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

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

[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

Abstract

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. 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 internalizing disorders. From these data, we computed 11 measures of pubertal timing at both time points, as well as 7 measures of internalizing psychopathology, and entered these in a Specification Curve Analysis. Overall, earlier pubertal timing was associated with increased internalizing psychopathology. Associations were stronger prospectively than concurrently, suggesting that timing of early pubertal processes might be especially important for later risk of mental illness. Associations were strongest 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, might be more meaningful determinants of internalizing psychopathology than biological ones in adolescent girls. Future research should precisely examine these psychosocial mechanisms.

Corresponding author: Marjolein Barendse, 1444 Franklin Blvd, 97403 Eugene, OR, USA, barendse@uoregon.edu.
[*]Equal contributors. Note: family names are underlined

General Scientific Summary:

Early pubertal timing is a risk factor for the development of anxiety and depression in adolescence. The findings of this paper show that this link depends on how pubertal timing is measured and that timing of early pubertal processes might be especially important for *later* risk of mental illness in adolescent girls. It suggests that psychosocial mechanisms, captured by timing measures of visible physical characteristics, might be more meaningful predictors of depression and anxiety than hormonal ones.

Keywords

Introduction

Adolescence is a sensitive period of life for neurobiological development and risk for psychopathology (Ladouceur et al., 2012; Paus et al., 2008). Girls are two to three times more likely to experience depression than boys from puberty onwards (Angold et al., 1998). The substantial changes in social, physical, and hormonal development that occur during pubertal development can be related to mental health outcomes (Patton & Viner, 2007). 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 (Angold & Costello, 2006; Kaltiala-Heino et al., 2003; Mendle et al., 2010). 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 (Graber, 2013), although some studies have found that this effect is small (Stice et al., 2001) or not statistically significant (Angold et al., 1998). A recent meta-analysis (Ullsperger & Nikolas, 2017) 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 (Rudolph, 2014). Biological processes include sensitivity of the brain to pubertal hormones, for example (Ge & Natsuaki, 2009). 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 (Petersen et al., 1988)) or the Tanner Staging Line Drawings (LD (Morris & Udry, 1980))) 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 (Counts et al., 1987). Compared to gonadarche, there is much less research on adrenarcheal processes predicting

psychopathology. Adrenarcheal processes include increases in hormones produced in the adrenal gland (DHEA and, in girls, testosterone), whereas gonadarche is driven by gonadal hormones like estradiol. Animal and human studies have shown that adrenal hormones can influence brain function and development through a range of mechanisms, including antiglucocorticoid effects of DHEA (Campbell, 2011; Maninger et al., 2009), but the direct evidence for associations between timing of adrenacheal processes of puberty and internalizing psychopathology is inconsistent (Byrne et al., 2017).

Differences in measurement or definition of internalizing psychopathology may also contribute to inconsistencies in associations with pubertal timing (for review, see (Negriff & Susman, 2011)). The aforementioned meta-analysis found a significant association between pubertal timing and both "distress" and "fear" psychopathology (Ullsperger & Nikolas, 2017). 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 (Alloy et al., 2016; Graber et al., 1997; Platt et al., 2017).

Critically, that same meta-analysis (Ullsperger & Nikolas, 2017) 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 (Conley et al., 2012; Copeland et al., 2010; Graber et al., 2004; Marceau et al., 2011) have been prospectively associated with internalizing psychopathology in later adolescence and sometimes through young adulthood, although not always (Lee et al., 2017). There is also conflicting evidence whether early timing is related to internalizing psychopathology when controlling for history of psychopathology (Crockett et al., 2013; Hamlat et al., 2020). This has implications for identifying the best time window for prevention and early intervention efforts that mitigate the risk of internalizing psychopathology.

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 strength 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 (Barendse et al., 2020). 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 (Angold et al., 1998; Patton et al., 1996). 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 (Simonsohn et al., 2020). SCA thereby prevents selective reporting and p-hacking in the context of multiple, justifiable measurement and analytic approaches, and it has built-in (usually bootstrapping) approaches to handle multiple comparisons problems. 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 BMI, 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 (Ullsperger & Nikolas, 2017), 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, including adrenal and gonadal hormones, 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; fluent in English; no developmental disability or autism, psychotic disorder, or behavioral disorder (ODD/CD); 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; M timespan =1.57 years, SD = 0.12 years). We administered all measures below at both time points. 66% of participants were White and out of the remaining participants, most were multiracial or Latina/Hispanic. 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 (Barendse et al., 2020). We received ethics approval from the Institutional Review Board of the University of Oregon (protocol # 03232015.027). Parents provided informed consent and adolescents assented to participate.

Measures of Pubertal Timing

Subjective Timing—We used the question in the Pubertal Development Scale (Petersen et al., 1988) 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. 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.

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. In the remainder of the paper, we refer to these timing variables as "residualized [name of pubertal development measure]", e.g. residualized self-report PDS stage.

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 (Morris & Udry, 1980) using validated conversion methods (Shirtcliff et al., 2009).

Tanner Stage Line Drawings (LD)—The Tanner stage LD (Morris & Udry, 1980), 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).

Gonadal and Adrenal Composites—For the gonadal and adrenal 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. 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. (2020) 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, to obtain one basal hormone level per person per time point.

Measures of internalizing psychopathology

Depressive symptoms—We measured depressive symptoms with the Center for Epidemiologic Studies Depression Scale for Children (CES-DC) (Faulstich et al., 1986; Weissman et al., 1980). 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 (Faulstich et al., 1986) and DSM diagnoses, as well as good discriminant validity (Fendrich et al., 1990).

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 (Birmaher et al., 1999). 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 (Birmaher et al., 1999).

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) (Kaufman et al., 1997). 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 (Cohen, 1960; Fleiss, 1971). At Time 1, the average κ was .806, and at Time 2 it was .782, which are considered to be in the "excellent" range (Kaufman et al., 1997). Diagnoses were determined by the trained interviewers in consultation with a clinician-researcher and the whole process was monitored by a professor in developmental clinical psychology (author NBA). 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.

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 (Kotov et al., 2017), which produced additional 'distress disorder' (including depressive disorders, GAD and PTSD) and 'fear disorder' (the remaining anxiety disorders) variables.

Control variables

We considered three control variables: Time 1 internalizing psychopathology, early life stress and Time 1 BMI. 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 (Bernstein et al., 2003) at Time 1. Previous research has demonstrated that the association between ELS and pubertal timing is limited to threat-related ELS (Colich et al., 2020). 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. We included BMI as another control variable because it likely leads to earlier timing of puberty (Chen et al., 2019) and BMI is positively associated with risk for internalizing psychopathology (Ames et al., 2015). BMI was calculated from experimenter-measured height and weight (for detailed procedures, see Barendse et al., 2020) and converted to age-and-sex-specific z-scores based on the 2000 CDC growth charts.

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 (Honaker et al., 2011), since we considered these variables to be missing at random (Honaker et al., 2011). For a table of percentages of missing data and details on imputation, see Supplemental Material.

Specifications Considered—We considered 11 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, since this occurs only once. 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 control variables. This led to a total of 2352 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.

Bootstrapping and Inferential Statistics—We performed bootstrapping to examine whether the associations across specifications were significant (Simonsohn et al., 2020). To this end, we created datasets in which we knew the null hypothesis was true using the method suggested by Simonsohn et al (2020) 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 1000 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. 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 Przybylski (2019) 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. Total CES-DC scores ranged from 0 to 49 at Time 1 and 0 to 57 at Time 2 (the screening cut-off is 15 (Fendrich et al., 1990)). Mean SCARED-R scores ranged from 0 to 2 at both time points (the screening cut-off is a mean score of 0.6 (Birmaher et al., 1999)). 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.

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) −0.10, 95% confidence interval −0.11 to −0.09, $p < .001$; share of results in the negative direction 1827/2352, $p < .001$; share of significant results in the negative direction 193/2352, $p = <.001$). As seen in Table 2, the strongest associations were found for residualized LD (Tanner stage) and residualized gonadal composite scores (i.e. LD and PDS combined). Table 3 demonstrates that pubertal timing has the strongest association with risk for depressive disorders.

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 (both −0.10, bootstrapped $p = .94$; 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 (120/1120 versus 73/1120, bootstrapped p <.001).

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 $p = .61$; models with imputed data had median point estimate $= -0.10$ and median SE $= 0.19$; for models with complete data it was -0.10 and 0.21, respectively). The median point estimate was somewhat weakened by including BMI as a control variable, but not by the other controls (bootstrapped $p = .01$ for specifications with BMI compared to those without; bootstrapped $p = .52$ for with versus without Time 1 psychopathology; bootstrapped $p = .47$ for with versus without ELS). The share of significant models in the negative direction was 30/294 for no controls and 25/294 for all controls.

Discussion

This study applied a novel approach, 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. Importantly, the current study is one of the few that has allowed comparative examination of these relationships within the same sample, allowing us to both draw an overall conclusion about the association of timing with internalizing psychopathology (regardless of measure), as well as identify which measurement and analysis decisions were impactful.

Cross-sectional vs. Prospective Associations

One of the most valuable contributions of both this dataset and statistical approach is the revelation that 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 occur early in pubertal development and are relevant to internalizing mental health. However, 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. Furthermore, we were unable to examine the effect of pubertal timing on non-linear trajectories of internalizing symptoms because we only had two time points.

Moreover, including the equivalent Time 1 psychopathology measure did not eliminate or weaken the results. Few previous studies have controlled for the history of psychopathology when examining how early timing is related to internalizing psychopathology, and the two studies that did this showed conflicting findings (Crockett et al., 2013; Hamlat et al.,

2020). Our specification curve analysis included a decision point of whether or not to control for Time 1 psychopathology and thus showed that associations between pubertal timing and internalizing psychopathology remained after controlling for this variable. It is possible, therefore, that the direction of effect is pubertal timing predicting later internalizing psychopathology, but our methods do not allow us to infer causality. Future studies should consider examining whether psychopathology earlier in life may predict pubertal timing.

Age at Menarche

In contrast with several studies looking at age at menarche in relation to depression and anxiety, we did not find significant associations with age at menarche (Patton et al., 1996; Platt et al., 2017; Rierdan & Koff, 1991; Stice et al., 2001). 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 capture individual variation in timing of earlier maturational milestones (prior to menarche), may matter 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 prospectively 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, despite using best-practice methods for assessing hormone levels. 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 significantly improves upon what could be inferred from the meta-analysis (Ullsperger & Nikolas, 2017), 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 (Alloy et al., 2016; Graber et al., 1997; Stice et al., 2001).

The bootstrapped inferential statistics point to no significant association between pubertal timing and anxiety 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 (Lijster et al., 2017).

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 offers indirect 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. We acknowledge that we have not directly tested mechanisms, and therefore recommend future research to test mediation models that include specific psychosocial measures, such as self-perception and treatment by others (i.e., adultification). 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 (Skoog et al., 2016). 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 (Pfeifer & Allen, 2020) 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 themselves 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, 14% 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 did not give us enough power to test whether the examined associations hold for specific racial/ethnic groups. Pubertal development occurs, on average, earlier for Black/African American adolescents (Chumlea et al., 2003; Herting et al., 2021), and there are inconsistent findings as to whether the link between early timing and internalizing psychopathology is similar for Black/African American adolescents compared to White adolescents or those of other races (Carter et al., 2013; Deardorff et al., 2021). Therefore, examining associations between pubertal timing and mental health within each racial/ethnic group is an important future direction that can be addressed using large, representative samples (such as the ABCD study for the US).

Further, we found that including threat-related early life stress (before age 7) 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 (Meyer et al., 1991; Mustanski et al., 2004).

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 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, and were more often significant prospectively than cross-sectionally. 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Alloy LB, Hamilton JL, Hamlat EJ, & Abramson LY (2016). Pubertal Development, Emotion Regulatory Styles, and the Emergence of Sex Differences in Internalizing Disorders and Symptoms in Adolescence. Clinical Psychological Science, 4(5), 867–881. 10.1177/2167702616643008 [PubMed: 27747141]

Ames ME, Wintre MG, & Flora DB (2015). Trajectories of BMI and internalizing symptoms: Associations across adolescence. Journal of Adolescence, 45, 80–88. 10.1016/j.adolescence.2015.08.016 [PubMed: 26397879]

Angold A, & Costello EJ (2006). Puberty and depression. Child and Adolescent Psychiatric Clinics of North America, 15(4), 919–937, ix. 10.1016/j.chc.2006.05.013 [PubMed: 16952768]

Angold A, Costello EJ, & Worthman CM (1998). Puberty and depression: The roles of age, pubertal status and pubertal timing. Psychological Medicine, 28(1), 51–61. [PubMed: 9483683]

Barendse MEA, Vijayakumar N, Byrne M, Flannery J, Cheng T, Flournoy J, Nelson B, Cosme D, Mobasser A, Chavez S, Hval L, Brady B, Nadel H, Helzer A, Shirtcliff EA, Allen N, & Pfeifer JH (2020). Study Protocol: Transitions in Adolescent Girls (TAG). Frontiers in Psychiatry, 10. 10.3389/fpsyt.2019.01018

Bernstein DP, Stein JA, Newcomb MD, Walker E, Pogge D, Ahluvalia T, Stokes J, Handelsman L, Medrano M, Desmond D, & Zule W (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. Child Abuse & Neglect, 27(2), 169–190. 10.1016/S0145-2134(02)00541-0 [PubMed: 12615092]

Birmaher B, Brent DA, Chiappetta L, Bridge J, Monga S, & Baugher M (1999). Psychometric Properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A Replication Study. Journal of the American Academy of Child & Adolescent Psychiatry, 38(10), 1230–1236. 10.1097/00004583-199910000-00011 [PubMed: 10517055]

Byrne ML, Whittle S, Vijayakumar N, Dennison M, Simmons JG, & Allen NB (2017). A systematic review of adrenarche as a sensitive period in neurobiological development and mental health. Developmental Cognitive Neuroscience, 25. 10.1016/j.dcn.2016.12.004

Campbell B (2011). Adrenarche in comparative perspective. American Journal of Human Biology, 23(1), 44–52. 10.1002/ajhb.21111 [PubMed: 21140467]

Carter R, Silverman WK, & Jaccard J (2013). Race and Perceived Pubertal Transition Effects on Girls' Depressive Symptoms and Delinquent Behaviors. Journal of Youth and Adolescence, 42(8), 1155–1168. 10.1007/s10964-012-9885-1 [PubMed: 23275063]

Chen Y-C, Fan H-Y, Yang C, Hsieh R-H, Pan W-H, & Lee YL (2019). Assessing causality between childhood adiposity and early puberty: A bidirectional Mendelian randomization and longitudinal study. Metabolism - Clinical and Experimental, 100. 10.1016/j.metabol.2019.153961 [PubMed: 30690037]

Chumlea WC, Schubert CM, Roche AF, Kulin HE, Lee PA, Himes JH, & Sun SS (2003). Age at Menarche and Racial Comparisons in US Girls. PEDIATRICS, 111(1), 110–113. 10.1542/peds.111.1.110 [PubMed: 12509562]

Cohen J (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37–46.

Colich NL, Rosen ML, Williams ES, & McLaughlin KA (2020). Biological aging in childhood and adolescence following experiences of threat and deprivation: A systematic review and meta-analysis. Psychological Bulletin. 10.1037/bul0000270

Conley CS, Rudolph KD, & Bryant FB (2012). Explaining the longitudinal association between puberty and depression: Sex differences in the mediating effects of peer stress. Development and Psychopathology, 24(2), 691–701. 10.1017/S0954579412000259 [PubMed: 22559140]

Copeland W, Shanahan L, Miller S, Costello EJ, Angold A, & Maughan B (2010). Outcomes of Early Pubertal Timing in Young Women: A Prospective Population-Based Study. American Journal of Psychiatry, 167(10), 1218–1225. 10.1176/appi.ajp.2010.09081190 [PubMed: 20478880]

Counts DR, Pescovitz OH, Barnes KM, Hench KD, Chrousos GP, Sherins RJ, Comite F, Loriaux DL, & Cutler GB (1987). Dissociation of Adrenarche and Gonadarche in Precocious Puberty and in Isolated Hypogonadotropic Hypogonadism. The Journal of Clinical Endocrinology & Metabolism, 64(6), 1174–1178. 10.1210/jcem-64-6-1174 [PubMed: 3571422]

Crockett LJ, Carlo G, Wolff JM, & Hope MO (2013). The role of pubertal timing and temperamental vulnerability in adolescents' internalizing symptoms. Development and Psychopathology, 25(2), 377–389. 10.1017/S0954579412001125 [PubMed: 23627951]

Deardorff J, Marceau K, Johnson M, Reeves JW, Biro FM, Kubo A, Greenspan LC, Laurent CA, Windham GC, Pinney SM, Kushi LH, & Hiatt RA (2021). Girls' Pubertal Timing and Tempo and Mental Health: A Longitudinal Examination in an Ethnically Diverse Sample. Journal of Adolescent Health, 68(6), 1197–1203. 10.1016/j.jadohealth.2021.01.020

Faulstich ME, Carey MP, Ruggiero L, Enyart P, & Gresham F (1986). Assessment of depression in childhood and adolescence: An evaluation of the Center for Epidemiological Studies Depression Scale for Children (CES-DC). The American Journal of Psychiatry, 143(8), 1024–1027. 10.1176/ajp.143.8.1024 [PubMed: 3728717]

Fendrich M, Weissman MM, & Warner V (1990). Screening for Depressive disorder in Children and Adolescents: Validating the Center for Epidemiologic Studies Depression Scale for Children. American Journal of Epidemiology, 131(3), 538–551. 10.1093/oxfordjournals.aje.a115529 [PubMed: 2301363]

Fleiss JL (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76, 378–382.

Ge X, & Natsuaki MN (2009). In Search of Explanations for Early Pubertal Timing Effects on Developmental Psychopathology. Current Directions in Psychological Science, 18(6), 327–331. 10.1111/j.1467-8721.2009.01661.x

Graber JA (2013). Pubertal timing and the development of psychopathology in adolescence and beyond. Hormones and Behavior, 64(2), 262–269. 10.1016/j.yhbeh.2013.04.003 [PubMed: 23998670]

Graber JA, Lewinsohn PM, Seeley JR, & Brooks-Gunn J (1997). Is Psychopathology Associated With the Timing of Pubertal Development? Journal of the American Academy of Child & Adolescent Psychiatry, 36(12), 1768–1776. 10.1097/00004583-199712000-00026 [PubMed: 9401339]

Graber JA, Seeley JR, Brooks-Gunn J, & Lewinsohn PM (2004). Is Pubertal Timing Associated With Psychopathology in Young Adulthood? Journal of the American Academy of Child & Adolescent Psychiatry, 43(6), 718–726. 10.1097/01.chi.0000120022.14101.11 [PubMed: 15167088]

Hamlat EJ, McCormick KC, Young JF, & Hankin BL (2020). Early pubertal timing predicts onset and recurrence of depressive episodes in boys and girls. Journal of Child Psychology and Psychiatry. 10.1111/jcpp.13198

Herting MM, Uban KA, Gonzalez MR, Baker FC, Kan EC, Thompson WK, Granger DA, Albaugh MD, Anokhin AP, Bagot KS, Banich MT, Barch DM, Baskin-Sommers A, Breslin FJ, Casey BJ, Chaarani B, Chang L, Clark DB, Cloak CC, … Sowell ER. (2021). Correspondence Between Perceived Pubertal Development and Hormone Levels in 9-10 Year-Olds From the Adolescent Brain Cognitive Development Study. Frontiers in Endocrinology, 11. 10.3389/fendo.2020.549928

Honaker J, King G, & Blackwell M (2011). Amelia II: A Program for Missing Data. Journal of Statistical Software, 45(7), 1–47.

Hossain A, Diaz-Ordaz K, & Bartlett JW (2017). Missing continuous outcomes under covariate dependent missingness in cluster randomised trials. Statistical Methods in Medical Research, 26(3), 1543–1562. 10.1177/0962280216648357 [PubMed: 27177885]

Kaltiala-Heino R, Marttunen M, Rantanen P, & Rimpel?? M. (2003). Early puberty is associated with mental health problems in middle adolescence. Social Science and Medicine, 57(6), 1055–1064. 10.1016/S0277-9536(02)00480-X [PubMed: 12878105]

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, & Ryan N (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial Reliability and Validity Data. Journal of the American Academy of Child & Adolescent Psychiatry, 36(7), 980–988. 10.1097/00004583-199707000-00021 [PubMed: 9204677]

Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, Brown TA, Carpenter WT, Caspi A, Clark LA, Eaton NR, Forbes MK, Forbush KT, Goldberg D, Hasin D, Hyman SE, Ivanova MY, Lynam DR, Markon K, … Zimmerman M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. Journal of Abnormal Psychology, 126(4), 454–477. 10.1037/abn0000258 [PubMed: 28333488]

Ladouceur CD, Peper JS, Crone EA, & Dahl RE (2012). White matter development in adolescence: The influence of puberty and implications for affective disorders. Developmental Cognitive Neuroscience, 2(1), 36–54. 10.1016/j.dcn.2011.06.002 [PubMed: 22247751]

Lee C-T, Tsai M-C, Lin C-Y, & Strong C (2017). Longitudinal Effects of Self-Report Pubertal Timing and Menarcheal Age on Adolescent Psychological and Behavioral Outcomes in Female Youths from Northern Taiwan. Pediatrics & Neonatology, 58(4), 313–320. 10.1016/j.pedneo.2016.04.004 [PubMed: 27600751]

Lijster J. M. de, Dierckx B, Utens EMWJ, Verhulst FC, Zieldorff C, Dieleman GC, & Legerstee JS (2017). The Age of Onset of Anxiety Disorders: A Meta-analysis. The Canadian Journal of Psychiatry, 62(4), 237–246. 10.1177/0706743716640757 [PubMed: 27310233]

Maninger N, Wolkowitz OM, Reus VI, Epel ES, & Mellon SH (2009). Neurobiological and neuropsychiatric effects of dehydroepiandrosterone (DHEA) and DHEA sulfate (DHEAS). Frontiers in Neuroendocrinology, 30(1), 65–91. 10.1016/j.yfrne.2008.11.002 [PubMed: 19063914]

Marceau K, Ram N, Houts RM, Grimm KJ, & Susman EJ (2011). Individual differences in boys' and girls' timing and tempo of puberty: Modeling development with nonlinear growth models. Developmental Psychology, 47(5), 1389–1409. 10.1037/a0023838 [PubMed: 21639623]

Matta TH, Flournoy JC, & Byrne ML (2017). Making an unknown unknown a known unknown: Missing data in longitudinal neuroimaging studies. Developmental Cognitive Neuroscience. 10.1016/j.dcn.2017.10.001

Mendle J, Harden KP, Brooks-Gunn J, & Graber JA (2010). Development's tortoise and hare: Pubertal timing, pubertal tempo, and depressive symptoms in boys and girls. Developmental Psychology, 46(5), 1341–1353. [PubMed: 20822243]

Meyer JM, Eaves LJ, Heath AC, & Martin NG (1991). Estimating genetic influences on the age-at-menarche: A survival analysis approach. American Journal of Medical Genetics, 39(2), 148–154. 10.1002/ajmg.1320390207 [PubMed: 2063916]

Morris NM, & Udry JR (1980). Validation of a self-administered instrument to assess stage of adolescent development. Journal of Youth and Adolescence, 9(3), 271–280. 10.1007/BF02088471 [PubMed: 24318082]

Mustanski BS, Viken RJ, Kaprio J, Pulkkinen L, & Rose RJ (2004). Genetic and Environmental Influences on Pubertal Development: Longitudinal Data From Finnish Twins at Ages 11 and 14. Developmental Psychology, 40(6), 1188–1198. 10.1037/0012-1649.40.6.1188 [PubMed: 15535766]

Negriff S, & Susman EJ (2011). Pubertal Timing, Depression, and Externalizing Problems: A Framework, Review, and Examination of Gender Differences: PUBERTAL TIMING, DEPRESSION, AND EXTERNALIZING PROBLEMS. Journal of Research on Adolescence, 21(3), 717–746. 10.1111/j.1532-7795.2010.00708.x

Orben A, & Przybylski AK (2019). The association between adolescent well-being and digital technology use. Nature Human Behaviour, 3(2), 173–182. 10.1038/s41562-018-0506-1

Patton GC, Hibbert ME, Carlin J, Shao Q, Rosier M, Caust J, & Bowes G (1996). Menarche and the onset of depression and anxiety in Victoria, Australia. Journal of Epidemiology & Community Health, 50(6), 661–666. [PubMed: 9039386]

Patton GC, & Viner R (2007). Pubertal transitions in health. Lancet, 369(9567), 1130–1139. 10.1016/S0140-6736(07)60366-3 [PubMed: 17398312]

Paus T, Keshavan M, & Giedd JN (2008). Why do many psychiatric disorders emerge during adolescence? Nature Reviews Neuroscience, 9(12), 947–957. 10.1038/nrn2513 [PubMed: 19002191]

Petersen AC, Crockett L, Richards M, & Boxer A (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. Journal of Youth and Adolescence, 17(2), 117–133. 10.1007/BF01537962 [PubMed: 24277579]

Pfeifer JH, & Allen NB (2020). Puberty initiates cascading relationships between neurodevelopmental, social, and internalizing processes across adolescence. Biological Psychiatry. 10.1016/j.biopsych.2020.09.002

Platt JM, Colich NL, McLaughlin KA, Gary D, & Keyes KM (2017). Transdiagnostic psychiatric disorder risk associated with early age of menarche: A latent modeling approach. Comprehensive Psychiatry, 79, 70–79. 10.1016/j.comppsych.2017.06.010 [PubMed: 28757148]

Rierdan J, & Koff E (1991). Depressive symptomatology among very early maturing girls. Journal of Youth and Adolescence, 20(4), 415–425. 10.1007/BF01537183 [PubMed: 24263398]

Rudolph KD (2014). Puberty as a Developmental Context of Risk for Psychopathology. In Lewis M & Rudolph KD (Eds.), Handbook of Developmental Psychopathology (pp. 331–354). Springer US. 10.1007/978-1-4614-9608-3_17

Shirtcliff EA, Dahl RE, & Pollak SD (2009). Pubertal development: Correspondence between hormonal and physical development. Child Development, 80(2), 327–337. 10.1111/j.1467-8624.2009.01263.x [PubMed: 19466995]

Simonsohn U, Simmons JP, & Nelson LD (2020). Specification curve analysis. Nature Human Behaviour. 10.1038/s41562-020-0912-z

Skoog T, özdemir SB, & Stattin H (2016). Understanding the Link Between Pubertal Timing in Girls and the Development of Depressive Symptoms: The Role of Sexual Harassment, Journal of Youth and Adolescence, 45(2) 316–327. 10.1007/sl0964-015-0292-2 [PubMed: 25971216]

Stice E, Presnell K, & Bearman SK (2001). Relation of early menarche to depression, eating disorders, substance abuse, and comorbid psychopathology among adolescent girls. Developmental Psychology, 37(5), 608–619. 10.1037//0012-1649.37.5.608 [PubMed: 11552757]

Ullsperger JM, & Nikolas MA (2017). A meta-analytic review of the association between pubertal timing and psychopathology in adolescence: Are there sex differences in risk? Psychological Bulletin, 143(9), 903–938. 10.1037/bul0000106 [PubMed: 28530427]

Weissman MM, Orvaschel H, & Padian N (1980). Children's symptom and social functioning self-report scales. Comparison of mothers' and children's reports. The Journal of Nervous and Mental Disease, 168(12), 736–740. 10.1097/00005053-198012000-00005 [PubMed: 7452212]
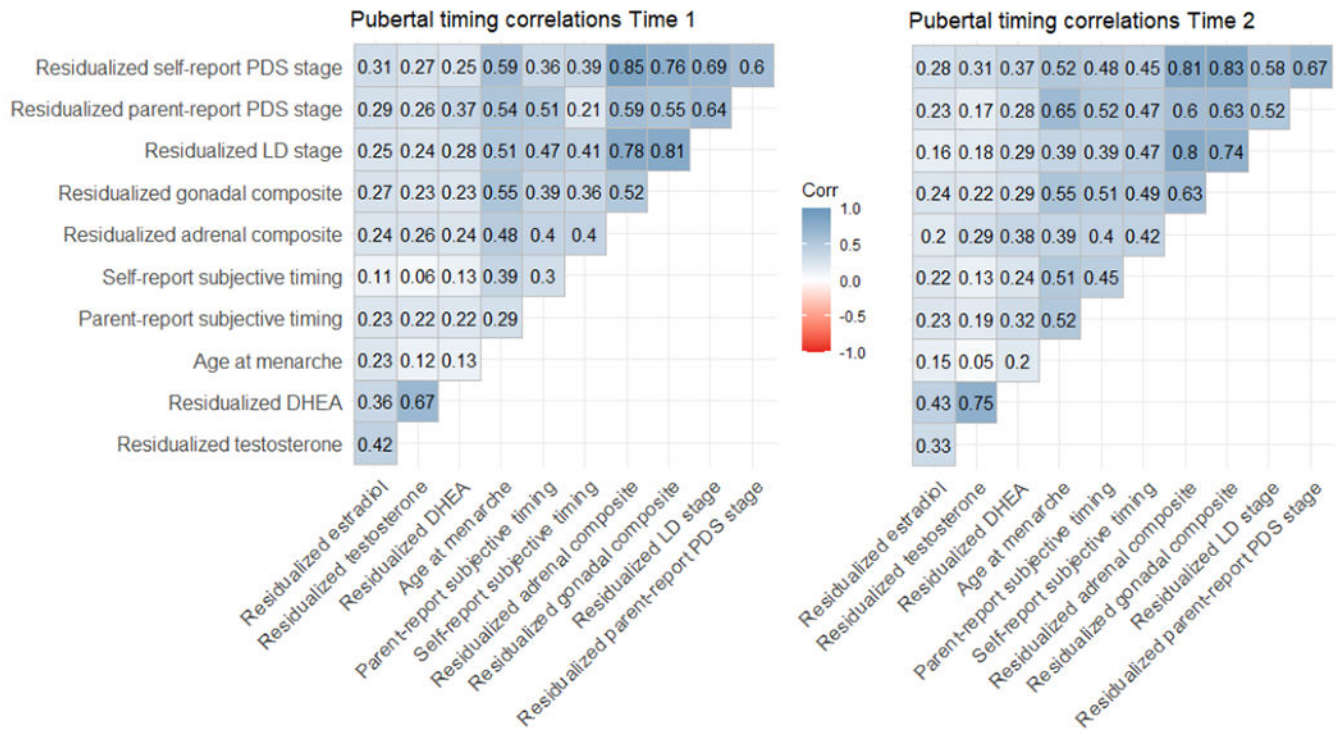
Figure 1.
Correlations (Spearman's rho) between measures of pubertal timing at Time 1 and Time 2. Note that residual-based variables have been multiplied by −1 so that for all pubertal timing variables higher values indicate later timing.
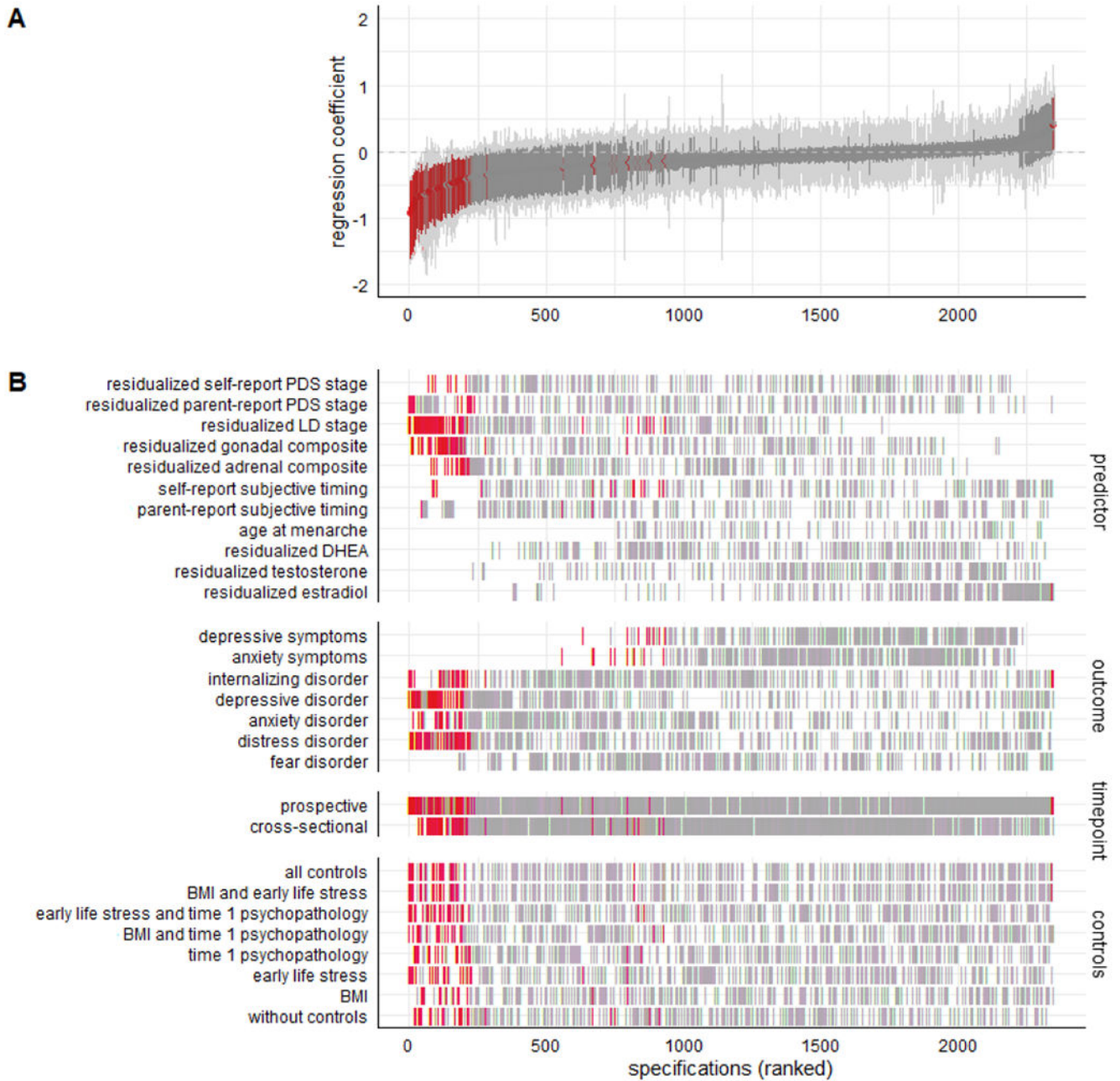
Figure 2.
A) Specification curve for the association between pubertal timing and internalizing psychopathology. Each dot represents one specification, with red coloured dots representing significant models (p<.05). The red area around each dot is the bootstrapped 95% confidence interval. Specifications 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 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
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 timing	Much earlier	3.0%	4.4%	.50
	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).

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	−0.14 [−0.17; −0.07]	<.001	91/112	<.001	8	.650	−0.08 [−0.13; −0.06]	<.001	100/112	<.001	1	1	−0.10 [−0.13; −0.08]	<.001	191/224	<.001	9	.998
Residualized parent-report PDS stage	−0.18 [−0.24; −0.13]	<.001	96/112	<.001	14	.012	−0.09 [−0.12; −0.05]	<.001	95/112	<.001	0	1	−0.13 [−0.16; −0.09]	<.001	191/224	<.001	14	.752
Residualized IIO Tanner stage	−0.29 [−0.35; −0.22]	<.001	112/112	<.001	38	<.001	−0.31 [−0.33; −0.22]	<.001	112/112	<.001	39	<.001	−0.30 [−0.32; −0.23]	<.001	224/224	<.001	77	<.001
Residualized gonadal composite	−0.20 [−0.26; −0.16]	<.001	111/112	<.001	36	<.001	−0.15 [−0.21; −0.11]	<.001	99/112	<.001	16	<.001	−0.18 [−0.22; −0.15]	<.001	210/224	<.001	52	<.001
Residualized adrenal composite	−0.19 [−0.21; −0.11]	<.001	100/112	<.001	17	.001	−0.19 [−0.20; −0.13]	<.001	112/112	<.001	4	.997	−0.19 [−0.19; −0.13]	<.001	212/224	<.001	21	.057
Self-report subjective timing	0.04 [−0.03; 0.05]	.009	64/112	.156	4	.997	−0.17 [−0.19; −0.11]	<.001	112/112	<.001	13	.063	−0.08 [−0.11; −0.05]	<.001	160/224	<.001	17	.529
Parent-report subjective timing	−0.14 [−0.17; −0.09]	<.001	80/112	<.001	3	.999	−0.19 [−0.17; −0.09]	<.001	99/112	<.001	0	1	−0.16 [−0.16; −0.10]	<.001	179/224	<.001	3	1
Age at menarche	NA		NA		NA		NA		NA		NA		−0.05 [−0.05; 0.01]	.055	66/112	.088	0	1
Residualized DHEA	−0.15 [−0.13; −.05]	<.001	94/112	<.001	0	1	−0.04 [−0.08; 0]	.001	82/112	<.001	0	1	−0.08 [−0.09; −0.03]	<.001	176/224	<.001	0	1
Residualized testosterone	−0.03 [−0.03; 0.05]	.058	69/112	.018	0	1	−0.03 [−0.06; 0.01]	.060	80/112	<.001	0	1	−0.03 [−0.04; 0.01]	.012	149/224	<.001	0	1
Residualized estradiol	0.11 [0.08; 0.15]	<.001	6/112	1	0	1	−0.01 [−0.05; 0.02]	.371	63/112	.221	0	1	0.05 [0.03; 0.09]	<.001	69/224	1	0	1

	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
All predictors combined	−0.10 [−0.11;−0.08]	**<.001**	807/1120	**<.001**	120	**<.001**	−0.10 [−0.12;−0.10]	**<.001**	954/1120	**<.001**	73	.594	−0.10 [−0.11;−0.09]	**<.001**	1827/2352	**<.001**	193	**<.001**

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. DHEA=dehydroepiandrosterone, CI=confidence interval, LD=Line Drawings, PDS=Pubertal Development Scale.

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	−0.02 [−0.03; −0.01]	**<.001**	214/336	**<.001**	9	.484
Anxiety symptoms	−0.04 [−0.05; −0.03]	**<.001**	225/336	**<.001**	14	**.021**
Internalizing disorder	−0.15 [−0.19; −0.13]	**<.001**	291/336	**<.001**	25	.916
Depressive disorder	−0.25 [−0.28; −0.21]	**<.001**	273/336	**<.001**	63	**<.001**
Anxiety disorder	−0.19 [−0.23; −0.16]	**<.001**	290/336	**<.001**	21	.993
Distress disorder	−0.17 [−0.22; −0.15]	**<.001**	259/336	**<.001**	61	**<.001**
Fear disorder	−0.11 [−0.15; −0.08]	**<.001**	275/336	**<.001**	0	1

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