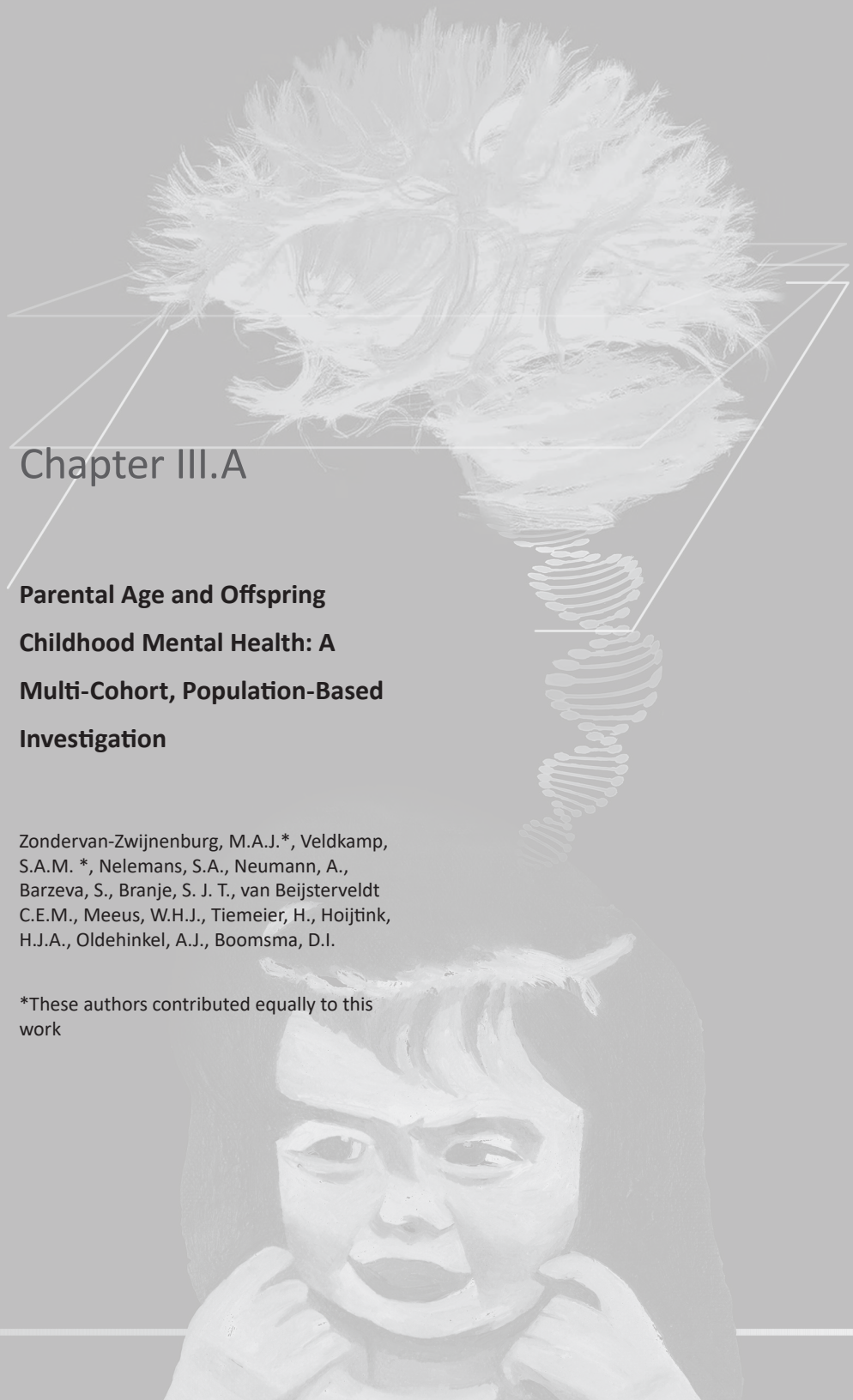

Chapter III.A

Parental Age and Offspring Childhood Mental Health: A Multi-Cohort, Population-Based Investigation

Zondervan-Zwijnenburg, M.A.J.*, Veldkamp, S.A.M. *, Nelemans, S.A., Neumann, A., Barzeva, S., Branje, S. J. T., van Beijsterveldt C.E.M., Meeus, W.H.J., Tiemeier, H., Hoijtink, H.J.A., Oldehinkel, A.J., Boomsma, D.I.

*These authors contributed equally to this work

ABSTRACT

To examine the contributions of maternal and paternal age on offspring externalizing and internalizing problems, this study analyzed problem behaviors at age 10-12 years from four Dutch population-based cohorts (N = 32,892) by a multiple informant design. Bayesian evidence synthesis was used to combine results across cohorts with 50% of the data analyzed for discovery and 50% for confirmation. There was evidence of a robust negative linear relation between parental age and externalizing problems as reported by parents. In teacher-reports, this relation was largely explained by parental socio-economic status. Parental age had limited to no association with internalizing problems. Thus, in this large population-based study, either a beneficial or no effect of advanced parenthood on child problem behavior was observed.

Since 1995, the mean maternal age at first birth has increased at a rate of 0.10 years per year in OECD countries, and in 2017 exceeded 30 years in the vast majority of these countries (Organisation for Economic Co-operation and Development, 2017). Only in Mexico was the mean age of women at childbirth lower than 28 years, and only in eight countries was it between 28 and 30 years of age. Women's reproductive years generally range from about 15 to 45 years.[1] Within this wide age range some periods are generally considered more suitable to have children than others, but which parental reproductive ages are optimal for offspring physical and mental health has been a matter of debate ever since individuals have engaged in active birth control. Whereas having children at an advanced age was quite common historically, when families tended to be larger[2], the current trend to delay childbearing has given rise to public health concerns

.

Concerns Regarding Delayed Childbearing

Concerns regarding delayed childbearing are understandable, as a large number of research reports highlight that increased maternal age at childbirth is associated with several adverse consequences, ranging from physical problems such as increased BMI, blood pressure and height[3] to psychiatric conditions such as autism[4,5], bipolar disorder[6], symptoms of depression, anxiety and stress[7], and poor social functioning[8]. More recently, increased paternal age at birth has also been associated with adverse child outcomes, such as stillbirth and cleft palate.[9] In over 40 million live births between 2007 and 2016, having an older father increased the risk of low birthweight, apgar score, and premature birth.[10] A study of the Danish population, which included 2.8 million persons, found that older fathers are at risk of having offspring with intellectual disabilities, autism spectrum disorders and schizophrenia.[11,12]

Several, not mutually exclusive, mechanisms have been proposed to explain the increased physical and mental health risks in offspring of older parents. First, age-related deterioration of the functioning of women's reproductive organs, such as DNA damage in germ cells, and worse quality of oocytes and placenta, can increase the risk of obstetric and perinatal complications.[13] Second, male germline cells undergo cell replication cycles repeatedly during aging, with de novo point mutations accumulating over time[14] and the number of de novo mutations in the newborn increasing with higher age of the father at the time of conception[15,16]. Although weaker than with paternal age, de novo mutations in offspring correlate with maternal age as well.[17,18] Third, genomic regions in the male germline may become less methylated with increasing age and alter the expression of health-related genes.[19] Fourth, age effects can be due to selection, with older parents differing from younger ones in characteristics that are relevant for developmental outcomes in their offspring, such as poor social skills. The influence of selection effects can be exacerbated by assortative mating.[20] Fifth, being the child of older parents carries the risk of having to cope with parental frailty or losing a parent at a relatively young age,[21] and the stress evoked by these experiences may trigger health problems. Most of these mechanisms involve consequences of biological ageing. Par-

enthood at an advanced age is disadvantageous from a biological perspective; except for very young, physiologically immature mothers, younger parents are in a better physical condition.

Possible Benefits of Delayed Childbearing

Whereas the effects of older parental age on children's physical health and psychiatric disorders tend to be predominantly negative, the effects of older parental age on mental health problems with a stronger psychosocial component, such as externalizing and internalizing problems, tend to be more inconsistent. An indication that the negative consequences of high parental age may stretch beyond clinical diagnosis is provided by Tearne and colleagues,[7,22] who found that high maternal age predicted symptoms of depression, anxiety and stress in daughters, and by Janecka and colleagues[23] who reported a negative association between advanced paternal age and social development. In contrast, in several population-based studies, offspring of older parents, particularly of older mothers, perform better at school and work, score higher on intelligence tests, report better health and higher well-being, use fewer drugs, and have fewer behavioral and emotional problems than offspring of younger parents.[3,11,21,22,24,25]

While the biology of ageing seems to put older parents in an unfavorable position with regard to their offspring's physical and mental health, these contradictory effects of parental age on offspring mental health outcomes might be explained by a psychosocial perspective. Being a child of older parents can have substantial benefits,[26] as older parents not only are often in a better socioeconomic position than young parents,[27] thereby providing a more favorable environment for children, they also have greater life experience. Furthermore, older parents display more hardiness[28] and tend to have less substance use and fewer mental health problems,[29] hence score higher on parenting factors that promote health and development.[29,30] In part, positive associations of advanced parental age could be related to selection effects. In young people, substance abuse and related externalizing problems go together with earlier sexual activity,[31] which increases the probability that intergenerational transmission of externalizing problems occurs at an early parental age.[32] Like age-related parental characteristics that may have negative effects on offspring outcomes, the influence of such selection effects can be exacerbated by assortative mating.[20]

In sum, whereas advanced parenthood, particularly advanced paternal age, has primarily been associated with physical health and neurodevelopmental outcomes, such as autism and schizophrenia, advanced parenthood, particularly advanced maternal age, rather seems to predict mental health problems with a stronger psychosocial component, such as externalizing problems. Although it seems plausible that parental age interferes with subclinical problems and traits underlying these conditions, comprehensive evidence from population-based cohorts is scarce and inconsistent, and more empirical evidence is desirable. Moreover, prior population-based studies that used continuous measures of mental health problems usually focused on cognitive or behavioral

problems[3,25] and, with a few exceptions that require replication in other cohorts[7,22,23] rarely included internalizing problems. A final reason to extent the research conducted thus far with the present study is the wide variety of populations, designs and outcomes used, which makes it hard to distinguish between substantive variation in association patterns and sample-specific artefacts. In short, there is a need for studies that investigate both maternal and paternal age effects on continuously assessed core dimensions of offspring mental health (including internalizing problems) and use robust analytical methods that allow the possibility of increased risk for both young and old parenthood.

The Present Study

We investigated parental age effects on offspring externalizing and internalizing problems around age 10-13 years in four Dutch population-based cohorts: Generation R (Gen-R), the Netherlands Twin Register (NTR), the Research on Adolescent Development and Relationships-Young cohort (RADAR-Y), and the Tracking Adolescents' Individual Lives Survey (TRAILS) (see Table 1). The Netherlands is characterized by a high maternal age at birth, and relatively few teenage pregnancies. In 1950, 1.6% of the children were born to mothers younger than 20 years of age, with a comparable percentage (1.7%) in 1990. In 2016 this number had decreased to 0.6%. In contrast, the percentages of women who gave birth at an age above 40 years were 8.5% in 1950, 1.5% in 1990, and 4.3% in 2016.[33]

As the perception of childhood problems may differ for different informants,[34,35] we aimed to obtain a comprehensive set of outcome measures of internalizing and externalizing problems through a multiple informant design. The four cohorts provided reports from mothers, fathers, the children themselves, and the children's teachers. The addition of reports from teachers is particularly valuable, because their reports are unlikely to be affected by parental age-related report biases. We tested both linear and nonlinear effects, to be better able to distinguish effects of older parenthood versus younger parenthood. We tested effects with and without adjusting for child gender and socio-economic status. Socio-economic status was included as a covariate to get an impression of the relative importance of socio-economic factors in explaining parental age effects.

Bayesian evidence synthesis was used to summarize the results over the cohorts. The current era is one of increased awareness of the need for replication research before making scientific claims.[36] Therefore, in this study, the datasets of the four cohort studies were used to evaluate the same set of hypotheses with respect to the relation

between parental age and offspring mental health problems. This approach is called Bayesian evidence synthesis.[37]

METHOD

Participants

The participants in this study came from the Gen-R, NTR, RADAR-Y, and TRAILS population cohort studies. Table 2 gives the total sample size and information on parental age for each cohort. The total number of children in each cohort was 4,769 for Gen-R, 25,396 for NTR, 497 for RADAR-Y, and 2,230 for TRAILS.

Gen-R mothers were recruited in the city of Rotterdam during pregnancy. Their partners and later their children were also invited to participate. For Gen-R, participants from the child age-10 study wave (born between 2002 and 2006) were included if they had complete information on maternal age and a child behavioral problems sum score by at least one informant. When multiple children from one family were present, one sibling was randomly removed (N = 397) to create a sample of unrelated individuals. Mean child age for mother report: 9.72 (SD = 0.32), father report: 9.77 (SD = 0.32), and child self-report: 9.83 (SD = 0.36). 71.2% of the Gen-R sample is Dutch or European. Other groups are Suriname (6.4%), Turkish (5.3%), and Moroccan (4.2%). Mother's educational level is low (i.e., no education or primary education) for 9%, intermediate (i.e., secondary school, vocational training) for 42%, and high (i.e., bachelor's degree, university) for 49%. Based on CBCL T-scores for mother reports, 93.2% of the children had non-clinical scores for internalizing problems, 4.7% scored in the borderline category, and 2.1% scored in the clinical category. With respect to externalizing problems, 97.0% scored in the non-clinical category, 1.9% in the borderline category, and 1.0% in the clinical category.

The NTR study recruits new-born twins from all regions in the Netherlands. Here we included the data on 10-year-olds who were born between 1986 and 2007. Children were not included if they had a severe handicap which interfered with daily functioning. Mean child age for mother report was 9.95 (SD = 0.51), father report 9.94 (SD = 0.50) and teacher report 9.80 (SD = 0.58). The children in NTR were mostly born in the Netherlands (99.5%). The remaining 0.5% consisted mainly of other West European nationalities (0.4%). Parents in the NTR were mostly born in the Netherlands (95.7% of fathers and 96.7% of mothers). 3.1% of mothers had a low skill occupation (primary education), 11.4% had an occupation that required lower secondary education, 40.3% had an upper secondary educational level, 30.6% had a higher vocational occupation level, and 14.6% worked at the highest (i.e. scientific) level. According to mother reports for internalizing problems, 86.1% of children had a non-clinical score, 5.9% had a borderline score,

and 8.0% scored in the clinical range. For externalizing problems, 85.7% scored in the non-clinical range, 6.5% scored in the borderline range, and 7.8% in the clinical range.

The RADAR-Y sample was recruited in the provence of Utrecht and four large cities in the mid–west of the Netherlands. Because the RADAR-Y study had a focus on delinquency development, children with borderline externalizing behavior problems at age 12 were oversampled. All participants from the first wave of data collection, born between 1990 and 1995, were selected. The mean age of the children at this wave was 13.03 years (SD = 0.46). The sample consisted mainly of native Dutch (87.9%) children. Remaining participants belonged to the following groups: Surinam (2.4%), Indonesian/ Moluccan (2.4%), Antillean (1.8%), Turkish (0.4%), and other (4.8%). The majority of children came from families with a medium or high socio-economic status (89.2%). According to the children's reports for externalizing problems, 81.6% of the participants had a non-clinical score, 7.2% had a borderline score, and 11.2% scored in the clinical range. Using the cutoff scores for the depression scale as described by Reynolds,[38] 4.0% of the children scored in the subclinical or clinical range of depressive symptoms. Using the cutoff scores for the anxiety scale of Birmaher et al.,[39] 5.3% of the children scored in the subclinical or clinical range for anxiety symptoms.

The TRAILS sample was recruited in the Northern regions of the Netherlands. All participants from the first wave of data collection (born between 1990 and 1991) were selected. The mean age of the children at the first wave was 11.09 (SD = 0.56). The large majority of participants were Dutch (86.5%), with other participants being Surinam (2.1%), Indonesian (1.7%), Antillean (1.7%), Mo roccan (0.7%), Turkish (0.5%), and other (6.9%). Based on mother-reported sum-scores for the internalizing and externalizing scales, TRAILS participants were categorized in a non-clinical, borderline, or clinical category. For internalizing problems, 67.3% of the participants had a non-clinical score, 13.9% had a borderline score, and 18.8% had a clinical score. For externalizing problems, 74.5% had a non-clinical score, 10.2% a borderline score, and 15.4% had a score in the clinical range.

To summarize, the cohorts represented the entire Dutch geographic region across all strata from society. They had a similar distribution of SES. The percentage of participants with parents born in the Netherlands was relatively high in NTR (>95%), around 87% in Radar-Y and TRAILS and relatively low in Gen-R (<72%). The percentage of non-clinical behavioral problems was highest in Gen-R and lowest in TRAILS.

All studies were approved by central or institutional ethical review boards. The participants were treated in compliance with the Declaration of Helsinki, and data collection was carried out with their adequate understanding and parental consent. All measures in RADAR-Y were self-reports. In the other cohorts, children were rated by any combination of: their parents, themselves, or their teachers. Table 3 shows the total number of children in each cohort, and the number of participants with an externalizing

Table 1: General Cohort Information

Full cohort name	Short name	Website	Birthyears	References (DOI)
Generation R	Gen-R	generationr.nl	2002-2006	10.1007/s10654-016-0224-9 10.1016/j.jaac.2012.08.021
Netherlands Twin Register	NTR	tweelingenregister.org	1986-2017	10.1017/thg.2012.118 10.1016/j.jaac.2012.10.009
Research on Adolescent Development And Relationships – Young Cohort	RADAR-Y	www.uu.nl/onderzoek/radar	1990-1995	10.1111/cdev.12547 10.17026/dans-zrb-v5wp
TRacking Adolescents' Individual Lives Survey	TRAILS	trails.nl	1989-1991	10.1093/ije/dyu225

Table 2: Cohort Descriptive Statistics of Total Sample Size and Parental Age in Current Study

Cohort	*N*	Maternal age at birth child		Paternal age at birth child	
		Range	*M (SD)*	Range	*M (SD)*
Gen-R	4,769	16.56 – 46.85	31.68 (4.79)	17.61 – 68.67	34.24 (5.58)
NTR	25,396	17.36 – 47.09	31.35 (3.95)	18.75 – 63.61	33.76 (4.71)
RADAR-Y	497	17.80 – 48.61	31.38 (4.43)	20.34 – 52.52	33.70 (5.10)
TRAILS	2,23	16.34 – 44.88	29.32 (4.58)	18.28 – 52.09	31.99 (4.71)

Table 3: Total Sample Size and Sample Sizes per Informant per Cohort

(Total Sample Size)		Gen-R (*N*= 4,769)		NTR (*N*=25,396)		RADAR-Y (*N*=497)		TRAILS (N=2,230)	
Variable	**Informant**								
Externalizing behavior problems	Child	BPM[a]	4,01	-	-	YSR[b]	491	YSR[b]	2,188
	Mother	CBCL[c]	4,549	CBCL[c]	21,921	-	-	CBCL[c]	1,965
	Father	CBCL[c]	3,259	CBCL[c]	14,715	-	-	-	-
	Teacher	-	-	TRF[d]	12,573	-	-	TCP[e]	1,925
Internalizing behavior problems	Child	BPM[a]	4,018	-	-	RADS-2[f]+ SCARED[g]	266	YSR[b]	2,171
	Mother	CBCL[c]	4,55	CBCL[c]	021,731	-	-	CBCL[c]	1,955
	Father	CBCL[c]	3,259	CBCL[c]	14,626	-	-	-	-
	Teacher	-	-	TRF[d]	12,389	-	-	TCP[e]	1,924

[a]Brief Problem Monitor (BPM; Achenbach, 2011).
[b]Youth Self Report (YSR; Achenbach, 1991).
[c]Child Behavior Checklist (CBCL; Achenbach, 1991; Achenbach, 2001).
[d]Teacher Report Form (TRF; Achenbach, 2001).
[e]Teacher Checklist of Psychopathology (TCP). Vignette questionnaire on the basis of the Achenbach Teacher Report Form developed by TRAILS.
[f]Reynolds Adolescent Depression Scale – 2nd edition (RADS-2; Reynolds, 2000). Excluding anhedonia scale. Standardized before averaged with SCARED.
[g]Screen for Child Anxiety Related Disorders (SCARED; Birmaher, et al., 1997). Standardized before averaged with RADS-2.

Table 4: Mean and SD for Externalizing and Internalizing Problems

Informant	Cohort	Externalizing	Internalizing	*N*-Ext/*N*-Int
Child	Gen-R	1.94 (1.92)	2.15 (2.09)	4,010/4,018
	RADAR-Y	10.61 (7.15)	-0.04 (0.86)	491/266
	TRAILS	8.68 (6.25)	11.28 (7.41)	2,188/2,171
Mother	Gen-R	3.92 (4.91)	4.86 (5.05)	4,549/4,550
	NTR	5.61 (6.12)	4.68 (5.07)	11,086/10,986
	TRAILS	8.40 (7.03)	7.85 (6.20)	1,965/1,955
Father	Gen-R	3.99 (4.91)	4.58 (4.72)	3,259/3,259
	NTR	4.66 (5.41)	3.56 (4.24)	7,420/7,374
Teacher	NTR	3.28 (5.88)	4.41 (4.96)	6,536/6,446
	TRAILS	0.44 (0.77)	0.99 (1.12)	1,925/1,924

For instruments, see Note Table 3.

and internalizing behavior problem score, as a function of informant (father, mother, teacher and self).

Measures

Predictors. Maternal and Paternal Age at Birth. The age of the biological parents at birth of the child was measured in years up to two decimals for each cohort.

Outcomes. Externalizing and Internalizing Problems. In most cohorts, internalizing and externalizing problems were assessed by the parent-rated Child Behavior Checklist (CBCL; Achenbach, 1991; Achenbach & Rescorla, 2001),[40,41] the Youth Self-Report (YSR),[40] and the Teacher Report Form (TRF)[41]. These questionnaires contain a list of around 120 behavioral and emotional problems, which can be rated as 0 = *not true,* 1 = *somewhat* or *sometimes true*, or 2 = *very or often true in the past 6 months*. The broadband scale Internalizing problems includes the syndromes anxious/depressed behavior, withdrawn/depressed behavior, and somatic complaints; the broadband scale Externalizing problems involves aggressive and rule-breaking behavior. In TRAILS, the Teacher Checklist of Psychopathology (TCP) was developed to be completed by teachers. The TCP contains descriptions of problem behaviors corresponding to the syndromes of the TRF. Teachers rated the TCP on a 5-point scale.[42] In Gen-R, the YSR was replaced by the Brief Problem Monitor (BPM), containing six items for internalizing and seven items for externalizing behavior problems from the YSR. All items were scored on a 3-point scale. In RADAR-Y, internalizing behavior problems were assessed by a combined score of the Reynolds Adolescent Depression Scale-2nd edition (RADS-2)[38] and the Screen for Child Anxiety Related Emotional Disorders (SCARED)[39] questionnaires. The RADS-2 contained 23 items (the subscale anhedonia was deleted) and the SCARED contained 38 items, which were rated on a 4-point scale (1 = *almost never*, 2 = *hardly ever*, 3 = *sometimes*, 4 = *most of the time*) and 3-point scale (1 = *almost never*, 2 = *sometimes*, 3 = *often*), respectively.

Table 3 gives an overview of the rating instruments, the informants for each of the cohorts and the number of children in each cohort for each informant/instrument combination. A sum score was calculated per informant/instrument for the relevant items for externalizing and internalizing problems respectively. Table 4 shows the mean scores for externalizing and internalizing problems per cohort. The scores for girls and boys are given in Tables S1 and S2 of the supplementary materials, respectively.

Covariates. Socio-Economic Status (SES) and child gender. In Gen-R, SES was defined as a continuous variable (principal component) based on parental education and household income. In NTR, SES was a 5-level ordinal variable based on occupational level. In TRAILS, SES was a 3-level ordinal variable based on parental education, parental occupational status and household income. In RADAR-Y SES was a dichotomous variable based on parents' occupational level. Child gender was coded as male = 0 and female = 1.

Missing Data and Data Imputation

Missing Data. For externalizing problem behavior, 15.9% of the child self-reports were missing for Gen-R, while for RADAR-Y and TRAILS these percentages were 1.2% and 1.9%, respectively. For mother reported data, 4.6% were missing for Gen-R, 13.7% for NTR and 11.9% for TRAILS. For father reported data, 31.7% were missing for Gen-R and 42.1% for NTR. For teacher reported data, 50.5% were missing for NTR and 13.7% for TRAILS. For internalizing problem behavior, the percentages were similar, except for child-reported data in RADAR-Y, where 46.4% was missing. For the predictor variables, age mother and age father, 0.3% and 1.3%, were missing for NTR, 0.0% and 14.4% for Gen-R, 0.4% and 9.7% for RADAR-Y, and 5.1% and 25.0% for TRAILS, respectively. For SES, the percentage of missing values was always below 3.0%, except for Gen-R where 22.3% was missing. For child gender, all cohorts had complete information.

Please note that the higher percentage for missing teacher- and father-reported data of NTR is due to the fact that NTR did not collect teacher-reported data at the initiation of the study and that NTR had not collected father-reported data in multiple birth years due to financial constraints. The higher percentage of missing self-reported data of internalizing problem behavior for RADAR-Y is caused by the fact that not all subscales on which the internalizing problem behavior score was based were collected from all participants.

Data Imputation. Missing data was handled by means of multiple imputation (Schafer & Graham, 2002; Van Buuren, 2012). When multiple imputation is used, the missing values are repeatedly (in this study 100 times) imputed, that is, replaced by values that are plausible given the child's scores that are not missing, resulting in 100, so-called, completed data sets. Subsequently, each completed data set is analyzed (for example, using a multiple regression) and the 100 analyses are summarized such that the fact that "artificial data" are created by imputation is properly accounted for. Multiple imputation proceeds along three steps:

1. *Determine which variables are to be used for imputation*. The variables used for imputation have to be chosen such that conditional on these variables the missing data are believed to be missing at random (MAR),[43] that is, whether or not a score is missing does not depend on the missing value.[44] Unless missingness is planned, the variables causing the missingness are unknown to the researcher. What is often done in practice is that variables are chosen that are expected to be good predictors of the variables containing missing values. One can argue with respect to which and how many variables to use, but there is no way to test whether MAR is achieved, and MAR is an assumption.

The imputation model included the outcome variables externalizing and internalizing behavioral problems per informant, total behavioral problems, SES, child gender, age of the child, age of the father and age of the mother. In some cohorts, other variables were present that could also contribute to the imputation. Specifically, parent

psychopathology (in Gen-R) and total number of siblings (in NTR) contributed to the imputation model. Variables functioned only as predictors when a correlation of at least .10 with the imputed variable was present. Since the NTR dataset contained twins, the imputation process differed from that of the other cohorts. The imputation for NTR was done for each family instead of each participant, so that the same value for SES, age father and age mother was obtained for both twins. The imputation of missing data was done for informants available in each cohort. So, for example, when a cohort had no teacher-reported data, teacher data were not imputed.

2. *Generate imputed data matrices.* The R package MICE (Multiple Imputation by Chained Equations)[43] was used to create 100 imputed data matrices. MICE uses an iterative procedure in which sequentially each variable is imputed conditional on the real and imputed values of the other variables. Continuous variables were imputed by predictive mean matching. Categorical variables were imputed using logistic regression.[45] Success of the imputation was evaluated by checking the events logged by the software, and by checking convergence plots for a lack of trends and proper mixing of the imputation chains.

3. *Analyze each imputed data set as desired and pool the results*. In the current study each of the 100 imputed data sets was analyzed using multiple regression or cluster linear regression. The results, for each regression coefficient, were 100 estimates and 100 standard errors of the estimate. As may be clear, each of the standard errors was too small because they are partly based on artificial imputed data. This was accounted for by properly pooling the results using Rubin's rules.[43] The variance over the 100 estimates reflects the uncertainty in the estimate due to missing values (in each of the 100 completed data sets different values are imputed). In Rubin's rules the variance of the 100 estimates is used to increase the standard errors such that they properly account for the fact that part of the data is imputed. Gen-R, TRAILS and RADAR-Y used the 'pool' function of MICE in R for summarizing the effects of the 100 separate imputed datasets, whereas NTR used the pooling option of Mplus instead of R, to appropriately take into account the family clustering of the twins in the same analysis. Both pooling methods are based on the principles as explained here. The pooled estimates and standard errors were the main outcomes of the analyses after imputation.

Analytical Strategy: Bayesian Evidence Synthesis

The process of Bayesian evidence synthesis consists of four steps: (1) creating exploratory and confirmatory data sets; (2) generating competing hypotheses using exploratory analysis; (3) quantifying the support for each of the competing hypotheses using Bayesian hypothesis evaluation; and (4) Bayesian evidence synthesis, that is, sum-

marizing the support resulting from each study into the overall support for the competing hypotheses in the data from the four cohort studies.

Exploration and Confirmation

As was elaborated in the introduction, diverse results regarding the relation between parental age and child problem behavior have been found in the literature, with increased parental age both positively and negatively related to child problem behavior. In the same vein, there may be a quadratic effect and if there is, the change in child problem behavior may be accelerating or decelerating across parental age. Since research is indecisive, especially for the non-clinical studies reviewed in this paper, the data resulting from each of the cohorts were split randomly into two parts containing the same number of children: an exploratory part, which was used to generate a set of competing hypotheses; and a confirmatory part, which was used to quantify the support in the data for each of the hypotheses considered. Since the NTR dataset consisted of twins, the cross-validation datasets were split based on family ID for this cohort, to ensure independent datasets. Multiple imputation was applied separately to the exploratory and confirmatory part of the data. Having an exploratory and confirmatory dataset avoids the so-called "double dipping", that is, using *the same* data to generate and evaluate hypotheses. Here a hypothesis survived if it: 1) emerged from the exploratory analyses and 2) was supported by the confirmatory analyses. The process of generating hypotheses is explained below.

Generating Hypotheses using Exploratory Analyses

The exploratory half of the data resulting from each of the four cohorts was used to generate hypotheses with respect to the relation between child problem behavior and parental age. First, for each cohort seperately, linear regression analyses were conducted to regress internalizing and externalizing problem behavior as evaluated by child, mother, father, and teacher (See Table 3 for the informants that were present per cohort) on paternal and maternal age and age squared (both with and without child gender and social economic status as covariates). Parental age was mean-centered to obtain the linear effect at the mean age of the samples and to reduce the correlation between the linear and quadratic term. For Gen-R, RADAR-Y and TRAILS, the analyses were conducted in R (R Core Team, 2017). For the NTR twin-data, cluster linear regression analyses were conducted in Mplus version 8.0.[46] All analyses were repeated with SES and child gender as covariates. This rendered, for each combination (e.g., predicting externalizing problems as rated by the mother from mother age and age squared) an estimate of both the linear and quadratic effect for each of the cohorts that included the informant of interest. These estimates and the corresponding p-values provided information with respect to whether the linear and non-linear effects were expected to be negative, zero, or positive. To interpret the strength of relations, the variables in the exploratory analyses were all standardized. The results of the regression analyses were

translated into so-called informative hypotheses,[47] that is, hypotheses that represent expectations with respect to the state of affairs in the populations from which the data of the four cohorts were sampled. One example of such an informative hypothesis is: H1: $\beta < 0$. That is, the regression coefficient is negative. Informative hypotheses go beyond the traditional null hypothesis (here H0: $\beta = 0$) by stating explicitly which relations between variables are expected. Often the null is added to the set of hypotheses under consideration to protect against unjustified claims that the effect specified by an informative hypothesis exists. Another hypothesis that can be added besides the informative hypotheses is the alternative hypothesis Ha: β. That is, there are no restrictions on the regression coefficient. The alternative hypothesis is used to protect against choosing the best of a set of inadequate informative hypotheses. For example, H0: $\beta = 0$, and H1: $\beta < 0$ constitute the set of hypotheses supported by the exploratory parts of the data, but both are inadequate in the confirmatory data. Instead, another unspecified hypothesis ($\beta > 0$) describes the confirmatory data best. In this case the Bayesian approach (specified below) will prefer the alternative hypothesis, Ha: β, over both informative hypotheses. By using informative hypotheses, the exact same hypotheses could be evaluated in all cohorts, even when cohorts used different measurement instruments for the same concepts. Not requiring the exact same measurement instruments is an important benefit of Bayesian evidence synthesis over classical meta-analyses.

Confirmatory Bayesian Hypotheses Evaluation

Once a set of competing informative hypotheses had been formulated (including the traditional null and alternative hypotheses), the empirical support for each pair of hypotheses was quantified using the Bayes factor (BF).[48] The Bayes factor is the ratio of the marginal likelihood of two competing hypotheses. Loosely spoken, the marginal likelihood of a hypothesis is the probability of that hypothesis given the data. Consequently, a Bayes factor comparing H1 with Ha of, for example, 5 indicates that the support in the data for H1 is five times larger than for Ha. The BF as the ratio of two marginal likelihoods implies that the fit (how well does a hypothesis describe the data set at hand) and the specificity (how specific is a hypothesis) of the hypotheses involved are accounted for.49 To give an example, if $\beta = -2$, H1: $\beta < 0$, and Ha: β, both have an excellent fit, but H1: $\beta < 0$ is more specific than Ha: β (anything goes), and as a result, the BF will prefer H1 over Ha. Note that the size of the Bayes factor is related to sample size. If the precision of the evidence in the data for a hypothesis increases as a result of a larger sample, the Bayes factor for that hypothesis will increase as well. The Bayes factor implemented in the R package Bain[49] was used to evaluate informative hypotheses in the context of (cluster) multiple linear regression models.

Assuming that a priori each hypothesis is equally likely to be true, the Bayes factors were transformed in so-called posterior model probabilities (PMPs), that is, the support in the data for the hypothesis at hand given the set of hypotheses under evaluation. PMPs have values between 0 and 1 and sum to 1 for the hypotheses in the set under

consideration. For example, if PMP H0 = .05, PMP H1 = .85, and PMP Ha = .10, then it is clear that H1 receives the most support from the data, because it has by far the largest PMP. Thus, the result of the confirmatory Bayesian hypotheses evaluation were PMPs for each hypothesis and for each informant by each of the cohorts that had ratings by this informant. The next step was to apply Bayesian evidence synthesis.

Bayesian Evidence Synthesis

Bayesian evidence synthesis was used to summarize the support for the hypotheses of interest over the four cohort studies. Bayesian evidence synthesis[37] can be illustrated using the set of hypotheses: H0: $\beta = 0$, H1: $\beta < 0$, and Ha: β. In the context of this paper, these hypotheses are incompletely specified. The complete specification would be H0: $\beta 1 = 0$ for NTR, H1: $\beta 1 < 0$ for NTR and Ha: $\beta 1$ for NTR, and analogously for the other three cohort studies. This specification highlights that the support for the hypotheses depends on the cohort study at hand. Bayesian evidence synthesis can then be used to determine support for a set of hypotheses:

H0: H0 for NTR & H0 for TRAILS & H0 for Gen-R & H0 for Radar-Y
H1: H1 for NTR & H1 for TRAILS & H1 for Gen-R & H1 for Radar-Y
Ha: Ha for NTR & Ha for TRAILS & Ha for Gen-R & Ha for Radar-Y

that is, the regression coefficient is zero in the populations corresponding to each of the four cohort studies, the regression coefficient is smaller than zero in the populations corresponding to each of the four cohort studies, and there is not prediction with respect to the regression coefficient in the populations corresponding to each of the four cohort studies. If for a specific set of hypotheses only two or three cohorts contain the necessary variables, the hypotheses can be adjusted accordingly. Like for each individual study, the support for these composite hypotheses was quantified using posterior model probabilities (PMPs).

If a hypothesis emerges from the exploratory analyses of the data corresponding to the cohort studies and is supported by the confirmatory analyses of the data corresponding to the cohort studies, then there is evidence that this hypothesis provides an adequate description of the relation between child problem behavior and parental age, that is, in general, independent of the specific cohort studies used to evaluate this hypothesis. With the methodological approach elaborated in this section and applied in the remainder of this paper, the increased awareness of the need for replication studies before making scientific claims is explicitly addressed.

RESULTS

Exploratory Analyses

The results of the exploratory analyses (see Supplementary Materials) generally showed a negative relation between mean-centered parental age and externalizing problems accompanied by a positive quadratic coefficient, implying that the negative relation with age at the mean declined across age (see Table S3 of the Supplemental Materials). This model explained about 1.9% of total variance in externalizing problems with maternal age and 1.2% with paternal age. For internalizing problems, the relation with parental age was less apparent: about 0.5% of the total variance was explained by mothers' age, and about 0.2% was explained by fathers' age. In analyses including the covariates SES and gender, the relation with age diminished, but remained significant (Tables S4, and S5, of the supplementary materials). Higher SES was related to fewer externalizing problems, and boys showed more externalizing problems than girls. In general, no relation between parental age and internalizing problems was observed (see Tables S6, S7, and S8 of the Supplemental Materials).

Our interpretation of the exploratory results led to the following set of competing informative hypotheses with respect to the relation between parental age (mean-centered), as indicated by a linear (i.e., $\beta 1$) and quadratic (i.e., $\beta 2$) coefficient, and child problem behavior:

- H1: $\beta 1 = 0, \beta 2 = 0$. That is, age does not have a linear or quadratic relation.
- H2: $\beta 1 < 0, \beta 2 = 0$. That is, age has a negative linear relation, there is no quadratic relation.
- H3: $\beta 1 < 0, \beta 2 > 0$. That is, age has a negative linear relation, and a positive quadratic relation.
- Ha: $\beta 1, \beta 2$. That is, none of the above.

Based on the exploratory results, we expected most evidence for H2 or H3 in analyses with parental age predicting externalizing problems, and most evidence for H1 in analyses with parental age predicting internalizing problems. Since the exploratory results did not show a positive linear or a negative quadratic relation between age and behavioral problems, the hypotheses do not include these features. However, we remained open to other options by including the alternative hypothesis Ha that imposes no constraints on the parameters, and accordingly claims that anything can be true. Ha receives the most support if none of the specified informative hypotheses provides an adequate description of the confirmatory part of the data from each of the four cohorts. In this manner, we avoided that the best hypothesis out of the set of H1, H2, and H3, is an implausible hypothesis.

Confirmatory Analyses

Similarly to the exploratory data, the results showed negative relations across cohorts between parental age and externalizing problems. However, in the confirmatory

Table 5: Posterior Model Probabilities for Parental Age Predicting Externalizing Problems

Informant	Cohort	Age Father				Age Mother			
		H_1	H_2	H_3	H_a	H_1	H_2	H_3	H_a
Child	Gen-R	.23	*.56*	.16	.05	.22	.18	*.49*	.13
	RADAR-Y	.28	.02	*.49*	.22	*.43*	.07	*.38*	.12
	TRAILS	*.86*	.13	.00	.01	*.83*	.15	.02	.01
	All	**.98**	.02	.00	.00	**.93**	.02	.04	.00
Mother	Gen-R	*.90*	.07	.02	.01	*.82*	.04	.10	.05
	NTR	.00	.02	*.74*	.24	.00	*.89*	.09	.03
	TRAILS	.18	*.74*	.06	.02	.00	*.88*	.09	.03
	All	.00	**.53**	**.45**	.00	.00	**.97**	.03	.00
Father	Gen-R	*.65*	.22	.10	.03	*.60*	.19	.17	.04
	NTR	.00	*.49*	.38	.13	.00	*.93*	.05	.02
	All	.00	**.73**	.25	.02	.00	**.95**	.05	.00
Teacher	NTR	*.55*	.41	.03	.01	.29	*.60*	.09	.02
	TRAILS	*.48*	*.31*	.16	.05	.00	*.73*	.21	.06
	All	**.67**	.32	.01	.00	.00	**.96**	.04	.00

Numbers in italic font represent the highest posterior model probability per cohort. Numbers in bold font represent the highest meta-analytic results

Table 6: Posterior Model Probabilities for Parental Age Predicting Externalizing Problems after Correction for Impact Covariates

Informant	Cohort	Age Father				Age Mother			
		H_1	H_2	H_3	H_a	H_1	H_2	H_3	H_a
Child	Gen-R	*.62*	.33	.04	.01	*.83*	.10	.05	.02
	RADAR-Y	*.36*	.02	*.42*	.19	*.53*	.08	.29	.10
	TRAILS	*.88*	.11	.00	.01	*.89*	.09	.02	.01
	All	**1.00**	.00	.00	.00	**1.00**	.00	.00	.00
Mother	Gen-R	*.96*	.03	.00	.00	*.97*	.02	.00	.01
	NTR	.00	.31	*.52*	.17	.00	*.95*	.04	.01
	TRAILS	*.67*	.31	.01	.01	.30	*.63*	.05	.02
	All	.03	**.99**	.00	.00	.00	**1.00**	.00	.00
Father	Gen-R	*.88*	.10	.02	.00	*.92*	.06	.01	.00
	NTR	.02	*.84*	.11	.04	.00	*.96*	.03	.01
	All	**.**15	**.84**	.02	*.00*	.00	**.99**	.01	.00
Teacher	NTR	*.79*	.20	.01	.00	*.68*	.28	.03	.01
	TRAILS	*.87*	.11	.02	.00	*.60*	.32	.07	.02
	All	**.97**	.03	.00	.00	**.81**	.18	.00	.00

Table 7: Posterior Model Probabilities for Parental Age Predicting Internalizing Problems

Informant	Cohort	Age Father				Age Mother			
		H_1	H_2	H_3	H_a	H_1	H_2	H_3	H_a
Child	Gen-R	*.91*	.08	.01	.00	*.86*	.09	.04	.01
	RADAR-Y	*.84*	.09	.05	.03	*.81*	.16	.02	.01
	TRAILS	*.96*	.04	.00	.00	*.93*	.06	.01	.00
	All	**1.00**	.00	.00	.00	**1.00**	.00	.00	.00
Mother	Gen-R	*.58*	.25	.14	.04	*.35*	.25	.33	.08
	NTR	*.69*	.26	.04	.01	.26	*.72*	.01	.01
	TRAILS	*.94*	.05	.00	.00	*.81*	.17	.02	.01
	All	**.99**	.01	.00	.00	**.71**	.29	.00	.00
Father	Gen-R	*.43*	*.42*	.11	.03	*.48*	.36	.13	.03
	NTR	*.96*	.04	.00	.00	*.95*	.05	.00	.00
	All	**.96**	.04	.00	.00	**.97**	.03	.00	.00
Teacher	NTR	*.99*	.01	.1	.00	*.99*	.01	.00	.00
	TRAILS	*.85*	.06	.07	.02	.24	.15	*.49*	.12
	All	**1.00**	.00	.00	.00	**.99**	.01	.00	.00

Table 8: Posterior Model Probabilities for Parental Age Predicting Internalizing Problems after Correction for Impact Covariates

Informant	Cohort	Age Father				Age Mother			
		H_1	H_2	H_3	H_a	H_1	H_2	H_3	H_a
Child	Gen-R	*.77*	.21	.02	.01	*.82*	.09	.07	.02
	RADAR-Y	*.86*	.07	.04	.03	*.86*	.11	.02	.01
	TRAILS	*.97*	.03	.00	.00	*.95*	.04	.00	.00
	All	**1.00**	.00	.00	.00	**1.00**	.00	.00	.00
Mother	Gen-R	*.88*	.11	.01	.00	*.93*	.05	.01	.00
	NTR	*.88*	.11	.01	.00	*.70*	.29	.00	.00
	TRAILS	*.96*	.04	.00	.00	*.91*	.08	.01	.00
	All	**1.00**	.00	.00	.00	**1.00**	.00	.00	.00
Father	Gen-R	*.88*	.09	.02	.01	*.90*	.08	.01	.00
	NTR	*.96*	.03	.00	.00	*.96*	.04	.00	.00
	All	**1.00**	.01	.00	.00	**1.00**	.01	.00	.00
Teacher	NTR	*.99*	.01	.00	.00	*.99*	.01	.00	.00
	TRAILS	*.94*	.04	.02	.01	*.83*	.06	.08	.03
	All	**1.00**	.00	.00	.00	**1.00**	.00	t.00	.00

Numbers in italic font represent the highest posterior model probability per cohort. Numbers in **bold** font represent the highest meta-analytic results.

Table 7: Posterior Model Probabilities for Parental Age Predicting Internalizing Problems

Informant	Cohort	Age Father				Age Mother			
		H_1	H_2	H_3	H_a	H_1	H_2	H_3	H_a
Child	Gen-R	*.91*	.08	.01	.00	*.86*	.09	.04	.01
	RADAR-Y	*.84*	.09	.05	.03	*.81*	.16	.02	.01
	TRAILS	*.96*	.04	.00	.00	*.93*	.06	.01	.00
	All	**1.00**	.00	.00	.00	**1.00**	.00	.00	.00
Mother	Gen-R	*.58*	.25	.14	.04	*.35*	.25	.33	.08
	NTR	*.69*	.26	.04	.01	.26	*.72*	.01	.01
	TRAILS	*.94*	.05	.00	.00	*.81*	.17	.02	.01
	All	**.99**	.01	.00	.00	**.71**	.29	.00	.00
Father	Gen-R	*.43*	*.42*	.11	.03	*.48*	.36	.13	.03
	NTR	*.96*	.04	.00	.00	*.95*	.05	.00	.00
	All	**.96**	.04	.00	.00	**.97**	.03	.00	.00
Teacher	NTR	*.99*	.01	.1	.00	*.99*	.01	.00	.00
	TRAILS	*.85*	.06	.07	.02	.24	.15	*.49*	.12
	All	**1.00**	.00	.00	.00	**.99**	.01	.00	.00

Numbers in **bold** font represent the posterior model probability per cohort. Numbers in Italic font represent the meta-analytic results.

data, the quadratic coefficients from the cohorts were less often significantly different from zero than in the exploratory data. The model with a linear and quadratic coefficient for parental age explained on average about 1.1% of total variance in externalizing problems with maternal age and 0.9% with paternal age as a predictor. With respect to internalizing behavior problems, the model with maternal age explained on average about 0.4% of the total variance, and paternal age explained on average about 0.3%.

Parental Age and Externalizing Behavior Problems

The posterior model probabilities (PMPs) concerning the relation between parental age and externalizing problems are presented in Tables 5. The table only shows PMP scores for those cohorts that included the associated informants (See Table 3 for an overview of informants per cohort). As shown by Table 5, for parent-reported externalizing behavior problems, Gen-R yielded most evidence for H1 (i.e., no relation with parental age). NTR supports H3 (i.e., the relation with parental age follows a negative linear trend including a positive quadratic factor) for mother-reported externalizing behavior problems, while TRAILS provided most support for H2 (i.e., the relation with parental age is linear and negative). The combined results for mother-reported externalizing behavior problems predicted by father age showed substantial support (PMP = .53 and .45 respectively) for H2 and H3. For father reported externalizing behavior

Table 8: Posterior Model Probabilities for Parental Age Predicting Internalizing Problems after Correction for Impact Covariates

Informant	Cohort	Age Father				Age Mother			
		H_1	H_2	H_3	H_a	H_1	H_2	H_3	H_a
Child	Gen-R	*.77*	.21	.02	.01	*.82*	.09	.07	.02
	RADAR-Y	*.86*	.07	.04	.03	*.86*	.11	.02	.01
	TRAILS	*.97*	.03	.00	.00	*.95*	.04	.00	.00
	All	**1.00**	.00	.00	.00	**1.00**	.00	.00	.00
Mother	Gen-R	*.88*	.11	.01	.00	*.93*	.05	.01	.00
	NTR	*.88*	.11	.01	.00	*.70*	.29	.00	.00
	TRAILS	*.96*	.04	.00	.00	*.91*	.08	.01	.00
	All	**1.00**	.00	.00	.00	**1.00**	.00	.00	.00
Father	Gen-R	*.88*	.09	.02	.01	*.90*	.08	.01	.00
	NTR	*.96*	.03	.00	.00	*.96*	.04	.00	.00
	All	**1.00**	.01	.00	.00	**1.00**	.01	.00	.00
Teacher	NTR	*.99*	.01	.00	.00	*.99*	.01	.00	.00
	TRAILS	*.94*	.04	.02	.01	*.83*	.06	.08	.03
	All	**1.00**	.00	.00	.00	**1.00**	.00	.00	.00

Numbers in **bold** font represent the posterior model probability per cohort. Numbers in Italic font represent the meta-analytic results.

problems predicted by father age and for parent-reported externalizing behavior problems predicted by mother age, the combined results provided most support for H2: the relation with parental age is linear and negative, in other words, higher parental age is associated with less externalizing behavioral problems. For teacher-reported externalizing behavior problems, TRAILS and NTR yielded most evidence for H1 when paternal age was used as a predictor in the linear regression model. When maternal age was included, most support was yielded for H2: the relation with parental age is linear and negative. For child-reported externalizing behavior problems, the results were mixed over cohorts (Gen-R prefers H2, RADAR-Y H3, and TRAILS H1). After combining the results from the three cohorts, however, most support was obtained for H1, that is, no relation with parental age.

Table 6 shows the results after inclusion of the covariates as predictors of externalizing problems. After adjusting for the effect of SES and gender, all cohorts yielded substantial evidence for H1 with respect to child- and teacher-reported externalizing problem behavior. This meant a shift especially for the child-reported problem behavior by Gen-R, and the teacher-reported problem behaviors by both NTR and TRAILS. For

parent reported problem behavior some cohorts provided most support for H1 (Gen-R for all parent-reports, and TRAILS for paternal age predicting mother-reported problem behavior), others for H2 (TRAILS and NTR), and NTR for H3 in mother-reported problem scores related to paternal age. By including covariates in the model Gen-R and TRAILS mainly handed in support on H2 while in NTR the support for H2 increased at the expense of support for H3. When combining evidence for the parent reports most support was still found for H2, that is, there is a linearly decreasing relation between age and externalizing problem behavior.

Parental Age and Internalizing Behavior Problems

With regard to internalizing problems (the results are presented in Table 7), the cohorts generally found most evidence for H1 for multiple informants, except for mother-reported internalizing problems reported by mother age in NTR. All combinations of studies rendered most support for H1, which means that the hypothesis that there is no relation between parental age and internalizing problems is best supported by the data.

After correction for SES and gender (Table 8), all findings still suggested H1 for the impact of parental age on internalizing problem behavior, irrespective of the cohort and informant, and, consequently, combining the results from the various cohorts provided overwhelming support for H1, that is, no evidence for a relation bteween parental age and child internalizing problem behavior.

DISCUSSION

Parental Age and Externalizing Problems

We found evidence for a negative linear relation between parental age and externalizing problems as reported by parents. That is, across increasing maternal and paternal age there was a decrease in offspring externalizing problems. There was also evidence for a negative linear relation between maternal age and externalizing problems as reported by teachers. For teachers, this finding was partly explained by socio-economic status, but the relation between parental age and parent-reported externalizing problems persisted after adjusting for SES, so the favorable effect of parental age is not solely due to socio-economic status.

Parental Age and Internalizing Problems

Parental age seemed unrelated to child internalizing problem behavior, especially when taking SES into account. Tentatively, older parenthood might be associated with both high and low vulnerability to develop internalizing problems. On the one hand, older parents may have a lower probability of internalizing problems because they are less likely to have a background characterized by deprivation and social instability,[50] known to be related to internalizing problems such as anxiety and depression. On the

other hand, internalizing problems can increase the probability of older parenthood, by stimulating engagement in and consolidation of romantic relationships.[51,52] Possibly, both processes play a role, and their joint influence results in a lack of net result.

Sociodemographic Factors as a Potential Explanation

The relatively consistent beneficial effect of advanced parenthood for childhood externalizing problems may seem unexpected, given mixed findings from earlier research on more common mental health problems (De Kluiver, Buizer-Voskamp, Dolan, & Boomsma, 2017; McGrath et al., 2014).[11,12] The beneficial effect of advanced parental age could have more than one explanation. Older and younger parents have different parenting styles. For example, there is evidence that older mothers use less frequent sanctions towards their children, are more sensitive to the child's needs and provide more structure.[53] Older parents may also tend to appraise a specific problem level as less disturbing than younger parents, and older parents might be more patient and are capable of setting limits, thus feeling more equipped to handle externalizing behaviors. The positive impact of higher quality parenting by older parents is expected to be more relevant to externalizing problem behavior than to autism and schizophrenia where a disadvantageous impact of increased parental age is established.

Previous studies provided evidence indicating that offspring of older parents are, in several respects, more affluent than those with younger parents.[3,7,11,21,22,25] The fact that the negative relation of parental age and externalizing problems became weaker when SES was taken into account, indicates that the relatively high socio-economic status of older parents, or SES-related selection effects,[50] at least partly explain why their children have a decreased probability of externalizing problems. Myrskylä, Barclay and Goisis[24] argued that there are indeed important socio-demographic pathways associated with delayed parenthood in more recent birth cohorts. Older mothers tend to have better health behaviors during pregnancy, for example with respect to smoking during pregnancy, which is an established risk factor for offspring externalizing problems.[54]

Furthermore, parents who have externalizing behavior problems themselves may be higher in risk taking and may have children at a younger age. Hence, externalizing behavior problems may be transmitted especially by younger parents and less by older parents. This idea is in line with the unclarity about a relation between ADHD and advanced paternal age.[11,12]

From a biological point of view, advanced parenthood seems mostly disadvantageous, but sociodemographic factors might compensate or even more than compensate for the biological disadvantages related to reproductive ageing when it comes to mental health problems. Older mothers from more recent birth cohorts are more socioeconomically advantaged, and happier after childbearing. The observation that older parents have offspring with fewer externalising problems, tended to disappear when

SES was taken into account, shows that demographic factors can indeed compensate for the biological disadvantages.

Earlier Versus Later Birth Cohorts

In the 1950's and 60's the number children born to mothers over 40 was larger than in 2016. For offspring born during the 1960s, Saha et al.[55] found a negative association between maternal age and externalizing behavior problems, but in contrast to our results, they observed a positive association between maternal age and internalizing problems, and a positive association between paternal age and externalizing behavior problems. The study differed in several important aspects from the current one. All offspring were born during the 1960s, whereas in our study, all offspring were born after 1980. The age at which fathers and mothers have children has increased in the last 20 years. In the Saha et al. study average maternal and paternal ages were 24.8 and 28.4, respectively, while in our samples average maternal- and paternal ages were around 31 and 33 years. Older mothers from earlier birth cohorts tended to have low levels of education and their offspring had many older siblings.[24] In later birth cohorts, older mothers have higher education than younger mothers and their offspring have fewer older siblings. Thus, the family resources are spread less thinly across siblings than in earlier times. This may be the reason that our results differ from some of the findings of Saha, Barnett, Buka and McGrath.[55] As argued by Myrskylä, Barclay & Goisis,[24] as well, being a parent during the 1960s differs from being a parent in the 1980s and children born during the 1980s and later might benefit from positive changes in the macro-environment.

Informant Effect

We used a multi-informant design (i.e., mother, father, teacher, child) to investigate parental age effects on behavioral problems. Most questionnaires belonged to the same system (ASEBA), but they do not necessarily capture the exact same construct, as different informants observe the children in different contexts. Consistent with the notion that different informants provide partly non-overlapping information, the results in this study depended on the choice of informant, since, as opposed to parent-reported problems, child-reported externalizing problems were not predicted by parental age. Conceivably, this different outcome for child-reported problems is due to a limited ability of 10-year-old children to report reliably and validly on their externalizing behaviors. It is less likely that the associations with parent-reports are caused by report bias, as teacher-reports also provide support for an association with maternal age. Thus, the choice of informant is not an arbitrary one, and may influence the associations that are found. Obviously, the parent and teacher sample sizes were also substantially larger than the sample size for child-reports. Additionally, the largest study with child reports

(i.e., TRAILS) used a shortened version of the YSR, which could cause lower reliability and validity of child-reports.

Strengths of the Current Study

This study adopted an analysis strategy that used the data of multiple cohort studies to evaluate the same set of hypotheses. First, the data of each cohort study were divided into two parts: an exploratory part and a confirmatory part. Second, the exploratory part was used to generate a set of competing informative hypotheses. Third, the confirmatory part was used to compute the support in each cohort for the hypotheses entertained and to combine studies by means of Bayesian updating to compute overall results.[37] This analysis strategy had a number of advantages. In the exploratory analyses data snooping or even p-hacking is allowed, because this part of the data is only used to generate a set of competing informative hypotheses and not to evaluate these hypotheses. In contrast, the confirmatory part of each data set is only used to evaluate this set of informative hypotheses to the traditional null and alternative hypotheses, which should, especially in ages of replication crisis, publication bias and questionable research practices, increase the credibility of our results. The interested reader is referred to the Supplementary Materials where we highlight, why exploratory analyses may lead to incorrect interpretations, even with large samples, and that cross-validation can prevent this from happening. In addition, with traditional null hypothesis significance testing, we would not have been able to quantify the support for the null hypothesis (p-values cannot be used to "accept" the null-hypothesis), which appeared an important hypothesis in our study. Bayes factors and posterior model probabilities are not used to reject or not reject the null-hypotheses, they are used to quantify the support in each of the cohorts for the hypotheses entertained. Furthermore, combining studies using Bayesian updating enabled us to quantify the relative evidence with respect to multiple hypotheses using the data from multiple cohorts. Again, in ages of replication crisis, it is valuable to base conclusions on data from multiple cohorts that can all be used to address the same research question.

Limitations

Although the study has a number of methodological strengths, there are also limitations. First, the study focused on children's externalizing and internalizing behavior problems and did not examine other outcomes that may be positively associated with parental age, such as physical health problems and neurodevelopmental conditions. Second, children's behavior problems were only assessed during early adolescence. Thus, the study could not investigate the possibility that the direction or magnitude of the associations may vary at different points in development. For example, previous research suggesting a negative association between parental age and individuals' well-being has focused on late adolescents and young adults.[7,8] Third, a tiny percentage of the parents were under the age of 20 at the time of the child's birth. Although this

reflects societal changes in Netherlands, it would be important to note that some results may not replicate in other populations that have a higher percentage of teenage pregnancies. This may be especially relevant when interpreting the lack of an association between parental age and children's internalizing behavior problems in this study.

Conclusion

The strategy applied to large cohorts showed us a beneficial association between advanced parental age and externalizing problem behavior, while for internalizing problem behavior there is no beneficial association. We found no evidence for a harmful effect of advanced parenthood.

REFERENCES

1. te Velde ER. The variability of female reproductive ageing. Hum Reprod Update. 2002;8(2):141-154. doi:10.1093/humupd/8.2.141
2. Desjardins B, Bideau A, Brunet G. Age of mother at last birth in two historical populations. J Biosoc Sci. 1994;26(04). doi:10.1017/s0021932000021635
3. Carslake D, Tynelius P, van den Berg G, Smith GD, Rasmussen F. Associations of parental age with health and social factors in adult offspring. Methodological pitfalls and possibilities. Sci Rep. 2017;7(1). doi:10.1038/srep45278
4. Sandin S, Hultman CM, Kolevzon A, Gross R, MacCabe JH, Reichenberg A. Advancing Maternal Age Is Associated With Increasing Risk for Autism: A Review and Meta-Analysis. J Am Acad Child Adolesc Psychiatry. 2012;51(5):477-486.e1. doi:10.1016/j.jaac.2012.02.018
5. Lee BK, McGrath JJ. Advancing parental age and autism: multifactorial pathways. Trends Mol Med. 2015;21(2):118-125. doi:10.1016/j.molmed.2014.11.005
6. Menezes PR, Lewis G, Rasmussen F, et al. Paternal and maternal ages at conception and risk of bipolar affective disorder in their offspring. Psychol Med. 2010;40(03):477. doi:10.1017/s003329170999064x
7. Tearne JE, Robinson M, Jacoby P, et al. Older maternal age is associated with depression, anxiety, and stress symptoms in young adult female offspring. J Abnorm Psychol. 2016;125(1):1-10. doi:10.1037/abn0000119
8. Weiser M, Reichenberg A, Werbeloff N, et al. Advanced Parental Age at Birth Is Associated With Poorer Social Functioning in Adolescent Males: Shedding Light on a Core Symptom of Schizophrenia and Autism. Schizophr Bull. 2008;34(6):1042-1046. doi:10.1093/schbul/sbn109
9. Andersen A-MN, Urhoj SK. Is advanced paternal age a health risk for the offspring? Fertil Steril. 2017;107(2):312-318. doi:10.1016/j.fertnstert.2016.12.019
10. Khandwala YS, Baker VL, Shaw GM, Stevenson DK, Lu Y, Eisenberg ML. Association of paternal age with perinatal outcomes between 2007 and 2016 in the United States: population based cohort study. BMJ. 2018:k4372. doi:10.1136/bmj.k4372
11. McGrath JJ, Petersen L, Agerbo E, Mors O, Mortensen PB, Pedersen CB. A Comprehensive Assessment of Parental Age and Psychiatric Disorders. {JAMA} Psychiatry. 2014;71(3):301. doi:10.1001/jamapsychiatry.2013.4081
12. de Kluiver H, Buizer-Voskamp JE, Dolan C V, Boomsma DI. Paternal age and psychiatric disorders: A review. Am J Med Genet Part B Neuropsychiatr Genet. 2017;174(3):202-213. doi:10.1002/ajmg.b.32508
13. Goisis A, Schneider DC, Myrskylä M. The reversing association between advanced maternal age and child cognitive ability: evidence from three {UK} birth cohorts. Int J Epidemiol. 2017;46(3):850-859. doi:10.1093/ije/dyw354
14. Jónsson H, Sulem P, Kehr B, et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. Nature. 2017;549(7673):519-522. doi:10.1038/nature24018
15. Kong A, Frigge ML, Masson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature. 2012;488(7412):471-475. doi:10.1038/nature11396
16. Francioli LC, and Paz P. Polak, Koren A, et al. Genome-wide patterns and properties of de novo mutations in humans. Nat Genet. 2015;47(7):822-826. doi:10.1038/ng.3292

17. Goldmann JM, Seplyarskiy VB, Wong WSW, et al. Germline De Novo Mutation Clusters Arise During Oocyte Aging in Genomic Regions With High Double-Strand-Break Incidence. Obstet Gynecol Surv. 2018;73(9):531-532. doi:10.1097/ogx.0000000000000604
18. Wong WSW, Solomon BD, Bodian DL, et al. New observations on maternal age effect on germline de novo mutations. Nat Commun. 2016;7(1). doi:10.1038/ncomms10486
19. Jenkins TG, Aston KI, Pflueger C, Cairns BR, Carrell DT. Age-Associated Sperm {DNA} Methylation Alterations: Possible Implications in Offspring Disease Susceptibility. Greally JM, ed. {PLoS} Genet. 2014;10(7):e1004458. doi:10.1371/journal.pgen.1004458
20. Gratten J, Wray NR, Peyrot WJ, McGrath JJ, Visscher PM, Goddard ME. Risk of psychiatric illness from advanced paternal age is not predominantly from de novo mutations. Nat Genet. 2016;48(7):718-724. doi:10.1038/ng.3577
21. Myrskylä M, Fenelon A. Maternal Age and Offspring Adult Health: Evidence From the Health and Retirement Study. Demography. 2012;49(4):1231-1257. doi:10.1007/s13524-012-0132-x
22. Tearne JE, Robinson M, Jacoby P, Li J, Newnham J, McLean N. Does Late Childbearing Increase the Risk for Behavioural Problems in Children? A Longitudinal Cohort Study. Paediatr Perinat Epidemiol. 2015;29(1):41-49. doi:10.1111/ppe.12165
23. Janecka M, Haworth CMA, Ronald A, et al. Paternal Age Alters Social Development in Offspring. J Am Acad Child Adolesc Psychiatry. 2017;56(5):383-390. doi:10.1016/j.jaac.2017.02.006
24. Myrskylä M, Barclay K, Goisis A. Advantages of later motherhood. GynaOrlebeke JF, Knol DL, Boomsma DI, Verhulst FC. Frequency of Parental Report of Problem Behavior in Children Decreases with Increasing Maternal Age at Delivery. Psychol Rep. 1998;82(2):395-404. doi:10.2466/pr0.1998.82.2.395
25. Orlebeke JF, Knol DL, Boomsma DI, Verhulst FC. Frequency of Parental Report of Problem Behavior in Children Decreases with Increasing Maternal Age at Delivery. Psychol Rep. 1998;82(2):395-404. doi:10.2466/pr0.1998.82.2.395
26. Lawlor DA, Mortensen L, Andersen A-MN. Mechanisms underlying the associations of maternal age with adverse perinatal outcomes: a sibling study of 264 695 Danish women and their firstborn offspring. Int J Epidemiol. 2011;40(5):1205-1214. doi:10.1093/ije/dyr084
27. Bray I, Gunnell D, Smith GD. Advanced paternal age: How old is too old? J Epidemiol Community Heal. 2006;60(10):851-853. doi:10.1136/jech.2005.045179
28. McMahon CA, Gibson FL, Allen JL, Saunders D. Psychosocial adjustment during pregnancy for older couples conceiving through assisted reproductive technology. Hum Reprod. 2007;22(4):1168-1174. doi:10.1093/humrep/del502
29. Kiernan KE. Becoming a Young Parent: A Longitudinal Study of Associated Factors. Br J Sociol. 1997;48(3):406. doi:10.2307/591138
30. Janecka M, Rijsdijk F, Rai D, Modabbernia A, Reichenberg A. Advantageous developmental outcomes of advancing paternal age. Transl Psychiatry. 2017;7(6):e1156. doi:10.1038/tp.2017.125
31. Crockett LJ, Bingham CR, Chopak JS, Vicary JR. Timing of first sexual intercourse: The role of social control, social learning, and problem behavior. J Youth Adolesc. 1996;25(1):89-111. doi:10.1007/bf01537382
32. Bailey JA, Hill KG, Oesterle S, Hawkins JD. Parenting practices and problem behavior across three generations: Monitoring, harsh discipline, and drug use in the intergenerational transmission of externalizing behavior. Dev Psychol. 2009;45(5):1214-1226. doi:10.1037/a0016129
33. Centraal Bureau voor de Statistiek. Geboorte; kerncijfers. 2018. https://opendata.cbs.nl/stat-

line/#/CBS/nl/dataset/37422ned/table?ts=1522410899684.

34. 34. Rescorla LA, Ginzburg S, Achenbach TM, et al. Cross-Informant Agreement Between Parent-Reported and Adolescent Self-Reported Problems in 25 Societies. J Clin Child Adolesc Psychol. 2013;42(2):262-273. doi:10.1080/15374416.2012.717870
35. Hudziak JJ, van Beijsterveldt CEM, Bartels M, et al. Individual differences in aggression: Genetic analyses by age, gender, and informant in 3-, 7-, and 10-year-old Dutch twins. Behav Genet. 2003;33(5):575-589. doi:10.1023/a:1025782918793
36. Open Science Collaboration. Estimating the reproducibility of psychological science. Science (80-). 2015;349(6251). doi:10.1126/science.aac4716
37. Kuiper RM, Buskens V, Raub W, Hoijtink H. Combining Statistical Evidence From Several Studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. Sociol Methods Res. 2012;42(1):60-81. doi:10.1177/0049124112464867
38. Reynolds WM. Reynolds Adolescent Depression Scale. 2000. doi:10.1002/9780470479216.corpsy0798
39. Birmaher B, Khetarpal S, Brent D, et al. The screen for child anxiety related emotional disorders (SCARED): Scale construction and psychometric characteristics. J Am Acad Child Adolesc Psychiatry. 1997;36(4):545-553. doi:10.1097/00004583-199704000-00018
40. Achenbach TM, Edelbrock C. Child behavior checklist. 1991:546-552.
41. Achenbach TM, Rescorla LA. Manual for the ASEBA School-Age Forms and Profiles. . (University of Vermont Youth, and Families RC for C, ed.). Burlington, VT; 2001.
42. de Winter AF, Oldehinkel AJ, Veenstra R, Brunnekreef JA, Verhulst FC, Ormel J. Evaluation of non-response bias in mental health determinants and outcomes in a large sample of pre-adolescents. Eur J Epidemiol. 2005;20(2):173-181. doi:10.1007/s10654-004-4948-6
43. Van Buuren S. Flexible Imputation of Missing Data. Chapman & Hall/CRC; 2012. doi:10.1201/b11826
44. Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychol Methods. 2002;7(2):147-177. doi:10.1037//1082-989x.7.2.147
45. Van Buuren S, Groothuis-Oudshoorn K. {MICE}: Multivariate imputation by chained equations in {R}. J Stat Softw. 2011;45(3):1-67. doi:10.18637/jss.v045.i03
46. Muthén LK, Muthén BO. How to use a {M}onte {C}arlo study to decide on sample size and determine power. Struct Equ Model. 2002;9:599-620. doi:10.1207/S15328007SEM0904_8
47. Hoijtink H. Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists. CRC Press; 2012. doi:10.1201/b11158
48. Kass RE, Raftery AE. Bayes Factors. J Am Stat Assoc. 1995;90(430):773-795. doi:10.1080/01621459.1995.10476572
49. Gu X, Mulder J, Hoijtink H. Approximated adjusted fractional {B}ayes factors: A general method for testing informative hypotheses. Br J Math Stat Psychol. 2018;71(2):229-261. doi:10.1111/bmsp.12110
50. Robson K, Pevalin DJ. Gender differences in the predictors and socio-economic outcomes of young parenthood in Great Britain. Res Soc Stratif Mobil. 2007;25(3):205-218. doi:10.1016/j.rssm.2007.08.002
51. Manning WD, Trella D, Lyons H, Toit NC Du. Marriageable Women: A Focus on Participants in a Community Healthy Marriage Program. Fam Relat. 2010;59(1):87-102. doi:10.1111/j.1741-

3729.2009.00588.x

52. Sandberg-Thoma SE, Dush CMK. Indicators of Adolescent Depression and Relationship Progression in Emerging Adulthood. J Marriage Fam. 2014;76(1):191-206. doi:10.1111/jomf.12081
53. Trillingsgaard T, Sommer D. Associations between older maternal age, use of sanctions, and children's socio-emotional development through 7, 11, and 15 years. Eur J Dev Psychol. 2018;15(2):141-155. doi:10.1080/17405629.2016.1266248
54. Dolan C V, Geels L, Vink JM, et al. Testing Causal Effects of Maternal Smoking During Pregnancy on Offspring's Externalizing and Internalizing Behavior. Behav Genet. 2016;46(3):378-388. doi:10.1007/s10519-015-9738-2
55. Saha S, Barnett AG, Buka SL, McGrath JJ. Maternal age and paternal age are associated with distinct childhood behavioural outcomes in a general population birth cohort. Schizophr Res. 2009;115(2-3):130-135. doi:10.1016/j.schres.2009.09.012

SUPPLEMENTARY MATERIAL

Table S1: Mean and SD for Externalizing and Internalizing Problems of Girls

Rater	Cohort	Externalizing	Internalizing	N-Ext/N-Int
Child	Gen-R	1.77 (1.81)	2.35 (2.15)	2,054/2,058
	RADAR-Y	10.19 (6.75)	0.15 (0.93)	213/120
	TRAILS	7.55 (5.40)	12.06 (7.51)	1,115/1,106
Mother	Gen-R	3.34 (4.29)	4.90 (4.96)	2,305/2,305
	NTR	4.79 (5.33)	4.81 (5.14)	5,626/5,577
	TRAILS	7.26 (6.21)	7.89 (6.25)	1,006/1,002
Father	Gen-R	3.38 (4.20)	4.54 (4.67)	1,655/1,656
	NTR	4.03 (4.84)	3.62 (4.25)	3,764/3,734
Teacher	NTR	2.12 (4.28)	4.29 (4.85)	3,314/3,268
	TRAILS	0.26 (0.59)	0.96 (1.09)	992/993

Table S2: Mean and SD for Externalizing and Internalizing Problems of Boys

Rater	Cohort	Externalizing	Internalizing	N-Ext/N-Int
Child	Gen-R	2.12 (2.02)	1.95 (2.00)	1,955/1,959
	RADAR-Y	10.93 (7.44)	-0.20 (0.77)	278/146
	TRAILS	9.85 (6.83)	10.47 (7.23)	1,073/1,065
Mother	Gen-R	4.51 (5.41)	4.81 (5.14)	2,244/2,245
	NTR	6.46 (6.73)	4.55 (4.99)	5,460/5,409
	TRAILS	9.59 (7.62)	7.80 (6.16)	959/953
Father	Gen-R	4.61 (5.48)	4.63 (4.77)	1,604/1,603
	NTR	5.32 (5.87)	3.50 (4.22)	3,656/3,640
Teacher	NTR	4.48 (6.96)	4.52 (5.08)	3,222/3,178
	TRAILS	0.63 (0.88)	1.03 (1.14)	933/931

Table S3: Parental Age Predicting Externalizing Problems from Exploratory Results

Rater	Cohort	Age F.	Age² F.	r^2	Age M.	Age² M.	r^2
		β (*p*-value)	β (*p*-value)		β (*p*-value)	β (*p*-value)	
Child	Gen-R	t-.06 (<.007)	.08 (<.001)	.01	-.05 (.02)	.05 (.03)	.01
	RADAR-Y	-.05 (.44)	.14 (.05)	.02	-.08 (.22)	.18 (.01)	.04
	TRAILS	-.01 (.83)	-.01 (.77)	.00	-.03 (.39)	-.03 (.36)	.00
Mother	Gen-R	-.10 (<.001)	.09 (<.001)	.01	-.10 (<.001)	.02 (<.001)	.02
	NTR	-.12 (<.001)	.08 (<.001)	.01	-.11 (<.001)	.06 (<.001)	.02
	TRAILS	.09 (.02)	.08 (.04)	.01	-.13 (<.001)	.06 (.06)	.02
Father	Gen-R	-.10 (<.001)	.08 (.003)	.01	-.08 (.001)	.07 (<.001)	.01
	NTR	-.13 (<.001)	.07 (<.001)	.02	-.12 (<.001)	.06 (<.001)	.02
Teacher	NTR	-.05 (<.001)	.03 (.047)	.00	-.04 (.001)	.04 (.009)	.00
	TRAILS	-.08 (.03)	.06 (.11)	.01	-.11 (<.001)	.04 (.20)	.01

F. = Father.
M. = Mother

Table S4: Age Father and Covariates Predicting Externalizing Problems from Exploratory Results

Rater	Cohort	Age	Age²	SES	Gender Child	r^2
		β (*p*-value)	β (*p*-value)	β (*p*-value)	β (*p*-value)	
Child	Gen-R	-.05 (.03)	.07 (<.001)	-.07 (<.001)	-.08 (<.001)	.02
	RADAR-Y	-.05 (.50)	.13 (.07)	-.06 (.39)	-.06 (.39)	.02
	TRAILS	-.01 (.88)	-.01 (.77)	-.01 (.67)	-.18 (<.001)	.03
Mother	Gen-R	-.09 (<.001)	.07 (.004)	-.08 (<.001)	-.15 (<.001)	.04
	NTR	-.10 (<.001)	.07 (<.001)	-.08 (<.001)	-.13 (<.001)	.05
	TRAILS	-.04 (.27)	.06 (.10)	-.17 (<.001)	-.16 (<.001)	.06
Father	Gen-R	-.10 (<.001)	.06 (.01)	-.06 (.03)	-.15 (<.001)	.04
	NTR	-.11 (<.001)	.06 (<.001)	-.13 (<.001)	-.14 (<.001)	.05
Teacher	NTR	-.04 (.006)	.02 (.125)	-.10 (<.001)	-.17 (<.001)	.04
	TRAILS	-.05 (.20)	.05 (.19)	-.13 (<.001)	-.25 (<.001)	.09

Table S5: Age Mother and Covariates Predicting Externalizing Problems from Exploratory Results

Rater	Cohort	Age	Age^2	SES	Gender Child	r^2
		β (*p*-value)	β (*p*-value)	β (*p*-value)	β (*p*-value)	
Child	Gen-R	-.04 (.11)	.04 (.12)	-.06 (.009)	-.08 (<.001)	.02
	RADAR-Y	-.07 (.25)	.17 (.01)	.-05 (.43)	-.05 (.41)	.04
	TRAILS	-.02 (.51)	-.02 (.58)	-.01 (.79)	-.18 (<.001)	.03
Mother	Gen-R	-.08 (<.001)	.06 (.004)	-.06 (.006)	-.14 (<.001)	.04
	NTR	-.09 (<.001)	.06 (<.001)	-.12 (<.001)	-.14 (<.001)	.05
	TRAILS	-.08 (.02)	.06 (.06)	-.15 (<.001)	-.16 (<.001)	.07
Father	Gen-R	-.07 (.009)	.06 (.02)	-.05 (.09)	-.15 (<.001)	.04
	NTR	-.10 (<.001)	.05 (<.001)	-.12 (<.001)	-.14 (<.001)	.05
Teacher	NTR	-.03 (.035)	.03 (.019)	-.10 (<.001)	-.17 (<.001)	.04
	TRAILS	-.07 (.03)	.05 (.11)	-.12 (<.001)	-.25 (<.001)	.09

Table S6: Exploratory Results for Parental Age Predicting Internalizing Problems

Rater		Age F.	Age² F.	r^2	Age M.	Age² M.	r^2
		β (*p*-value)	β (*p*-value)		β (*p*-value)	β (*p*-value)	
Child	Gen-R	-.03 (.001)	.05 (.020)	.00	-.02 (.32)	.04 (.07)	.00
	RADAR-Y	-.03 (.69)	.03 (.76)	.01	-.04 (.64)	.06 (.41)	.01
	TRAILS	.00 (.98)	-.01 (.78)	.00	-.02 (.55)	.03 (.40)	.00
Mother	Gen-R	-.04 (.12)	.06 (.02)	.00	-.06 (.01)	.05 (.05)	.01
	NTR	-.06 (<.001)	.05 (<.001)	.00	-.06 (<.001)	.03 (.022)	.00
	TRAILS	.01 (.81)	.05 (.17)	.00	-.05 (.12)	.04 (.26)	.00
Father	Gen-R	-.05 (.06)	.06 (.02)	.00	-.03 (.21)	.03 (.28)	.00
	NTR	-.07 (<.001)	.04 (.013)	.01	-.07 (<.001)	.02 (.116)	.01
Teacher	NTR	-.01 (.538)	.02 (.301)	.00	-.01 (.719)	.01 (.299)	.00
	TRAILS	-.02 (.56)	.01 (.89)	.00	-.04 (.21)	.04 (.20)	.00

F. = Father.
M. = Mother.

Table S7: Age Father and Covariates Predicting Internalizing Problems from Exploratory Results

Rater	Cohort	Age	Age²	SES	Gender Child	r^2
		β (*p*-value)	β (*p*-value)	β (*p*-value)	β (*p*-value)	
Child	Gen-R	-.02 (.47)	.04 (.08)	-.04 (.07)	.10 (<.001)	.02
	RADAR-Y	-.03 (.75)	.05 (.53)	-.05 (.61)	.27 (.001)	.09
	TRAILS	.01 (.84)	-.01 (.72)	-.02 (.53)	.11 (<.001)	.01
Mother	Gen-R	-.02 (<.001)	.03 (.20)	-.10 (<.001)	.00 (.90)	.01
	NTR	-.05 (<.001)	.04 (.001)	-.06 (<.001)	.02 (.081)	.01
	TRAILS	.03 (.48)	.04 (.24)	-.06 (.06)	.04 (.25)	.01
Father	Gen-R	-.04 (.14)	.04 (.09)	-.05 (.04)	-.02 (.44)	.01
	NTR	-.06 (<.001)	.03 (.034)	-.07 (<.001)	-.01 (.495)	.01
Teacher	NTR	-.00 (.846)	.01 (.386)	-.055 (<.001)	-.03 (.007)	.00
	TRAILS	.02 (.49)	-.02 (.54)	-.16 (<.001)	-.01 (.78)	.03

Table S8: Age Mother and Covariates Predicting Internalizing Problems from Exploratory Results

Rater	Cohort	Age	Age²	SES	Gender Child	r^2
		β (*p*-value)	β (*p*-value)	β (*p*-value)	β (*p*-value)	
Child	Gen-R	-.01 (.64)	.04 (.11)	-.04 (.09)	.11 (<.001)	.02
	RADAR-Y	-.02 (.78)	.09 (.23)	-.04 (.62)	.27 (.001)	.09
	TRAILS	-.02 (.63)	-.03 (.27)	-.01 (.66)	.11 (<.001)	.01
Mother	Gen-R	-.03 (.18)	.03 (.25)	-.09 (<.001)	.00 (.88)	.01
	NTR	-.04 (<001)	.02 (.049)	-.06 (<.001)	.02 (.085)	.01
	TRAILS	-.04 (.30)	.03 (.34)	-.04 (.19)	.03 (.28)	.01
Father	Gen-R	-.02 (.56)	.02 (.52)	-.06 (.03)	.02 (.47)	.01
	NTR	-.05 (<.001)	.02 (.220)	-.07 (<.001)	-.01 (.489)	.01
Teacher	NTR	.00 (.936)	.01 (.434)	-.05 (<.001)	-.03 (.009)	.00
	TRAILS	.01 (.72)	.03 (.33)	-.15 (<.001)	-.01 (.71)	.03