Behavior Problems and Timing of Menarche: A Developmental Longitudinal Biometrical Analysis Using the NLSY-Children Data

Joseph Lee Rodgers · Carol Van Hulle ·
Brian D'Onofrio · Paul Rathouz · Will Beasley ·
Amber Johnson · Irwin Waldman · Benjamin B. Lahey

Abstract A powerful longitudinal data source, the National Longitudinal Survey of Youth Children data, allows measurement of behavior problems (BP) within a developmental perspective linking them to menarcheal timing (MT). In a preliminary analysis, we evaluate the bivariate relationships between BP measured at different developmental periods and the timing of menarche. Correlations were not consistent with any correlational/causal relationship between BP and MT. In the major part of our study, MT was used to moderate the developmental trajectory of BP, within a genetically-informed design. Girls reaching menarche early had behavior problem variance accounted for by the shared environment; those reaching menarche with average/late timing had behavior problem differences accounted for by genetic variance. Our findings match previous empirical results in important ways, and also extend those results. A theoretical interpretation is offered in relation to a theory linking genetic/shared environmental variance to flexibility and choices available within the family in relation to BP.

Keywords Behavior problems · Conduct disorder · Timing of menarche · Cholesky models · Moderation · NLSY-Children data

J. L. Rodgers (✉)
Department of Psychology and Human Development, Vanderbilt University, Nashville, TN 37204, USA
e-mail: joseph.l.rodgers@vanderbilt.edu

C. Van Hulle · P. Rathouz
University of Wisconsin – Madison, Madison, USA

B. D'Onofrio
University of Indiana, Bloomington, USA

W. Beasley
University of Oklahoma, Norman, USA

A. Johnson
Portland State University, Portland, USA

I. Waldman
Emory University, Atlanta, USA

B. B. Lahey
University of Chicago, Chicago, USA

Introduction

The current paper presents a developmental study of behavior problems (BP) and their relationship to timing of menarche. The unique measurement feature of our data is the availability of measures of behavior/conduct problems at three different age periods (two of which use comparable instrumentation): childhood, young adolescence, and older adolescence. There are two unique design features. First, in relation to timing of menarche, BP are measured before, approximately concurrently, and after menarche. Second, we use kinship patterns to decompose the multivariate relationship across our developmental measures into biometrical components, with timing of menarche as a moderator.

The timing of menarche marks an important developmental moment in the life of an early adolescent girl. Social scientists have proposed that it may also mark, directly or indirectly, other developmental processes as well. Recent research has focused on the relation between BP and menarcheal timing (MT). But the causal directionality involved in the relation between MT and BP is still unspecified. We define a taxonomy of six possible

causal directions: (a) MT has some causal influence on BP; (b) BP have an influence on MT; (c) both causal processes occur simultaneously; (d) MT moderates the developmental process associated with BP; (e) some other mediating or moderating set of processes cause them to covary with no direct causal influence of any kind; (f) there is no relationship. Social scientists confidently have stated each interpretation above (but seldom have accounted for the possibility of simultaneity).

Both theoretical and empirical research has marked MT as one of the lynchpins in a complex and interrelated set of processes. Although timing of menarche clearly contains important information, the causes and correlates of MT are not easily modeled. In particular, timing of menarche cannot be manipulated experimentally. Therefore, quasi-experimental designs are used to study its causes and consequences, and such designs are replete with threats to internal validity (see Shadish et al. 2002).

To investigate alternative a–c above, we use one of the strongest leverage points within quasi-experimental designs, the logic of temporal precedence. Simply stated, it is logically impossible for something later in time to directly cause something that occurred earlier in time. Pearl (2000) discussed "causal symmetry," which implies that a natural threat to the validity of the assertion that A causes B is the possibility that B may cause A instead of (or in addition to) A causing B. Pearl noted (p. 250), "By imposing the constraint that an effect never precede its cause, the symmetry is broken and causal inference can commence." A prerequisite for a causal link from BP to MT would be a correlation between BP and MT, with BP measured before menarche. A prerequisite for a causal link from MT to BP would be a correlation between MT and BP, with BP measured after menarche. Ge and Natsuaki (2009) identify several theoretical reasons to expect such a correlation, and focus on this latter directionality. Or, both correlations of MT with BP measured before *and* after menarche could be significant, supporting investigation of reciprocal causes. A cross-sectional twin study (Burt et al. 2006) found that MT moderated biometrical structure in early adolescent conduct disorder. It is possible for there to be no causal relation between BP and MT (i.e., none of a–c are correct), and yet MT could moderate developmental patterns (alternative d above). Or, causal patterns could occur in addition to the moderating patterns. Obviously, the complexity of the causal questions have only begun to be addressed.

The ultimate methodological and substantive goals of the current study are to compare and interpret patterns in developmental behavior problem data in relation to the explanations in a–f above (with a particular focus, motivated by previous literature, on (d)). We begin with a review of the literature relating menarche and BP. Next, we describe our data and analyses, and fit phenotypic and biometrical models to the correlations/covariances reflecting multivariate relationships between our critical variables. Finally, we discuss our results, and place them in the context of past findings in the literature.

Literature review

Research on timing of menarche and behavior problems

Several scholars have developed theories of childhood and adolescent BP (e.g., Conradt et al. 2013; Ellis 2004; Lahey and Waldman 2003; Moffitt 1993; Repetti et al. 2002). Belsky, Steinberg, and Draper (1991; also see Belsky 2000; Draper and Harpending 1982) suggested that stress/violence in a female's early family environment could stimulate a reproductive strategy of early menarche, early sexual involvement, and early reproduction, compared to females raised in less-stressful environments, an adaptive evolutionary response to insure reproductive success. The mechanisms stimulating this response include father absence, and family aggression and stress, a process Ellis (2004) referred to as "psychosocial acceleration theory;" the role of father absence he called "paternal investment theory." Researchers evaluating this theory have found correlations between father absence and MT (Chisolm et al. 2005; Deardorff et al. 2011; Doughty and Rodgers 2000; Hoier 2003; Mustanski et al. 2004; Rowe 2002; see Brooks-Gunn 1988 for an exception).

Others have challenged the Belsky et al. hypothesis. Maccoby (1991) and Rowe (2000) proposed an alternative suggesting that genetic and/or environmental contingencies from previous generations are the causes of both early family stress and later child outcomes. This argument specifies a selection model, in which the causal interpretation is not a direct path from family stress to child outcomes, but rather suggests that other background factors select parents for whom family stress and giving birth to children with certain child outcomes will co-occur. Molecular/behavior genetic studies have empirically evaluated this argument (Comings et al. 2002; Jorm et al. 2004; Rowe 2002); each supported the cross-generational selection model. Mendle et al.'s (2006) analyses of children-of-twins designs "strongly suggest selection, and not causation, accounts for the relationship between stepfathering and early menarche" (p. 533). Belsky has expanded his perspective to partially accommodate cross-generational selection, by defining the differential susceptibility of children to both "negative effects of risky environments," as well as "positive effects of supportive environments" (Belsky et al. 2007 p. 300). The theoretical origins of differential susceptibility extend back before the child's specific family environment (as in the selection argument), but have evolutionary origins.

This larger research agenda is relevant to the current paper because the link between BP and MT has also been used to distinguish cross-generational selection from direct causation. Moffitt et al. (1992) tested genetic influence on problem behaviors as the basis for selection (also see Pickles et al. 1998). Of particular relevance for the current study, they found a nonsignificant correlation (r = .04) between BP at age seven and self-reported age at menarche among 378 adolescent girls in New Zealand, and this correlation remained small when BP were separated into internalizing and externalizing dimensions, findings supportive of the selection interpretation. Caspi et al. (1993) used the same data to link MT and later adolescent delinquency. As in the current study, they obtained ratings from different development periods, conduct problems from childhood, norm-breaking behaviors in early adolescence, and self-reported delinquency at middle adolescence. However, rather than analyzing the trajectory of conduct problems, they compared girls in all-girls schools to those in mixed-sex schools. They found a moderating effect of MT on conduct problems at age 13, but only for girls in mixed-gender schools.

Burt et al. (2006) used data from 354 twin pairs (218 MZ and 136 DZ) from the Minnesota Twin Family Study to evaluate the link between MT (defined into three categories: early, average, and late) and conduct disorder (CD; measured using a DSM-III-R symptom-count) during the age range 13–16. They found a significant, though small, correlation between CD and MT, r = .09 (with MT reverse-coded). Further, mean age at menarche did not differ for girls with early-onset CD versus those with average timing of CD, further supporting the absence of a MT and CD link, measured both earlier in childhood, and during early adolescence. The results of their biometrical analysis will be reported in the next section. Several researchers have focused on shared genes as the source of the selection process. Empirical work by Van Hulle et al. (2009) showed overlapping genetic variance in antisocial behavior between late childhood and early adolescence (though MT was not a part of this framework). Button et al. (2005) found that a "risk genotype conferring susceptibility to family dysfunction is responsible for most of the variance in antisocial symptoms in childhood and adolescence" (p. 115). The findings reviewed above treat MT as an outcome in the causal framework (though no explicit link from childhood problem behaviors to MT has been proposed). But MT also has been treated as endogenous to child outcomes, and to problem/antisocial behaviors in particular.

Several different theories directly link MT and BP. First, the accentuation hypothesis (Caspi and Moffitt 1991) posits that disruptive transitions in the life course, such as puberty, can potentially magnify already-present behaviors (e.g., problem behaviors). They used data from the Dunedin Multidisciplinary Health and Development Study, and their empirical results studying timing of menarche in relation to BP supported an "early timing hypothesis." Second, Ge and Natsuaki (2009) reviewed four theories (including the accentuation hypothesis). The maturation disparity hypothesis focuses on the difference between physical and social/psychological development to explain how early puberty could cause problem behaviors. They suggested that this hypothesis has seldom been directly tested, and noted that "the higher rates of psychopathology among early maturers are expected because their slow developing neurocognitive systems are mismatched with the fast-approaching social and affective challenges at the onset of puberty" (p. 329). The other theoretical orientations they review—contextual amplification and hormonal influence—are also designed to explain the directional link from MT to BP. These past results provide motivation for looking at disparate sources and directions for the causal links among MT and problem behaviors.

To summarize, there is a consistent and important empirical finding that has not been given its proper attention in past research. Because of the natural interest in the link between father's absence and early family stress with subsequent timing of menarche, the *absence* of a correlation between problem behaviors/conduct problems and MT is notable. But MT also played a moderating role in Burt et al. (2006). Caspi et al. (1993) and others support that the relation between MT and behavior/conduct problems can be subtle (see Celio et al. 2006; Posner 2006).

Biometrical research on behavior/conduct problems and timing of menarche

Considerable biometrical research exists studying behavior/conduct problems in univariate and multivariate contexts. These studies support moderate heritability in BP/delinquency, and low levels of shared environmental variance (Ellis 2004). Rhee and Waldman (2002) conducted a meta-analysis of biometrical studies of conduct disorder, and estimated the proportion of additive genetic variance = .32, slight evidence of non-additive genetic variance (dominance) = .09, the proportion of shared environmental variance = .16, and the proportion of nonshared environmental variance/measurement error = .43. Using data from the Children of the National Longitudinal Survey of Youth (NLSYC), (Rodgers et al. 1994a) documented heritability in the Behavior Problem Inventory (BPI) in the NLSYC, and (Rodgers et al. 2001a, b) reported delinquency heritability patterns in the original adolescent NLSY79 sample (which includes the mothers of the NLSYC).

Menarche has also been studied in univariate biometrical analyses, and consistent patterns have been obtained.

Fig. 1 a Design structure for phenotypic correlational analysis (note: no correlations reported in Caspi et al.). **b** Design structure for biometrical moderation analysis
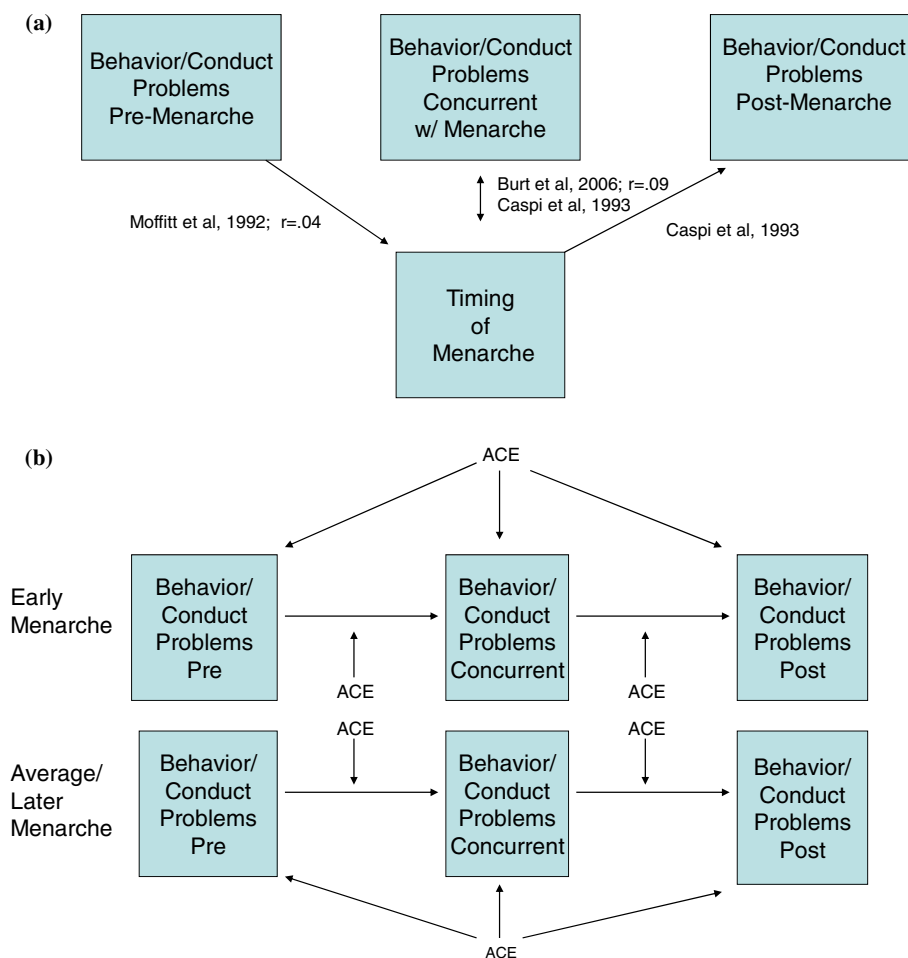
Twin studies showing heritability in MT include Pickles et al. (1998) and Kaprio et al. (1995). Rowe (2000) used the Add Health twin data, and reported a menarche heritability (estimating proportion of menarche variance that is genetic) of $h^2 = .44$ and shared environmental variance of $c^2 = .08$. Doughty and Rodgers (2000) used the NLSY79 family data to examine the etiology of MT and obtained $h^2 = .51$ and $c^2 = .01$. Mustanski et al. (2004) used longitudinal pubertal development data from Finnish female twins, and found $h^2 = .42$ and $c^2 = .21$ (obtained by squaring the standardized menarche coefficients in their Fig. 1). Ellis (2004) summarized the twin literature as having "converged on the conclusion that genotypic effects account for 50–80 % of the phenotypic variation in MT, and that the remaining variance is attributable to nonshared environmental effects and measurement error" (p. 922). Meyer et al. (1991) suggested that there exists some non-additive genetic variance in MT, which could hide potential shared environmental influence.

The previous subsection identified a correlational finding that appears inconsistent with direct causation from the early family to later BP, the absence of a bivariate correlation between timing of menarche and conduct problems. Past arguments suggest that if the family environment affects menarche directly, significant c^2 estimates would emerge (Rowe 2000). But whether smaller c^2 values are supportive of selection models is questionable, because of the potential for the psychosocial acceleration theory processes to operate through nonshared environmental as well as shared environmental sources (Ellis 2004).

Examining the menarche-behavior problems link within a developmental biometrical design

The review above identifies a gap in otherwise extensive past research, which the current research will help fill. Few studies exist that link timing of menarche and behavior/conduct problems using genetically-informative data. Burt et al. (2006) reported a moderator analysis in which they found that heritabilities for conduct problems were higher for females who reached menarche with average timing than for females reporting either early or late menarche, and the shared environmental patterns were the reverse. This same study found relatively low phenotypic correlation between

menarche and conduct disorder. No bivariate structure linking either genetic or shared environmental variance in conduct disorder and menarche was identified.

Harden and Mendle ([2012](#)) used the biometrical structure in 924 twin and sibling pairs from the Add Health data to study the link between pubertal timing and later delinquency measures. They found significant heritability in both female pubertal timing and delinquency measures. In biometrical analysis, they found a moderating effect of pubertal timing on the heritability of both violent and nonviolent delinquency. For non-violent delinquency, h^2 increased and c^2 decreased for increasing age at menarche; for violent delinquency, the reverse patterns was obtained. When they modeled the overlapping variance, they found that pubertal timing and both types of delinquency shared common genetic etiology, but did not share any environmental variance. They concluded that "our results challenge a purely socioenvironmental account of why early maturing girls show increased involvement in delinquent behavior" (p. 81–82).

A study by Dick et al. ([2000](#)) used a Finnish twin sample to link pubertal timing and adolescent substance use. They used within-family correlations to control for between-family selection bias. If the apparent causal link were indeed through background selection processes, the relation between pubertal timing and substance use would be fully explained by between family confounds. Instead, a direct link between pubertal timing and adolescent substance use remained. Other studies linking early puberty and increased adolescent substance use include Lanza and Collins ([2002](#)) and Stice et al. ([2001](#)). Both of these studies found small to moderate correlations between MT and subsequent substance abuse, but Stice et al. found no correlation between MT and comorbidity among outcomes such as substance abuse and eating disorders.

To summarize, there is consistent heritability in behavior/conduct problems, at various ages. There is heritability in menarche. But few multivariate designs have linked the two, and none developmentally. Menarche may function as a moderator, rather than as a primary causal variable itself. In an earlier review, (Moffitt [2005](#)) noted the value of genetically-informed data in studying family influences on antisocial behavior: "…causation is not well understood. Studies that cannot disentangle genetic and environmental influences cannot help" (p. 533).

Method

The data and the kinship links

The data for the current study come from the National Longitudinal Survey of Youth Children (NLSYC) dataset.

On December 31, 1978, a probability sample of US households containing 14–21 year old adolescents was drawn, the NLSY79, which included an oversample of minority and poor white households and a small military sample of older adolescents, with N = 12,686. Since 1979, the 5,827 females in the non-military sample have provided complete pregnancy and birth histories. By 2004, when NLSY79 females were 39–46, childbearing was almost completed. 11,428 biological children had been born, well over 90 % of the projected ultimate childbearing within this cohort (Center for Human Resource Research, 2006, p. 5).

Since 1986, a biennial survey effort has collected in-depth data from all biological children born to the NLSY79 mothers, the NLSYC sample. In 2004, those children ranged in age from infants to adults in their mid-30's (older NLSYC respondents had mothers who had already given birth before the first year of the NLSY79 survey), and are representative of the US population of children born to women in this age cohort in the US (except for attrition, nonresponse, and design features that will be discussed).

There is a well-known selection bias that characterizes the NLSYC data, which necessarily corresponds to children born to younger mothers. Although this selection bias is disappearing over time, it is still a feature of most NLSYC research (Chase-Lansdale et al. [1991](#)). Methods to measure and model this selection bias have been developed (see Rodgers and Rowe [1988](#), among others). We further discuss this form of selection bias in our Discussion section. Another potential bias in the retrospective reports in the NLSY data is the potential for telescoping, whereby the time since an event causes a time-dependent bias in reporting the event; telescoping has also been investigated in past research (Rodgers et al. [1992](#)).

Our design dictated that we trade-off some advantages of the probability sampling process against practical requirements of our design. Our sample still supports higher external validity than most similar studies, because it originated as a probability sample. However, a complex screening process precludes valid use of the sampling weights that adjust for attrition, nonresponse, and oversampling. Though our study is focused on internal validity, we note that using the NLSYC automatically provides more external validity than many other comparison samples that did not originate from a probability sample (many of which are based on twin registries, with well-known problems of generalizability; see Lykken et al. [1987](#); Stoolmiller [1999](#)).

Our research design used longitudinal information on behavior/conduct problems for ages 6–9, 12–14, and 15–17, as well as information from intact female kinship pairs in the NLSYC. Although the NLSY79 females have virtually completed childbearing, some children have not

aged past late adolescence, and many families do not contain two female children. When we screened for these requirements, we obtained 890 females in 445 female–female kinship pairs, representing five different types of genetic relatedness across pairs. This is a reduced sub-sample of the NLSYC database, though it is comparable in size to many of the larger genetically informed databases used in biometrical studies (many of which are based on twin samples of MZ and DZ twins). For example, the Burt et al. (2006) study reviewed above used 354 female twin pairs, with conduct disorder measured at a single point in time.

Although the 2004 NLSYC database did not contain any information to directly ascertain genetic relatedness, a kinship linking algorithm to ascertain kinship relatedness was developed in 1994, with an update completed in 2005. This algorithm identified genetic relatedness for the NLSYC kinship pairs, and also identified cousin pairs (whose NLSY79 mothers are full siblings). The kinship pair categories (and associated genetic coefficients) included MZ twins ($R = 1.0$), DZ twins/full siblings ($R = .50$), half siblings ($R = .25$), cousins ($R = .125$), and an ambiguous full/half sibling category ($R = .375$; see below). The sample sizes within these categories are approximately representative of those in the population; there are 104 cousin pairs, 168 half-sibling pairs, 110 ambiguous half/full siblings pairs, 502 full sibling/DZ twin pairs, and 6 MZ twin pairs. These sample sizes are disproportionate because they are structured that way in the population. We have run many sensitivity analyses, to assess the impact of the small MZ twin category, the presence of cousins (who did not share a family environment), and the effect of the ambiguous sibling category. Results were relatively robust to these adjustments, except for a few cases which are discussed in detail below.

By design, all sibling pairs in the NLSYC are biological offspring of an NLSY79 mother, and thus all kinship pairs share a common biological mother (or a biological mother and an aunt, in the case of cousins). Whether siblings share a father is not directly indicated, but is logically implied by maternal responses. If two children have lived in the same household with their biological mother and father for some time, they are assigned full sibling status. If two siblings lived in the same household with their biological mother, but only one lived with her biological father, they are assigned half-sibling status. If neither lived with their biological father, additional location questions are used for classification. If a sibling pair cannot be classified with available NLSY information, then they are assigned a mid-range genetic coefficient of $R = .375$ (halfway between the sibling $R = .50$ and the half-sibling $R = .25$). A number of research studies have been published using the 1994 kinship links (e.g., Rodgers et al. 1994a, b; van den Oord and Rowe 2000), and more recent studies using the updated

kinship links have been conducted (D'Onofrio et al. 2008; Goodnight et al. 2012; Rodgers et al. 2008a, b; Van Hulle et al. 2007; Van Hulle et al. 2009). The algorithm is documented in (Rodgers et al. 2005), which contains a validity analysis showing that the biometrical structure of adult height is similar to that obtained in meta-analyses of height published in the behavior genetic literature (see e.g., Plomin 1990; Mittler 1971). The kinship links are publicly available from the first author.[1]

Measurement

We used four critical variables for which we obtained complete longitudinal information for a relatively large sample size of NLSYC female kinship pairs: BP in young childhood (ages 6–9), BP in early adolescence (ages 12–14), BP (delinquency) in late adolescence (ages 15–17), and timing of menarche. Descriptive statistics for MT and the three behavior problem scales are presented in Table 1.

Behavior problem scale scores for the first two age ranges were obtained from the NLSYC BPI, which is obtained by summing maternal rating of 28 behavior

---

[1] In 2006, direct indicators of sibling relatedness were included within both the NLSY79 and NLSYC surveys. Our research team recently completed updating and released both sets of kinship links using these explicit indicators in combination with previous links defined by implicit processes as described within the current paper. Public release of these new kinship links occurred in November, 2013. Although the new links improve on the 2006 links that are used in the current paper, the older links have been used effectively in past research, and validity checks show that the two sets of links function in very similar ways. There are a number of indicators of the quality of the 2006 links, and reasons to believe that they are at least as good, and in some ways superior, to direct ascertainment of sibling relatedness. First, concurrent validity has been established in relation to a number of different phenotypes, in which NLSYC results are very similar to those obtained using other datasets with "direct" ascertainment (see Rodgers et al. 2008a, b, for a summary of matching studies that establishes this type of validity). Second, information provided by mothers about biological fathers is in several senses a more logically direct indicator of sibling relatedness than asking the siblings for their own perceived relatedness. Issues of honesty aside, the mother potentially knows details about her children's relatedness of which the child might not even be aware. Further, the mother is more likely to understand the concept of biological relatedness than young children. Finally, the mother is quite likely to be more aware of location information about the biological father—which is used to distinguish some half siblings in the 2006 NLSYC linking algorithm—than are the children. Comparison between our 2006 implicit kinship links and newer explicit links among the female–female pairs have identified well over 90 % agreement. The potential value of the new links is substantial in resolving the "ambiguous sibling" category, and in increasing the sample sizes, but there is very little disagreement among the more recent direct indicators and the large number of earlier links that were ascertained by the linking algorithm. In future research efforts, we recommend use of the most recently updated kinship links, which can be obtained by e-mailing the first author or online at: cran.r-project.org/web/packages/NlsyLinks/.

Table 1 Overall descriptive statistics for the four constructed measures

| | Mean | Standard deviation | Median | IQR | Skewness | Univariate kinship correlations | |
						R = .125/.25/.375	R = .50
BP_6-9	107.2	13.9	107.0	18.5	.30	.33	.49
BP_12-14	108.3	15.2	107.0	22.0	.20	.25	.28
BP_15-17 (stdized)	.0	1.0	−.3	1.5	1.20	.04	.21
MT	12.3	1.3	12.3	1.5	.15	.24	.39

N = 890; *BP* behavior problems, *MT* menarcheal timing; note that BP_15-17, which was constructed as described in the paper, was standardized in the sample; see the text for scores before standardization

problem indicators, with mothers indicating "often true," "sometimes true," and "not true" in regards specific BP exhibited within the previous 3 months among children age 4 to 14. The NLSYC BPI instrument is a short form of the Child Behavior Check List (Achenbach 1978; Peterson and Zill 1986). We used ages 6–9 for the younger age category (specifically, between 6 years, 0 months and 9 years, 11 months at interview), and an age range of 12–14 for the older. We used a four-year age range for the younger age category to increase sample sizes; there were no systematic patterns across age in standardized female BPI scores for either age range. The overall BPI is already normed for age and gender. When respondents had two BPI measures within an interval, we averaged those scores. The age-and-gender standardized versions of the BPI scores were used, which are scaled to have a population mean of 100 and a standard deviation of 15; in our sample, the range was 71–146 (with higher values indicating higher BP), with means around half a standard deviation above the population mean (at least partially reflective of the selection bias discussed earlier of this sample being children born to younger mothers). The overall BPI reliability is .87.

Because the BPI scales contained non-conduct trait-like items (measures of dependence, anxiety, etc.), we also created a subscale specifically measuring conduct problems, based on the following seven items: cheats or tells lies; has trouble getting along with teachers; disobedient at home; disobedient at school; bullies or is cruel or mean to others; breaks things on purpose or deliberately destroys his/her own or another's things; does not seem sorry after misbehaving. Across ages, the median coefficient alpha reliability of the seven-item scale was $\alpha = .74$. This measure has been used in other NLSYC studies focused on conduct problems (e.g., Lahey et al. 2006). We will report correlations, heritabilities, and models for both complete BPI (BP_6-9 and BP_12-14) and the conduct problem subscale (CP_6-9 and CP_12-14). Results were slightly cleaner using the conduct problem scores, but the BPI models had notably better goodness of fit indicators. Final interpretations are similar across the two measure.

The later-adolescent behavior problem scale was based on delinquency scores (BP_15-17; note that there was no corresponding conduct problem subscale associated with this measure). This scale was difficult to construct, because of changes in the items used to measure delinquency across time.[2] The BP_15-17 scale was a count of the number out of seven delinquency items that the respondent indicated they had performed within the last year. We standardized the overall NLSYC female sample at each single age (using, e.g., all respondents between 15 years 0 months to 15 years 11 months to do the age 15 norming), to adjust for age changes in delinquency levels. Finally, we used separate standardization samples before and after 2000, to account for delinquency changes over time. The pre-2000 sample standardization means (and variances/sample sizes) before standardization were M = 1.52 (V = 2.49, N = 474) for age 15, M = 1.52 (V = 2.58, n = 419) for age 16, and M = 1.56 (V = 2.14, N = 307) for age 17. The equivalent post-2000 statistics were M = 1.03 (V = 1.46, N = 384) for age 15, M = 1.03 (V = 1.21, N = 404) for age 16, and M = 1.08 (V = 1.37, N = 412) for age 17. Because we used only female respondents in this study, gender norming was not an issue. As before, we averaged when there were multiple standardized delinquency measures within the three-year age range from 15–17 years.

---

[2] The construction involved the following set of features and steps. A common set of items was used in 1994, 1996, and 1998; an overlapping but smaller item subset was used in 2000; and in 2002 and 2004, several items were added back in, but two of them— "damaged property" and "skipped school"—used slightly different wording that changed the baseline response. We identified seven items that were approximately common across the five different survey dates 1994, 1996, 1998, 2002, and 2004 (the 2000 survey was not used, because there were not enough items). Those seven items asked respondents if they had ever hurt someone, lied to parents, damaged/destroyed property, stolen something from a store, skipped a day of school without an excuse, brought a parent to school because of doing something wrong, and run away from home. One other shift between the two time periods occurred; in the later time period, instead of asking "have you ever …" about each item, a four-item ordinal scale was provided: 1=never; 2=once; 3=twice; 4=more than twice, which we recoded into a binary yes/no response matching the response format from the earlier time period.

Virtually no skewness existed in the BP_6-9, BP_12-14, and MT measures; means, medians, and skewness coefficients are presented in Table 1. Even for BP_15-17, which had a moderate skewness coefficient, the mean and median differed by less than a third of a standard deviation. We ran a few analyses using a square root transformation (pre-standardization) of the BP_15-17 scale, and results were virtually identical to those without the transformation. The raw and square root transformed BP_15-17 scales correlated r = .99. After this investigation of skewness, we ran all analysis on the constructed variables.

The final variable to support our design was timing of menarche (MT), which was self reported by each girl in two ways, as age and as month and year of menarche. Such retrospective reports are considered reliable and valid (Brooks-Gunn et al. 1987; Damon et al. 1969; Moffitt et al. 1992), especially when reporting occurs shortly after menarche. We constructed measures for stated age, and calculated age, and used the more precise calculated age based on month/year reports. In a few cases with differences greater than 2 years between calculated and stated age, we imputed stated age as the more likely correct response (N = 37 cases), or if calculated age was missing, stated age was used (the two measures correlated r = .91). Age at menarche responses ranged from 0–19; 14 respondents were disqualified for giving responses of less than 7 years of age. Finally, when stated age was imputed in place of calculated age, .5 was added to correct for the step-function feature of reported age (e.g., stated age 12 would refer to an average of 12.5 from the interval from age 12.00 to 12.99). Most respondents were only asked once, but for 604 respondents the question was re-asked at a second time (and their responses were averaged). The correlation between these—a test–retest reliability estimate—is r = .65, a somewhat low reliability for this type of response. However, 75 % of the responses were within 1 year of one another.

Scores on all four critical variables were obtained for N = 1651 NLSYC females. However, we only used the 890 respondent subset of these respondents who also had a female twin, sister, or cousin within the dataset. In a very few cases, one NLSYC female contributed to more than one pair (e.g., if there were three daughters in the family, or if an NLSYC female had both a sister and cousin in the sample). We reran analyses by dropping duplicate kinship links, and results were virtually identical to those for the full sample.

Several demographic control variables were included within the analysis. The standardization procedures described above already adjusted for age and test-administration period differences between female respondents within each of our age categories. Race was accounted for within the design, although only the Black and Non-white Non-Hispanic (Caucasian) categories were large enough to support separate analyses.

Data analysis and modeling

The phenotypic analysis helps address the question, "Do behavior problems and timing of menarche covary?"(see Fig. 1a), the first time a data source has allowed simultaneous investigation of the correlation between measures of BP at three ages to timing of menarche (Caspi et al. 1993, could have evaluated this question, but had different goals).

The biometrical modeling emerged from the part of our design that allows the separation of environmental and genetic variance for each of the three time periods in which BP were measured. This is a straightforward developmental analysis, except with the data conditioned in two interesting and useful ways. First, timing of menarche is treated as a moderator within the biometric analysis of BP. Second, relationships across the developmental periods are separated by kinship category, supporting partitioning of variance and covariances into genetic and environmental components. This analysis is motivated empirically as a developmental extension of the Burt et al. (2006) study, and is motivated theoretically in relation to the causal goals a–f defined in the introduction.

Whether genetic and environmental factors that influence BP among young children are the same (at least in part) as those still causing BP in older adolescents is an important research problem. Van Hulle et al. (2009) addressed this question using a different subset of the same data, and found significant overlap, especially for the latter two age categories. Our design allows this question to be addressed simultaneously in relation to timing of menarche as a moderator of these patterns. We partition timing of menarche in two different ways, as early, average, and late menarche (e.g., Burt et al. 2006), by using the cut points of MT < 11.5, MT between 11.5 and 13, and MT > 13; and also as a two-category indicator, separating females who matured early, MT < 11.5, from those who do not, MT ≥ 11.5 (see Fig. 1b). The biometrical analysis models we use are based on correlation/regression procedures and structural equation models (SEMs). The regression procedure is based on a biometrical approach called DF analysis (named for DeFries and Fulker 1985). This approach uses simple OLS regression to obtain unbiased estimates of the proportion of total variance attributable to genetic and shared environmental variance (Rodgers and McGue 1994). Many DF analysis innovations have been published (Cherny et al. 1992; Kohler and Rodgers 1999; Kohler and Rodgers 2001; LaBuda et al.1992; Purcell and Koenen 2005; Rodgers and Kohler 2005; Waldman et al. 1992; Waller 1994). More complex univariate moderator analyses, and multivariate analyses, rely on Mx, a SEM program

originally developed as a dedicated biometrical software system (Neale and Cardon 1992). Our analytic approach—using DF analysis models in combination with SEM methods—has been presented in methodological detail in past NLSY studies (see Rodgers et al. 1994a for a DF analysis of NLSYC BPI measures; Rodgers et al. 2001a, Rodgers et al. 2001b, for a study of NLSY79 fertility patterns using DF analysis, SEM, and a comparison of the two; and Rodgers, Buster, and Rowe, Rodgers et al. 2001a, for a DF analysis of delinquency in the NLSY79).

In both DF analysis and SEM, the biometrical model is an ACE model; A refers to additive genetic variance, C refers to common (or shared) environmental variance, and E refers to nonshared variance/measurement error. Using the different genetic categories, and the methods and assumptions of quantitative genetics (Falconer 1979; Jinks and Fulker 1970), these sources of variance can be partitioned into the ACE variances (Neale and Cardon 1992). The C component was assumed constant across all categories. Assumptions of the additive genetic model are ignorable assortative mating, equal shared environments, and an additive genetic model (Plomin et al. 1990). Within our study, there is likely to be some assortative mating on BP, although the level of that assortment is unlikely to substantially bias our results (Maes et al. 1998; Taylor et al. 2000; but also see Krueger et al. 1998). The equal environments assumption (EEA) is usually considered a larger threat to internal validity in relation to MZ twins than to other kinship categories; we have only six MZ twin pairs in our study, and these make only a minor contribution to our outcomes. However, we do include cousins, who grew up in different households, and for them the EEA is conceptually tenuous. Half-siblings may also differ from full siblings in relevant ways (e.g., they have typically experienced divorce), although the half siblings within our study virtually all reside in a common household (with their mothers). We evaluate the EEA within our analytic approach in sensitivity analysis by deleting cousins, who do not share the family environment. Violations of the additive model are diagnosed through negative variance estimates (though small negative estimates can occur by sampling error).

Results

Menarcheal timing-behavior problem relationships

In the first set of analyses we present bivariate estimates of the relationships between BP and timing of menarche (see Table 2; Fig. 1a). Low correlations occur for the link between BP and menarche with BP measured before, approximately concurrently, and after menarche. In many

Table 2 Correlations between menarcheal timing (MT) and behavior problems (BP), overall and separately by levels of kinship relatedness (R is the coefficient of genetic relatedness); MZ Twins are not included because of small sample size; conduct problems subscale correlations after slashes

	Within-individual Overall MT	By Kinship categories			
		R = .125 MT	R = .25 MT	R = .375 MT	R = .50 MT
BP_6-9	.04/.06	.00/.01	.06/.09	.06/−.04	−.04/−.04
BP_12-14	.00/−.06	.05/−.04	.02/−.03	−.10/−.21*	−.03/−.04
BP_15-17	−.05	−.09	.05	.03	−.08
N (pairs)	445	52	84	55	251

* $p < .05$

studies these patterns would simply be the end of the discussion, because of the consistently low correlations. The first column of Table 2 shows a lack of relationship between menarche and BP within an individual, using all 890 individuals within our study. In the second through fifth columns, we show the kinship pair correlations for different genetic categories. There is only one significant correlation among the 25 presented in Table 2; the median across these 25 correlations is $r = -.03$. Earlier work (e.g., Burt et al. 2006; Caspi et al. 1993; Posner 2006) provided theoretical motivation to view precocious/early menarche as of particular relevance. Thus, we also defined a dummy variable accounting for timing of menarche before the age of 11.5 or not, and recomputed the correlations; they were very similar to those in Table 2.

Burt et al. (2006) also found low correlations, but then found an interesting moderating role for timing of menarche in studying conduct disorder. Thus, we redefined menarche into categories reflecting early, average, and delayed menarche in a set of univariate and multivariate biometrical analyses.

Longitudinal relationships across behavior problems, conditional on menarcheal timing

Although there are no relationships between behavior/conduct problems and MT, there are moderate-to-strong relationships between the behavior/conduct problems measures across time (the developmental patterns in Fig. 1b). Using the overall BPI-based measures, correlations within individuals (N = 890) are r(BP_6-9,BP_12-14) = .53, r(BP_6-9,BP_15-17) = .14, and r(BP_12-14,BP_15-17) = .24 (all $p < .0001$). Not surprisingly, the BPI-based measures correlated more highly than either measure with late-adolescent delinquency; this pattern likely reflects similar instrumentation and rater effects (BPIs at the first two ages were based on maternal measures, delinquency scores at the third age were adolescent self reports), but also reflects that the nature of BP shifts

Table 3 Biometrical results (ACE Models) using DF analysis, behavior problems before menarche (ages 6–9)—BP_6-9; approximately concurrent with menarche (ages 12–14)—BP_12-14; after menarche (ages 15–17)—BP_15-17 (conduct problem subscale analysis results after slashes)

	Overall	Early menarche (<11.5, N = 240)	Average/late menarche (>11.5, N = 650)
BP_6-9	h^2 = .59*/.72*	h^2 = .12/−.21	h^2 = .73*/1.02*
	c^2 = .19/.01	c^2 = .25/.26	c^2 = .17/−.07
BP_12-14	h^2 = .39*/.55*	h^2 = .19/.54*	h^2 = .50*/.33
	c^2 = .11/−.01	c^2 = .19/.05	c^2 = .06/.02
BP_15-17	h^2 = .62*	h^2 = .47	h^2 = .66*
	c^2 = −.13	c^2 = .01	c^2 = −.18
Menarcheal	h^2 = .48*		
Timing	c^2 = .14		

* $p < .05$

from childhood to adolescence. There were similar correlations for the conduct problem subscales.

The three correlations in the paragraph above are within-individual correlations. The critical relationships for our biometrical analysis are the kinship correlations. Kinship pair correlations are presented for the overall dataset in Tables 4 and 5. In Table 4 the correlations are separated into kinship pair categories using the five different kinship categories described in the Methods section (cousins, genetic coefficient R = .125; half-siblings, R = .25; ambiguous siblings, R = .375; full siblings, R = .50; and MZ twins, R = 1.0). In Table 4, we further partition these by timing of menarche. In a preliminary analysis, we defined three MT categories (MT ≤ 11.5, 11.5 < MT ≤ 13, and 13 < MT), which creates approximate 25, 50, and 25 % categories and approximately matches Burt et al. (2006). However, correlation patterns were virtually identical across average and late menarche, and these were combined into a single category in Table 5. The correlations in Table 5 suggest that there may be lower univariate heritabilities for BP in the early menarche group (compare corresponding on-diagonal correlations across the kinship categories). For the correlations reflecting multivariate relationships, the correlations for early menarche between BP_6-9 and BP_12-14 suggest shared environmental variance, and for average/later menarche possible heritability; few of the cross-variable correlations involving BP_15-17 are high enough to suggest any patterning (compare corresponding off-diagonal correlations across kinship categories). These results will be formalized in the modeling sections.

Univariate biometrical analysis

The univariate DF Analyses (ACE Models) are based on the full double-entered dataset of 890 individual

Table 4 Overall behavior problem kinship correlations (note:1 refers to the first member of the kinship pair, 2 refers to the second member; these correlations are based on double-entered data, and therefore are intra-class correlations; correlations after slashes are for Conduct Problem subscale scores; N = 890)

	BP_6-9_2	BP_12-14_2	BP_15-17_2
Cousins, R = .125, N = 104			
BP_6-9_1	.22/.11	.18/.16	.06/.03
BP_12-14_1	.18/.16	.15/.13	.27/.02
BP_15-17_1	.06/.03	.27/.02	.07
Half-siblings, R = .25, N = 168			
BP_6-9_1	.35/.21	.28/.08	−.08/−.11
BP_12-14_1	.28/.08	.26/.05	−.08/−.10
BP_15-17_1	−.08/−.11	−.08/−.10	−.02
Ambiguous half/full siblings, R = .375, N = 110			
BP_6-9_1	.28/.13	.11/.17	.05/.03
BP_12-14_1	.11/.17	.21/.36	.06/.07
BP_15-17_1	.05/.03	.06/.07	.02
Full siblings, R = .50, N = 502			
BP_6-9_1	.49/.40	.33/.24	.07/.09
BP_12-14_1	.33/.24	.26/.14	.12/.12
BP_15-17_1	.07/.09	.12/.12	.20

observations, representing 445 kinship pairs across the five categories (see Rodgers et al. 2001a, b, for discussion of double entry). Kinship correlations are presented in Tables 1, 4, and 5, and univariate heritabilities ($h^{2'}$s) and shared environmental variances ($c^{2'}$s) are presented in Table 3, for the overall dataset and for the two separate menarche categories. For the first two age categories, analyses are reported using both the overall BPI and the CP measures (from the BPI conduct problem subscales). The older two menarche categories were similar in h^2 and c^2 patterns (as they were for their correlations), and thus were combined. A statistical test of the similarity between the two older MT categories was run for each BP measure using a dummy variable accounting for category differences; each test failed to reject equivalence of the two categories, providing statistical support for combining these two categories.

Within Table 3, a clear and statistically distinguishable difference emerges that differentiates the females who reached early menarche from those who were average/late in their MT (based on testing dummy variables coding for menarche categories). All heritabilities for the average/ older menarche category for the BP_6-9 and BP_12-14 measures were significant, and of substantial magnitude; for the younger menarche category, all were small and nonsignificant. None of the c^2's from this table were significant, although several were moderate in size. The patterns for the CP_6-9 and CP_12-14 conduct problem subscales were identical to those for the overall BPI results,

Table 5 Behavior problem kinship correlations, separate for early menarche (≤11.5 years old) and combined average/late menarche (>11.5 years) (Note: 1 refers to the first member of the kinship pair, 2 refers to the second member; these correlations matrices are based on double-entered data, and therefore are intra-class correlations; matrices are asymmetric because both members of a kinship pair will not necessarily be in the same menarche category; these can be interpreted as "proband—co-kin" correlations—see text for further explanation; correlations after slashes are for conduct problems subscale)

	BP_6-9_2	BP_12-14_2	BP_15-17_2
Early menarche			
Cousins, R = .125, N = 32			
BP_6-9_1	.22/−.06	.25/.25	−.10/−.11
BP_12-14_1	.21/.14	.17/.20	−.13/−.01
BP_15-17_1	.01/−.05	−.03/−.03	.19
Half siblings, R = .25, N = 48			
BP_6-9_1	.27/.17	.22/.01	−.09/−.12
BP_12-14_1	.24/.24	.38/.19	.16/.18
BP_15-17_1	−.07/−.16	−.08/−.17	−.05
Ambiguous half/full siblings, R = .375, N = 36			
BP_6-9_1	.00/−.02	.08/.03	.19/.22
BP_12-14_1	.08/.36	.25/.39	.31/.32
BP_15-17_1	.39/.36	.23/.23	.30
Full siblings, R = .50, N = 121			
BP_6-9_1	.37/.17	.29/.12	.13/.20
BP_12-14_1	.18/.08	.11/.05	.13/.17
BP_15-17_1	−.02/−.06	.14/.16	.26
Average and Late Menarche			
Cousins, R = .125, N = 72			
BP_6-9_1	.22/.11	.15/.15	.11/.07
BP_12-14_1	.17/.16	.14/.11	.07/.04
BP_15-17_1	.08/.05	.03/.05	.02
Half siblings, R = .25, N = 120			
BP_6-9_1	.39/.22	.30/.12	−.07/−.10
BP_12-14_1	.30/.01	.20/−.04	−.18/−.22
BP_15-17_1	−.08/−.08	−.09/−.08	−.02
Ambiguous half/full siblings, R = .375, N = 74			
BP_6-9_1	.40/.23	.18/.29	.01/−.01
BP_12-14_1	.09/−.04	.10/.26	−08/−.08
BP_15-17_1	−.03/−.10	.01/.03	−.06
Full siblings, R = .50, N = 381			
BP_6-9_1	.53/.47	.35/.28	.05/.05
BP_12-14_1	.38/.29	.30/.17	.11/.10
BP_15-17_1	.09/.13	.10/.10	.17

with one exception. The CP_12-14 measure for early menarche showed significant heritability of approximately the same magnitude as the overall and average/later menarche categories. However, there is reason to doubt the stability of this result, based on sensitivity analyses of these patterns. When we dropped the six MZ twin pairs from the

analysis (which accounted for 1 % of the total pairs), the Table 3 results were virtually identical, with the exception of the early menarche CP_12-14 h^2 and c^2, which were somewhat unstable in relation to the MZ twin correlations; the other 14 categories showed little sensitivity to the presence/absence of the MZ twin pairs.

The menarche measure itself had a significant heritability ($h^2 = .48$), and nonsignificant shared environmental variance ($c^2 = .12$). The kinship correlations underlying these patterns were r = .38 for cousins (N = 52 pairs), r = .06 for half siblings (N = 84 pairs), r = .37 for ambiguous siblings pairs (N = 55 pairs), and r = .39 for full sibling pairs (N = 251 pairs). In previous research, similar ACE models were estimated for mothers of the NLSY females, with $h^2 = .52$ and $c^2 = .01$ (Doughty and Rodgers 2000), and for the NLSYC females themselves, with $h^2 = .66$ and $c^2 = .00$ (Bard and Rodgers 2006).

To assess the EEA in relation to cousin pairs we can study correlation patterns along the diagonals in Table 5. The cousin correlations do not appear inconsistent with other patterns. When analyses were replicated without cousins, results were similar to Table 3. Heritabilities increased slightly for the BP measures in the early menarche category, but none changed in statistical significance. Heritabilities for the average/late menarche category remained large and statistically significant. The c^2 estimates remained small and non-significant.

We also replicated the analyses above by race using three categories, Hispanic, Black, and Non-Black/Non-Hispanic; only the latter two were large enough to run DF Analysis (although barely, and power was uniformly low). For Black females, there were N = 114 in the early menarche and N = 226 in the average/late menarche category. For the non-Black/non-Hispanic (primarily Caucasian) female category, there were N = 64 in the early menarche, and N = 312 in the average/late menarche category (N's reflect earlier puberty for Black females; Rowe and Rodgers 1994, document and discuss these patterns in the NLSY79 females). For BP_6-9 and BP_12-14, the heritabilities were uniformly higher for the average/older menarche category than for the younger menarche category, consistent with the basic finding in Table 3. For the BP_15-17, heritability patterns were idiosyncratic and unstable.

Multivariate biometrical analysis

The purpose of performing multivariate biometrical modeling is to estimate shared or overlapping variance with genetic and/or environmental etiology (Fig. 1b). Thus, for example, significant heritabilities for the average/late menarche category identified in the previous section could derive from genetic influences shared in common across the longitudinal measures of BP or from different genetic

sources. In Table 5 there are no obvious multivariate patterns associated with BP_15-17 that are suggestive of overlapping genetic or shared environmental variance (compare the 1,3 and 3,1 elements and the 2,3 and 3,2 elements in Table 5 across genetic categories; there are low correlations, and no genetically-related pattern). However, early menarche bivariate correlations linking BP_6-9 and BP_12-14 are suggestive of overlapping shared environmental variance (note consistently moderate correlations in 1,2 elements across genetic categories), and the average/late menarche correlations are suggestive of overlapping genetic variance, and also some shared environmental variance overlap (note generally increasing patterns in 1,2 elements across genetic categories).

The univariate ACE models estimated and reported in the previous section were moderated by menarcheal category, including a statistically significant test using dummy variables. We performed a similar multivariate test using the Box M statistic to evaluate the moderating effect of menarcheal category on the covariance matrices that will be subjected to biometrical modeling. We defined covariance matrices for the overall dataset using the four behavior problem variables BP_6-9_1, BP-_12-14_1, BP_6-9_2, and BP_12-14_2, defined separately for early-menarche respondents and those who reached menarche on time or late. A standard statistical test of the null hypothesis that two independent covariance matrices came from the same population was developed by Box (1950), based on the Box M statistic (see Timm 1975, p. 251 for explanation and elaboration). When we compared the separate covariance matrices, we obtained $M = 31.6$, $p < .001$, rejecting the null of equal covariances and supporting that the early and on-time/late menarche group correlation patterns came from different populations. (Note that it is well-known that the Box M statistic is highly sensitive to non-normal distributions; all four of the BP measures defining this correlation matrix had skewness coefficients between .1 and .3, indicating only small deviations from normality. To give some perspective, the mean BP_6-8 was 107.2, with a median of 107.0. The mean BP_12-14 was 108.3, with a median of 107.0.)

We defined a cholesky model (Fig. 2) for BP_6-9 and BP_12-14 for each member of the kinship pair (i.e., four variables), in which BP_6-9 logically preceded BP_12-14, according to the temporal precedence (these cholesky models are biometrical models with temporal ordering). The model was fit to variances/covariances separately for the two menarche categories. To obtain stable covariance estimates, we combined genetic categories into two kinship levels. The $R = .125$, $R = .25$, and $R = .375$ kinship pairs (whose kinship correlations are defined in Table 5) were combined, and assigned their average genetic coefficient of $R = .25$ (note the approximately equal sample sizes in

these three categories). The 502 full sibling $R = .50$ pairs defined their own partition (and we also ran analyses that included the six MZ twin pairs in this category; the number of MZ twins was small enough that assigning them to the $R = .50$ category should not affect the estimates; sensitivity analyses verified this).

For each set of covariances, Mx estimated the cholesky model in Fig. 2, which optimizes a maximum likelihood (ML) function to estimate model parameters. Mx estimated 90 % confidence intervals (CI) for each parameter (choosing 90 % instead of 95 % results in erring on the side of including variables). When a parameter estimate was not significantly different from zero (i.e., the CI contained zero within the interval), we dropped that parameter and re-estimated the model. Results using the BP_6-9 and BP_12-14, and the CP_6-9 and CP_12-14 measures were slightly different, and will be reported separately.

Using BP_6-9 and BP_12-14, for early menarche, the best-fitting model is shown in Fig. 3a. The ML $\chi^2 = 29.5$ with 15 df's, RMSEA = .083, and the AIC = −.532 (note that these fit statistics are defined using the Mx definitions; see Neale and Cardon 1992). These are suggestive of adequate fit of this model (compare to the original fits for the overall cholesky model shown in Fig. 2 of $\chi^2 = 23.2$ with 11 df's, RMSEA = .083 and AIC = 1.21; the original model and the reduced model were statistically indistinguishable using a χ^2 difference test, $\chi^2 = 6.3$ with 4 df's, $p > .15$, leading to adoption of the reduced model). This best-fitting model did not include genetic variance but did include common shared environmental variance (like the univariate model). The majority of nonshared environmental variance/measurement error was estimated to be shared between the two BP measures.

For the average/late menarche category, the best-fitting model is shown in Fig. 3b. The fit indices for this model were excellent; ML $\chi^2 = 3.40$ with 15 df's, RMSEA = .000, and AIC = −23.15 (compare to fits for the overall cholesky of $\chi^2 = 2.32$ with 11 df's, RMSEA = .000, AIC = −17.00; the two models are indistinguishable using a χ^2 differences test—$\chi^2 = 1.08$ with 4 df's, $p > .80$, leading to adoption of the reduced model). The best fitting model included common genetic variance, and common shared environmental variance for the two BP variables. This model suggests that essentially all of the genetic and environmental variance in BP is shared in common across childhood and early adolescence. It is notable that, although the model contained within it the ability to account for temporal precedence—i.e., the potential for BP_6-9 to be causal in relation to BP_12-14—the model simply identified common factors underlying these two measures.

This analysis was replicated using CP_6-9 and CP_12-14, and the result was only slightly different, though in interesting ways. In both cases, the model became simpler
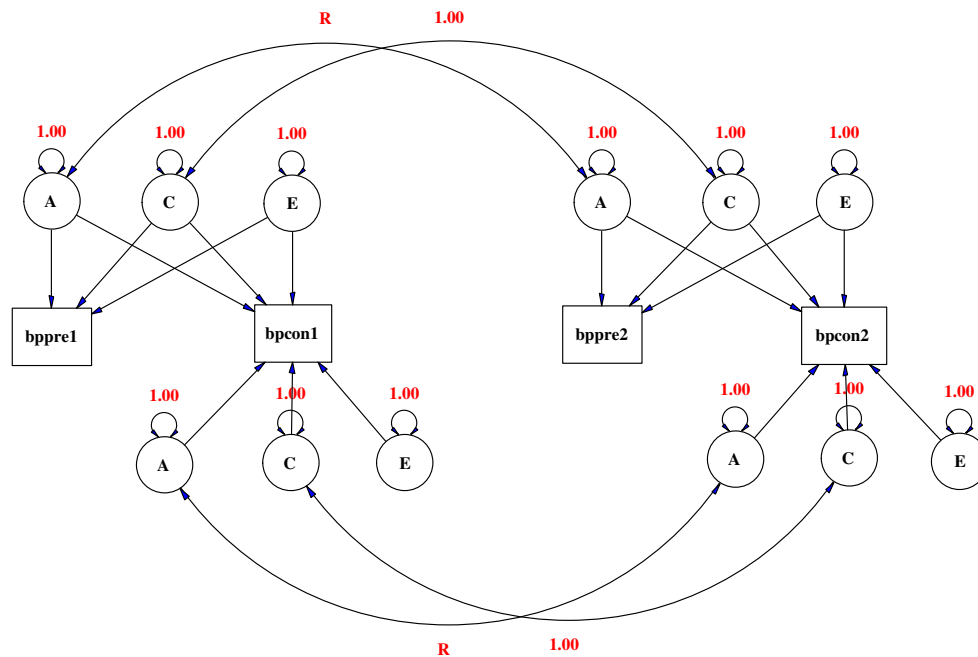
Fig. 2 Cholesky model of bivariate relationship between behavior problems pre-menarche, ages 6–9 (bppre) and behavior problems approximately concurrent with menarche, ages 12–14 (bpcon); bppre1 and bpcon1 refer to the first member of the kinship pair, and bppre2 and bpcon2 refer to the second member of the kinship pair (bppre is BP6-9/CP6-9, bpcon is BP12-14/CP12-14); A is an additive genetic source of variance, C is a shared environmental source of variance, and E is a nonshared environmental source of variance/measurement error

and more interpretable (though the fit measures were not as good). For the early menarche category, the genetic factors dropped out of the model, and only the common shared environmental factor (and both shared and unique e^2) remained in the model (Fig. 4a). The fit was not especially good, $\chi^2 = 9.38$ with 15 df's, RMSEA = .09 and AIC = 4.16 (compare to $\chi^2 = 7.59$ with 10 df's, RMSEA = .12 and AIC = 10.16 in the overall model; a χ^2 difference test leads to adoption of the reduced model). For the later menarche category, only the common shared genetic factor remained in the model, along with e^2 for the BP_6-9 measure (Fig. 4b). The fit was slightly better, though marginal; $\chi^2 = 29.80$ with 17 df's, RMSEA = .08, AIC = 18.65 (compare to $\chi^2 = 17.93$ with 10 df's, RMSEA = .11 and AIC = 24.34 in the overall model; the χ^2 difference test led to adoption of the reduced model).

Discussion

Summary of findings

The interpretation of our results is leveraged by a set of two critical empirical findings. First (and consistent with several previous studies), we found no correlation between BP/conduct disorder and timing of menarche. This result should be interpreted in the context of two recent review articles (Mendle et al. 2007; Negriff and Susman 2011), who developed a framework for the role of pubertal timing (in both sexes for Negriff and Sussman) and subsequent outcomes. Even those articles, in reviewing consistent patterns, note inconsistencies in these relationships (e.g., Negriff and Sussman summarized their findings by noting that "The relationship between pubertal timing and psychosocial adjustment is not straightforward, as indicated by the inconsistencies noted in the previous sections" p. 733). Our findings raise doubt as to the legitimacy of an effort to untangle the causal alternatives a–c in our introduction. There is little justification to develop models to explain the absence of correlational patterns. Second (and expanding on previous findings), timing of menarche plays a strong moderating role when developmental variance is partitioned biometrically; alternative d in our introduction is supported, and is even more interesting in the absence of direct causal paths between problem behaviors and MT. Specifically, NLSYC respondents measured at childhood or young adolescence who reach menarche at typical or later ages have the majority (for BP) or all (for conduct disorder) of the variance accounted for by shared genetic variance. NLSYC respondents who reach menarche early have all of that variance (for both BP and CD) accounted for by the shared environment.
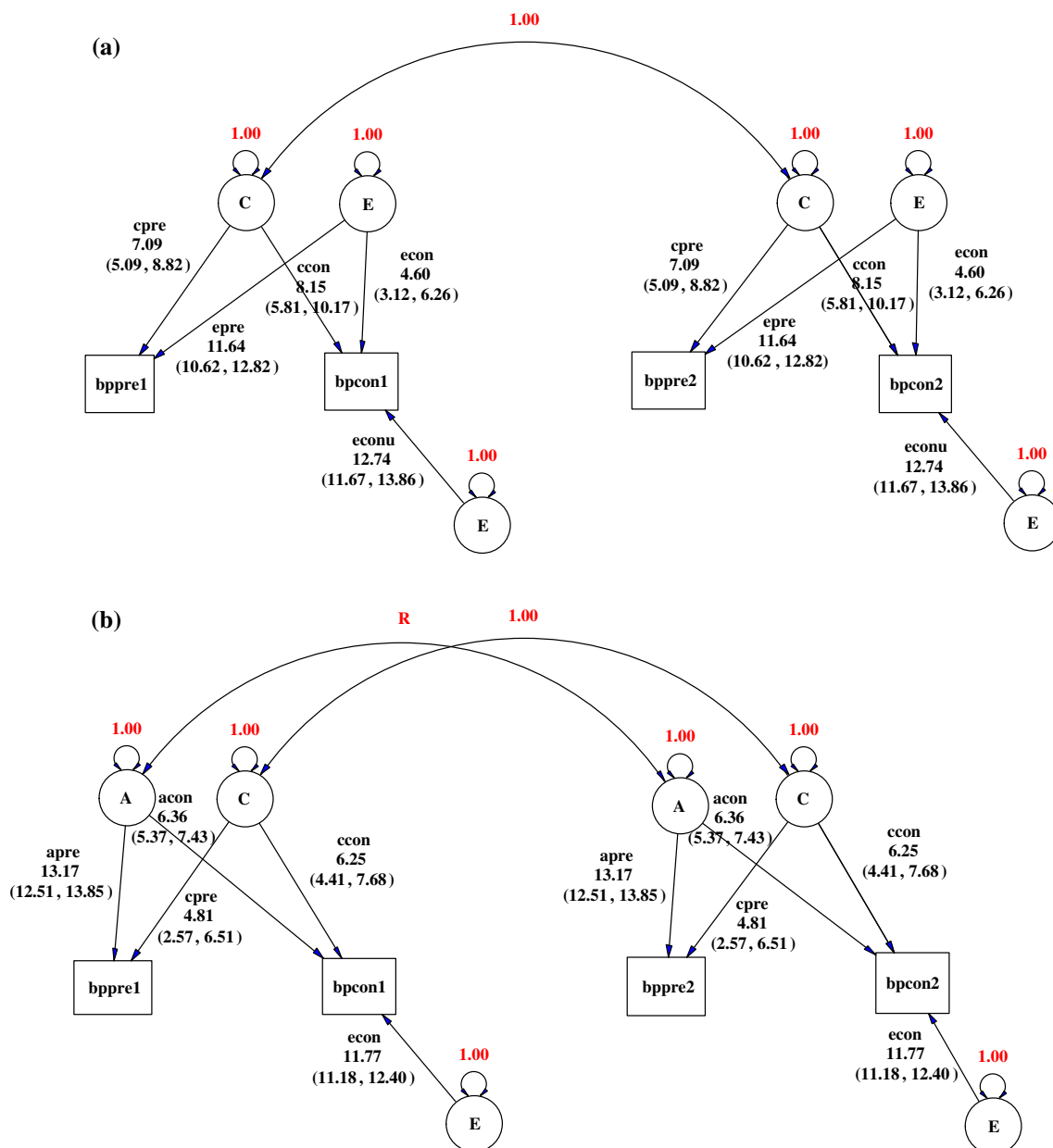
Fig. 3 a Best submodel, early-menarche subgroup, using overall BPI to measure behavior problems before and concurrently with menarche (bppre is BP6-9, bpcon is BP12-14; path coefficients that begin with a are additive genetic coefficients, with c are shared environmental coefficients, and with e are nonshared environmental/measurement error coefficients). **b** Best submodel, average/late-menarche subgroup, using overall BPI to measure behavior problems before and concurrently with menarche (bppre is BP6-9, bpcon is BP12-14; path coefficients that begin with *a* are additive genetic coefficients, with *c* are shared environmental coefficients, and with *e* are nonshared environmental/measurement error coefficients)

Relationship between our results and the Burt results

Our results are a substantial extension of those from Burt et al. (2006). Each set of findings are validated by the close match with one another (though their delayed menarche finding did not replicate in our data). Their design was not longitudinal, and so ours expands the developmental interpretation of the findings. Further, our dataset

originated as a national probability sample, with better external validity and a more diverse sample. Finally, the family design of our study, compared to Burt et al.'s twin design, supports expanded interpretation.

Conduct disorder in the Burt et al. (2006) study was measured approximately concurrently with menarche, so that their findings can be compared to those in our Tables 2 and 3 associated with our BP_12-14 and CP_12-14
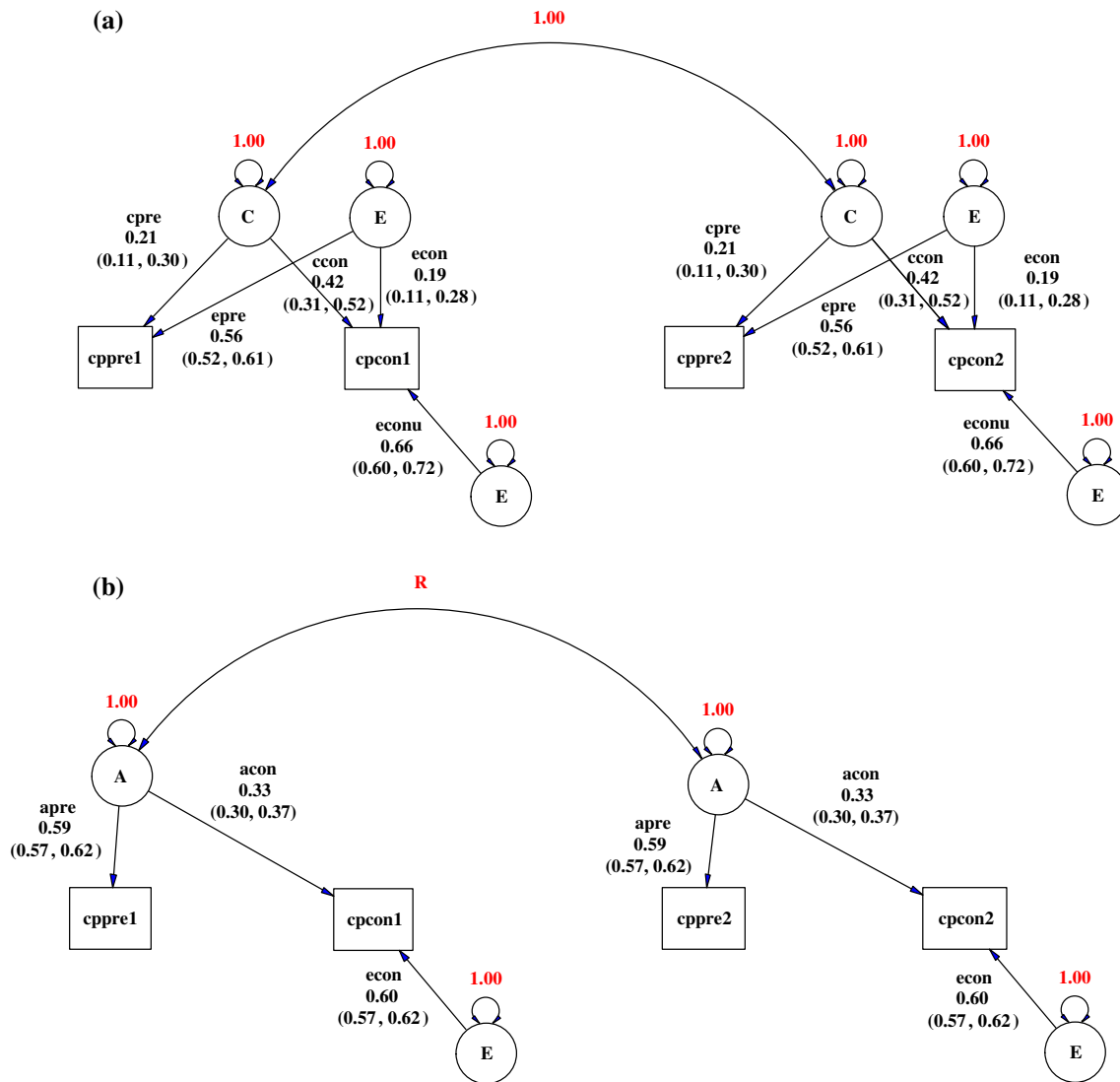
Fig. 4 a Best submodel, early-menarche subgroup, using conduct problems subscale to measure behavior problems before and concurrently with menarche (cppre is CP6-9, cpcon is CP12-14; path coefficients that begin with *a* are additive genetic coefficients, with *c* are shared environmental coefficients, and with *e* are nonshared environmental/measurement error coefficients). **b** Best submodel, average/late-menarche subgroup, using conduct problems subscale to measure behavior problems before and concurrently with menarche (cppre is CP6-9, cpcon is CP12-14; path coefficients that begin with *a* are additive genetic coefficients, with *c* are shared environmental coefficients, and with *e* are nonshared environmental/measurement error coefficients)

variables. Our finding of univariate shared environmental variance underlying early menarche and genetic variance underlying typical MT exactly matches their pattern of results. We also replicate the univariate young adolescent results for childhood (but not for late adolescence). We also find shared variance in a multivariate developmental design as well as in a univariate design.

Threats to validity

Before we provide theoretical interpretation, we review several threats to internal validity within our quasi-

experimental design. The first threat is the problem of selection bias. Because the NLSYC females in this study necessarily were born to younger mothers than the NLSYC females who are still too young to have their menarche and adolescent delinquency measured, our findings apply to daughters of mothers younger than the population average. In and of itself, this appears to be more of a threat to external than to internal validity (and many studies have made greater sacrifice to obtain a convenience sample; e.g., virtually all recruited twin samples fall into this category). However, because we are studying the distinction between early menarche and average/later menarche, if the selection

bias in the NLSYC impacts the distributions of these two groups, internal validity would be threatened. This threat is partially answered by age-standardizing using national norms. In addition, the similarity between the Burt et al. and our findings helps address this concern. In a few years when all of the NLSYC respondents reach late adolescence, the selection bias can be addressed empirically.

The second and third threats to internal validity that we address involve issues of statistical conclusion validity. Correlations are attenuated under range restriction. If the effect of defining the early menarche group creates restriction on the behavior problem distributions, then there simply may not be as much variance to explain, and thus the reduced heritabilities could be an artifact. We offer three responses. First, delayed menarche was defined with approximately the same range restriction as the early group, and did not show a reduction in heritability (though it did in Burt et al. 2006). Second, the shared environmental variances, which would potentially be affected by range-restricted correlations, were higher in the early menarche category. Third, we can observe the phenotypic variances directly. The early menarche group had standard deviations of $SD_{BP\_6-9} = 13.6$, $SD_{BP\_12-14} = 16.3$, and $SD_{BP\_15-17} = .96$. The larger (N = 650 vs. N = 240) and less restricted average/delayed menarche group had standard deviations of $SD_{BP\_6-9} = 14.0$, $SD_{BP\_12-14} = 14.7$, and $SD_{BP\_15-17} = 1.01$. The menarche category definitions do not appear to be creating substantial problems with range restriction.

The second statistical issue is a problem defined in relation to estimating biometrical cholesky models. A challenge has been raised involving boundary conditions, such that the ML χ^2 statistic may not be distributed as a χ^2 under the null (Carey 2005; Dominicus et al. 2006; also Self and Liang 1987). We note that using DF Analysis avoids this problem, because negative variances can be estimated (and inform as to violations of the model), and these models are fit with least squares instead of ML. Our cholesky models were estimated using ML optimization, but we used two additional fit statistics to evaluate model fit. The sequence of model adjustments was equivalent using ML, the AIC, and the RMSEA.

Next, we note a minor concern over the timing of our measures. The measure of BP approximately concurrently with menarche is obviously later in relation to puberty for girls who reached menarche early than those who reached menarche on time or late. This problem, however, is at least partially addressed by our developmental design, and by the 3 year age window. We did not, for example, observe differences in standardized BP/conduct disorder within years across the three-year windows.

Finally, we address the assumptions of the biometrical model as threats to internal validity. We have treated the EEA problem that involved including cousins by evaluating results with the cousins excluded; findings were equivalent. Second, the several c^2 estimates that were negative were not large enough negative to arouse concern that dominance or some other nonlinear genetic confound (e.g., epistasis) is a problem. Third, the assumption of vanishing assortative mating, which is probably incorrect (though perhaps not substantially so), has a known effect on patterns of findings. Significant heritabilities in the presence of non-ignorable assortative mating would be partially attributable to this genetic source. However, in our critical comparison, there would have to be differential assortative mating in relation to BP for parents of girls who reached early menarche versus average/delayed menarche. Although this is not fully implausible, we also do not have any information suggesting that this is the case. We urge caution in relation to this issue. The NLSY links between mothers and daughters may be useful to help address this type of cross-generational issue in future research (see Rodgers et al. 2008a).

Implication of these results for theories of early family environment influences

In this last section, we interpret our findings in relation to earlier theories linking the family environment, BP, and timing of menarche. We believe that the literature signals the compelling findings that we report in this paper. The absence of a correlation between MT and child BP obviates any interpretation that the family environment is a direct causal agent in the relationship between the development of problem behaviors and early menarche, because there is no direct relationship. If a common causal source did influence both (family stress, parental discipline, or longer-term family culture), moderate correlations would have been observed. Our results do not appear consistent with the Belsky et al. (1991) direct-causal link hypothesis (though we note that we did not measure family stress directly), and continue a longstanding and growing literature supporting an indirect selection model (also see Comings et al. 2002; Jorm et al. 2004; Mendle et al. 2006; Rowe 2002). It is difficult to develop a behavioral model of influence on MT—evolutionary or otherwise—within which family stress was critical, but which would leave childhood BP and CP entirely out of the explanation. However, there exists an intriguing apparent inconsistency, waiting for an explanation. Father's absence consistently correlates with MT, whereas childhood BP do not. As one hint in understanding this causal chain, we note that father's absence—related to divorce—has a more proximal link to MT, because both are related to sexual and reproductive behavior. (We note, however, this link is not the topic of our paper.)

One of our most intriguing findings—and one that required a developmental design to detect—is that the difference between girls who will eventually experience early menarche is already defined in childhood BP. But this difference is not carried through a straightforward link from individual differences in BP to timing of menarche. Rather, it required a biometrical decomposition to detect it, and then it is quite striking. The girls who will eventually go through early menarche are different, at an early age, from those who will not, because their shared family environment affects their BP. For those who will eventually have typical or late MT, their shared genes affect their BP. These differences are still detectable in early adolescence, and have virtually complete shared variance with the childhood patterns. They substantially disappear by later adolescence, when individual differences for both categories are largely genetic.

This finding has a relationship to a theory proposed to explain human fertility patterns. This theory explains why genetic influences, though potentially expressible, lie latent and unexpressed because of the dominance of family environmental influences. Udry (1996); (Rodgers et al. 2007); and (Rodgers et al. 2008a, b) discuss this interpretation in relation to fertility, and research has documented that timing of menarche relates to adolescent sexual and fertility behavior (Flannery et al. 1993; Phinney et al. 1990). In natural fertility settings, in which universal childbearing is socially normative, genetically-based individual differences in fertility motivation cannot be expressed. In contracepting societies, however, these genetically-based sources can emerge.

We posit that this same explanation may exist in relation to behavior/conduct problems in families with both a genetic and cultural tradition of BP. If such BP are normative, consistently expressed, and consistently evoked, typical individual differences emerging from genetic variance might not be easily expressed. In such families, even relatively quiet and docile children may express BP in relation to their volatile social environment, even if only as a protective mechanism against actual physical harm. In contrast, in a less volatile family, their natural genetic tendencies toward problem behaviors could be expressed. Thus, only when children and adolescents have a broader behavioral repertoire—the ability to choose whether to engage in BP or not, with a breadth of both potential behavior and potential responses—can the genetic variance be expressed.

The question of whether such differences arise directly from the specific family environment (Belsky et al. 1991), or through a selection process (e.g., Mendle et al. 2006) is difficult to fully resolve, because there are many confounded features of the two explanations. Mendle et al's Children-of-Twins design is a powerful method to control many of these confounds, however, and their support for the selection model must be taken seriously. Our findings are also consistent with a selection model—one in which long-term cross-generational patterns are primarily environmental for one type of family, and genetic for another. In fact, we note that these two different models do not necessarily compete. It is both plausible and logical that the selection model could be occurring, and some residual direct influence of the family environment could be occurring as well. However, though plausible, the consistent zero correlations between BP in children/adolescents and timing of menarche is difficult to reconcile with the direct interpretation that emerges from the Belsky et al. theory.

Rather, we propose that a theory in which genetic variability can be potentially expressed in some families, but not others, is consistent with both sets of findings. The theory also provides a mechanistic explanation to at least partially explain the Belsky et al. (2007) differential susceptibility hypothesis. This theory is consistent with our zero correlations, because there is no reason to expect correlations between behavior problem and timing of menarche within the categories. But differences between categories can be explained by the difference in diversity of choice available across the two categories.

Conflict of interest The authors declare that they have no conflict of interest.

References

Achenbach TM (1978) The child behavior profile: I. Boys aged 6–11. J Consult Clin Psychol 46:478–488

Bard D, Rodgers J L (2006) Use of discrete-time survival analysis for modeling multivariate ACE models of fertility precursors from the children of the NLSY. Poster presented at the June (2006) meetings of the Behavior Genetic Association. Storrs, CN

Belsky J (2000) Conditional and alternative reproductive strategies: Individual differences in susceptibility of rearing experiences. In: Rodgers JL, Rowe DC, Miller WB (eds) Genetic Influences on Human Fertility and Sexuality. Kluwer, Boston, pp 127–146

Belsky J, Steinberg L, Draper P (1991) Childhood experience, interpersonal development, and reproductive strategy: an evolutionary theory of socialization. Child Dev 62:647–670

Belsky J, Bakermans-Kranenburg, van IJzendoorn MH (2007) For better or for worse: differential susceptibility to environmental influences. Curr Dir Psychol Sci 16:300–304

Box GEP (1950) Problems in the analysis of growth and linear curves. Biometrika 6:362–389

Brooks-Gunn J (1988) Antecedents and consequences of variations in girls maturational timing. J Adolesc Health Care 9:365–373

Brooks-Gunn J, Warren MP, Rosso J, Gargiulo J (1987) Validity of self-report measures of girls' pubertal status. Child Dev 58:829–841

Burt SA, McGue M, DeMarte JA, Krueger RF, Iacono WG (2006) Timing of menarche and the origins of conduct disorder. Arch Gen Psychiatry 63:890–896

Button TMM, Scourfield J, Martin N, Purcell S, McGuffin P (2005) Family dysfunction interacts with genes in the causation of antisocial symptoms. Behav Genet 35:115–132

Carey G (2005) Cholesky problems. Behav Genet 35:653–665

Caspi A, Moffitt TE (1991) Individual differences are accentuated during periods of social change: the sample case of girls at puberty. J Pers Soc Psychol 61:157–168

Caspi A, Lynam D, Moffitt TE, Silva PA (1993) Unraveling girls' delinquency: biological, dispositional, and contextual contributions to adolescent misbehavior. Dev Psychol 29:19–30

Celio M, Karnik NS, Steiner H (2006) Early maturation as a risk factor for aggression and delinquency in adolescent girls: a review. Int J Clin Pract 10:1254–1262

Center for Human Resources Research (2006) NLSY79 Child & Young Adult Data Users Guide. Columbus, OH, Center for Human Resource Research

Chase-Lansdale PL, Mott FL, Brooks-Gunn J, Phillips DA (1991) Children of the National Longitudinal Survey of Youth: a unique research opportunity. Dev Psychol 27:918–931

Cherny SS, DeFries JC, Fulker DW (1992) Multiple regression of twin data: a model-fitting approach. Behav Genet 22:489–497

Chisolm JS, Quinlivan JA, Petersen RW, Coall DA (2005) Early stress predicts age at menarche and first birth, adult attachment, and expected lifespan. Hum Nat 16:233–265

Comings DE, Muhleman D, Johnson JP, MacMurray JP (2002) Parent-daughter transmission of the androgen receptor gene as an explanation of the effect of father absence on age of menarche. Child Dev 73:1046–1051

Conradt E, Measelle J, Ablow JC (2013) Poverty, problem behavior, and promise: differential susceptibility among infants reared in poverty. Psychol Sci 24:235–242

Damon A, Damon ST, Reed RB, Valadian I (1969) Age at menarche of mothers and daughters with a note on accuracy of recall. Hum Biol 38:255–273

Deardorff J, Kushi L, Ekwaru JP, Ellis BJ, Greenspan LC, Mirabedi BA, Landaverde MA, Hiatt RA (2011) Father absence, body mass index, and pubertal timing in girls: differential effects by family income and ethnicity. J Adolesc Health 48:441–447

DeFries JC, Fulker DW (1985) Multiple regression analysis of twin data. Behav Genet 15:467–473

Dick DM, Rose R, Viken RJ, Kaprio J (2000) Pubertal riming and substance use: associations between and within families across late adolescence. Dev Psychol 36:180–189

Dominicus A, Skrondal A, Gjessing HK, Pedersen NL, Palmgren J (2006) Likelihood ratio tests in behavioral genetics: problems and solutions. Behav Genet 36:331–340

D'Onofrio BM, Van Hulle CA, Waldman ID, Rodgers JL, Harden KP, Rathouz PJ, Lahey BB, Dick DM (2008) Smoking during pregnancy and offspring externalizing problems: an exploration of genetic and environmental confounds. Dev Psychopathol 20:139–164

Doughty D, Rodgers JL (2000) Behavior genetic modeling of menarche in US females. In: Rodgers JL, Rowe DC, Miller WB (eds) Genetic Influences on Human Fertility and Sexuality. Kluwer, Boston, pp 169–181

Draper P, Harpending H (1982) Father absence and reproductive strategy: an evolutionary perspective. J Anthropol Res 38:255–273

Ellis BJ (2004) Timing of maturation in girls: an integrated life history approach. Psychol Bull 130:920–958

Falconer DS (1979) Introduction to quantitative genetics. Longman, New York

Flannery DJ, Rowe DC, Gulley BL (1993) Impact of pubertal status, timing, and age on adolescent sexual experiences and delinquency. J Adolesc Res 8:21–40

Ge X, Natsuaki MN (2009) In search of explanations for early pubertal timing effects on developmental psychopathology. Curr Dir Psychol Sci 18:327–331

Ge X, Brody GH, Conger RD, Simons RL, Murry VM (2002) Contextual amplification of pubertal transition effects on deviant peer affiliation and externalizing behavior among African American children. Dev Psychol 38:42–54

Goodnight JA, Lahey BB, Van Hulle CA, Rodgers JL, Rathouz PJ, Waldman ID, D'Onofrio BM (2012) A quasi-experimental analysis of the influence of neighborhood disadvantage on child and adolescent conduct problems. J Abnorm Psychol 121:95–108

Harden KP, Mendle J (2012) Gene-environment interplay in the association between pubertal timing and delinquency in adolescent girls. J Abnorm Psychol 121:73–87

Hoier S (2003) Father absence and age at menarche: a test of four evolutionary models. Hum Nat 14:209–233

Jinks JL, Fulker DW (1970) A comparison of the biometrical-genetical, MAVA and classical approaches to the analysis of human behavior. Psychol Bull 73:311–349

Jorm AF, Christensen H, Rodgers B, Jacomb PA, Easteal S (2004) Association of adverse childhood experiences, age of menarche, and adult reproductive behavior: does the androgen receptor gene play a role? Am J Med Gen 125(B):105–111

Kaprio J, Rimpela A, Winter T, Viken RJ, Rimpela M, Rose RJ (1995) Common genetic influence on BMI and age at menarche. Hum Biol 67:739–753

Kohler H-P, Rodgers JL (1999) DF analysis of binary, ordered, and censored variables using probit and tobit approaches. Behav Genet 29:221–232

Kohler H-P, Rodgers JL (2001) DF analysis of heritability with double-entry twin data: asymptotic standard errors and efficient estimation. Behav Genet 31:179–191

Krueger RF, Moffitt TE, Caspi A, Bleske A, Silva PA (1998) Assortative mating for antisocial behavior. Behav Genet 28:173–186

LaBuda MC, DeFries JC, Fulker DW (1986) Multiple regression analysis of twin data obtained from selected samples. Genet Epidemiol 3:425–433

Lahey BB, Waldman I (2003) A developmental propensity model for the origins of conduct problems during childhood and adolescence. In: Lahey BB, Moffitt TE, Caspi A (eds) Causes of Conduct Disorder and Serious Delinquency. Guilford Press, New York, pp 76–117

Lahey BB, Van Hulle CA, Wadman ID, Rodgers JL, D'Onofrio BM, Pedlow S, Rathouz P, Keenan K (2006) Testing descriptive hypotheses regarding sex differences in the development of conduct problems and delinquency. J Abnorm Psychol 34:737–755

Lanza ST, Collins LM (2002) Pubertal timing and the onset of substance abuse in females during early adolescence. Prev Sci 3:69–82

Lykken DT, McGue M, Tellegen A (1987) Recruitment bias in twin research: the rule of two-thirds reconsidered. Behav Genet 17:343–362

Maccoby EE (1991) Different reproductive strategies in males and females. Child Dev 62:676–681

Maes HH, Neale MC, Kendler KS, Hewitt JK, Silberg UL, Foley DL, Meyer JM, Rutter M, Simonoff E, Pickles A, Eaves LJ (1998) Assortative mating for major psychiatric diagnoses in two population-based samples. Psychol Med 28:1389–1401

Mendle J, Turkheimer E, D'Onofrio BM, Lynch SK, Emery RE, Slutske WS, Martin NG (2006) Family structure and age at menarche: a children-of-twins approach. Dev Psychol 42:533–542

Mendle J, Turkheimer E, Emery RE (2007) Detrimental psychological outcomes associated with early pubertal timing in adolescent girls. Dev Rev 27:151–171

Meyer JM, Eaves LJ, Heath AC, Martin NG (1991) Estimating genetic influences on the age-at-menarche: a survival analysis approach. Am J Med Genet 39:148–154

Mittler P (1971) The Study of Twins. Penguin, Harmondsworth, Middlesex, England

Moffitt TE (1993) Adolescent limited and life-course persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Moffitt TE (2005) The new look of behavior genetics in developmental psychpathology: gene-environment interplay in antisocial behaviors. Psychol Bull 131:533–554

Moffitt TE, Caspi A, Belsky J, Silva PA (1992) Childhood experience and the onset of menarche: a test of a sociobiological model. Child Dev 63:47–58

Mustanski BS, Viken RJ, Kaprio J, Pulkkinen L, Rose RJ (2004) Genetic and environmental influences on pubertal develpoment: longitudinal data from Finnish twins at ages 11 and 14. Dev Psychol 40:1188–1198

Neale MC, Cardon LR (1992) Methodology for Genetic Studies of Twins and Families. Kluwer, Dordrecht, The Netherlands

Negriff S, Susman EJ (2011) Pubertal timing, depression, and externalizing problems: a framework, review, and examination of gender differences. J Res Adolesc 21:717–746

Pearl J (2000) Causality: Models, Reasoning, and Inference. Cambridge Press, Cambridge

Peterson JL, Zill N (1986) Marital disruption, parent-child relationships, and behavior problems in children. J Marriage Fam 48:295–307

Phinney VC, Jenson LC, Olsen JA, Cundick B (1990) The relationship between early development and psychosexual behaviors in adolescent females. Adolescence 25:321–332

Pickles A, Crouchley R, Simonoff E, Eaves L, Meyer J, Rutter M, Hewitt J, Silberg J (1994) Survival models for developmental genetic data: age of onset of puberty and antisocial behavior in twins. Genet Epidemiol 11:155–170

Pickles A, Pickering K, Simonoff E, Silberg J, Meyer J, Maes HJ (1998) Genetic "clocks" and "soft" events: a twin model for pubertal develpment and other recalled sequences of develpmental milestones, transitions, or ages at onset. Behav Genet 28:243–253

Plomin R (1990) Nature and Nurture: An Introduction to Human Behavioral Genetics. Brooks/Cole, Pacific Grove

Plomin R, DeFries JC, McClearn GE (1990) Behavioral Genetics: A Primer. W. H, Freeman and Company, New York

Posner RB (2006) Early menarche: a review of research on trends in timing, racial differences, etiology, and psychosocial consequences. Sex Roles 54:315–322

Purcell S, Koenen KC (2005) Environmental mediation and the twin design. Behav Genet 35:491–498

Repetti RL, Taylor SE, Seeman TE (2002) Risky families: family social environments and the mental and physical health of offspring. Psychol Bull 128:330–366

Rhee SH, Waldman ID (2002) Genetic and environmental influences on antisocial behavior: a meta-analysis of twin and adoption studies. Psychol Bull 128:490–529

Rodgers JL, Kohler H-P (2005) Reformulating and simplifying the DF analysis model. Behav Genet 35:211–217

Rodgers JL, McGue M (1994) A simple algebraic demonstration of the validity of DeFries–Fulker analysis in unselected samples with multiple kinship levels. Behav Genet 24:259–263

Rodgers JL, Rowe DC (1988) The influence of siblings on adolescent sexual behavior. Dev Psychol 24:722–728

Rodgers JL, Rowe DC, Harris D (1992) Sibling differences in adolescent sexual behavior: inferring process models from family composition patterns. J Marriage Fam 54:142–152

Rodgers JL, Rowe DC, Li C (1994a) Beyond nature versus nurture: DF analysis of nonshared influences on problem behaviors. Dev Psychol 30:374–384

Rodgers JL, Rowe DC, May K (1994b) DF analysis of NLSY IQ-achievement data: nonshared environmental influences. Intelligence 19:157–177

Rodgers JL, Buster M, Rowe DC (2001a) Genetic and environmental influences on delinquency: DF analysis of NLSY kinship data. J Quant Criminol 17:145–168

Rodgers JL, Kohler H-P, Kyvik K, Christensen K (2001b) Behavior genetic modeling of human fertility: findings from a contemporary Danish twin study. Demography 38:29–42

Rodgers JL, Johnson AB, Bard DE (2005) NSLY-Children/Young Adult (1986–2000) kinship linking algorithm. Unpublished manuscript. University of Oklahoma, Norman

Rodgers JL, Bard DE, Miller WB (2007) Multivariate cholesky models of human female fertility patterns in the NLSY. Behav Genet 37:345–361

Rodgers JL, Bard DE, Johnson A, D'Onofrio B, Miller WB (2008a) The cross-generational mother-daughter-aunt-niece design: establishing validity of the MDAN design with NLSY fertility variables. Behav Genet 38:567–578

Rodgers JL, Kohler H-P, McGue M, Behrman JR, Petersen I, Bingley P, Christensen K (2008b) Education and cognitive ability as direct, mediating, or spurious influences on female age at first birth: behavior genetic models fit to Danish twin data. Am J Sociol 114(Supplement):S202–S232

Rowe DC (2000) Environmental and genetic influences on pubertal development: Evolutionary life history traits? In: Rodgers JL, Rowe DC, Miller WB (eds) Genetic Influences on Human Fertility and Sexuality. Kluwer, Boston, pp 147–168

Rowe DC (2002) On genetic variation in menarche and age at first sexual intercourse: a critique of the Belsky–Draper hypothesis. Evol Hum Behav 23:365–372

Rowe DC, Rodgers JL (1994) A social contagion model of adolescent sexual behavior: explaining race differences. Soc Biol 41:1–18

Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 8:605–610

Shadish WR, Cook TD, Campbell DT (2002) Experimental and Quasi-experimental Designs for Generalized Causal Inference. Houghton-Mifflin, Boston

Stice E, Presnell K, Bearman SK (2001) Relation of early menarche to depression, eating disorders, substance abuse, and comorbid psychopathology. Dev Psychol 37:608–619

Stoolmiller M (1999) Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior-genetic adoption studies. Psychol Bull 125:392–409

Taylor J, McGue M, Iacono WG (2000) Sex differences, assortative mating, and cultural transmission effects on adolescent delinquency: a twin family study. J Child Psychol Psychiatry 41:433–440

Timm N (1975) Multivariate Analysis with Applications in Education and Psychology. Brooks/Cole, Monterrey, CA

Udry JR (1996) Biosocial models of low-fertility societies. In: Casterline JB, Lee RB, Foote KA (eds) Fertility in the United States: New Patterns, New Theories. The Population Council, New York

van den Oord EJCG (2001) Estimating effects of latent and measured genotypes in multilevel models. Stat Methods Med Res 10:393–407

van den Oord EJCG, Rowe DC (2000) Racial differences in birth health risk: a quantitative genetic approach. Demography 37: 285–298

Van Hulle CA, Rodgers JL, D'Onofrio BM, Waldman ID, Lahey BB (2007) Sex differences in the causes of self-reported adolescent delinquency. J Abnorm Psychol 116:236–248

Van Hulle CA, Waldman ID, D'Onofrio BM, Rodgers JL, Rathouz PJ, Lahey BB (2009) Developmental structure of genetic influences on antisocial behavior across childhood and adolescence. J Abnorm Psychol 118:711–721

Waldman ID, DeFries JC, Fulker DW (1992) Quantitative genetic analysis of IQ development in young children: multivariate multiple regression with orthogonal polynomials. Behav Genet 22:229–238

Waller NB (1994) A DeFries and Fulker regression model for genetic nonadditivity. Behav Genet 24:149–153