adrenarche and Gonadarche in Precocious Puberty and in Isolated Hypogonadotropic Hypogonadism. *J Clin Endocrinol Metab*. 1987;64(6):1174-1178. doi:10.1210/jcem-64-6-1174
16. Campbell B. Adrenarche and the evolution of human life history. *Am J Hum Biol Off J Hum Biol Counc*. 2006;18(5):569-589. doi:10.1002/ajhb.20528
17. Campbell B. Adrenarche in comparative perspective. *Am J Hum Biol*. 2011;23(1):44-52. doi:10.1002/ajhb.21111
18. Byrne ML, Whittle S, Vijayakumar N, Dennison M, Simmons JG, Allen NB. A systematic review of adrenarche as a sensitive period in neurobiological development and mental health. *Dev Cogn Neurosci*. 2017;25. doi:10.1016/j.dcn.2016.12.004
19. Negriff S, Susman EJ. Pubertal Timing, Depression, and Externalizing Problems: A Framework, Review, and Examination of Gender Differences: PUBERTAL TIMING, DEPRESSION, AND EXTERNALIZING PROBLEMS. *J Res Adolesc*. 2011;21(3):717-746. doi:10.1111/j.1532-7795.2010.00708.x
20. Platt JM, Colich NL, McLaughlin KA, Gary D, Keyes KM. Transdiagnostic psychiatric disorder risk associated

with early age of menarche: A latent modeling approach. *Compr Psychiatry*. 2017;79:70-79. doi:10.1016/j.comppsych.2017.06.010
21. Alloy LB, Hamilton JL, Hamlat EJ, Abramson LY. Pubertal Development, Emotion Regulatory Styles, and the Emergence of Sex Differences in Internalizing Disorders and Symptoms in Adolescence. *Clin Psychol Sci*. 2016;4(5):867-881. doi:10.1177/2167702616643008
22. Graber JA, Lewinsohn PM, Seeley JR, Brooks-Gunn J. Is Psychopathology Associated With the Timing of Pubertal Development? *J Am Acad Child Adolesc Psychiatry*. 1997;36(12):1768-1776. doi:10.1097/00004583-199712000-00026
23. Graber JA, Seeley JR, J. B-G, Lewinsohn PM. Is pubertal timing associated with psychopathology in young adulthood? *J Am Acad Child Adolesc Psychiatry*. 2004;43:718-726.
24. Copeland W, Shanahan L, Miller S, Costello EJ, Angold A, Maughan B. Outcomes of Early Pubertal Timing in Young Women: A Prospective Population-Based Study. *Am J Psychiatry*. 2010;167(10):1218-1225. doi:10.1176/appi.ajp.2010.09081190
25. Marceau K, Ram N, Houts RM, Grimm KJ, Susman EJ. Individual differences in boys' and girls' timing and tempo of puberty: modeling development with nonlinear growth models. *Dev Psychol*. 2011;47(5):1389-1409. doi:10.1037/a0023838
26. Conley CS, Rudolph KD, Bryant FB. Explaining the longitudinal association between puberty and depression: Sex differences in the mediating effects of peer stress. *Dev Psychopathol*. 2012;24(2):691-701. doi:10.1017/S0954579412000259
27. Lee C-T, Tsai M-C, Lin C-Y, Strong C. Longitudinal Effects of Self-Report Pubertal Timing and Menarcheal Age on Adolescent Psychological and Behavioral Outcomes in Female Youths from Northern Taiwan. *Pediatr Neonatol*. 2017;58(4):313-320. doi:10.1016/j.pedneo.2016.04.004
28. Crockett LJ, Carlo G, Wolff JM, Hope MO. The role of pubertal timing and temperamental vulnerability in adolescents' internalizing symptoms. *Dev Psychopathol*. 2013;25(2):377-389. doi:10.1017/S0954579412001125
29. Hamlat EJ, McCormick KC, Young JF, Hankin BL. Early pubertal timing predicts onset and recurrence of depressive episodes in boys and girls. *J Child Psychol Psychiatry*. Published online February 3, 2020. doi:10.1111/jcpp.13198
30. Barendse MEA, Vijayakumar N, Byrne M, et al. Study Protocol: Transitions in Adolescent Girls (TAG). *Front Psychiatry*. 2020;10. doi:10.3389/fpsyt.2019.01018
31. Patton GC, Hibbert ME, Carlin J, et al. Menarche and the onset of depression and anxiety in Victoria, Australia. *J Epidemiol Community Health*. 1996;50(6):661-666.
32. Simonsohn U, Simmons JP, Nelson LD. Specification curve analysis. *Nat Hum Behav*. Published online July 27, 2020. doi:10.1038/s41562-020-0912-z
33. Shirtcliff EA, Dahl RE, Pollak SD. Pubertal development: correspondence between hormonal and physical development. *Child Dev*. 2009;80(2):327-337. doi:10.1111/j.1467-8624.2009.01263.x
34. Faulstich ME, Carey MP, Ruggiero L, Enyart P, Gresham F. Assessment of depression in childhood and adolescence: an evaluation of the Center for Epidemiological Studies Depression Scale for Children (CES-DC). *Am J Psychiatry*. 1986;143(8):1024-1027. doi:10.1176/ajp.143.8.1024
35. Weissman MM, Orvaschel H, Padian N. Children's symptom and social functioning self-report scales. Comparison of mothers' and children's reports. *J Nerv Ment Dis*. 1980;168(12):736-740. doi:10.1097/00005053-198012000-00005
36. Fendrich M, Weissman MM, Warner V. Screening for Depressive disorder in Children and Adolescents: Validating the Center for Epidemiologic Studies Depression Scale for Children. *Am J Epidemiol*. 1990;131(3):538-551. doi:10.1093/oxfordjournals.aje.a115529
37. Birmaher B, Brent DA, Chiappetta L, Bridge J, Monga S, Baugher M. Psychometric Properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A Replication Study. *J Am Acad Child Adolesc Psychiatry*. 1999;38(10):1230-1236. doi:10.1097/00004583-199910000-00011

38. Kaufman J, Birmaher B, Brent D, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial Reliability and Validity Data. *J Am Acad Child Adolesc Psychiatry*. 1997;36(7):980-988. doi:10.1097/00004583-199707000-00021
39. Kotov R, Krueger RF, Watson D, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol*. 2017;126(4):454-477. doi:10.1037/abn0000258
40. Bernstein DP, Stein JA, Newcomb MD, et al. Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse Negl*. 2003;27(2):169-190. doi:10.1016/S0145-2134(02)00541-0
41. Colich NL, Rosen ML, Williams ES, McLaughlin KA. Biological aging in childhood and adolescence following experiences of threat and deprivation: A systematic review and meta-analysis. *Psychol Bull*. Published online August 3, 2020. doi:10.1037/bul0000270
42. Honaker J, King G, Blackwell M. Amelia II: A Program for Missing Data. *J Stat Softw*. 2011;45(7):1-47.
43. Orben A, Przybylski AK. The association between adolescent well-being and digital technology use. *Nat Hum Behav*. 2019;3(2):173-182. doi:10.1038/s41562-018-0506-1
44. Rierdan J, Koff E. Depressive symptomatology among very early maturing girls. *J Youth Adolesc*. 1991;20(4):415-425. doi:10.1007/BF01537183
45. Lijster JM de, Dierckx B, Utens EMWJ, et al. The Age of Onset of Anxiety Disorders: A Meta-analysis. *Can J Psychiatry*. 2017;62(4):237-246. doi:10.1177/0706743716640757
46. Skoog T, Özdemir SB, Stattin H. Understanding the Link Between Pubertal Timing in Girls and the Development of Depressive Symptoms: The Role of Sexual Harassment. *J Youth Adolesc*. 2016;45(2):316-327. doi:10.1007/s10964-015-0292-2
47. Pfeifer JH, Allen NB. Puberty initiates cascading relationships between neurodevelopmental, social, and internalizing processes across adolescence. *Biol Psychiatry*. Published online September 2020. doi:10.1016/j.biopsych.2020.09.002
48. Chumlea WC, Schubert CM, Roche AF, et al. Age at Menarche and Racial Comparisons in US Girls. *PEDIATRICS*. 2003;111(1):110-113. doi:10.1542/peds.111.1.110
49. Meyer JM, Eaves LJ, Heath AC, Martin NG. Estimating genetic influences on the age-at-menarche: A survival analysis approach. *Am J Med Genet*. 1991;39(2):148-154. doi:10.1002/ajmg.1320390207
50. Mustanski BS, Viken RJ, Kaprio J, Pulkkinen L, Rose RJ. Genetic and Environmental Influences on Pubertal Development: Longitudinal Data From Finnish Twins at Ages 11 and 14. *Dev Psychol*. 2004;40(6):1188-1198. doi:10.1037/0012-1649.40.6.1188

Supplementary Material

Date of Menarche Data Handling

To obtain age at menarche from as many participants as possible, we also included data beyond Time 2 (for n = 46 the first report of date of menarche was after Time 2; Time 3 data collection is ongoing and occurs 18+ months after Time 2). If participants reported date of menarche during multiple study time points and dates were inconsistent, we used the first reported date (closest to the actual event). If participants did not remember the exact date, we imputed the middle of the range they reported (e.g. June 2018 became 15 June 2018). Age at menarche was available for 81% of participants, 14% was pre-menarcheal at their latest participation date and the remaining 5% was post-menarche but did not remember or report the date.

Hormone Protocol and Analysis

We assessed hormones in saliva samples four times, one week apart each. This allowed us to obtain a more stable estimate of the hormone level, considering momentary, diurnal and monthly fluctuations. We instructed participants not to eat or brush their teeth before collecting the sample. Families stored the samples in their home freezer until bringing it to their lab session on ice in a cooler bag. At the lab, we stored samples in a -80°C freezer until they were shipped (overnight on dry ice) to the Stress Physiology Investigative Team at the Iowa State University. There they were assayed in duplicate for dehydroepiandrosterone (DHEA), testosterone, and estradiol using Salimetrics Enzyme-Linked Immunosorbent Assay (ELISA) kits. Samples were rerun if the optical density coefficient of variation (CV) was greater than 7% and enough sample was left over to do so. The intra-assay coefficients of variation (CVs) at Time 1 were 10.48 % for dehydroepiandrosterone (DHEA), 1.80 % for

testosterone (T), and 7.76 % for estradiol (E2). The intra-assay CVs at Time 2 were 2.07% for DHEA, 2.89% for T, and 1.84% for E2. We processed the samples in two batches per time point. The interassay CVs at Time 1 for Batch 1 (13 plates) were 20.62 % for DHEA, 10.23 % for T, and 11.53 % for E2, and for Batch 2 (7 plates) were 21.43 % for DHEA, 8.34 % for T, and 15.55 % for E2. The interassay CVs at Time 2 for Batch 1 (15 plates) were 11.9% for DHEA, 7.11% for T, and 17.7% for E2, and for Batch 2 (2 plates) were 5.85% for DHEA, 19.6% for T, and 15.4% for E2. All CVs reported are for the optical density wavelengths. See Barendse et al.[30] for our procedures for handling outliers and undetectable hormone levels.

Hormone levels were log-transformed and they were adjusted for confounds by running mixed effects models predicting the levels of each sample from the time difference between waking and starting collection, whether the sample was collected on a weekday or weekend day, whether the participant felt sick, and use of glucocorticoid sprays/inhalers, contraceptives, and antibiotics/antifungals. We selected these confounds because they predicted the levels of at least one hormone significantly. We fit separate models for both time points and extracted random intercepts for each participant correcting for these confounds.

KSADS Inter-Rater Reliability

Approximately 20% of the interviews were double scored by a second rater, and we calculated inter-rater reliability at the item level, including all screening symptoms and supplemental symptoms if applicable, using the kappa (κ) statistic.[51,52] For the KSADS diagnostic interview, a κ above .77 is considered to be in the "excellent" range.[38] At Time 1, the average κ was .806, and at Time 2 it was .782. At Time 1, the interviewers inquired about current symptoms and lifetime history, and at Time 2, about current symptoms and those occurring *after* Time 1.

Missing Data

We averaged all variables across the 25 multiple imputation datasets. Only age at menarche was not imputed because most missing data was from pre-menarcheal girls (had not experienced onset of menarche by the latest time point in which they participated, N=24) and therefore assumed missing not at random. However, the parent-reported PDS (including parent-reported subjective timing) at Time 1 was not introduced at the very beginning of Time 1 (N = 97 participants were missing Time 1 Parent subjective timing and N = 98 were missing the Time 1 Parent PDS score). Younger girls were recruited later because partway through recruitment we expanded the inclusion criteria to allow both 5th and 6th graders to enroll, as the participants initially recruited from grade 6 were older than expected. Therefore girls who were recruited earlier (and tended to be older) were more likely to be missing this measure at Time 1 only. Because older girls are likely at later pubertal stages, we considered this to be covariate-dependent missingness (i..e, the missingness is dependent on the covariates). Although this assumed missing data mechanism is suitable for multiple imputation methods,[53,54] we nevertheless included a decision point in the SCA to use imputed data or complete-case analysis.

Supplemental Table 1. Percentages of missing data by variable type (Total N = 174)

Variable	**Wave 1 – Missing N (%)**	**Wave 2* – Missing N (%)**
Pubertal timing measures		
PDS stage	12 (6.90%)	15 (8.62%)
Parent-report PDS stage	98 (56.32%)	20 (11.49%)

LD stage	19 (10.92%)	16 (9.20%)
Subjective timing	9 (5.17%)	15 (8.62%)
Parent-report subjective timing	97 (55.75%)	21 (12.07%)
Age at menarche	Total missing across waves: 33 (18.97%)	
Puberty composite	10 (5.75%)	16 (9.20%)
Adrenal composite	8 (4.60%)	16 (9.20%)
Gonadal composite	9 (5.17%)	16 (9.20%)
DHEA level	7 (4.02%)	26 (14.94%)
Testosterone level	7 (4.02%)	26 (14.94%)
Estradiol level	7 (4.02%)	26 (14.94%)
Internalizing measures		
Depressive symptoms (CESDC total)	10 (5.75%)	17 (9.77%)
Anxiety symptoms (short SCARED-R mean)	15 (8.62%)	17 (9.77%)
Depressive disorder diagnosis	0	11 (6.32%)
Anxiety disorder diagnosis	0	11 (6.32%)

Internalizing disorder diagnosis	0	11 (6.32%)
Distress disorder diagnosis	0	11 (6.32%)
Fear disorder diagnosis	0	11 (6.32%)

**This includes missing data due to 11 participants that did not participate in Wave 2.*

Bootstrapping and Inferential Statistics

We performed bootstrapping to examine whether the associations across specifications were significant.[32] To this end, we created datasets in which we knew the null hypothesis was true and examined the median point estimate (median regression coefficient), number of specifications in the dominant direction and number of significant specifications in the dominant direction. To create the datasets in which the null hypothesis was true, we used the method suggested by Simonsohn et al.[32] for continuous outcomes: extract the regression coefficient of the predictor, multiply it by the predictor, and subtract it from the outcome. For binary outcomes, we first calculated probabilities of the outcome with the effect of the pubertal timing predictor set to zero. Subsequently, we generated a binary variable with this probability at every bootstrap sample. We then used the resulting variable as the outcome and ran 500 bootstrapped (with replication) specification-curve analyses with this null-hypothesis data. To obtain a p-value, we divided the number of bootstraps with more (significant) specifications in the dominant direction or more extreme median point estimates than the original dataset by the overall number of bootstraps.

Post-hoc Age at Menarche Analyses

In the main analyses of this manuscript, age at menarche was treated as a continuous variable and linear associations with mental health were tested. This is in contrast to many previous studies of age at menarche and mental health which have dichotomized age at menarche into early and late categories. We conducted *post hoc* analyses to examine whether the discrepancy between our findings and those of previous studies is due to the above-mentioned difference in handling of the age at menarche variable or the model applied. First, we examined associations of continuous age at menarche with depressive symptoms and anxiety symptoms in an exponential model [i.e., log(T2 depressive/anxiety symptoms) = age at menarche + T1 depressive/anxiety symptoms + CTQ threat]. However, neither association was significant. Second, we conducted an analysis where we dichotomized age at menarche into early and average/late categories. We set the cutoff for early timing at 11.73 years based on a large epidemiological study which found that 25% of girls had reached menarche by age 11.73[48]. Missing values were imputed using multiple imputation with Amelia in R and differences between the menarche timing groups were examined with two-sample t-tests for continuous outcomes (symptom levels) and chi-square tests for binary outcomes (diagnoses). Symptoms and diagnoses did not differ between early and average/late menarche girls (all p's >.05).

Supplementary References

51. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37-46.
52. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76:378-382.
53. Hossain A, Diaz-Ordaz K, Bartlett JW. Missing continuous outcomes under covariate dependent missingness in cluster randomised trials. *Stat Methods Med Res*. 2017;26(3):1543-1562. doi:10.1177/0962280216648357
54. Matta TH, Flournoy JC, Byrne ML. Making an unknown unknown a known unknown: Missing data in longitudinal neuroimaging studies. *Dev Cogn Neurosci*. Published online 2017. doi:10.1016/j.dcn.2017.10.001