Supplementary Material

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 (Hossain et al., 2017; Matta et al., 2017), we nevertheless included a decision point in the SCA to use imputed data or complete-case analysis.

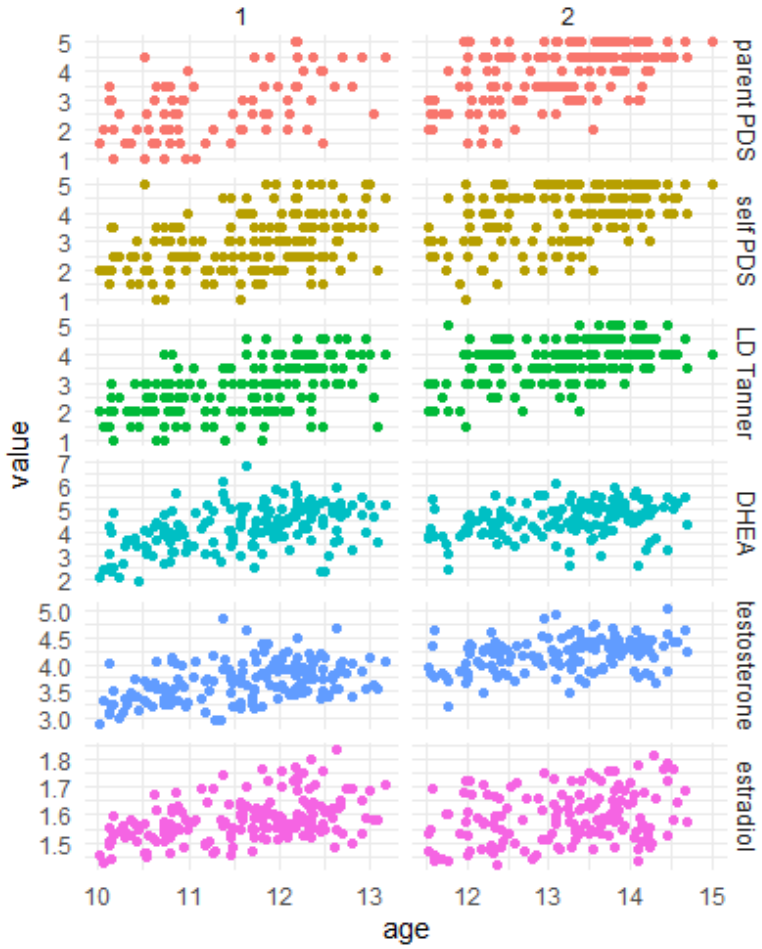
Supplemental Table 1. Percentages of missing data by variable type (Total N = 174)

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%)	
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%)

Control variables		
Early life stress (CTQ threat score)	35 (20.11%)	N/A
BMI z-score	1 (0.57%)	N/A

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



Supplemental Figure 1. Distribution of Pubertal Development Scale (PDS) scores, Tanner Stage from Line Drawings (LD), and log-transformed DHEA, testosterone, and estradiol levels at Time 1 (left) and Time 2 (right).

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 (Chumlea et al., 2003). 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).