

The gift of time? School starting age and mental health

Thomas S. Dee^{1,2} | Hans Henrik Sievertsen^{3,4} ¹Graduate School of Education, Stanford University, Stanford, CA, USA²NBER, Cambridge, MA, USA³University of Bristol, Bristol, United Kingdom⁴VIVE, Copenhagen, Denmark**Correspondence**Hans Henrik Sievertsen, University of Bristol, Priory Road Complex, Priory Road, Bristol BS8 1TU, United Kingdom.
Email: h.h.sievertsen@bristol.ac.uk**Funding information**

The Danish Council for Strategic Research, Grant/Award Number: DSF-09-070295

JEL Classification: I21; I28

Abstract

Using linked Danish survey and register data, we estimate the causal effect of age at kindergarten entry on mental health. Danish children are supposed to enter kindergarten in the calendar year in which they turn 6 years. In a "fuzzy" regression-discontinuity design based on this rule and exact dates of birth, we find that a 1-year delay in kindergarten entry dramatically reduces inattention/hyperactivity at age 7 (effect size = -0.73), a measure of self-regulation with strong negative links to student achievement. The effect is primarily identified for girls but persists at age 11.

KEYWORDS

mental health, school starting age

1 | INTRODUCTION

Over the last half century, the age at which children in the United States initiate their formal schooling has slowly increased. Historically, U.S. children attended kindergarten as 5-year olds and first grade as 6-year olds. However, roughly 20% of kindergarten students are now 6 years old (e.g., *The Boston Globe*, 2014; *The New York Times*, 2010). This "lengthening of childhood" reflects in part changes in state laws that moved forward the cutoff birth date at which 5-year olds were eligible for entering kindergarten (Deming & Dynarski, 2008). However, most of the increase in school starting ages is due to academic "redshirting"; an increasingly common decision by parents to seek developmental advantages for their children by delaying their school entry (i.e., the "gift of time"). Redshirting is particularly common for boys and in socioeconomically advantaged families (Bassok & Reardon, 2013).¹ Delayed school starts are also common in other developed countries. For example, in Denmark, one out of five boys and one out of 10 girls have a delayed school start.² The conjectured benefits of starting formal schooling at an older age reflect two broad mechanisms. The first is relative maturity; students may benefit when they start school at an older age simply because they have, on average, a variety of developmental advantages relative to their classroom peers. The second mechanism, absolute maturity, reflects the hypothesis that formal schooling is more developmentally appropriate for older children.

The decision of whether to delay a child's formal schooling is a recurring topic in the popular press (e.g., *The New Yorker*, 2013) with most coverage suggesting that there are educational and economic benefits to delayed school entry. However, the available research evidence largely suggests otherwise. A number of early studies (e.g., Bedard & Dhuey, (2006)) did indeed show that children who start school later have, on average, higher performance on in-school tests (i.e.,

¹For example, according to the U.S. National Center for Education Statistics, 14% of the children who delayed school entrance in 2010 were children of parents in the lowest quintile of socioeconomic status, whereas 24% were children of parents in the highest quintile. The measure of socioeconomic status is based on parental education, occupation, and household income at the time of data collection.

²Based on the 2003 and 2004 birth cohorts. Throughout this paper, we refer to school starting age as the age at which a child enters kindergarten, which in Danish is called grade zero or "Børnehaveklasse."

even after adjusting for the endogenous decision to redshirt). However, more recent studies suggest that these findings simply reflect the fact that children who start school later are older when the test is given.³

In this study, we examine the causal effect of higher school starting age on different dimensions of mental health among similarly aged Danish children. We exploit a unique data source: a large-scale survey of Danish children (the Danish National Birth Cohort or DNBC). Denmark constitutes a relevant setting for studying the effects of school starting age on mental health as existing evidence suggests that Denmark might be a special case. Dalsgaard, Humlum, Nielsen, and Simonsen (2012) find no evidence of a causal link between age at school entry and attention deficit hyperactivity disorder (ADHD) diagnoses in Denmark, although such a link has been documented for Canada (Morrow et al., 2012), Taiwan (Chen et al., 2016), and the United States (Elder & Lubotsky, 2009). However, as Dalsgaard et al. (2012) point out, ADHD diagnoses are considerably less common in Denmark, compared to the United States. The open question is thus whether the findings by Dalsgaard et al. (2012) are due to differences in standards for ADHD diagnoses between countries or because school starting age has no effect on mental health in Denmark. To assess this question, we use a widely validated mental health screening tool (the Strengths and Difficulties Questionnaire or SDQ). The SDQ was explicitly designed for children and generates measures of several distinct psychopathological constructs based on evaluations by the child's mother.

We are able to identify the effects of a delayed school start through a “fuzzy” regression-discontinuity (RD) design based on the day of birth. We identified the day of birth and school starting age of children in the DNBC by matching these data to population data available in the Danish administrative registry and Ministry of Education records. In Denmark, children are supposed to enter school in the calendar year in which they turn 6 years. Using data on children's exact date of birth, we find that school starting age does indeed “jump” discontinuously for children born January 1 or later relative those born December 31 or earlier.

Our results indicate that a 1-year increase in the school starting age leads to significantly improved mental health (i.e., reducing the “total difficulties” scores at age 7 by 0.6 *SD*). Interestingly, we find that these effects are largely driven by a large reduction (effect size = -0.73) in a single SDQ construct: the SDQ's inattention/hyperactivity score. Consistent with a literature that emphasizes the importance of self-regulation for student outcomes, we find that this construct is most strongly correlated with the in-school performance of Danish children. We are also able to examine whether these short-term effects persist using the most recently available data that tracks students to age 11. We find that the large and concentrated effects largely persist to later childhood (i.e., an effect size for inattention/hyperactivity of -0.69). The treatment effect is primarily identified for girls, as we have very little identifying variation in school starting age for boys, because they are less likely to comply with the school starting age cutoff. We also find evidence that these effects are heterogeneous. Using an approach introduced by Bertanha and Imbens (2014), we present evidence on the heterogeneity that distinguishes the “compliers” from the “never takers” to “always takers” in our “intent-to-treat” (ITT) design.

This paper proceeds as follows: Section 2 provides brief discussions of the theoretical relationships between school-starting delays and child outcomes and a description of the institutional setting. Section 3 introduces this study's data, particularly the DNBC and the SDQ measures. Section 4 presents the empirical framework. Section 5 presents the results. Section 6 relates our findings to the findings in the literature. Section 7 concludes this paper.

2 | BACKGROUND: SCHOOL STARTING AGE AND MENTAL HEALTH

One rationale for the growing number of parents who choose to delay their children's school starting age involves the perceived benefits of relative maturity for young children. This conjecture, popularized by Malcolm Gladwell's 2008 book, *Outliers*, turns on the claim that children who are slightly older than their peers experience early successes that are then followed by recursive processes of reinforcement and support.⁴ A second class of rationales for delayed school starting age

³Angrist and Pischke (2008) offer this as an example of a “fundamentally unidentified” research question. A student's school starting age by definition equals their current age minus their time in school. So for measures of in-school performance, the effects of school starting age cannot be disentangled from age-at-test and time-in-school effects. Some settings provide potential solutions for this issue. For example, using Norwegian data, Black, Devereux, and Salvanes (2011) find that a higher school starting age implies a small, negative effect on an IQ test taken outside of school at age 18. Another strategy to assess school starting age effects in school systems with several cutoff days throughout the year, as, for example, in the UK (Crawford, Dearden, & Meghir, 2007).

⁴Though parents' belief in the gains from relative maturity may be widespread, the empirical evidence on the direct educational benefits from a higher relative age is at best equivocal. In particular, a random-assignment study by Cascio and Schanzenbach (2016) finds that students who are old for their cohort may have poorer outcomes because of peer-group effects. To the extent that such effects exist in our Danish data, it implies that we are understating the targeted mental-health benefits of a higher school starting age.

turns on the perceived benefits of increasing the absolute maturity of children when they begin formal schooling. That is, a delay in formal schooling may benefit student outcomes because slightly older children are more developmentally aligned with the demands and opportunities of formal schooling.

Before they begin formal schooling, most children in Denmark (i.e., over 95%) are in childcare that is publicly provided and organized at the municipal level. Childcare consists of center-based nurseries and family daycare for children aged 1 to 3 years and daycare for children aged 3 to 6 years. The standards required of center-based daycare and their staff are high compared to other Organization for Economic Cooperation and Development countries (Datta Gupta & Simonsen, 2010). Compulsory education in Denmark begins in “grade zero” (i.e., kindergarten) in August of the year in which the child turns 6 years. Until 2009, kindergarten was not mandatory, but 98% of children attended anyway (Browning & Heinesen, 2007). According to the Danish Ministry for Education, the objective of kindergarten is to provide a bridge between “play-based activities” in preschool and formal “classroom teaching” in school. In contrast to preschool, there is a minimum number of hours in kindergarten (1,200 per year/30 per week) and at least 600 of these hours should be used for teaching within six centrally decided topics. A later school start is thus related to a later departure from “play-based activities.” A recent report documented that Danish preschools provide good support for children’s development of “socio-emotional skills,” but less strong support for the development of cognitive skills (Rambøll Management Consulting, Aarhus University, & University of Southern Denmark, 2016). Children who enroll in kindergarten later thus spent more time in a setting with more play-based activities that support the development of skills that presumably are important for mental health (i.e., socio-emotional skills).

3 | DATA

We create our analysis samples by matching children included in the DNBC to data available for the full Danish population from the national administrative registers. The DNBC provides detailed measures of children’s mental health at ages 7 and 11 years. The national administrative registers provide information on the child’s birthday (i.e., the forcing variable in our RD design) as well as data on child and family traits at baseline. We describe each of these datasets below.

3.1 | The Danish National Birth Cohort

The DNBC is a Danish nationwide cohort study based on a large sample of women who were pregnant between 1996 and 2002 (i.e., roughly 10% of the births in the population during this period). Nearly, 93,000 woman participated in the baseline interviews (i.e., during pregnancy). In this paper, we use data from the fifth and sixth survey wave, when the focal child was respectively 7 and 11 years old. During the fifth survey wave, the respondent was asked to identify when the child started kindergarten, which we use to identify their school starting age. Critically, the fifth and sixth survey waves also included the 25-item SDQ, which we describe in more detail below.⁵

3.2 | The Strengths and Difficulties Questionnaire

The SDQ is a mental-health screening tool designed specifically for children and teens and is in wide use internationally both in clinical settings and in research on child development. The questionnaire, which was developed by English child psychiatrist Robert N. Goodman in the mid-1990s, consists of 25 items (Goodman, 1997) that may describe the child in question.⁶ Examples of the items include “restless, overactive, cannot stay still for long” and “good attention span, sees work through to the end.” For each item, the rater (in our case, the mother) is asked to “consider the last 6 months” and to mark the description of the child in one of three ways: Not True, Somewhat True, and Certainly True. The established scoring procedure for the SDQ links each of the 25 items to one (and only one) of five distinct subscores: emotional symptoms, conduct problems, inattention/hyperactivity, peer problems, and a pro-social scale (measured with the opposite sign, compared to the other dimensions). Each subscore has five uniquely linked items, and the response to each item is scored as 0, 1, or 2. The value for the subscore is simply the sum of the ratings for its five linked items. So each subscore

⁵Each survey wave was fielded on a rolling basis so as to get child data at roughly the same age. Differential response times necessarily create some variation in the age at observation. However, we control for each child’s age at the time of interview and find that this age is well balanced around the threshold in our RD design.

⁶The complete SDQ questionnaire and aggregation scheme can be found on the website <http://www.sdqinfo.org/>.

has a range of 0 to 10. The total “difficulties” score is the sum of the subscales, excluding the pro-social score, and can range from 0 to 40. For this difficulties score, values between 0 and 13 are regarded as normal, whereas scores 14–16 are borderline and scores from 17 to 40 are regarded as abnormal. For the pro-social scale, 6–10 is normal, 5 is borderline, and 0–4 is abnormal. In our main analyses, we standardize each score (i.e., using the full population in each survey wave) so that our coefficients of interest can be interpreted as effect sizes. However, we also present linear probability models for the probability of an abnormal rating.

The development of the SDQ items (and their scaling) was conducted with reference to the main categories of child mental-health disorders recognized by contemporary classification systems like the *Diagnostic and Statistical Manual of Mental Disorders, Fourth edition* (American Psychiatric Association, 1994). Psychometric studies have generally confirmed the convergent and discriminant validity of the five-factor structure of the SDQ in a variety of populations (Achenbach et al., 2008), though some studies suggest there should be fewer subscores.⁷ Furthermore, in both the parent and teacher versions, the SDQ has demonstrated satisfactory internal consistency, test–retest reliability, and interrater agreement (e.g., Achenbach et al., 2008; Stone, Otten, Engels, Vermulst, & Janssens, 2010). The SDQ produces scores that are highly correlated with those from earlier prominent screening devices, the Rutter questionnaire and the Child Behavior Checklist (Goodman, 1997; Goodman & Scott, 1999).

To understand the properties of the SDQ subscores in our particular research context, we also examined how the SDQ scores of children in the DNBC predicted their in-school test performance on the Danish National Tests in two subjects (reading and mathematics). Specifically, we separately regressed the test score on the five SDQ subscores measured at age 7. Although there are also somewhat anomalous results, for example, pro-social scores predict lower test scores in both subjects and all grades (i.e., effect sizes of 0.04 and 0.05), our main finding is that the two constructs associated with “externalizing behavior”—the conduct and inattention/hyperactivity constructs—strongly predict lower test performance across all grades and subjects. A 1 *SD* increase in the inattention/hyperactivity score predicts a reduction in future test performance ranging from 0.14 *SD* to 0.16 *SD*.⁸

The uniquely strong link between the inattention/hyperactivity subscore and future student performance is noteworthy but not necessarily surprising. The inattention/hyperactivity construct is effectively synonymous with the concept of self-regulation (i.e., the voluntary control of impulses in service of desired goals; Blake, Piovesan, Montinari, Warneken, & Gino, 2015). And an extensive literature has documented the importance of such self-regulation for student success (e.g., Duckworth & Carlson, 2013).⁹ Interestingly, one of the theorized mechanisms through which higher school starting ages are thought to be developmentally beneficial, involves self-regulation. In particular, the extended periods of pretend play available to children who delay their school start may enhance their capacity for this important psychological adaption.

3.3 | The Danish administrative registers

The Danish administrative data actually consist of several individual registers including the birth records, the income registers, and the education registers. All datasets are hosted by Statistics Denmark and linked by a unique personal identifier. The critical variable we draw from the registers forms the basis for the forcing variable in our RD design (i.e., the exact date of birth). However, we also use the registers to construct a variety of other family- and child-specific control variables. For the children, we use information from the registers on birth weight, origin, gender, and gestational age. For the parents, we use information on gross annual income, educational attainment, and age. We also record the number of siblings (living in the household) when the child is 2 years old using register data. Before we link the children to their parents and siblings, we adjust the birth year to run from July to June instead of January to December. For example, all children born in the period July 2000 to June 2001 are merged to parents' characteristics for the calendar year January to December 1999.

3.4 | Sample selection and summary statistics

In the analyses, we use the 8,092 children born in the 30-day window around the cutoff date January 1, with information on school starting age and a completed SDQ questionnaire either at age 7 or at age 11, or at both ages.¹⁰ In Table 1, we

⁷We independently examined the item-level responses in our DNBC data using a principal component analysis. The principal component analysis revealed the same five dimensions as the standardized procedure.

⁸Results are available on request.

⁹The concept of self-regulation is also widely thought to be equivalent to the “Big 5” construct of conscientiousness, another highly outcome-relevant personality trait. Heckman and Kautz (2012) note that “conscientiousness—the tendency to be organized, responsible, and hardworking—is the most widely predictive of the commonly used personality measures.”

¹⁰Analyses based on other windows give qualitatively similar results as we show in Figures A2 and A3.

TABLE 1 Descriptive statistics

	Population data			Survey			p value
	Mean	SD	N	Mean	SD	N	
Born after January 1 cutoff	0.52	0.50	54,213	0.53	0.50	8,092	.21
School starting age				6.48	0.67	8,092	
School starting age > 6 years				0.78	0.41	8,092	
Female	0.48	0.50	54,213	0.50	0.50	8,092	.07
Birthweight (g)	3,474.63	622.11	53,292	3,528.83	601.20	8,050	.00
Non-Western origin	0.15	0.35	54,213	0.02	0.13	8,092	.00
Parents' years of schooling	14.13	2.82	53,121	15.43	2.01	8,080	.00
Parents gross income	85.87	60.45	53,121	98.63	85.72	8,080	.00
Mother's age at childbirth	29.88	4.93	53,002	30.72	4.31	8,079	.00
Father's age at childbirth	32.71	5.90	51,187	33.02	5.25	7,872	0.00

Note. Birth weight is measured in grams. Educational length is measured in years. Parents are defined as non-Western if they are immigrants to Denmark from a non-Western country according to the classification by Statistics Denmark. The mother's single status is one if the child is living with the mother, and the mother is not married or cohabiting. The gross income is measured in 1,000 DKK and adjusted to the 2010 level using the consumer price index. The parents' employment is for November in the lagged year.

show descriptive statistics for the key variables from our linked DNBC and register data compared to the full population of children born within the same window. The first row of the table shows that slightly more children are born after the cutoff date in our sample (53%), but that this rate is not significantly different from the rate in the general population. Given these rates and the school starting age rule that implies that children born before (after) the cutoff are less (more) than 6 years old at school start, we would expect that 53% of the children are older than 6 years at school start, but the third row shows that this is not the case, as almost four out of five children are older than 6 years at school start. This indicates that a substantial number of children born before the cutoff do not comply to the school starting age rule and instead postpone enrollment.

Table 1 also shows that compared to the population, the children in our survey data have a higher birth weight, are less likely to be of non-Western origin, and their parents have completed more years of education, have a higher gross income, and were older at child birth. Our survey data are thus not representative for the population of children born in these cohorts. Although this nonrandom selection into our survey data implies an external-validity caveat to our study, it does violate the internal-validity of our RD design as Figure A1 shows no sign of a jump in attrition around the cutoff. Participation in these surveys is balanced around the birthday threshold.

4 | EMPIRICAL FRAMEWORK

4.1 | The Danish context

As children are supposed to enroll in school the year they turn 6 years, school starting age should jump discontinuously as birthdays change from December 31 to January 1. Children who are born on January 1 and who comply with the rules will have a school starting age that is 1 year higher (and one extra year of daycare) relative to the children born just 1 day earlier. However, compliance with this rule is not strict. That is, it is possible to postpone enrollment in school. However, this requires some effort of the parents, including meeting with representatives from the future school and the municipality administration. Contingent on individual evaluations, children may also enroll in grade zero 1 year earlier (i.e., if their birthday is before October 1). Kindergarten class is part of the primary school and free of charge in the public schools.

4.2 | RD design

Our broad question of interest involves how school starting age influences the SDQ-based measures of mental health (Y) for individual i with covariates X_i . We represent this by the following linear specification:

$$Y_i = \beta_0 + \beta_1 SSA_i + \phi' X_i + e_i. \quad (1)$$

Credibly identifying the causal effect of school starting age on these outcomes is challenging because parents are likely to make decisions about when their child begins school based on information unobserved by researchers. In particular, parents who know their children face developmental challenges may be more likely to delay their child's initiation of

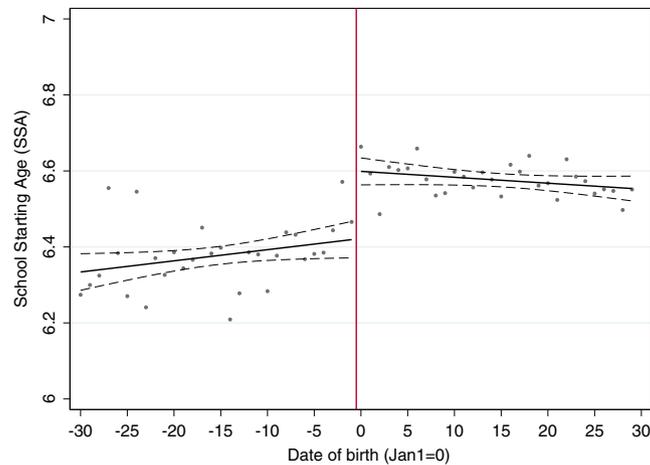


FIGURE 1 Date of birth and school starting age. Thirty days bandwidth and 1 day bins [Colour figure can be viewed at wileyonlinelibrary.com]

formal schooling (i.e., negative selection into treatment). Ordinary least squares (OLS) estimates of (1) are consistent with this concern. For example, OLS estimates suggest that children who start school late have substantially higher levels of inattention/hyperactivity.

We seek to identify the causal effect of school starting age by leveraging the variation created by the Danish rule that children are supposed to enroll in school the year they turn 6 years. That is, we implement an RD design that exploits the “jump” in school starting age that occurs for children born January 1 or later relative to those born earlier. So the forcing variable in this RD design (i.e., day_i) is the child’s exact birth date relative to the January 1 cutoff.¹¹ Our reduced-form equation of interest models the SDQ-based outcomes as a flexible function of this forcing variable and a “jump” at the policy-induced threshold:

$$Y_i = \gamma_0 + \gamma_1 \mathbf{1}(day_i \geq 0) + g(day_i) + \rho' \mathbf{X}_i + \epsilon_i. \quad (2)$$

Our parameter of interest is γ_1 , which identifies the discrete change in subsequent child outcomes for those born January 1 or later, controlling for a smooth function of their day of birth and other observed traits. Although Lee and Lemieux (2010) suggest that standard errors should be clustered on the running variable (i.e., date of birth), we show conventional heteroscedasticity-consistent standard errors, as these are slightly larger than clustered standard errors in our case, and we prefer to show the more conservative approach. We also report and discuss the corresponding IV estimates of β_1 from (1). These estimates are equivalent to the ratio of our reduced-form estimates to the first-stage estimates we describe below. In general, the causal warrant of such an RD design turns on whether the conditional change at the January 1 cutoff implies (a) variation in school starting age and (b) that this variation is “as good as randomized” (Lee & Lemieux, 2010). We now turn to evidence on both questions.

4.3 | Assignment to treatment

We first show that school starting age increases significantly for children whose birthdays are at the January 1 cutoff or later. One straightforward and unrestrictive way to show this is graphically as in Figure 1. The figures illustrate the conditional mean school starting age by day of birth relative to the cutoff (i.e., January 1=0, January 2=1). Children born January 1 or later generally comply and begin school in August of the year they turn 6 years. However, the compliance among children born in December is only partial. We examine some of the issues raised by this noncompliance with respect to our “ITT” analysis.

We present the results from estimating the first-stage relationship in Table 2. All the point estimates across these specifications (with and without covariates) and waves indicate that school starting age jumps by 0.18 to 0.20 years at the birth date cutoff.

¹¹That is, this forcing variables take on values of 0, 1, 2, and so forth for children born on January 1, 2, and 3, respectively. For children born on December 31, December 30, December 29, and so forth, the forcing variable takes on values of -1, -2, -3, and so forth.

TABLE 2 Regression-discontinuity estimates, first-stage regressions

	Age 7 wave		Age 11 wave	
	(1)	(2)	(3)	(4)
$day_i > 0$	0.18*** (0.03)	0.20*** (0.03)	0.18*** (0.04)	0.19*** (0.04)
F-stat	30.724	45.418	22.693	27.158
Observations	7,642	7,642	5,226	5,226
Controls	No	Yes	No	Yes

Note. Robust standard errors in parenthesis. Each cell shows the estimate from a single regression based on the local sample of children born 30 days before and after the cutoff. Covariates included are birth weight, origin, gender, parental education, parents' age, parental income, age at test, and birth year fixed effects. Missing values in covariates are replaced with zeros and indicators for missing variables are included.

* $p < .1$; ** $p < .05$; *** $p < .01$.

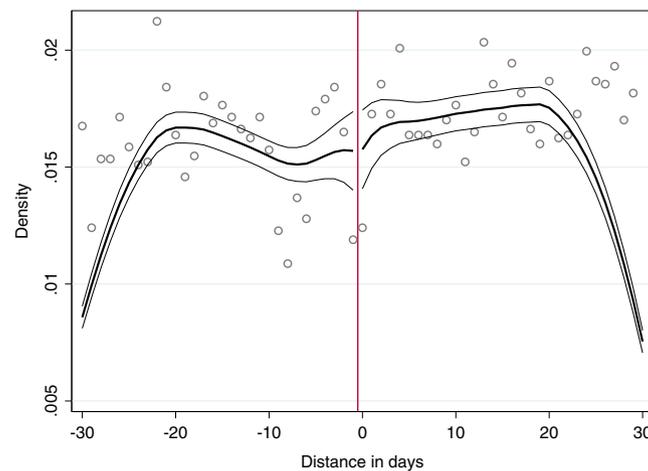


FIGURE 2 McCrary density test—Observations by date of birth. The jump is estimated to be -0.018 with a standard error of 0.086 [Colour figure can be viewed at wileyonlinelibrary.com]

4.4 | Validity of the RD design

The prior evidence demonstrates that there is a statistically significant jump in school starting age for children born January 1 and later. However, there are a number of reasons to be concerned that this relationship may not constitute a valid quasi-experiment. For example, a fundamental concern in any RD design is that the value of the forcing variable relative to the threshold may be systematically manipulated by those with a differential propensity for the relevant outcomes. In this setting, we might wonder whether expectant mothers either advance or delay the timing of their birth around the January 1 threshold and that the personal and family traits influencing this choice also influence child outcomes. We present two types of evidence that are consistent with the maintained hypothesis that there is no empirically meaningful manipulation of birth dates among our respondents.

First, we evaluate the distribution of births over the cutoff. Figure 2 shows the number of births around the cutoff date. The number of births are smoothly distributed around the threshold. We cannot reject the null hypothesis of no jump at this threshold. Interestingly, there appears to be a small drop in births around the new year (i.e., both December 31 and January 1), which may reflect some effort to avoid giving birth during a holiday (i.e., no planned c-sections). To consider possible issues related to undiagnosed “heaping” of the forcing variable, we also show in Figure A1, a histogram of birth dates local to the threshold. These data also suggest that the frequency of observations is continuous through the threshold that defines our ITT.

TABLE 3 Auxiliary regression-discontinuity estimates, balancing of the covariates

	Age 7 (1)	Age 11 (2)
\hat{Y} (total difficulties)	-0.01 (0.01)	-0.01 (0.01)
\hat{Y} (emotional symptoms)	-0.01 (0.01)	-0.01 (0.01)
\hat{Y} (conduct problems)	-0.01 (0.01)	-0.01 (0.01)
\hat{Y} (inattention/hyperactivity)	-0.01 (0.01)	-0.01 (0.01)
\hat{Y} (peer problems)	-0.01 (0.01)	-0.01 (0.01)
\hat{Y} (pro-social behavior)	-0.01 (0.01)	-0.01 (0.01)

Note. Robust standard errors in parenthesis. Each cell shows the estimate from a single regression based on the local sample of children born 30 days before and after the cutoff. We first regress the outcome variables (in parenthesis) of the following set of covariates: indicators for birth year, age at interview, parents' years of schooling, parents' gross income, parents' age at childbirth, birth weight, gender, and origin. We regress the predicted variable on an indicator for being born on January 1 or later, as well as the linear splines.

* $p < .1$; ** $p < .05$; *** $p < .01$.

Second, we use auxiliary regressions (i.e., the same specification as our RD design but with baseline covariates as the dependent variables) to examine the balance of observed traits of children and their families around the threshold. If the variation in school starting ages around this threshold is “as good as randomized,” we would expect the predetermined and observed traits of survey respondents to be similar on both sides of the threshold (i.e., no “jump” indicated by the RD estimates). Table A1 shows these results for each of the covariates. There is no clear sign of jumps in the covariates. An alternative strategy for testing covariate balance is to first regress the outcome variable on all covariates and compute the predicted values. These predicted value represents an index of all the covariates that are weighted by their OLS-estimated outcome relevance. In Table 3, we show the outcome of regressing this weighted average on the cutoff and time trends for each of the six dependent variables. As with the single-covariate regressions, there is no sign of a jump in any of these specifications.¹² The balance of outcome-relevant covariates around the January 1 threshold not only suggests a lack of manipulation of birth dates, but it is also general evidence for the validity of the RD design. We should also note that we also compared the balance of several developmental variables defined for the DNBC respondents before they attended kindergarten (e.g., making word sounds at 18 months). We found that these traits were balanced around the threshold (Table A2).

Another fundamental concern with any RD design involves the appropriate choice of functional form and bandwidth are. A visual inspection of our results provides one important and unrestrictive way to assess this concern. However, to examine the empirical relevance of functional-form issues and the choice of bandwidth more directly, we report results from various specifications in Figures A2 and A3.

An internal-validity concern unique to our application is that our treatment contrast necessarily conflates higher school starting ages with fewer years of schooling at the time of observation. That is, our ITT (i.e., a birth date of January 1 or later) implies both a higher school starting age and fewer years of formal schooling at the time parents rate their children on the SDQ. However, there are several reasons to deprecate the role of years of schooling in our analysis. For example, our pattern of results (i.e., effects on only one SDQ construct and not on the other measures of psychological adaptation) is not easily reconcilable with effects due to years of schooling but are consistent with the theorized effects of higher school

¹²Note that both Tables A1 and 3 show uncorrected standard errors and significance levels. Any corrections for multiple testing will make the conclusions of no correlation even stronger.

starting ages, as we would expect years of schooling to affect more dimensions. Furthermore, we find that our results are quite similar in size and significance among children at age 7 as at age 11 when the differences in years of schooling are relatively smaller. This pattern would only be consistent with effects due to years of schooling if a year has an additive effect without fade-out. Also, given that years of schooling are likely to have a positive effect on our mental-health measures (at least in later childhood), the collinearity in these measures (higher school starting age and fewer years of schooling) would not imply a bias that is problematic for our main findings.¹³

A second internal-validity threat unique to our setting involves reference biases in the SDQ ratings. It may be that children whose schooling is delayed are more likely to be rated positively simply because they appear to have better psychological adaptations than their younger classroom peers. Indeed, there is provocative evidence among U.S. children (Elder, 2010) that teachers are significantly more likely to rate children who are young for their grade as having ADHD. However, Elder (2010) finds that parental assessments (i.e., like those in the DNBC) are not subject to these biases; in all likelihood, because they have different reference points than teachers. Moreover, if the parent reports in the DNBC were subject to such biases, we would also expect to find effects on SDQ constructs other than inattention/hyperactivity but do not. We return to this issue in Section 5.4.

In sum, we find broad support for the internal validity of our research design. However, our analysis, like most RD applications, is qualified by several caveats related to external validity. First, because our estimates are defined by variation around the January 1 threshold, they are necessarily local estimates. Whether our results generalize to those born at other times is uncertain. There is evidence that shows that season of birth is not random with respect to parental characteristics (Buckles & Hungerman, 2013) so the localness of our RD estimates may have some empirical salience. Second, our estimates are qualified by the nonrandom nonresponse to the last DNBC survey waves. In general, these respondents tended to be more affluent. A third concern is related to the “fuzzy” nature of our RD design.

If our treatment effects of interest are not homogeneous, the LATE theorem implies that our treatment estimates are defined for the subpopulation of “compliers” with their ITT (Imbens & Angrist, 1994). We speak to these concerns in two ways. One is to estimate our treatment effects separately for subsamples of the data defined by pretreatment characteristics (e.g., boys vs. girls). Second, using a straightforward technique recently introduced by Bertanha and Imbens (2014), we examine whether our complier population is distinctive.

5 | RESULTS

5.1 | Graphical evidence

We begin with an unrestricted, visual representation of our reduced-form results. First, Figure 3 shows, for each distinct SDQ measure observed at age 7, the conditional means by day of birth on each side of the January 1 threshold. The first panel of this figure shows a distinct drop in the total difficulties score (i.e., of roughly 0.1 *SD*) for children whose birthday is January 1 and later. The next four panels (i.e., b through e) suggest that this drop occurred for each of the four measures that constitute the difficulties score. However, the decrease in difficulties is uniquely large for the inattention/hyperactivity measure (i.e., the measure indicating a lack of self-regulation). Figure 3f suggests that there is a noticeable increase in the pro-social measure for children born January 1 or later.

These age-7 results provide clear evidence that quasi-random assignment to a delayed school start appears to improve mental health, particularly self-regulation, reported at age 7. However, one concern with these short-run findings is that they may be an artifact of the age at which parents report these data. In particular, the children for whom the ITT is one (i.e., those born January 1 or later) are more likely to be in kindergarten relative to the ITT=0 children who are more likely to be in first grade. So it is possible that these effects, although valid, reflect the current differences in the student's exposure to formal schooling rather than deeper developmental effects. The fact that the effects are concentrated in self-regulation rather than other constructs (as well as the evidence of positive effect on sociability) argues somewhat against this interpretation.

However, a more compelling way to address this concern is to consider outcomes at a later age when the children have long spells of formal schooling. In Figure 4, we show such evidence by illustrating the mean values of the SDQ measures by date of birth for children observed in the most recent age-11 wave of the DNBC. As with the age-7 data, these graphs

¹³A study by Leuven, Lindahl, Oosterbeek, and Webbink (2010) utilizes the unique rolling-admissions policies in the Netherlands and their interaction with school holidays and finds that earlier enrollment opportunities improve the test performance of disadvantaged students but have no or possibly negative effects of more advantaged students.

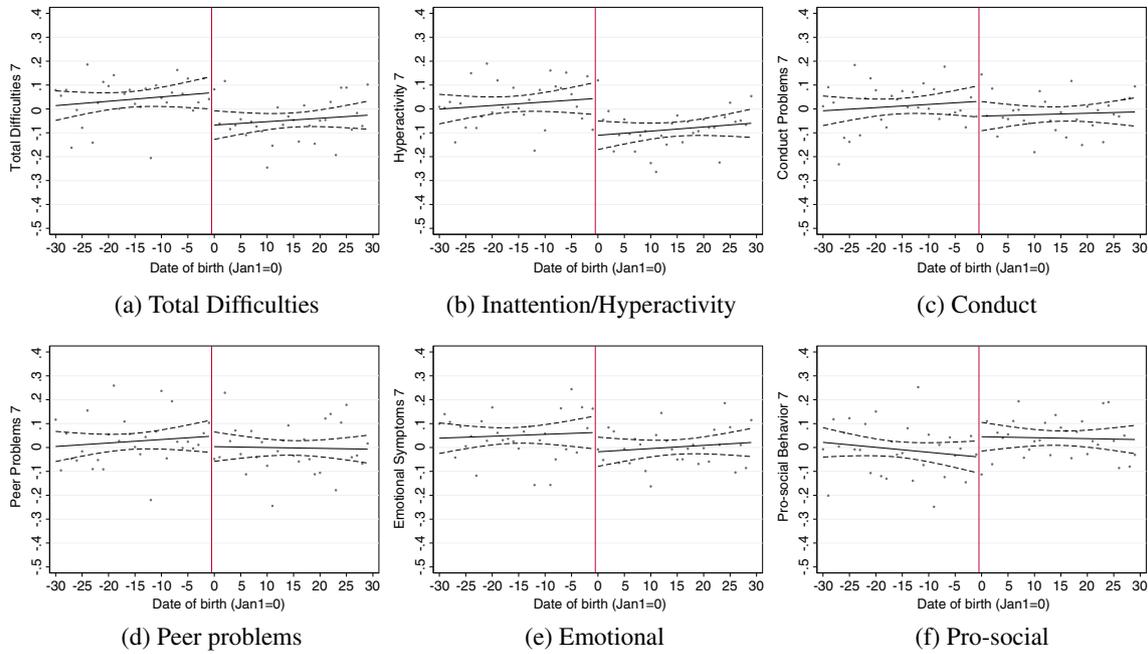


FIGURE 3 Reduced-form relationship, age 7. Bin width: 1 day [Colour figure can be viewed at wileyonlinelibrary.com]

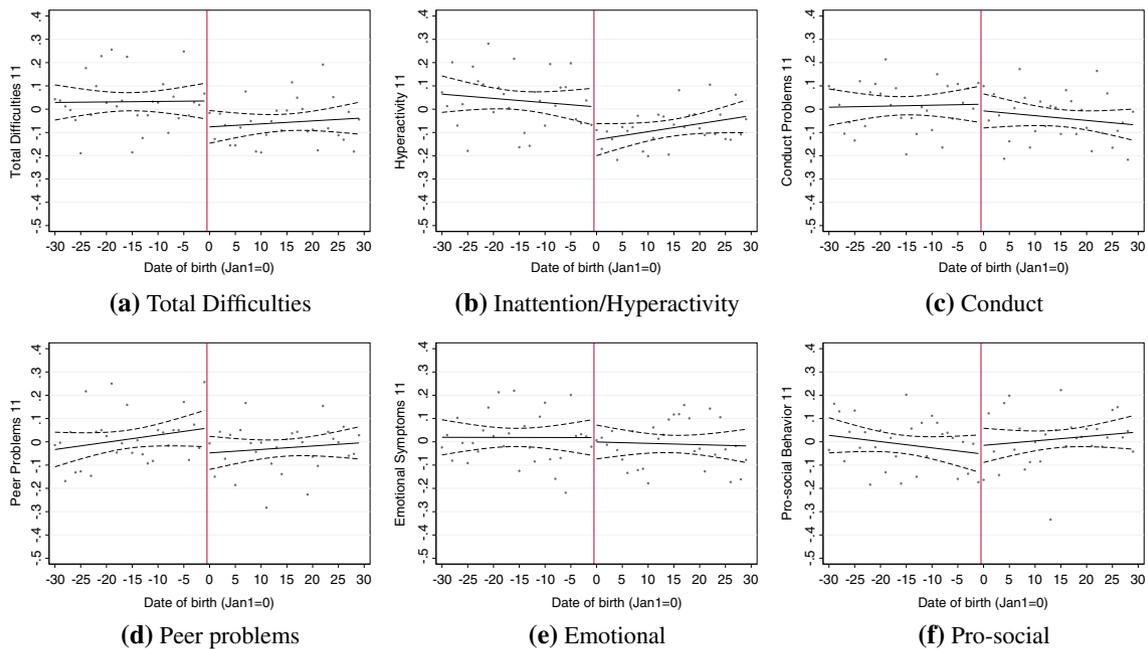


FIGURE 4 Reduced-form relationship, age 11. Bin width: 1 day [Colour figure can be viewed at wileyonlinelibrary.com]

suggest that those born on or after the cutoff (i.e., those with an ITT to delay their school start) have substantially lower levels of difficulties and a higher level of sociability. Again, we see (i.e., Figure 4b) that this effect is uniquely large with respect to the inattention/hyperactivity construct.

5.2 | Main estimates

Our graphical results provide highly suggestive evidence that a higher school starting age leads to an improvement in children's mental health, particularly with respect to inattention/hyperactivity. In this section, we present our key RD estimates. This regression framework allows to identify the point estimates of interest and, critically, test their statistical significance. However, this framework also allows us to explore the robustness of our visual evidence.

TABLE 4 Reduced-form regression-discontinuity estimates, the effect of $day_i > 0$ on Strengths and Difficulties Questionnaire

	Age 7		Age 11	
	(1)	(2)	(3)	(4)
Total difficulties	-0.14*** (0.05)	-0.12*** (0.05)	-0.11** (0.05)	-0.09* (0.05)
Emotional symptoms	-0.08* (0.05)	-0.07 (0.05)	-0.02 (0.06)	-0.00 (0.05)
Conduct problems	-0.06 (0.05)	-0.05 (0.05)	-0.03 (0.06)	-0.01 (0.05)
Inattention/hyperactivity	-0.15*** (0.05)	-0.15*** (0.05)	-0.14*** (0.05)	-0.13** (0.05)
Peer problems	-0.04 (0.05)	-0.04 (0.05)	-0.11* (0.06)	-0.10* (0.05)
Pro-social behavior	0.09* (0.05)	0.10** (0.05)	0.04 (0.06)	0.05 (0.06)
Observations	7,642	7,642	5,226	5,226
Controls	No	Yes	No	Yes

Note. Robust standard errors in parenthesis. Each cell shows the estimate from a single regression based on the local sample of children born 30 days before and after the cutoff. Covariates included are birth weight, origin, gender, parental education, parents' age, parental income, age at test, and birth year fixed effects. Missing values in covariates are replaced with zeros and indicators for missing variables are included.

* $p < .1$; ** $p < .05$; *** $p < .01$.

In Column 1 of Table 4, we present the reduced-form RD estimates for the SDQ measures at age 7. The results suggest that the ITT generates statistically significant reductions in total difficulties and marginally statistically significant increase in the pro-social construct at age 7. We find that the only consistently statistically significant reduction implied by the ITT is in the inattention/hyperactivity construct. Adding the full set of controls in Column 1 has almost no impact on the point-estimates. In Columns 3 and 4, we present similarly constructed reduced-form RD estimates for age-11 SDQ measures. The coefficients are very much in line with the age 7 outcomes. Overall, these point estimates indicate that a delayed school starting age causes a significant improvement in self-regulation that is sustained for at least several years and also qualitatively large. It should be noted that these ITT estimates identify the change in self-regulation implied by the change in school starting age from our first-stage equations (i.e., roughly 0.2 years).¹⁴

Our implied estimate of the effect of a full year increase in school starting age is five times as large as these reduced-form effects. For example, using the results conditional on controls, we find that increasing the school starting age by 1 year reduces inattention/hyperactivity at age 7 by 0.73 *SD* (i.e., -0.147/0.201). The corresponding two-stage least squares (2SLS) estimate for age 11 is -0.69 *SD* (i.e., -0.131/0.190). Arguably, these effect sizes are quite large, particularly for at-scale field settings.

Another potentially useful way to benchmark effects this large is to benchmark them against the mental-health gaps observed in the data. For example, children from families in the lowest decile of income have inattention/hyperactivity scores that are 0.61 *SD* higher at age 7 and 0.5 *SD* higher at age 11 relative to children in the top decile. Boys have inattention/hyperactivity SDQ levels that are about 0.7 *SD* higher than girls. Our finding indicates that a 1-year increase in school starting age produces an effect that is as large or larger than these mental-health gaps by income and gender.

5.3 | Treatment heterogeneity

Our main RD results provide robust evidence that a higher school starting age leads to a large and persistent increase in one particular dimension of children's mental health (i.e., self-regulation). However, there are several ways in which the generalizability of this evidence may be limited. For example, both local nature of an RD estimand and the nonrandom

¹⁴We note that we have not formally applied multiple-comparison adjustments to our inferences. However, our main results are estimated with sufficient precision that they would remain statistically significant after correcting for examining 12 core outcomes (i.e., six SDQ measures across two age groups).

participation of DNBC respondents to the last two survey waves raise external-validity concerns. Additionally, because we have a “fuzzy” RD design, the LATE theorem (Imbens & Angrist, 1994) implies that, in the absence of constant treatment effects, our point estimates are defined for the subpopulation of “compliers” (i.e., those who choose a treatment condition consistent with their ITT). To examine the empirical relevance of this concern, we follow the suggestion recently introduced by Bertanha and Imbens (2014). They recommend examining the continuity of outcomes, separately for children who took up the “treatment” and those who do not.

To apply this guidance in our setting, we defined the treatment as a binary indicator for older school starting age, SSO: first entering kindergarten at age 6.5 years or older. In Figure 5a, we show for the age-7 sample that this treatment “jumps” significantly at the threshold. Figure 5b illustrates the drop in the inattention/hyperactivity measure at this threshold. Figure 5c illustrates how the self-regulation measure changes at the threshold using only observations for which SSO = 0. Using these data, the threshold effectively separates “compliers” and “never takers” on the left from “never takers” on the right. The discrete jump in Figure 5c implies that the complier population has higher levels of inattention/hyperactivity than the never-takers (i.e., in the absence of treatment). Figure 5d presents a similarly constructed graph but using data only from those who took up the treatment (i.e., SSO = 1). This graph separates “always-takers” on the left from a population of always-takers and compliers on the right. The significant drop in the inattention/hyperactivity measure to the right of the threshold indicates that, even when all are taking the treatment, compliers have lower levels of inattention/hyperactivity than always-takers.

In Figure 6, we see effectively similar results when using the age-11 data. What do these results imply? We believe that they are consistent with the assertion that the complier subpopulation is a distinct one that may have treatment effects that differ from those for other parts of the population. For example, it is unsurprising that those who never choose to take up a delayed school start have low levels of inattention/hyperactivity (i.e., high degree of self-regulation) relative to the population that would comply when encouraged (Figure 6c). The never-takers may rightfully see little benefit in delaying a school start. Similarly, Figure 6d indicates that always-takers have uniquely high levels of inattention-hyperactivity and/or may have smaller treatment effects than compliers. This is consistent with the hypothesis that those who always seek a higher school starting age have unique developmental challenges that may be comparatively immune to the effects of a late start (i.e., relative to compliers).

To explore these issues in a more conventional and direct manner, we also examined how our key findings varied for subpopulations of the DNBC samples defined by baseline traits. Specifically, we estimated the effect of school starting age

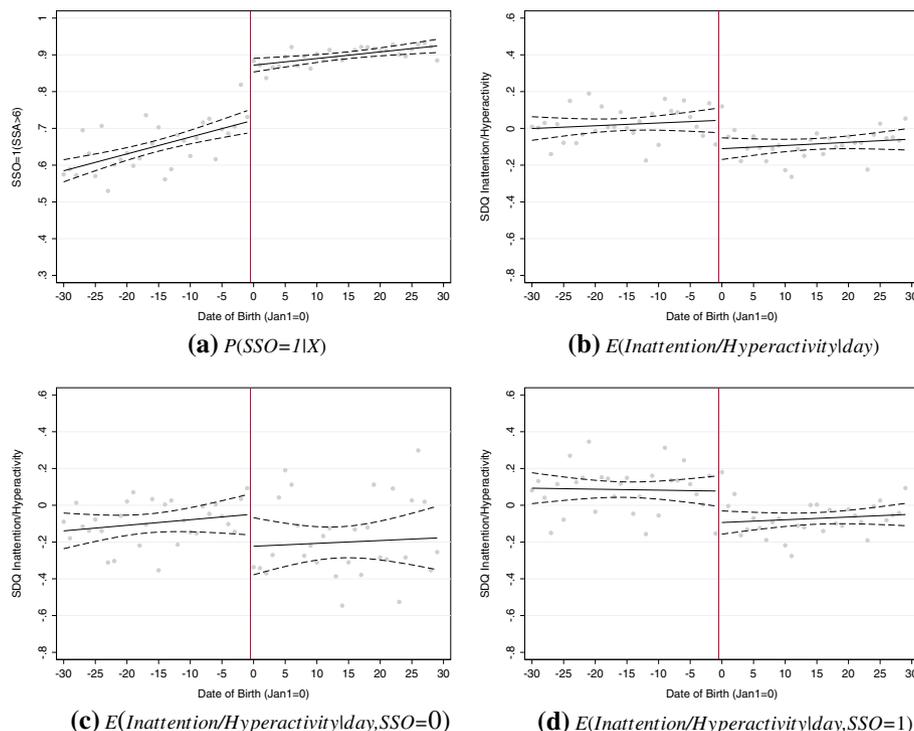


FIGURE 5 Inattention/hyperactivity at age 7, by treatment status [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

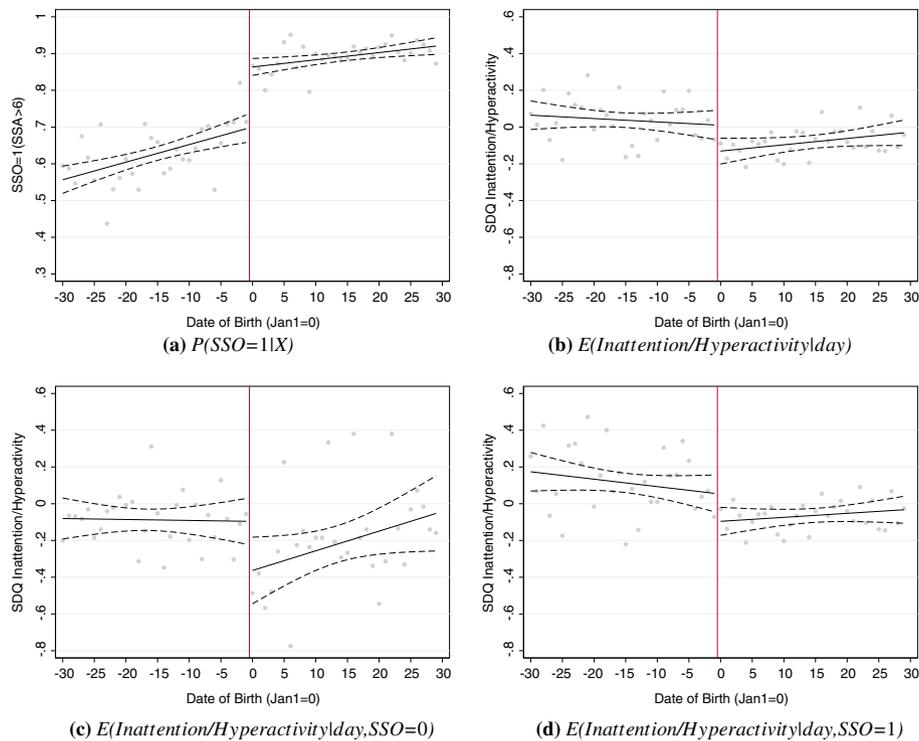


FIGURE 6 Inattention/hyperactivity at age 11, by treatment status

TABLE 5 Two-stage least squares estimates, the effect of school starting age on Strengths and Difficulties Questionnaire at age 7

	(1)	(2)	(3)	(4)	(5)	(6)	
	First stage	Emotional	Conduct	Inattention/ hyperact.	Peer prob.	Pro-social	
Main	0.20*** (0.03)	[45.48]	-0.34 (0.24)	-0.26 (0.23)	-0.73*** (0.25)	-0.20 (0.24)	0.48** (0.24)
Boys	0.12*** (0.04)	[9.35]	0.08 (0.56)	-0.47 (0.60)	-1.09 (0.70)	0.05 (0.60)	0.40 (0.60)
Girls	0.27*** (0.04)	[39.35]	-0.53** (0.26)	-0.21 (0.22)	-0.59*** (0.23)	-0.35 (0.22)	0.54** (0.23)
Highly educated	0.15*** (0.04)	[10.91]	-0.04 (0.42)	-0.59 (0.44)	-1.10** (0.53)	-0.46 (0.43)	0.59 (0.46)
Low educated	0.24*** (0.04)	[38.40]	-0.51* (0.30)	-0.12 (0.28)	-0.53* (0.28)	-0.05 (0.29)	0.41 (0.28)
High income	0.17*** (0.04)	[14.64]	-0.39 (0.37)	-0.40 (0.37)	-0.47 (0.37)	-0.54 (0.39)	0.46 (0.39)
Low income	0.22*** (0.04)	[32.21]	-0.34 (0.32)	-0.21 (0.31)	-0.96*** (0.35)	0.01 (0.32)	0.54* (0.31)
Low birthweight	0.17*** (0.04)	[17.03]	-1.00** (0.44)	-0.20 (0.36)	-0.84** (0.41)	-0.46 (0.38)	0.50 (0.37)
High birthweight	0.23*** (0.04)	[31.10]	0.26 (0.30)	-0.34 (0.30)	-0.62** (0.31)	0.05 (0.30)	0.49 (0.31)

Note. Robust standard errors in parenthesis. Regressions are based on the specification with the full set of covariates as well as linear time trends (separate for each side of the cutoff) based on the local sample of children born 30 days before and after the cutoff. Each cell shows the estimate from a single regression. Covariates included are birth weight, origin, gender, parental education, parents' age, parental income, age at test, and birth year fixed effects. Missing values in covariates are replaced with zeros, and indicators for missing variables are included. All sample splits are done at the median. Nonsingletons are always defined as having an older sibling. First-stage F -stats for the excluded instrument are shown in square-brackets.

* $p < .1$; ** $p < .05$; *** $p < .01$.

TABLE 6 Two-stage least squares estimates, the effect of school starting age on Strengths and Difficulties Questionnaire at age 11

	(1)	(2)	(3)	(4)	(5)	(6)	
	First stage	Emotional	Conduct	Inatt./ hyperact.	Peer Prob.	Pro-social	
Main	0.19*** (0.04)	[27.16]	-0.01 (0.29)	-0.06 (0.29)	-0.69** (0.31)	-0.50 (0.31)	0.24 (0.30)
Boys	0.09* (0.05)	[2.89]	0.87 (1.02)	0.35 (0.97)	-1.55 (1.37)	-1.23 (1.24)	0.10 (1.02)
Girls	0.28*** (0.05)	[30.12]	-0.23 (0.28)	-0.16 (0.26)	-0.43* (0.25)	-0.26 (0.25)	0.25 (0.25)
Highly educated	0.12** (0.06)	[4.35]	-0.18 (0.63)	-0.45 (0.64)	-0.97 (0.79)	-0.87 (0.76)	0.40 (0.71)
Low educated	0.27*** (0.05)	[33.97]	0.10 (0.30)	0.07 (0.32)	-0.55* (0.31)	-0.32 (0.32)	0.16 (0.31)
High income	0.15*** (0.06)	[7.17]	0.05 (0.50)	-0.55 (0.53)	-0.24 (0.49)	-0.87 (0.62)	0.74 (0.63)
Low income	0.22*** (0.05)	[22.06]	-0.03 (0.36)	0.14 (0.38)	-1.08** (0.43)	-0.31 (0.37)	-0.04 (0.35)
Low birthweight	0.18*** (0.05)	[11.84]	0.16 (0.44)	0.19 (0.43)	-0.34 (0.43)	-0.23 (0.43)	0.72 (0.49)
High birthweight	0.20*** (0.05)	[15.76]	-0.13 (0.37)	-0.28 (0.40)	-1.03** (0.46)	-0.75* (0.44)	-0.22 (0.41)

Note. Robust standard errors in parenthesis. Regressions are based on the specification with the full set of covariates as well as linear time trends (separate for each site of the cutoff) based on the local sample of children born 30 days before and after the cutoff. Each cell shows the estimate from a single regression. Covariates included are birth weight, origin, gender, parental education, parents' age, parental income, age at test, and birth year fixed effects. Missing values in covariates are replaced with zeros, and indicators for missing variables are included. All sample splits are done at the median. Nonsingletons are always defined as having an older sibling. First-stage *F*-stats for the excluded instrument are shown in square-brackets.

* $p < .1$; ** $p < .05$; *** $p < .01$.

on each SDQ measure using our RD design, first, for boys and girls separately and then for respondents who were above the sample median values for education, income, and birth weight. We report these 2SLS results in Tables 5 and 6 for the age-7 and age-11 samples, respectively.

Interestingly, these estimates indicate that a school starting age had statistically insignificant effects for boys across all measures and both ages. However, these findings reflect a considerable loss in precision for boys. In fact, we find that the first-stage effect for boys is smaller (0.12 compared to 0.20 for girls at age 7). As boys tend to be always-takers, the first stage is very weak for them. So our identifying variation is uniquely relevant for girls. And estimates based only on girls indicates that a high school starting age improves both self-regulation and emotional problems. Our remaining results do not show a clear pattern for specific subgroups having greater mental-health benefits of a higher school starting age. Although the effects are larger for low income children, the point-estimates are also larger for children of highly educated parents and children with a birth weight above 3.500 g. However, the estimates across these subgroups are quite imprecise, and we cannot reject that the coefficients are the same.

5.4 | The importance of reference group

One potential concern about using mother reported measures for the child's strength and difficulties is that these reports may be biased by reference group. Although Elder (2010) finds evidence that parents, in contrast to teachers, are not subject to these biases, we nevertheless assess this issue empirically using two distinct approaches. First, we split the sample by whether the child has an older sibling. For children with sisters or brothers, these siblings might constitute a natural reference point for the mother. Assuming that parents compare siblings at the same point in time, not at the same age, we would assume that effects would be stronger when there is no older sibling present, if the estimates suffer from rater biases. However, as the results in Table 7 show, we find evidence of the opposite. The effects are strongest if there is an older sibling present. Second, we control for the classmates' average school starting age in the regression. In this specification, we instrument the actual average school starting age with average assigned school starting age (i.e.,

TABLE 7 Two-stage least squares estimates, the effect of school starting age on Strengths and Difficulties Questionnaire

	(1)	(2)	(3)	(4)	(5)	(6)
	First stage	Emotional	Conduct	Inatt./ hyperact.	Peer Prob.	Pro-social
A. Age 7						
No older siblings	0.20*** [23.79] (0.04)	0.06 (0.34)	-0.02 (0.31)	-0.57* (0.34)	-0.05 (0.33)	0.26 (0.32)
Older siblings	0.20*** [23.12] (0.04)	-0.68* (0.36)	-0.49 (0.35)	-0.85** (0.37)	-0.34 (0.35)	0.69* (0.36)
Classmates SSA	0.36*** [142.42] (0.03)	-0.44 (0.40)	-0.51 (0.41)	-0.95* (0.53)	-0.13 (0.34)	0.48 (0.39)
B. Age 11						
No older siblings	0.20*** [16.25] (0.05)	0.32 (0.40)	0.54 (0.41)	-0.32 (0.40)	-0.28 (0.41)	-0.16 (0.40)
Older sibling	0.18*** [11.84] (0.05)	-0.33 (0.44)	-0.72 (0.48)	-1.09** (0.53)	-0.70 (0.48)	0.67 (0.49)
Classmates SSA	0.34*** [80.44] (0.04)	0.25 (0.64)	0.37 (0.66)	-0.82 (0.85)	-0.35 (0.63)	0.32 (0.68)

Note. Robust standard errors in parenthesis. Regressions are based on the specification with the full set of covariates as well as linear time trends (separate for each site of the cutoff) based on the local sample of children born 30 days before and after the cutoff. Each cell shows the estimate from a single regression. Covariates included are birth weight, origin, gender, parental education, parents' age, parental income, age at test, and birth year fixed effects. Missing values in covariates are replaced with zeros, and indicators for missing variables are included. Classmates SSA also conditions on classmates school starting age, which is instrumented by the average assigned school starting age. The *F*-statistics on the excluded instrument for peers' school starting age (assigned SSA) are 42 at age 7 and 23 at age 11. Nonsingletons are always defined as having an older sibling. First-stage *F*-stats for the excluded instrument for own school starting age are shown in square-brackets.

* $p < .1$; ** $p < .05$; *** $p < .01$.

the average school starting age if all peers complied to the cutoff). Table 7 shows that the coefficients on school starting age are slightly larger in magnitude but also less precise when we condition on classmates average school starting age, both at age 7 and 11.

In sum, the results in Table 7 show no sign of significant differences by reference group, which suggests that the results are not driven by reference group.

5.5 | Binary outcomes

Another relevant type of treatment heterogeneity concerns how the effects of a delayed school start may influence more severe levels of mental illness. Our prior estimates effectively identify the changes in mean SDQ measures, which are in diagnostically normal ranges. However, as noted earlier, each SDQ score can be classified as one of three levels: normal, borderline, and abnormal. To explore this form of heterogeneity, we estimated 2SLS models using our RD design and binary indicators for an abnormal rating (or for a borderline/abnormal rating) as the dependent variables. We report these RD estimates for the age-7 and age-11 samples in Table 8. We also report the mean value of these dependent variables. Diagnostically abnormal ratings on these scales are not common. For example, across both age 7 and age 11, only 5% to 8% of respondents had inattention/hyperactivity ratings that qualified as abnormal or borderline. Interestingly, the similar mean effects for age 7 and age 11 outcomes seem to be driven in different parts of the distribution. As these binary results show, at age 7, older school starting age reduces the likelihood of abnormal inattention/borderline values, but at age 11, the school starting age does not affect the likelihood of having extreme values on the inattention/hyperactivity scale.

The point estimates of -0.16 and -0.22 for, respectively, abnormal and borderline hyperactivity values at age 7 are large compared to the low prevalence of these outcomes (e.g., sample means of 0.05 and 0.08, respectively). However, it is worth noting that the relatively wide confidence bands around the point estimates include estimates as low and 4 and 8 percentage points, respectively. Furthermore, although these effect sizes are still substantial compared to the sample mean, it is worth remembering that this sample mean is based on the full sample, including never takers, who appear to have considerably lower levels of difficulties.

TABLE 8 Two-stage least squares estimates, the effect of school starting age on abnormal/borderline Strengths and Difficulties Questionnaire values

	Age 7		Age 11	
	(1) Abnormal	(2) Borderline	(3) Abnormal	(4) Borderline
Total difficulties	-0.12** (0.04) [0.03]	-0.08 (0.06) [0.06]	-0.05 (0.05) [0.04]	-0.08 (0.07) [0.07]
Emotional symptoms	-0.08 (0.07) [0.08]	-0.15* (0.09) [0.15]	-0.03 (0.09) [0.10]	0.05 (0.11) [0.16]
Conduct problems	0.05 (0.05) [0.05]	-0.04 (0.08) [0.13]	0.09 (0.06) [0.03]	0.14 (0.09) [0.09]
Inattention/hyperactivity	-0.16*** (0.06) [0.05]	-0.22*** (0.07) [0.08]	-0.09 (0.06) [0.05]	-0.11 (0.08) [0.08]
Peer problems	-0.03 (0.05) [0.04]	0.02 (0.07) [0.09]	-0.09 (0.08) [0.06]	-0.09 (0.10) [0.12]
Pro-social scale	-0.05 (0.03) [0.02]	-0.11** (0.06) [0.05]	-0.03 (0.04) [0.02]	-0.05 (0.07) [0.05]

Note. Means of the dependent variables in square-brackets. Robust standard errors in parenthesis. Regressions are based on the specification with the full set of covariates as well as linear time trends (separate for each site of the cutoff) based on the local sample of children born 30 days before and after the cutoff. Each cell shows the estimate from a single regression. Covariates included are birth weight, origin, gender, parental education, parents' age, parental income, age at test, and birth year fixed effects. Missing values in covariates are replaced with zeros, and indicators for missing variables are included.

* $p < .1$; ** $p < .05$; *** $p < .01$.

6 | DISCUSSION

Our findings of large and persistent effects of school starting age on children's inattention/hyperactivity give rise to two questions related to the existing evidence. One, why do we find effects of school starting age on mental health, whereas Dalsgaard et al. (2012) find no effect of school starting age on the likelihood of receiving an ADHD diagnosis in Denmark? Two, if the effects are persistent, why do they not show up in later life outcomes? In this section, we first relate our outcome measure, the SDQ, to ADHD diagnoses and then discuss the implications of our findings for long-run consequences of school starting age.

Although evidence for a causal relationship between age at school-entry and ADHD diagnoses has been documented for Canada (Morrow et al., 2012), Taiwan (Chen et al., 2016), and the United States (Elder & Lubotsky, 2009), Dalsgaard et al. (2012) find no evidence for such a relationship in Denmark. To relate our findings to this evidence, we link the DNBC data to hospital records on ADHD diagnoses. In Panel A of Table 9, we show the fraction of children with an ADHD diagnoses for all children born in the period 1997 to 2003, for all children in the DNBC sample and for all children born within 30 days of the cutoff in the DNBC sample. In the general population, the diagnosis rate is 0.5%.¹⁵ To shed light on how the ADHD diagnoses relate to the SDQ measures, we show the share of children with a diagnosis by SDQ score in Panel B of Table 9. Among children with normal SDQ values, 0.2% have a diagnosis. The diagnosis rate is more than 10 times higher among children with a borderline or abnormal SDQ score. However, even among children with an abnormal SDQ total difficulties score at age 7, only 6% have a diagnosis. These rates suggest that our outcome measure, the SDQ, captures a much less extreme outcome than ADHD diagnoses. In light of this conclusion, it seems natural to ask

¹⁵Dalsgaard et al. (2012) report a diagnosis rate of 1.3%. One reason for the discrepancy is that they use a different data source (the Danish Psychiatric Central Registry). A second potential explanation is that we use different birth cohorts.

TABLE 9 The SDQ and ADHD diagnoses

	Population	DNBC	DNBC Jan+Dec
A. ADHD diagnosis rates across samples			
	0.005	0.003	0.003
B. ADHD diagnosis rate across SDQ groups			
	Normal	Borderline	Abnormal
SDQ total difficulties age 7	0.002	0.036	0.059
SDQ total difficulties age 11	0.002	0.024	0.033
SDQ inattention/hyperactivity age 7	0.001	0.030	0.042
SDQ inattention/hyperactivity age 11	0.002	0.024	0.032
C. Test scores (standardized) across SDQ groups			
Danish/reading, Grade 2	0.032	-0.505	-0.544
Mathematics, Grade 3	0.029	-0.449	-0.503

Note. ADHD diagnoses are based the following ICD-10 codes: F900, F901, F908, F909, and F989. The test scores are divided into SDQ groups according to the inattention/hyperactivity score at age 7. The test scores are standardized to a mean of zero and a unit standard deviation. ADHD = attention deficit hyperactivity disorder; SDQ = Strengths and Difficulties Questionnaire; DNBC = Danish National Birth Cohort.

whether the variation in SDQ reflects important variation in mental health. To investigate this issue, we show the average test scores in reading and mathematics across the three ADHD groups in Panel C of Table 9. Children with a borderline or abnormal SDQ level have test scores that are more than 0.5 standard deviations below the mean. In sum, although we identify effects on a considerably less severe outcome than ADHD diagnoses, the variation in our outcomes is strongly correlated with outcomes on other domains, suggesting that the SDQ captures important behavioral differences.

As both empirical evidence and economic theory on skill formation suggest that early development of skills are important for later life outcomes Cunha, Heckman, Lochner, and Masterov, (e.g., Cunha et al., 2006), we would expect that our findings imply long-run effects for market and nonmarket long-run outcomes. However, Black et al. (2011) find small or even negative effects of being old at school enrollment on IQ at age 18 and very small effects on mental health. Fredriksson and Öckert (2013) assess the effect of school starting age on life time income and only find positive effects for the subgroup of children of low-educated parents. Lastly, Landersø, Nielsen, and Simonsen (2017) find that school starting age affects crime, but mainly by affecting the timing of the onset of the “criminal career.” Landersø et al. (2017) also find some evidence on college enrollment for girls, but limited evidence for an effect on completed years of schooling at age 27. Why do the large effects on children's mental health not translate into large long-run outcomes? One explanation is that skills are multidimensional, and although the effects captured by the SDQ may be important for well-being, they may not be the skill-dimension that is important for success on the labor market. One avenue to explore is how SDQ measures at young ages, especially related to school starting age, relate to well-being at older ages.

7 | CONCLUSIONS

The decision to delay the age at which children in developed nations begin formal schooling is increasingly common. These delays may confer developmental advantages through both relative and absolute-age mechanisms. However, an active research literature has generally found that these delays do not clearly result in longer-run educational or economic advantages. In this study, we examined the effect of school starting age on distinctive and more proximate outcomes: measures of mental health in childhood. One key feature of our study is the availability of data on several psychopathological constructs from a widely used and extensively validated mental-health screening tool fielded among children in the DNBC study. We are able to identify the causal effect of higher school starting ages by leveraging the Danish rule that children should begin kindergarten in the calendar year in which they turn 6 years. We match the children in the DNBC to the Danish administrative registries that include the exact day of birth and confirm that school starting age increases significantly for children born after the cutoff.

The results based on this “fuzzy” RD design indicate that delays in school starting age imply substantial improvements in mental health (e.g., reducing the overall “difficulties” score by at least 0.5 *SD*). The evidence for these effects is robust and, critically, persists in the latest wave of the DNBC when the children were aged 11. However, we also find that these

mental-health gains are narrowly confined to one particular construct: the inattention/hyperactivity score (i.e., a measure indicating a lack of self-regulation). Interestingly, this finding is consistent with one prominent theory of why delayed school starts are beneficial. Specifically, a literature in developmental psychology emphasizes the importance of pretend play in the development of children's emotional and intellectual self-regulation. Children who delay their school starting age may have an extended (and appropriately timed) exposure to such playful environments. Our findings are consistent with this absolute-age mechanism and suggest that there may be broader developmental gains to policies that delay the initiation of formal schooling (and that support playful early-childhood environments).

ACKNOWLEDGEMENTS

We thank Silke Anger, Paul Bingley, Björn Öckert, Kjell Salvanes, and seminar participants at SFI, Stanford University, the fourth SOLE/EALE World Conference, the Copenhagen Education Network, and the CESifo Economics of Education Conference for helpful comments and suggestions. This paper uses data from the Danish National Birth Cohort (DNBC). The Danish National Research Foundation has established the Danish Epidemiology Science Centre that initiated and created the DNBC. The DNBC is a result of a major grant from this foundation, as well as grants from the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Augustinus Foundation, and the Health Foundation. Sievertsen acknowledges support from The Danish Council for Strategic Research, Grant DSF-09-070295. Dee and Sievertsen have no financial or personal conflicts of interest related to this study. No ethics approval for this project was required; we have not collected data (other than survey) from human subjects

ORCID

Hans Henrik Sievertsen  <http://orcid.org/0000-0003-0126-828X>

REFERENCES

- Achenbach, T. M., Becker, A., Döpfner, M., Heiervang, E., Roessner, V., Steinhausen, H.-C., & Rothenberger, A. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: Research findings, applications, and future directions. *Journal of Child Psychology and Psychiatry*, *49*(3), 251–275.
- American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders (DCM). (*Technical report*). Washington, DC: American Psychiatric Association.
- Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton university press.
- Bassok, D., & Reardon, S. F. (2013). "Academic redshirting" in kindergarten prevalence, patterns, and implications. *Educational Evaluation and Policy Analysis*, *35*(3), 283–297.
- Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics*, *121*(4), 1437–1472.
- Bertanha, M., & Imbens, G. W. (2014). External validity in fuzzy regression discontinuity designs. (*Technical report*). Cambridge, MA: National Bureau of Economic Research.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *The Review of Economics and Statistics*, *93*(2), 455–467.
- Blake, P. R., Piovesan, M., Montinari, N., Warneken, F., & Gino, F. (2015). Prosocial norms in the classroom: The role of self-regulation in following norms of giving. *Journal of Economic Behavior & Organization*, *115*, 18–29.
- Browning, M., & Heinesen, E. (2007). Class size, teacher hours and educational attainment*. *The Scandinavian Journal of Economics*, *109*(2), 415–438.
- Buckles, K. S., & Hungerman, D. M. (2013). Season of birth and later outcomes: Old questions, new answers. *The Review of Economics and Statistics*, *95*(3), 711–724.
- Cascio, E. U., & Schanzenbach, D. W. (2016). First in the class? Age and the education production function. *Education Finance and Policy*, *11*(3), 225–250.
- Chen, M.-H., Lan, W.-H., Bai, Y.-M., Huang, K.-L., Su, T.-P., Tsai, S.-J., ... Hsu, J. W. (2016). Influence of relative age on diagnosis and treatment of attention-deficit hyperactivity disorder in taiwanese children. *The Journal of Pediatrics*, *172*, 162–167.
- Crawford, C., Dearden, L., & Meghir, C. (2007). When you are born matters: The impact of date of birth on child cognitive outcomes in england. (<http://discovery.ucl.ac.uk>): University College London.
- Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, *1*, 697–812.
- Dalsgaard, S., Humlum, M. K., Nielsen, H. S., & Simonsen, M. (2012). Relative standards in ADHD diagnoses: The role of specialist behavior. *Economics Letters*, *117*(3), 663–665.

- Datta Gupta, N., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics*, 94(1), 30–43.
- Deming, D., & Dynarski, S. (2008). The lengthening of childhood. *The Journal of Economic Perspectives*, 22(3), 71–92.
- Duckworth, A. L., & Carlson, S. M. (2013). Self-regulation and school success. *Self-regulation and autonomy: Social and developmental dimensions of human conduct*, 40, 208.
- Elder, T. E. (2010). The importance of relative standards in ADHD diagnoses: Evidence based on exact birth dates. *Journal of Health Economics*, 29(5), 641–656.
- Elder, T. E., & Lubotsky, D. H. (2009). Kindergarten entrance age and children's achievement impacts of state policies, family background, and peers. *Journal of Human Resources*, 44(3), 641–683.
- Fredriksson, P., & Öckert, B. (2013). Life-cycle effects of age at school start. *The Economic Journal*, 124(579), 977–1004.
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586.
- Goodman, R., & Scott, S. (1999). Comparing the strengths and difficulties questionnaire and the child behavior checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, 27(1), 17–24.
- Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4), 451–464.
- Imbens, G. W., & Angrist, J. D. (1994). No title. *Econometrica*, 62(2), 467–475.
- Landersø, R., Nielsen, H. S., & Simonsen, M. (2017). School starting age and the crime-age profile. *The Economic Journal*, 127(602), 1096–1118. <https://doi.org/10.1111/eoj.12325>
- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *The Journal of Economic Literature*, 48(2), 281–355.
- Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2010). Expanding schooling opportunities for 4-year-olds. *Economics of Education Review*, 29(3), 319–328.
- Morrow, R. L., Garland, E. J., Wright, J. M., Maclure, M., Taylor, S., & Dormuth, C. R. (2012). Influence of relative age on diagnosis and treatment of attention-deficit/hyperactivity disorder in children. *Canadian Medical Association Journal*, 184(7), 755–762.
- Rambøll Management Consulting, Aarhus University, & University of Southern Denmark (2016). Børns tidlige udvikling og læring i dagtilbud. (Technical report): Udviklingsprogrammet Fremtidens Dagtilbud.
- Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review*, 13(3), 254–274.
- The Boston Globe (2014). Holding kids back for kindergarten doesn't help. Evan Horowitz.
- The New York Times (2010). The littlest redshirts sit out kindergarten. Pamela Paul.
- The New Yorker (2013). *Youngest kid, smartest kid?* The New Yorker: Maria Konnikova.

How to cite this article: Dee TS, Sievertsen HH. The gift of time? School starting age and mental health. *Health Econ*. 2017;1–22. <https://doi.org/10.1002/hec.3638>

APPENDIX A: APPENDICES

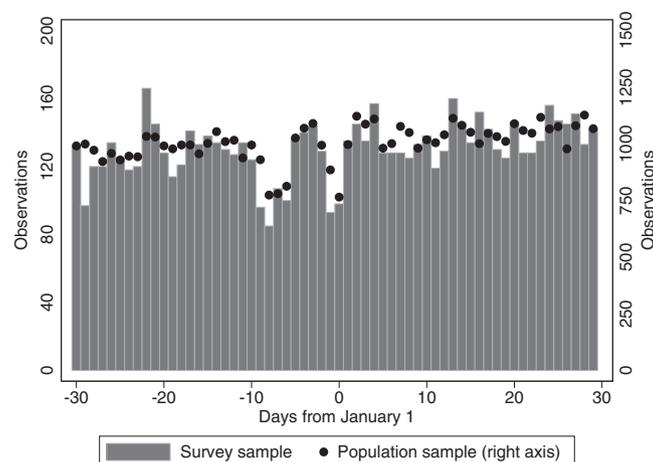


FIGURE A1 Observations by date of birth, survey data, and population data. The survey data are the data used in our analysis, and the population includes all children born in Denmark in the period 1997–2003

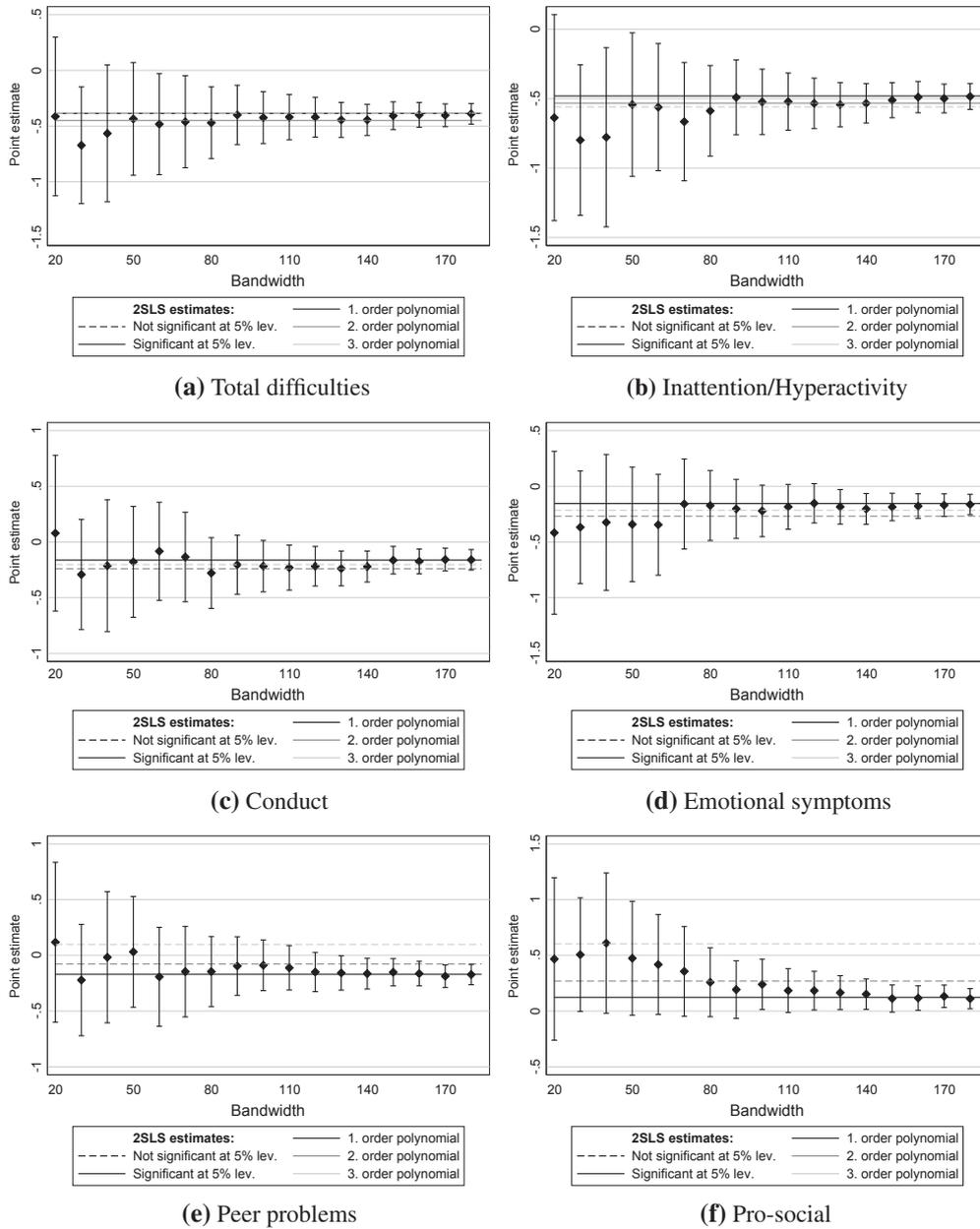


FIGURE A2 Bandwidth sensitivity, age 7. Each diamond marker is the 2SLS point estimate from a local regression with the bandwidth size denoted on the x-axis. The bandwidth size increases in steps of 10 days. A bandwidth of 10 implies a sample of children born 10 days before and after January 1st. The horizontal lines are the 2SLS point estimate from a regression using the full sample with separate trends on each side of the January 1st cutoff. The lines are solid if the estimate is significant on a five percent level, and dashed if it is not significant on a five percent level

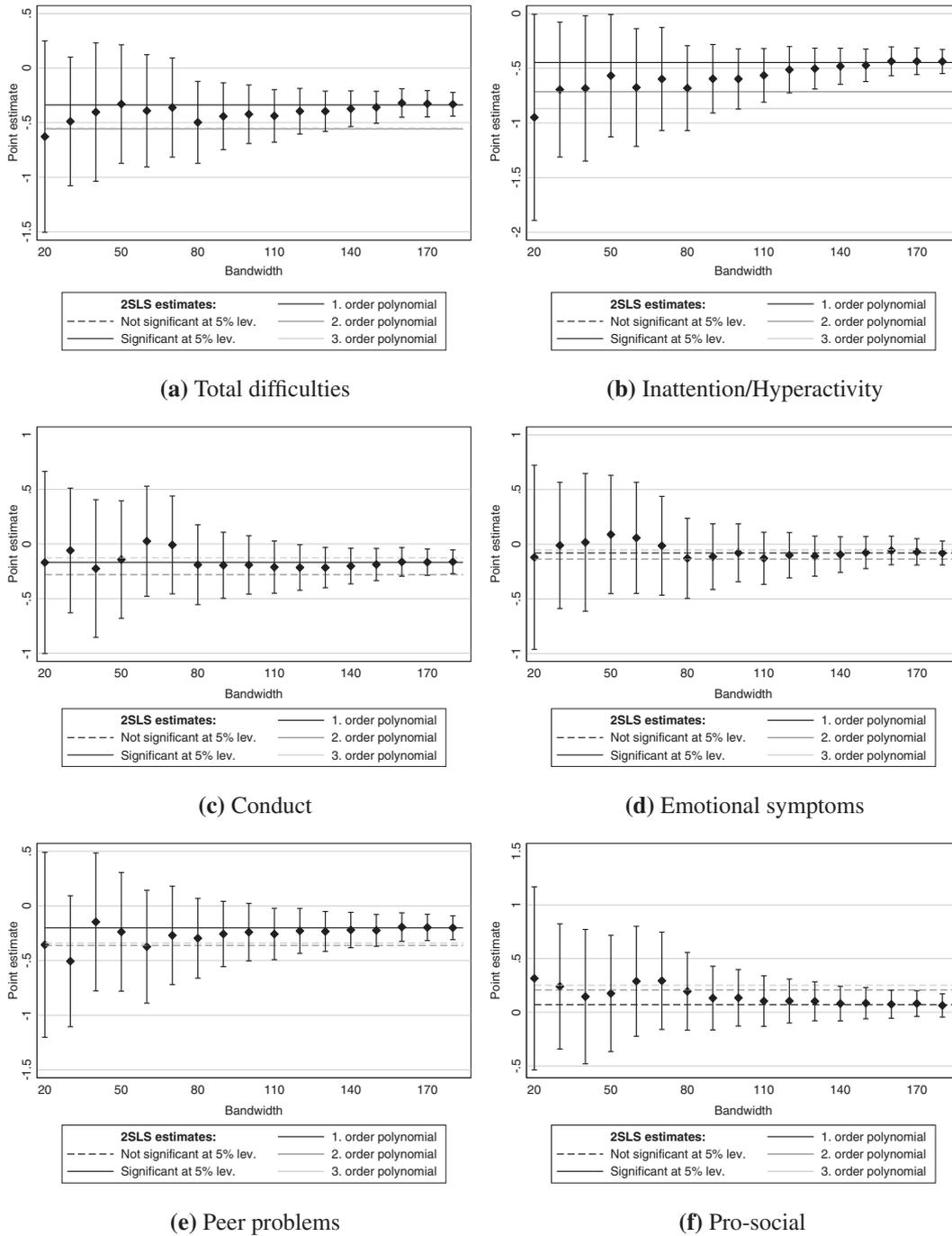


FIGURE A3 Bandwidth sensitivity, age 11. Each diamond marker is the two-stage least squares point estimate from a local regression with the bandwidth size denoted on the x-axis. The bandwidth size increases in steps of 10 days. A bandwidth of 10 implies a sample of children born 10 days before and after January 1. The horizontal lines are the two-stage least squares point estimate from a regression using the full sample with separate trends on each side of the January 1 cutoff. The lines are solid if the estimate is significant on a 5% level, and dashed if it is not significant on a 5% level level

TABLE A1 Auxiliary regression-discontinuity estimates, balancing of the covariates. Dependent variable: Born after cutoff

	(1)
Female	-0.003 (0.006)
Birthweight (gr.)	0.000 (0.000)
Non-Western origin	-0.003 (0.022)
Parents' years of schooling	0.002 (0.001)
Parents gross income	-0.000 (0.000)
Mother's age when child was born	0.000 (0.001)
Father's age when child was born	0.000 (0.001)
<i>F</i> -statistic for test of joint significance	0.778
<i>p</i> value for test of joint significance	0.606

Note. Robust standard errors in parenthesis. Regressions of the indicator for being born on January 1 or later on the covariates listed above as well as linear time trends (separate for each site of the cutoff) based on the local sample of children born 30 days before and after the cutoff.

* $p < .1$; ** $p < .05$; *** $p < .01$.

TABLE A2 Placebo regressions with pretreatment outcomes

	(1)	(2)
Can keep occupied for 15 min aged 18m	0.02 (0.09)	0.02 (0.09)
Turns pictures right aged 18m	0.26* (0.14)	0.25* (0.13)
Makes word sounds aged 18m	0.05 (0.04)	0.04 (0.04)
Can walk up stairs aged 18m	-0.00 (0.03)	-0.00 (0.03)
Can bring things aged 18m	-0.00 (0.04)	-0.00 (0.03)
Observations	5,946	5,945
Covariates	No	Yes

Note. Robust standard errors in parenthesis. Regressions are based on the specification with the full set of covariates as well as linear time trends (separate for each site of the cutoff) based on the local sample of children born 30 days before and after the cutoff. Covariates included are birth weight, origin, gender, parental education, parents' age, parental income, age at test, and birth year fixed effects.

* $p < .1$; ** $p < .05$; *** $p < .01$.