I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17091

Starting School and ADHD: When Is It Time to Fly the Nest?

Catia Nicodemo
Cheti Nicoletti
Joaquim Vidiella-Martin

JUNE 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17091

Starting School and ADHD: When Is It Time to Fly the Nest?

Catia Nicodemo
University of Oxford, IZA and University of Verona

Cheti Nicoletti
University of York, ISER University of Essex and IZA

Joaquim Vidiella-Martin
University of Oxford, Erasmus University Rotterdam and Tinbergen Institute

JUNE 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Starting School and ADHD: When Is It Time to Fly the Nest?*

Does deferring school entry for children born just before the enrollment cutoff date improve their mental well-being? We address this question using administrative data on prescriptions for attention deficit hyperactivity disorder (ADHD) in England. Higher ADHD rates among early school starters are often attributed to a peer-comparison bias caused by differences in relative age among classmates. However, previous studies do not consider other potential underlying mechanisms. By adopting a more comprehensive framework, we can confirm that relative age is the primary driver of the gap in ADHD rate in the long term. Furthermore, we find that such a long-term gap is driven by first-time prescriptions between ages 5 and 8, which is a critical period when the accuracy of ADHD diagnosis is most important. Based on these findings, our policy recommendations include sorting children by age and refining diagnostic decision-making in early primary school.

JEL Classification: I10, I20, J13

Keywords: children, mental health, school starting age, ADHD, England, NHS

Corresponding author:
Cheti Nicoletti
Department of Economics and Related Studies
University of York
Heslington, York YO10 5DD
United Kingdom
E-mail: cheti.nicoletti@york.ac.uk

* We are grateful to Sonia Bhalotra, Gabriella Conti, Thomas Cornelissen, Stuart Redding, and seminar participants at the Universitat Pompeu Fabra, the University of York, the University of Warwick and, the Heriot-Watt university for their valuable suggestions, and attendants of the American-European Health Economics Study Group, Universitat Pompeu Fabra, June 2022; the Workshop on Global Health, Environment, and Labour, Royal Holloway University of London, July 2022; the Workshop on Labour and Family Economics, University of York, September 2022; the Mental Health and Economic Status Workshop, University of Warwick, October 2022; the Economic Observatory and CEPEO Workshop 'What makes schools work for disadvantaged children?', University of Strathclyde, January 2023; and the International Health Economics Association, July 2023. The authors acknowledge funding from the Economic and Social Research Council (grant number ES/T008415/1). This study used non-public microdata provided by QResearch via a remote access facility and complies with the respective data agreement. QResearch pre-viewed the findings before publication to ensure that privacy-sensitive, individual-specific information is not revealed. We acknowledge the contribution of EMIS practices who contribute to QResearch and EMIS Health and the Universities of Nottingham and Oxford for expertise in establishing, developing or supporting the QResearch database. This project involves data derived from patient-level information collected by the NHS, as part of the care and support of cancer patients. The hospital, cancer and mortality data are collated, maintained and quality assured by the National Disease Registration Service which is part of NHS England. Access to the data was facilitated by the NHS England Data Access Request service. NHS England bears no responsibility for the analysis or interpretation of the data. We have no conflicts of interest to disclose; all errors are our own.

1 Introduction

Attention deficit and hyperactivity disorder (ADHD) is one of the most prevalent childhood disorder (Thomas et al., 2015), characterized by concentration problems, excessive activity, and impulsivity, which has been found to worsen educational attainment (Currie and Stabile, 2006; Fletcher and Wolfe, 2008; Ding et al., 2009; Currie et al., 2010) and income (Fletcher, 2014) and increase crime (Fletcher and Wolfe, 2009).

However, ADHD is difficult to diagnose, and there are increasing concerns about misdiagnoses. An issue that has been emphasized by previous literature is that children who start school at a relatively younger age than their school-grade peers are more likely to be diagnosed and to receive drugs for ADHD, and this difference has been attributed to their relative immaturity with respect to their older peers (see, e.g., Elder, 2010; Evans et al., 2010). While both the oldest and the youngest children in a grade are likely to receive drugs if they have extreme ADHD symptoms, only the youngest children are likely to be treated for milder symptoms (Persson et al., 2021). More marginal diagnoses for the relatively younger children imply an unfair and inefficient use of medical resources (e.g. Persson et al., 2021; Furzer et al., 2022).[1] Emerging new literature has highlighted that treatment of marginally diagnosed patients may produce modest or even negative effects on their health and can have long-term economic costs.[2]

Previous empirical papers have documented the gap in ADHD diagnoses and drug prescriptions for relatively younger children using quasi-experimental evidence, which exploits the discontinuity in the age at the start of school caused by the cutoff date for school entry.[3] For example, in England, the cutoff date is September 1, and there is a gap of almost one year in the age at which a child starts school if born just after September 1 (late school starters) rather than just before (early school starters). In this paper, we use rich administrative data from England to study

[1]If more marginal treatments for early starters have beneficial effects, then we would have an unfair under-treatment for late relative to early starters. On the contrary, if marginal treatments have negative effects, we would have an unfair over-treatment of early relative to late starters. In all cases, a gap in ADHD treatment between early and late starters would suggest a misallocation of resources.

[2]See, e.g., Brewer et al. (2007), Alalouf et al. (2019), Cuddy and Currie (2020), Einav et al. (2020), Persson et al. (2021), Currie and Zwiers (2023) and Bos et al. (2023).

[3]See Elder and Lubotsky (2009); Evans et al. (2010); Elder (2010); Layton et al. (2018) for the US, Morrow et al. (2012) and Furzer et al. (2022) for Canada, Zoëga et al. (2012) for Iceland, Root et al. (2019) and Fleming et al. (2022) for the UK, Krabbe et al. (2014) for the Netherlands, Schwandt and Wuppermann (2016) for Germany, Chen et al. (2016) for Taiwan, and Persson et al. (2021) for Sweden. A notable exception is the case of Denmark where Dalsgaard et al. (2012) and Dalsgaard et al. (2014) do not find strong evidence of increased rates of ADHD among early school starters, while Pottegård et al. (2014) find some modest evidence. A review of the relationship between school starting age and ADHD is provided in Whitely et al. (2018) and Schnorrbusch et al. (2020).

what explains the gap in ADHD between early and late school starters, and we suggest how to design interventions to reduce the unfair and inefficient use of medical resources for ADHD.

Similarly to previous studies, we find evidence that the relationship between ADHD drug prescription rate and children's date of birth has a discontinuity at the cutoff date for school entry, and we confirm this discontinuity for any age group between 5 and 15. However, when we look at the first-time prescription rate, i.e., when focusing on the incidence of ADHD prescriptions rather than prevalence, we find a discontinuity between ages 5 and 8 but no discontinuity from age 9 onward. These findings suggest that the long-term gap in ADHD between early and late starters is caused by prescriptions initiated in the first years of primary school, emphasizing the importance of correct diagnoses in this early critical period.

The ADHD gap between early and late starters has been interpreted in previous papers as resulting from a peer comparison bias by teachers. Early starters are the youngest in their grade, so teachers may misinterpret their more immature behavior relative to older classmates (for example, more inattention and more disruptive behavior in class) as ADHD symptoms.[4] Unlike previous literature that attributes the gap in ADHD against early school starters exclusively to their younger age relative to their classmates (see, e.g., Elder, 2010; Evans et al., 2010), we introduce a framework that considers a more comprehensive set of mechanisms. By doing so, we provide empirical evidence for the first time to confirm that relative age is the main mechanism explaining the long-term effect of being an early school starter on ADHD prescriptions and diagnosis rates.

While previous papers have not considered the full set of mechanisms that can explain the effect of being an early starter on ADHD, they have provided convincing evidence on the relevance of the effect of relative age and the comparison bias by teachers. Elder and Lubotsky (2009) show that the classmates' average age at school entry increases the probability of a child being diagnosed with ADHD even after controlling for the child's age at school entry, suggesting a potential relative age effect. Elder (2010) and Furzer et al. (2022) show that the child's ADHD prediction based on teachers' reports plotted against the date of birth has a discontinuity at the cutoff date for school entry, while the ADHD prediction based on parents' reports does not.

[4] See, e.g, Elder and Lubotsky (2009); Elder (2010); Evans et al. (2010); Morrow et al. (2012); Schwandt and Wuppermann (2016).

However, previous papers remain silent on other mechanisms that influence teachers' assessments besides the relative age effect. Teachers can have an amplified perception of ADHD symptoms in early starters not only because of their relatively younger age but also because their symptoms can become temporarily more severe when exposed to higher stress in school. Although school stress does not cause ADHD, which is primarily genetic (Riglin et al., 2016), it can worsen symptoms in class even if just temporarily,[5] leading teachers to notice more ADHD behaviors and hence to more marginal diagnoses and drug prescriptions.

Early starters may be exposed to higher stress levels of starting school due to their lower age and maturity at school entry – we refer to this as the *age at start of school* effect. They are also in a higher school grade than late starters observed at the same age, thus exposed to higher expectations and demand by teachers, which can cause an increase in their stress – *length of school exposure* effect. Young-for-grade children are more likely to experience an increase in stress caused by comparing themselves with classmates who tend to be older and more skilled (Kiessling and Norris, 2023) and by an increased likelihood of being targeted for bullying because their behavior can stand out more (Sarzosa and Urzúa, 2021) – *relative age* effect. Furthermore, the ADHD gap between early and later starters could be underestimated if we do not consider the *absolute age* effect. Because ADHD diagnoses and prescription rates increase with age, a non-age-adjusted rate of ADHD for early starters who are observed at a younger age than later starters in the same grade could underestimate the actual rate.

Our conceptual framework considers 4 potential mechanisms driving the early/late starter ADHD gap: the *age at the start of school*, the *length of school exposure*, the *relative age*, and the *absolute age*. Unlike previous papers focusing only on relative age, we adopt a model including these four mechanisms to explain the effect of early school start on ADHD.[6] Below, we explain how we address the issue of perfect collinearity between absolute age, age at the start of school, length of school exposure, and relative age.

As in Black et al. (2011), we account for the effect of absolute age by observing both early and

[5]Studies in behavioral genetics show that a stressful environment has an important role in explaining ADHD symptoms (see Livingstone et al., 2016; Björkenstam et al., 2018; Humphreys et al., 2019; Hartman et al., 2019). Focusing on adolescents with ADHD and using a qualitative approach, Öster et al. (2020) links stress and ADHD symptoms.

[6]See Cornelissen and Dustmann (2019) and Crawford et al. (2014) for similar models applied to other cognitive and socio-emotional skills.

late starters at the same age using administrative data on general practices, which enables us to determine whether a child has been diagnosed with ADHD or received prescriptions at any point in their life. To disentangle the three remaining mechanisms, we follow Crawford et al. (2014) and compare the effect of starting school early for children observed at the same age with the corresponding effect for children in the same school grade. This comparison allows us to isolate the effect of the absolute age and the length of school exposure and reveals that these two effects are short-lived and disappear by age 8. The only longer-term effects are thus the age at the start of school and the relative age. Starting school can cause more stress for early starters. However, we exclude stress as a potential channel explaining the long-term effects by showing that starting school early has no significant effects on anxiety diagnoses and prescriptions. Therefore, our results suggest that the long-term effect on ADHD is mainly driven by the relative age effect, as claimed in previous studies on ADHD (see, e.g., Elder, 2010; Evans et al., 2010; Persson et al., 2021; Furzer et al., 2022).

We leverage novel administrative data on general practices and hospitals in England, which provide information on health conditions, prescriptions, and treatments of children from birth onward. Focusing on a cohort of children born between 2002 and 2010, we estimate the effect of starting school early on diagnoses and prescriptions for ADHD and anxiety disorders at (i) each age between 5 and 15 and (ii) each grade from reception (kindergarten) to grade 10. Using the September 1 entry cutoff, we compare July-August to September-October births, controlling for observed child/family characteristics and medical practice fixed effects. The ADHD gap between early and late starters does not decrease with age. By the end of compulsory schooling, early starters still have a prescription rate for ADHD between 40% and 50% larger than late starters. This gap is mainly explained by differences in relative age and driven by first-time prescriptions initiated in the first years of primary school.

We rule out the possibility that our estimates are biased by red-shirting practices or grade retention because, unlike countries such as the United States, these practices are virtually nonexistent in England (Crawford et al., 2013). Even in the presence of red-shirting, our estimated effect of early starters on ADHD can interpreted as a lower bound on the true effect. This is because parents tend to delay entry into the school for more immature children. We also rule out that our results are driven by the season of birth, as gaps in ADHD between early and late starters have been consistently found across countries with varying cutoff dates and in both the south and north hemispheres (Bedard and Dhuey, 2006). Furthermore, our main findings are confirmed when considering the regression discontinuity approach and other sensitivity analyses.

Our first main contribution is a better understanding of the role of different mechanisms, besides the relative age effect, in explaining the gap in ADHD prescriptions between early and later starters. We analyze the effect of being an early starter on ADHD prescription and diagnosis rates both across age and school grades and provide a conceptual framework to explain how the comparison of these rates helps to understand the role of different mechanisms.

Our second main contribution is to the emerging literature on marginal treatments (e.g. Einav et al., 2020; Persson et al., 2021; Currie and Zwiers, 2023; Bos et al., 2023). A gap in marginal treatments between early and late starters suggests an inefficient and unfair use of medical resources. We provide guidelines on how to use the incidence and prevalence rates of ADHD prescriptions across ages and grades to identify the critical period in the child's life when gaps in marginal treatments between early and late starters are most likely to initiate. This is important because ADHD treatments, similar to treatments for other mental health disorders, tend to last for years. This means that initiating marginal treatments can have long-term effects on a child's health and public resources.

Finally, informed by our findings, we provide some advice on interventions to reduce the gap in ADHD between early and late starters. We warn against red-shirting indiscriminately for all early school starters and suggest two alternative strategies to reduce marginal diagnoses and treatments for early starters: sorting children into classes based on age and improving the diagnostic decision-making in early primary school.

The remainder of this paper unfolds as follows. In Section 2, we present our conceptual framework. Section 3 provides background on ADHD diagnosis and treatment and the education system in England. We describe our data in Section 4 and explain our strategy for identifying the effect of starting school early on ADHD in Section 5. We present our main results on gaps in prevalence and incidence rates of prescriptions between early and late starters and provide evidence on the underlying mechanisms and initiation period for such gaps in Section 6. We then move to explain the interpretation of these gaps in Section 7 and to the presentation of the heterogeneity and robustness checks in Section 8. Finally, Section 9 concludes with some policy recommendations.

2 Conceptual Framework

Twin studies have found that ADHD is highly heritable (see Jepsen and Michel, 2006), but differences in environment can alter the severity of ADHD symptoms (e.g., Livingstone et al., 2016; Björkenstam et al., 2018; Hartman et al., 2019). School stress is a school environmental factor that differs between children born just before and after the cutoff date for school entry (i.e., between early and late school starters), which is likely higher for early than late starters. Such school stress can temporarily amplify ADHD symptoms and explain the higher ADHD diagnosis and prescription rates, which are usually observed for early school starters. Besides stress, another factor that may explain the difference in ADHD rates between early and late starters is a peer-comparison bias. As emphasized by Elder and Lubotsky (2009), McEwan and Shapiro (2008) and Elder (2010), among others, early starters are younger and more immature than their classmates, so teachers may perceive ADHD symptoms in early starters as worse.

The temporary amplification and the magnified perception of ADHD symptoms caused by the heightened school stress and peer-comparison bias can lead to marginal diagnoses for early starters, i.e., to a higher probability of being diagnosed with ADHD even if the symptoms are moderate. As emphasized by Persson et al. (2021), these may lead to prescriptions with increased health care costs without health improvement. Even if marginal treatments for early starters were beneficial, we still would have an issue of misallocation of health resources with an unfair under-treatment of late starters with respect to early starters.

As explained in Section 1, there are 4 mechanisms through which being an early starter may affect ADHD diagnoses and prescriptions, which we describe in more detail in the following:

1. *Age at school entry*: Younger children have a lower level of emotional and cognitive maturity (low school readiness), causing a higher stress of starting school and making ADHD symptoms more evident, especially in the first school years.

2. *Length of school exposure*: Early starters observed at the same age as late starters will have been exposed to more schooling and less preschool childcare. Teachers' expectations and curriculum difficulty increase with school grades (length of school exposure), leading

to an increase in school stress for early starters, which can amplify the severity of ADHD symptoms, especially in the school environment.[7]

3. *Relative age*: Early starters are the youngest in their grade and can be perceived by teachers as having more severe symptoms due to peer-comparison bias. Furthermore, the stress caused by relative immaturity can worsen their mental health and temporarily aggravate their ADHD symptoms.[8]

4. *Absolute age* (also known as age at test, i.e., age when ADHD is measured): Early starters have a lower absolute age than late starters in the same grade. The effect of absolute age is explained neither by stress nor by the peer comparison bias but by the fact that ADHD diagnoses and prescriptions increase steadily from age 5 to 15 (Scahill and Schwab-Stone, 2000). If early and late starters are compared at the same time, late starters will be older and, hence, more likely to be diagnosed and treated.

If the main cause of the gap between early and late starters was the length of school exposure and, therefore, the stress caused by higher schools' and teachers' expectations by grade, then a solution could be to adjust the curriculum to the needs of each child's age within a grade. If the main driver of the gap was the age at school entry, i.e. the stress caused by lack of school readiness, then a solution could be to increase the school entry age. If, instead, the ADHD gap was mainly explained by relative age, then a better solution would be to group children into more refined classes within the same grade based on their age or improve diagnostic procedures to take account of the relative age. Finally, if the main mechanism behind the ADHD gap was the difference in absolute age between early and late starters, then there would be no need for interventions, as comparing children's ADHD at the same age would eliminate the ADHD gap. Below, we present a model that takes into account all four mechanisms.

We define a dummy variable $ADHD_{i,t}$ which takes value 1 if the child i received at least one prescription for ADHD in the 1-year period t and 0 otherwise, and we model the relationship between $ADHD_{i,t}$ and the four mechanisms discussed above adopting a linear probability model:[9]

[7]Note that the effect of this higher level of stress for early starters can be in part attenuated by the fact that a longer school exposure can improve children's socio-emotional skills (Cornelissen and Dustmann, 2019) and make them more accustomed to the school environment.

[8]Kiessling and Norris (2023) show that a worsening of the relative rank of children with respect to their schoolmates leads to a deterioration of mental health even after controlling for their actual skills.

[9]This model is similar to the model adopted by Cornelissen and Dustmann (2019) but it is extended to include

$$ADHD_{i,t} = \alpha_t + \gamma_{EXP,t}\, Exposure_{i,t} + \gamma_{AGEE,t}\, AgeEntry_{i,t} + \\ \gamma_{RELAGE,t}\, RelativeAge_{i,t} + \gamma_{ABSAGE,t}\, AbsAge_{i,t} + u_{i,t}, \tag{1}$$

where the subscript t denotes the specific one-year interval in the life of child i (either age or school grade). The length of school exposure ($Exposure$), absolute age ($AbsAge$), age at school entry ($AgeEntry$), and relative age ($RelativeAge$) are all measured in years and days. $u_{i,t}$ captures the effect of the covariates, which we will introduce in our empirical section, and a residual error term. For the reasons explained above, while the probability of having ADHD should increase with the length of school exposure and the absolute age, it decreases with the age at entry and relative age. In other words, we expect $\gamma_{EXP,t} > 0$, $\gamma_{ABSAGE,t} > 0$, $\gamma_{RELAGE,t} < 0$ and $\gamma_{AGEE,t} < 0$.

Note that we allow the coefficients of the model to change across different periods in the child's life, so our assumption of linearity in $Exposure$, $AbsAge$, $AgeEntry$, and $RelativeAge$, only needs to hold within the interval t. We also consider interactions between these mechanisms. For example, we allow the effects of relative age exposure and age at school entry to decrease as the child ages by allowing these effects to differ across t (age or school grade).

The effect of being an early school starter (born in July and August) with respect to late starters (born in September and October) is equivalent to the effect of increasing $Exposure$ by approximately 1 year and decreasing $AgeEntry$, $RelativeAge$, and $AbsAge$ by approximately 1 year, so the effect of being an early starter is given by:

$$\gamma_t = \gamma_{EXP,t} - \gamma_{AGEE,t} - \gamma_{RELAGE,t} - \gamma_{ABSAGE,t}. \tag{2}$$

The slope coefficients in equation (1) cannot be separately identified because of multicollinearity. However, by observing both early and late starters in the same academic year, i.e., defining the time t as school grade g, we keep $Exposure$ constant and identify:

$$\gamma_g = -\gamma_{AGEE,g} - \gamma_{RELAGE,g} - \gamma_{ABSAGE,g}, \tag{3}$$

the relative age as an additional mechanism. The assumption of a linear probability model will be relaxed in our sensitivity analysis.

where γ_g denotes the effect of being an early starter in academic grade g, which we call the same grade effect. Similarly, if we observe children at the same age $t = a$, we ensure that $AbsAge$ is the same for early and late starters, and we can identify the following effect of being an early starter at age a:

$$\gamma_a = \gamma_{EXP,a} - \gamma_{AGEE,a} - \gamma_{RELAGE,a}, \tag{4}$$

which we call the same age effect.

We compute γ_a and γ_g for age a going from 5 to 15 and for school grade g going from 0 (reception year/kindergarten)[10] to 10 and we compare the same school grade effect γ_g with the same age effect γ_a computed at $a = g + 5$, where 5 is the age that children turn during their first year in primary school in England, implying that $\gamma_{AGEE,a} = \gamma_{AGEE,g}$ and $\gamma_{RELAGE,a} = \gamma_{AGEE,g}$. Therefore, by taking the difference between γ_a and γ_g, we isolate the combined effect of school exposure and absolute age:

$$\gamma_a - \gamma_g = \gamma_{EXP,a} + \gamma_{ABSAGE,g}. \tag{5}$$

In our empirical analysis, we find that $(\gamma_{EXP,a} + \gamma_{ABSAGE,g})$ goes to zero from grade 3 (age 8) onward. If the effects $\gamma_{EXP,a}$ and $\gamma_{ABSAGE,g}$ are both positive, then we can conclude they both are short-lived. We expect the effect of absolute age to be positive because the rate of ADHD prescription increases with age. We also expect the effect of length of school exposure to be positive, mainly because early starters are in a higher school grade relative to late starters observed at the same age and, therefore, have heightened stress caused by a more challenging curriculum. Given these expectations, we conclude that $\gamma_{EXP,a}$ and $\gamma_{ABSAGE,g}$ are short-lived. Hence, from grade 3 onward, we are left with only two mechanisms explaining the effect of being an early starter: the effects of age at entry and relative age. Starting school early can cause an increase in stress, which could affect ADHD diagnoses and prescriptions. By looking at the effect of starting school early on anxiety diagnoses and prescriptions, we can exclude stress as a potential channel that explains the long-term effect of being an early starter on ADHD.

In our application, we also compute the same age effect γ_a and the same grade effect γ_g by considering first-time prescriptions, i.e., excluding prescriptions that were initiated before age

[10] Children in England start schooling with a reception year, which is equivalent to the US kindergarten year.

a and before grade g. This allows us to show that the gap in ADHD first-time prescriptions between early and late school starters vanishes as children grow older and relative age differences among classmates become less salient.

The gap in ADHD between children born just before and just after the cutoff date for school entry has been attributed by previous papers exclusively to a relative age effect (see, e.g., Elder and Lubotsky, 2009; Elder, 2010; Evans et al., 2010; Morrow et al., 2012; Schwandt and Wuppermann, 2016). Our conceptual framework helps to assess the role of other mechanisms to understand how to design better interventions that reduce the gap in ADHD.

3 Background

3.1 ADHD: Diagnosis and Treatment

ADHD is one of the most common neurodevelopmental disorders during childhood (Mannuzza and Klein, 2000). The symptoms of ADHD include inattentiveness, hyperactivity, and impulsiveness. Most children are diagnosed when they are between 6 and 12 years old, and ADHD often persists into adulthood.

In England, when parents or teachers notice ADHD symptoms in a child, they are advised to raise their concerns with their school's special educational needs coordinator or general practitioner (GP). GPs cannot diagnose ADHD; they can discuss parental concerns and refer the child to a specialist evaluation if necessary. Several specialists can conduct a formal assessment, including a psychiatrist, pediatrician, and community pediatrician,[11] learning disability specialist, social worker, or occupational therapist with experience with ADHD.

ADHD diagnosis is not determined through a single test. The process typically includes physical examinations and interviews with the child and significant others (such as parents and teachers). Symptoms need to be consistently displayed for a minimum of six months for a diagnosis to be made. These symptoms should be present in at least two different settings (which, in practice, involves showing them both at home and school) to ensure that they are not a reaction to certain teachers or parents. Therefore, the views of parents and teachers play a crucial role in

[11] Community pediatricians are child health doctors trained in hospital and community settings. They have experience working with children and their families with various conditions and needs (see https://www.nhft.nhs.uk/community-paediatrics/ for more information).

determining a diagnosis.

In most cases, ADHD is treated with a combination of behavioral therapy and medication. For children below the age of 5, behavioral therapy, particularly training for parents, is recommended as the first line of treatment before any medication is tried. There are five types of drugs licensed for the treatment of ADHD in England: methylphenidate, lisdexamfetamine, dexamfetamine, atomoxetine, and guanfacine. While these drugs may not provide a permanent cure for ADHD, they can provide relief from some of the associated symptoms. They have been known to improve concentration, reduce impulsivity, induce a sense of calmness, and facilitate the learning and applying new skills for individuals with this condition. This treatment has the potential to aid children with ADHD in focusing better in the classroom and minimizing risky behaviors outside of school. Aizer (2008); Dalsgaard et al. (2014); Chorniy and Kitashima (2016) show that children diagnosed with ADHD in pharmacological treatment have fewer hospital contacts if treated and that treatment, to some extent, protects against their engagement in criminal behavior. However, psychoactive medications also alter brain function and could have short- and long-term negative effects on the formation of human capital (Gould et al., 2009; Cascade et al., 2010; Currie et al., 2014).

The National Institute for Health and Care Excellence in the UK (NICE) provides guidelines for diagnosing and treating ADHD. Since 2008, the importance of considering relative age biases is explicitly included in these guidelines: “ADHD should be considered in all age groups, with symptom criteria adjusted for age-appropriate changes in behaviour” (NICE, 2018).

3.2 Early Schooling in England

In England, children are supposed to start their first primary school year (which is equivalent to the first grade in elementary school in the US) in September after they turn 5, with a large majority of children starting schooling one year earlier and attending reception year, which is equivalent to the US kindergarten year. As a result, most children start school full-time in the September after their fourth birthday, resulting in almost a year’s difference in age between children born just before and after September 1. Red-shirting – i.e., delaying a child’s entry to school – is uncommon in England, and virtually all children attend reception classes. We cannot observe the actual age at school entry in our administrative data, but Crawford et al. (2013)

show that over 99% of children were enrolled in the correct academic year for their age when considering the full population of children in state (public) schools in 2008 in England; while Cornelissen and Dustmann (2019) document that 96% of pupils attended a reception year when considering children born between 2000 and 2001 in England.

Although most primary schools in England have a single cutoff date for school entry (September 1), in the very early 2000s, a few primary schools in England allowed children to start at different times during the academic year (for more details, see Cornelissen and Dustmann 2019). Our sample covers children who started school between 2007 and 2015 (born between 2002 and 2010), and during this period, almost all minority schools switched to a single-entry cutoff date.[12]

Before starting reception class, children in England can be in formal childcare or be cared for at home. All 3- and 4-year-olds have the right to fully subsidized part-time preschool, which provides about 12.5 hours a week of free childcare (Blanden et al., 2016).

Children are assessed at the start and end of the reception year with an on-entry assessment (the reception baseline assessment) and a progress assessment (summary assessment). After completing the reception class, children attend primary (elementary) school from ages 5 to 11 and secondary school from 11 to 16. Most schools in England are state schools that do not charge tuition fees, and less than 10% of children are enrolled in independent schools requiring tuition fees. There are assessments called Standard Attainment Tests (SATs) administered at ages 7, 11, and 14, and an assessment called General Certificate of Secondary Education (GCSE) is typically taken at ages 15-16. Compulsory schooling ends at age 16.[13]

4 Data

4.1 Data Sources

We use data from QResearch, a large consolidated database derived from anonymized health records from general practices in England matched with hospital administrative data, the Hos-

[12] As our data does not allow us to identify or isolate children from the minority of schools with multiple entry points, we checked for any potential bias by comparing the effects of early school entry on ADHD rates across different cohorts of children born from 2002 to 2010. Our analysis shows that the effect of early school entry on ADHD rates remained consistent across these cohorts.

[13] As of 2015, children in England must do one of the following three until age 18: stay in full-time education; start an apprenticeship or traineeship; or spend 20 hours or more a week working or volunteering while in part-time education or training.

pital Episode Statistics (HES). In our analysis, we use individual-level information on general practice diagnostics, drug prescriptions, and maternity records from HES, which allows us to link children with their respective mothers. A commissioned report by Hippisley-Cox et al. (2005) found QResearch to closely match the national sample of practices age-sex distribution and geographical spread. General practices within the QResearch consortium exhibit a marginally larger size (i.e., more registered patients) than the average practice in England, a phenomenon likely indicative of the willingness of clinical practices to collaborate and exchange clinical data. Nevertheless, the patient demographic profiles within these QResearch-affiliated practices closely parallel the broader national patient demographics. QResearch has been used as a representative sample of the primary care population in well-published medical journals (Hippisley-Cox and Coupland, 2010; Gao et al., 2021; Aveyard et al., 2021).[14]

For all children in the data, we consider their prescriptions for ADHD-related disorders[15] and a vector of sociodemographic characteristics which includes the region of residence, ethnicity, general practice identifier, and socioeconomic status (SES). SES is measured at the postal code level using the Townsend deprivation index, reported in quintiles (Coupland et al., 2007). The Townsend deprivation index is an area-based measure of deprivation (Townsend et al., 1988) which is constructed from the following four census variables: households without a car, overcrowded households, households not owner-occupied, and unemployed individuals.[16] Our data also include information on the health at birth of children, including date of birth[17] and maternal age at birth.

4.2 Main Sample and Variables

Our main sample is given by all singletons born between July 1 and October 30 in any of the years between 2002 and 2010 and traceable in the administrative data up to age 10 ($N = 97,117$). As

[14]Furthermore, prediction algorithms based on QResearch data have been validated against data from the Clinical Practice Research Datalink (CPRD) (Hippisley-Cox et al., 2014) and the UK Office for National Statistics (Nafilyan et al., 2021).

[15]ADHD prescriptions in the database are coded using British National Formulary (BNF; a standard drug reference text in the UK) codes (see, e.g., Baker et al., 2017).

[16]Each of these four variables is divided by the appropriate count of households or persons to obtain a percentage score. The unemployment and overcrowding percentages are then log-transformed to normalize the raw values, which tend to be highly skewed. All four variables are then standardized using a Z-score. The Townsend deprivation index is the sum of these four scores.

[17]We observe the date of entry and exit to maternity care and proxy date of birth by taking the median point between these two dates. The average length of stay in maternity care is two days. In Section 6, we run a sensitivity analysis to show that this approximation of the date of birth does not bias our results.

detailed in Section 3, almost all children in England start the school reception (kindergarten) year in September after their fourth birthday. Restricting the sample to children born between July and October ensures that children are comparable in most dimensions, except for their school-starting age. We exclude individuals with missing information on socioeconomic variables (419 observations, less than 0.5% of the initial sample). This results in a final sample of 96,698 children. Since the children in our sample were born between 2002 and 2010 and our data includes information up to the end of 2020, we can track all children at least until age 10. After age 10, the sample gradually reduces. Table 1 shows each age's sample size.

In our empirical application, we examine the profile of ADHD by age and by school grade and consider two different definitions of the outcome variable: (1) a dummy variable for ADHD at age a, $ADHD_a$, that takes value 1 if a child received at least one prescription related to ADHD in the period $[a, a+1)$ where a is the child's age in years and ranges from 5 to 15, (2) a dummy variable for ADHD in school grade g, $ADHD_g$, that takes value 1 if a child received at least one prescription related to ADHD during the school grade g (September to August) with g starting from 0, the reception (kindergarten) year, and following with grades 1 to 10. Our key regressor is $Early$, a dummy variable that takes value 1 for children born in July-August and 0 if born in September-October.

4.3 Descriptive Statistics

The variables used in our analysis are summarized in column 1 of Table 2, where we report their averages using the full sample. As expected, approximately half of our sample is male, and half is female. Half of the children are of white ethnicity, and 15% are of nonwhite ethnicity. We do not have information about ethnicity for about one-third of the individuals in the sample, which we code as a separate category labeled 'unclassified'. Around half of the individuals in our sample live in postal codes with a deprivation index in the bottom two quintiles, representing the least deprived areas.[18] Over 7% of the children in our sample have a mother aged 20 or less. In the last two rows, we report the rate (in percentage points) of ADHD diagnosis and prescription averaged across the age range from 5 to 15.

In columns 2 and 3 of Table 2, we compare the characteristics of early school starters (children

[18] In later analyses, we label the bottom two quintiles as 'high SES' and the top two quintiles as 'low SES' (see Section 6).

born between July 1 and August 31) relative to late starters (children born between September 1 and October 30). We find statistically significant differences between early and late starters for some of these characteristics. For example, early starters are less likely to live in the least deprived postal codes (bottom quintile) and more likely to live in a postal code belonging to the top two quintiles. However, the only differences that are significant both statistically and in magnitude are the differences in the rate of diagnoses and prescriptions for ADHD. The rate of ADHD diagnosis increases by 50% for early starters relative to late starters, while the rate of drug prescriptions has an even larger increase of around 67%.

Next, we inspect the evolution of ADHD prescription rates across ages 0 to 15 for the entire sample considering the average of the dummy variable $ADHD_a$, which takes the value 1 if a child was prescribed at least one ADHD-related drug in the 1-year period $[a, a+1)$. The average of $ADHD_a$ multiplied by 100 is the rate of prescription at age a expressed in percentage points, which we plot for each age from 0 to 15 in Figure 1a. The percentage of children with an ADHD prescription is close to zero until the age of 5, which is in line with the medical guidelines described in Section 3, and gradually increases to 1.25 percent by the time children are 15 years old. Figure 1b again plots the profile of the ADHD prescription rate by age but separately for early and late starters. As expected, after age 5, early starters are more likely to receive a prescription for ADHD, and the disparity between them and late starters grows wider over time.

While the main focus of our analysis is to explain the differences between early and later starters, for completeness we also plot the ADHD prescription rate by age for middle starters (that is, children born between November and June) in Figure A.2. As expected, the prescription rates for middle starters are always below the early starters and above the late starters ones.

5 Empirical Strategy

In England, most children start school full-time in September after their fourth birthday by attending the reception year, equivalent to the US kindergarten year. Therefore, children born just before September start school almost 1 year earlier than children born on September 1 or soon after. In this section, we explain how we estimate the effect of being an early school starter (i.e., born in July-August) with respect to late school starters (i.e., born in September-October) on the probability of receiving a prescription for ADHD.

Let us consider the dummy variable ADHD observed at a specific period in the child's life t, which takes value 1 if a child has received at least one prescription for ADHD in the 1-year period t and 0 otherwise. Because the rate of ADHD prescriptions has been increasing across the years and varies across general medical practices, we control for the year of birth and general practice by estimating the following linear probability model:

$$ADHD_{i,t} = \alpha_t + \gamma_t \, Early_i + \beta_t \, X_i + \mu_{t,j} + \mu_{t,s} + \epsilon_{i,t}, \tag{6}$$

where the subscripts i, j, t, and s denote, respectively, the child, the general medical practice where the child is registered, the time period in the child's life when the ADHD prescription is observed, and the child's year of birth ($s = 2002, ..., 2010$). We observe children born between July and August, that is, within 2 months from the cutoff date for school entry, September 1. $Early_i$ takes value 1 if the child i was born in July-August and 0 if born in September-October. $\mu_{t,j}$ and $\mu_{t,s}$ are the general practice and birth year fixed effects, respectively, and $\epsilon_{i,t}$ is the idiosyncratic error term. α_t, γ_t, and β_t are the intercept, the effect of being an early starter as defined in equation (2) in Section 2, and the vector of coefficients corresponding to the vector of background controls X_i. The vector of controls includes sex, ethnicity, area deprivation index, and maternal age at birth (see Section 4 for a description of these variables). We account for heteroscedasticity by using robust standard errors using the Huber-White sandwich estimator.[19]

We estimate equation (6) separately for different values of t at which the child is observed. In our empirical application, we consider two ways to define the time t: age a and school grade g. To do so, we consider the following two outcome variables: (1) $ADHD_{i,a}$, that takes value 1 if a child received at least one prescription related to ADHD in the period $[a, a+1)$ with $a = 5, ..., 15$, and (2) $ADHD_{i,g}$, which takes the value 1 if child i received at least one prescription related to ADHD during the school grade g (September to August) with $g = 0, ..., 10$.

$(\gamma_t \cdot 100)$ measures the effect of being an early starter as a percentage point increase in ADHD rate, the probability of receiving an ADHD diagnosis at time t. Because the rate of ADHD prescriptions varies substantially across age (see Figure 1a), we present the effect of being an early starter in terms of percentage increase, rather than percentage points increase, in the

[19] In our robustness checks, we show that our results do not change when using a logistic regression specification rather than a linear probability model.

ADHD rate at time t, i.e.,

$$r_t \cdot 100 = \frac{Pr(ADHD_{i,t} = 1|Early_i = 1) - Pr(ADHD_{i,t} = 1|Early_i = 0)}{Pr(ADHD_{i,t} = 1|Early_i = 0)} \cdot 100, \qquad (7)$$

where the numerator can be replaced with γ_t and the denominator with the rate of ADHD prescription for late starters at time t.

To consistently estimate the effect of being an early starter using model (6), the following conditions need to hold: (1) parents do not delay the start of the school of their children based on potential gains (i.e., no red-shirting); and (2) the month of birth is not related to unobserved child and family characteristics that may affect ADHD. Evidence on the validity of condition (1) has been provided by Crawford et al. (2013), who find that over 99% of children are enrolled in the correct academic year for their age, considering the full population of children in state (public) schools in 2008 in England. If there was a substantial habit of red-shirting in our sample, our estimated effect could still be interpreted as a lower bound for the effect of early starters on ADHD. This is because red-shirting would lead parents to delay entry into the school, especially for more immature children, therefore attenuating the effect of being an early starter. The birth month can be correlated with child and family background characteristics (Buckles and Hungerman, 2013; Shigeoka, 2015). In our model, we address this endogeneity issue by controlling for the general practice and year of birth fixed effect, and a set of background characteristics.

Our rich administrative data on hospitals and general practices allow us to follow children across time and observe their ADHD diagnosis and prescriptions at each specific age between 5 and 15, so we can avoid recall and measurement issues, which usually affect parents' and teachers' reports on children's ADHD in sample surveys. However, we do not observe the child's age at school entry, so our estimate of the effect of starting school early should be interpreted as an intention-to-treat (ITT) estimate. We expect the ITT estimate to be close (if not identical) to the actual treatment effect in our sample given that more than 99% of children are enrolled in the correct school grade based on their age (Crawford et al., 2013).

To evaluate the robustness of our estimation method, we show that we obtain similar results when considering the following regression discontinuity model:

$$ADHD_{i,t} = \tilde{\alpha}_t + h_t(z_i) - \tilde{\gamma}_t\,\mathbb{1}(z_i \geq 0) + \tilde{\beta}_t\,X_i + \tilde{\mu}_{t,j} + \tilde{\mu}_{t,s} + \tilde{\epsilon}_{i,t}, \tag{8}$$

where the notation is the same as in equation (6) but we use the tilde˜sign to emphasize elements that may differ from model (6). z_i is the running variable, which is the birth date of child i centered around the cutoff date for school entry (September 1), $h_t(.)$ is a smooth function of the running variable, $\mathbb{1}(z_i \geq 0)$ is a dummy taking value 1 if the running variable is positive (late school starters) and 0 otherwise. Notice that $\tilde{\gamma}_t$ captures the effect of being an early starter with respect to being a late starter when comparing individuals born just before and just after the cutoff point of September 1. We allow $h_t(.)$ to be different to the right and left side of the cutoff point by fitting a local second-order polynomial on each side of the cutoff and restricting the bandwidth to 2 months (60 days) at both sides of the threshold.[20] Notice that the discontinuity at the cutoff can be considered a sharp discontinuity, given that almost all children start school in September after they turn 4.

While it is, in general, important to specify the correct function $h_t(.)$ for the relationship between ADHD and the running variable, the specification of $h_t(.)$ is less crucial if we focus on a narrow bandwidth, i.e., a sub-sample of children born close the cutoff point. The estimation of the model (6) is equivalent to the estimation of an RDD (regression discontinuity design) model with a bandwidth of 2 months on both sides of the cutoff date and a function $h_t(.)$ which is constant between July and October except for the jump (discontinuity) at the cutoff date. It is, therefore, unsurprising that the estimation of the model (6) produces results similar to the RDD model (8) as we will show in Section 8.2.

6 Main Results: Gaps in the Rates of ADHD Prescriptions between Early and Late Starters

In this section, we focus on the differences in the prevalence rates of ADHD prescriptions between early and late starters. That is, we include children who have either a first-time (new)

[20]When using covariate-adjusted bandwidth selection and robust bias-corrected inference (see Calonico et al., 2017), we find that the optimal bandwidth is between 57 and 89 depending on the child's age. The results remain similar to the ones produced using the 2-month bandwidth. This suggests that the 2-month bandwidth used to select our sample of children is close to the optimal choice.

prescription and children who have renewed prescriptions that were initiated earlier.

6.1 Descriptive Evidence on Our Identifying Variation

To evaluate the effect of starting school early on ADHD, our empirical analysis exploits the discontinuity in children's age at the start of school caused by the school entry cutoff. We compare ADHD rates between children born in July-August and in September-October – that is, children born before and after the school entry cutoff for England. Figure 2a plots the ADHD drug prescription rate against the week of birth centered around September 1 for all children aged between 5 and 15. In sub-figures (b) to (d), we repeat this exercise separately for children in groups 5-7, 8-10, and 11-15. All four sub-figures show a clear discontinuity in ADHD prescription rates at the cutoff date for school entry.

We check if the sorting into July-August and September-October births is related to the family background by comparing the average of background characteristics between early and late starters while controlling for the year of birth and general practice fixed effect as in our regression (6). Our results do not suggest any issue of endogeneity, as shown by the fact that the effect of the dummy *Early* on each of the control variables is never significantly different from zero at 5% level once we control for general practice and year-fixed effects (see Figure 3).[21] Even if the inclusion of these background variables in our regression is not needed for identification, we keep them in our model (equation (6)) to increase the precision of our estimation.

Finally, the regression discontinuity approach described in equation (8) relies on the assumption of random sorting of children and families at the cutoff date for school entry. There should be no bunching of the child's date of birth to the right or to the left of September 1. We evaluate this assumption's validity by plotting the birth date's histogram in days and a bandwidth of 60 days around the cutoff for school entry. Following recent developments in the literature on RDD, we also implement a non-parametric density estimator based on local polynomial techniques to evaluate the continuity of the density function around the cut-off date as proposed by Cattaneo et al. (2020). Figure A.1 summarizes the results of this exercise. We cannot reject the null hypothesis of no discontinuity in the density function at the cut-off point (conventional p-value

[21]The only exception we find is when looking at the probability of living in the most deprived postal codes, which has a difference between early and late starters that is statistically significant at 5% level but it is very small in magnitude – 0.02 percentage points.

= 0.587, robust p-value = 0.588), which suggests that there is no manipulation and supports the assumption of random sorting at the cutoff.

6.2 Gaps in Prevalence Rates by Age and School Grade: Exploring Mechanisms

Recall from Section 2 that the effect of being an early starter when observing both early and late starters at the same age a is given by $\gamma_a = \gamma_{EXP,a} - \gamma_{AGEE,a} - \gamma_{RELAGE,a}$, while the effect measured for children in the same grade g is $\gamma_g = -\gamma_{AGEE,g} - \gamma_{RELAGE,g} - \gamma_{ABSAGE,g}$, and we compare the same school grade effect γ_g with the same age effect γ_a computed at age $a = g + 5$. Recall also that we expect $\gamma_{EXP,t} > 0$, $\gamma_{ABSAGE,t} > 0$, $\gamma_{RELAGE,t} < 0$ and $\gamma_{AGEE,t} < 0$, which implies that γ_g will provide an under-estimation of the effect of being an early starter because biased by the absolute age effect, while γ_a will provide a larger estimation which includes the effect of length of school exposure (grade).

We begin by examining the effect of being an early starter on ADHD prescriptions, comparing early and late starters at the same age a, γ_a. To do so, we estimate equation (6) considering as outcome the dummy variable taking value one if a child was prescribed drugs related to ADHD in $[a, a+1)$, with age a going from 5 to 15. The results are presented in Figure 4, where we plot the proportional increase in the ADHD prescription rate caused by being an early starter (the coefficient r_a in equation (7)) with the corresponding 95% confidence interval. Our findings indicate that early starters are more likely to receive an ADHD prescription at all ages, from 5 to 15. The percentage increase in the ADHD prescription rate for early starters relative to late starters is statistically significant at 5% level at all ages and varies between 40% to 100%.[22]

These results are aligned with several other previous studies that have found that being an early starter leads to increased rates of ADHD regardless of the estimation technique or cutoff date, which eliminates the seasonal effect as an explanation for these results. For instance, Elder (2010) finds that in the US, early starters are about 60% more likely to be diagnosed with ADHD and twice as likely to use ADHD-related stimulants regularly in grades 5 and 8. Similarly, Evans et al. (2010) find that starting school earlier increases the ADHD rate by about 25% with respect to the mean rate in the US. Recent work by Persson et al. (2021) confirms

[22] The estimates are less precise at age 5 because the ADHD rate at this age is extremely low, as shown in Figure 1.

these findings in Sweden, where early starters are about 30% percent more likely to receive ADHD medications than late starters.

Next, we estimate the effect of being an early starter when measuring children in the same school grade γ_g. As shown in Figure 5, the percentage increase effect is negative in reception class and grades 1 and 2, while from grade 3 to grade 10, it is positive and ranges between about 30% and 50%. The negative effects are caused by the absolute age effect $\gamma_{ABSAGE,g}$, which contributes to γ_g with a negative sign. Since medications for ADHD are typically not recommended for children under the age of 5 and the use of medications increases with age, the effect of age for early starters $-\gamma_{ABSAGE,g}$ is negative, especially in reception class when early starters are not yet 5 and in grade 1 when they are just 5 years old. In Section 8.2, we use GP diagnosis to confirm that the negative and significant coefficient from Figure 5 is driven by the medical guidelines not recommending prescriptions to children below age 5.[23]

Most previous studies do not consider ADHD rates by grade. Two exceptions are Pottegård et al. (2014) and Schwandt and Wuppermann (2016), which focus on Denmark and Germany, respectively. They found patterns by grade for the gap in prescription rates similar to the pattern observed in our Figure 5. In both Denmark and Germany, the gaps by grade are small or even negative in the first two grades of primary school, increase up to grade 3-4, and then stabilize.[24]

Finally, we compare the effects of being an early starter on ADHD when considering children at the same age γ_a and in the same grade γ_g. Again, we represent these effects as proportional increases along with the 95% confidence intervals in Figure 6. The difference between the 'same age' effects and the 'same grade' effects identifies the sum of the effect of length of school exposure and absolute age effect, $(\gamma_{ABSAGE,a} + \gamma_{EXP,g})$, as shown in equation (5). Figure 6 reveals that $(\gamma_{ABSAGE,a} + \gamma_{EXP,g})$ tends to zero from grade 3 onward.[25]

We formally test for differences between the same age and the same grade effects in Table 3, and we find that these differences become statistically not significant at 5% level from age 8 (grade

[23] Note that the two negative coefficients in Figure 15 become statistically insignificant when we use ADHD diagnoses as an outcome variable, instead of prescriptions.

[24] These gaps by grade are reported in Figure 2 in Pottegård et al. (2014). The gaps by grades are not reported directly by Schwandt and Wuppermann (2016) but can be derived approximately by looking at their Figure 2.

[25] Figure 6 reports γ_a and γ_g divided by the probability of being prescribed for ADHD for late starters at age $a = g+5$ and at grade g. In Figure A.3, we show that these probabilities in the denominators are almost identical.

3) onward. This implies that we do not reject the assumption that $(\gamma_{ABSAGE,a} + \gamma_{EXP,g}) = 0$. The effect of absolute age $\gamma_{ABSAGE,a}$ is expected to be positive because the rate of ADHD prescription increases with age. We also expect the effect of length of school exposure $\gamma_{EXP,g}$ to be positive because of the higher stress for early starters caused by being in a higher grade. Under these expectations, $(\gamma_{ABSAGE,a} + \gamma_{EXP,g}) = 0$ implies that both $\gamma_{ABSAGE,a}$ and $\gamma_{EXP,g}$ tend to zero from age 8 (grade 3) onward and the effect of being an early starter ends up being driven by the effects of relative age and age at school entry.

Starting school early can cause stress for children, which may increase the likelihood of ADHD prescriptions in the short term. While stress itself is difficult to quantify directly, we leverage data on anxiety disorder diagnoses and prescriptions for children in our sample. Anxiety disorders can be driven by remarkable manifestations of stress (Rockhill et al., 2010). In Figure 7, we estimate the impact of school starting age on both anxiety prescriptions (Figure 7a) and diagnoses (Figure 7b). For prescriptions, we observe smaller effects than those documented for ADHD. Compared with Figure 6, the estimates on anxiety are smaller and often statistically insignificant. Because prescriptions for anxiety may only capture extreme cases of stress, we also estimate the effects of starting school early on diagnoses for anxiety. These results are also close to zero,[26] and suggest that stress is not a driver of the short and long term effects of being an early starter on ADHD. This implies that the age at the start of school is unlikely to be the mechanism explaining the long-term ADHD gap between early and late starters.

Further, support for this conclusion is provided by comparing our results with countries where primary school entry is at age 6 rather than 5. In these countries, there is still evidence for a strong effect of being an early starter on ADHD, which is comparable to ours.[27] Taken together with evidence from other countries, our findings on anxiety disorders reinforce the claim that the age at school entry is not the main driver of the differences in ADHD between early and late starters. This ultimately suggests that raising the age at entry into school for all children by 1 year would not be an effective policy to reduce the gap in ADHD between early and late starters.

In conclusion, the relative age effect seems to be the main driver of the long-term ADHD gap,

[26]In Tables 4 and 5, we evaluate the difference between the 'same age' and 'same grade' coefficients analogously to what we do in Table 3.

[27]See e.g., Elder (2010) for the US and Persson et al. (2021) for Sweden.

leading to more marginal diagnoses for early starters. While both early and late starters with high symptoms get treated, only early starters are likely to be treated for milder symptoms. Marginal diagnoses and treatments for ADHD for early but not late starters indicate an unfair and inefficient use of medical resources. On the one hand, if marginal treatments have beneficial effects, we would have an issue of under-treatment for later starters. On the other hand, if marginal treatments have negative effects, we would have an unfair over-treatment of early starters. In all cases, a gap in ADHD treatment between early and late starters would suggest a misallocation of resources.

A school policy that allows red-shirting (that is, allowing parents to delay their child's entry when born just before the cutoff date for school entry) may help reduce the gap in ADHD between children born just before and after the cutoff for school entry. However, it can also lead to a worse allocation of resources, for example, in the scenario where parents with children who most need some treatment for ADHD are the ones that delay the school entry and potentially also the diagnosis and treatment for ADHD. On the contrary, if parents from high socioeconomic status (SES) were more likely to delay their child's entry and children from low SES were the ones most in need of ADHD treatment, then allowing parents to delay school entry could redirect resources from high to low SES children and help to reduce the misallocation of resources and make it fairer. However, a gap would remain between low SES children born before and after the cutoff date, still suggesting an inefficient and unfair allocation of resources.

Since the long-term effects of being an early starter seem to be explained mainly by the relative age effect, a better solution than red-shirting could be sorting children into classes based on their months of birth to ensure that children have classmates of similar ages or improving ADHD diagnostic by making increasing awareness of the relative age issues.

6.3 Gaps in Incidence Rates of Prescriptions: When Do They Initiate?

Figure 6 suggests that the effect of being an early starter persists across ages. Why does the effect of starting school early not vanish with age? This may be caused by the fact that ADHD is a condition that continues into adulthood and that ADHD drug treatment requires adherence for long periods. To confirm this, we check the persistence of ADHD drug treatment between ages 9 and 15 for children who received at least one prescription between ages 5 and 8. We find that 86.16% of children are still prescribed drugs for ADHD at age 9, 83.98% at 10, 81.29% at 11, 82.75% at 12, 74.63% at 13, 70.21% at 14, 64.20% at 15.

Given this persistence in prescriptions, it is important to distinguish between incidence and prevalence rates of prescriptions. The incidence rate considers only first-time prescriptions, while the prevalence rate includes first-time and renewed prescriptions, i.e., prescriptions initiated at an earlier age.[28]

We plot the incidence rate of ADHD prescriptions against the week of birth (centered around September 1) and find a discontinuity between age 5 and 8 but no discontinuity from age 9 onward (see Figure 8). These findings suggest that the gap in the prevalence rate between early and late starters is generated by prescriptions initiated between the ages of 5 and 8 and not by new prescriptions.

We then re-estimate the main equation (6) with a new dependent variable $FirstPresc_{i,t}$ defined as a dummy taking value 1 if a child i receives his/her first prescription for ADHD in the 1-year period t and 0 otherwise. The estimated effects from this exercise are reported in Figure 9 and expressed as the difference in the probability of first-time prescriptions in period t (either age a or grade g) between early and late starters over the probability of any prescriptions in period t for later starters,

$$pr_t = \frac{Pr(FirstPresc_{i,t} = 1|Early_i = 1) - Pr(FirstPresc_{i,t} = 1|Early_i = 0)}{Pr(ADHD_{i,t} = 1|Early_i = 0)}, \tag{9}$$

where $ADHD_{i,t}$ in the denominator takes value 1 if a child receives either a new (first) or a renewed old prescription in period t. Note that the proportional increase in prevalence rate defined in (7) is

$$\begin{aligned} r_t &= \frac{Pr(ADHD_{i,t} = 1|Early_i = 1) - Pr(ADHD_{i,t} = 1|Early_i = 0)}{Pr(ADHD_{i,t} = 1|Early_i = 0)} \\ &= \frac{Pr(FirstPresc_{i,t} = 1|Early_i = 1) - Pr(FirstPresc_{i,t} = 1|Early_i = 0)}{Pr(ADHD_{i,t} = 1|Early_i = 0)} \\ &+ \frac{Pr(RenewPresc_{i,t} = 1|Early_i = 1) - Pr(RenewePresc_{i,t} = 1|Early_i = 0)}{Pr(ADHD_{i,t} = 1|Early_i = 0)}, \end{aligned} \tag{10}$$

[28] In Figure A.4, we show most children with ADHD receive their first-time prescription at the age of 9 or earlier.

so pr_t in (9) is the part of the proportional increase in prevalence rate explained by first prescriptions.

From age 9 (grade 4) onward, there is no effect of being an early starter on first prescriptions for ADHD, whereas there are effects up to age 8 (grade 3) similar in magnitude to the effects found in Figure 6. Differences in ADHD prescriptions between early and late starters at later ages are explained by the persistence of treatment for ADHD initiated before age 9. Given the persistence in prescriptions over time, it becomes clear that an adequate diagnosis in the early years is essential to reduce a recurrent unfair difference in marginal treatments for ADHD between early starters and late starters.

Our findings of no effect of starting school early on the rate of prescriptions initiated after age 8 are aligned with the results by Dalsgaard et al. (2012) and Dalsgaard et al. (2014), who find that there is no effect or very little effect on diagnoses established and prescriptions initiated after age 7 in Denmark. These findings again emphasize the importance of reducing potential biases in initiating prescriptions in the first years of primary school.

7 Interpretation of the Gap in ADHD Between Early and Late Starters

Every child can exhibit varying degrees of ADHD symptoms, including impulsiveness and inattention, and these symptoms follow a natural distribution within the population. Diagnoses and treatment for ADHD are given to children with symptoms in the top tail of such distribution. Considering the age-standardized distribution of symptoms, the existence of a gap in ADHD between early and late starters implies that the symptoms' threshold over which children get diagnosed and prescribed for ADHD differ between children born in July-August (early starters) and in September-October (late starters). This interpretation aligns with the one posited by Persson et al. (2021) and is graphically summarized in Figure 10. While children displaying pronounced ADHD symptoms are likely to receive diagnoses and be treated regardless of their date of birth, those with milder symptoms are more likely to undergo diagnosis and treatment if born in July-August than if born in September-October.

Why are there two different ADHD symptom thresholds used to identify ADHD in early and late starters? One reason that has been suggested in previous papers is that teachers interpret the worse behavior of early starters to late starters as symptoms of ADHD rather than attributing these behavioral differences to the relative age gap. Both Elder (2010) and Furzer et al. (2022) find that the level of ADHD symptoms reported by teachers has a discontinuity at the cutoff date for school entry while parents' level does not, therefore suggesting that the gap in ADHD diagnoses and prescriptions between early and late starters is likely to be driven by teachers. This implies that the teachers' perceived density distribution of symptoms is not adjusted for age and is inherently shifted to the right for early starters and the left for late starters, as stylized in Figure 11, where the density in blue represents the scientifically correct age-standardized density, while the green and red densities are non-standardized densities for early and later starters respectively. Evaluating ADHD symptoms through non-age-standardized distributions results in an inflation of the level of symptoms among early starters in contrast to their late-starting counterparts. Specialists are aware that ADHD symptoms vary by age,[29] so they are likely to consider an age-standardized distribution of ADHD symptoms. However, because their diagnosis process takes into consideration teachers' reports of ADHD symptoms, they end up adopting diagnostic thresholds that are lower for early starters and higher for late starters than their intended age-standardized threshold, as depicted in Figure 10. This ultimately leads to an increase in marginal diagnoses and prescriptions for early relative to late starters.

The lower symptom threshold for children born in July-August may imply that early starters are over-diagnosed and over-treated for ADHD. However, the higher symptoms threshold for children born in September-October could imply that late starters are under-diagnosed and under-treated. To determine whether there are issues of over- or under-diagnosis and treatment, we would need to know the threshold in the distribution of ADHD symptoms that is considered correct by the scientific community. As emphasized by Persson et al. (2021), this scientifically correct threshold is not observed in any data available to researchers.

Like previous studies, our study cannot determine whether the increased rate of ADHD prescriptions among early starters is due to under- or over-diagnosis of ADHD. Notwithstanding, we can interpret the increase in ADHD rates for early starters as the consequence of marginal diagnoses and prescriptions. Furthermore, we can attempt to contribute to this debate about

[29] In England, the guidelines from the National Institute for Health and Care Excellence recommend that "ADHD should be considered in all age groups, with symptom criteria adjusted for age-appropriate changes in behaviour".

over- and under-diagnosis and treatment by looking at different subgroups.

For subgroups such as girls that are at risk of under-diagnosis (Faraone et al., 2003), we expect the scientifically correct threshold to be much lower than the symptoms' thresholds used to identify ADHD in late-starter girls and possibly also lower than the threshold used for early starter girls. If so, late-starter girls would be much more under-diagnosed than early-starter girls. On the contrary, for subgroups such as high SES children that are at risk of over-diagnosis (see, e.g., Elder, 2010; Emma Degroote and Houtte, 2022; Elder and Zhou, 2021), we expect the scientifically correct threshold to be higher than the thresholds used to identify ADHD in early and late starters. In this case, high-SES children who are early starters are likely to be much more over-diagnosed than late starters.

In the following, we empirically explore the differences by gender and SES.

8 Heterogeneity and Robustness Analyses

8.1 Is the Effect of Starting School Early Different Across Subgroups?

In this section, we investigate whether the effect of being an early starter varies across subgroups, particularly those that have traditionally been under-diagnosed. Specifically, we examine the impact of early school entry on ADHD outcomes for two key characteristics: gender and socioeconomic status (SES).

Figure 12a shows the percentage of boys receiving at least one ADHD prescription is higher than girls across all ages between 5 and 15. One contributing factor to this discrepancy is that ADHD symptoms in girls are more concealed than in boys. Biederman et al. (2002) compare girls and boys diagnosed with ADHD and find that girls seem to have fewer in-school and out-of-school problems and a lower likelihood of experiencing learning disabilities, and their primary ADHD symptom is inattentiveness, which is a more covert symptom than hyperactivity and impulsivity usually observed in boys. These gender differences suggest that the lower ADHD prescription rate among girls may be explained by a risk of under-diagnosis and under-treatment for girls.

To examine how school-starting age affects ADHD differently across genders, we estimate equation (6) separately for boys and girls. The results are reported in Figure 13a in terms of the proportional increase in ADHD prescription rate by age. Interestingly, these effects are greater for girls than boys, although this difference is not statistically significant at all ages. Girls may benefit more from an extra preschool year than boys. This can lead to a larger gap in maturity at the start of school between early and late starters for girls than for boys and, hence, to a larger peer comparison bias.

To investigate socioeconomic disparities in ADHD prescriptions, we utilize the Townsend deprivation score, an area-level measure of SES. As explained in Section 4, our measure of SES is provided in quintiles. We designate areas in the bottom two quintiles (i.e., the two least deprived quintiles) as high SES, while areas in the top two quintiles are classified as low SES. In Figure 12b, we present the percentage of children with ADHD prescription by age, separately, for these two groups. Children living in low SES areas are more likely to be prescribed drugs related to ADHD. However, because of the much larger risks children from low SES are exposed to, these results do not necessarily imply that there are no issues of under-diagnosis among low SES.[30] Evidence on under-diagnosis of ADHD in low-SES children has been provided by Elder (2010), Emma Degroote and Houtte (2022), and Elder and Zhou (2021). Elder (2010) shows that for low-SES children (i.e., children in the bottom quartile of a composite SES index based on parental education, occupation, and income), the predicted probability of ADHD based on parents' reports is higher than the equivalent probability based on teachers' reports, which is higher than the probability based on actual diagnosis. The opposite relationship holds instead for high SES (top quartile) children. Emma Degroote and Houtte (2022) find that children with higher cognitive skills are more likely to be excused for their ADHD behavioral issues by providing them with an ADHD label. Because high-SES children tend to have higher cognitive skills, they are also probably more likely to be labeled as children with ADHD. Elder and Zhou (2021) suggest that children from low SES backgrounds, such as black children, are more likely to have school peers with lower skills, and because of the comparison bias, they are less likely to be diagnosed with ADHD. An over-diagnosis in children for high SES can also be explained by parents from privileged backgrounds being more aware of ADHD and more concerned about having their child treated for potential behavioral issues.

Subsequently, we separately estimate the effects of early school entry for low and high-SES

[30] E.g., low SES children are exposed to more ADHD risk factors such as low birth weight and maternal mental depression (Saigal et al., 2003; Froehlich et al., 2007).

children following the same procedure as for boys and girls. The results reported in Figure 13b illustrate that the effects of early school entry are more substantial for high SES children. However, the confidence intervals overlap for all ages except for age 10. The higher effect of being an early starter on high-SES children can be explained by the fact that investments in high-SES children during preschool are larger than for low-SES children. Therefore, they lead to a larger gap in maturity at the start of school between early and late starters for high than low-SES children (see Elder and Lubotsky, 2009). If high-SES children are more likely to be in school with high-SES children, then this maturity gap will be more visible to teachers and lead to an increase in diagnoses and prescriptions caused by a peer comparison bias. The results in Figure 13b suggest that high-SES children who are early school starters are more likely to receive marginal prescriptions for ADHD drugs than early starters with a low SES.

Suppose we believe that there are issues of under-prescription for girls and over-prescription for high-SES children, a theory supported by previous empirical papers. Then, we can go a step further and suggest the direction of the prescription error caused by the gap in the ADHD rate between early and late starters. We can thus infer that girls who are early starters are less at risk of under-prescription than late-starter girls;[31] while high-SES children who start school earlier are at higher risk of over-prescription than high-SES who are late starters. However, given the lack of precision in our results and the impossibility of knowing the scientifically correct threshold for the level of ADHD symptoms over which a prescription would be advisable, the issues of under and over-prescription should be assessed with a degree of caution.

For a complete analysis, we also explore whether the effect of starting school on first-time prescriptions varies by gender and SES. To do so, we define the first-time prescription rate and report the effects (see equation (9)) separately by gender and SES in Figure A.5. Once again, the findings suggest that the effects of early school entry on ADHD prescriptions are primarily driven by first-time prescriptions initiated before age 9, therefore emphasizing that there are no long-term effects caused by teachers' comparison bias but through the persistence of prescriptions initiated before age 9.

[31] This conclusion is aligned with recent work by Furzer et al. (2022), who find that teachers underestimate ADHD symptoms among late-starter females. In contrast, they are overestimated in males regardless of age.

8.2 Robustness Checks

We conduct several tests to evaluate the internal validity of our findings. First, we address concerns about differences in unobserved characteristics between early and late starters using the RDD approach described in Section 5. Being born just before or after the cutoff date for school entry is more plausibly exogenous than being born in July-August rather than in September-October, so differences in unobservable characteristics are less of a concern when comparing individuals born around the cutoff date using the RDD.[32] We report the effect of being born just before the cutoff date using the RDD approach at each age between 5 and 15 in Figure 14. The effects are expressed as a proportional increase in ADHD prescription rate for children born immediately before compared to those born right after the cutoff date, as we do in our baseline results. Our estimates are similar to those reported in Figure 4. If anything, the relative increase in ADHD prescriptions among early starters is larger when estimated using an RDD.

Second, we test the sensitivity of our results to the inclusion of family fixed effects instead of general practice fixed effects. Following Dhuey et al. (2019); Chen et al. (2015), we compare siblings who just met or missed the cutoff birthdate for school entry. The results of this exercise are presented in Figure A.6. As this exercise severely constrains the available sample, we extend our main estimates' +/− 60-day bandwidth (Figure A.6a) to 120 days (Figure A.6b). While imprecisely estimated, the results are qualitatively similar to our baseline results, particularly when we increase the sample in panel (b).

Third, we address concerns about measurement error in the date of birth and conduct a sensitivity analysis excluding children born close to the September 1 cutoff. Note that we do not observe the exact date of birth in our data; we only observe the start and end of the maternity care. The length of these maternity care spells is equal to or less than 3 days for 99% of our sample, with the average and median being around 2 days. We set the date of birth as the middle date between the start and end date of maternity care. We find virtually identical results to our baseline specification when we re-estimate our baseline model by excluding children born within +/− 3 days of the September 1 cutoff. The new results, shown in Figure A.7, are very similar to our baseline specification, as reported in Figure 4).

[32] The manipulation test does not reject of absence of discontinuity in the birth date at the cutoff (see Section 6.1).

Fourth, to test the sensitivity of our results to model specification, we adopt a logistic regression rather than a linear probability model. We report in Figure A.8 the estimated odds ratios, i.e., the ratio of the odds of receiving a prescription for early starters to the odds for late starters at each age between 5 and 15. These odds ratios are all statistically significantly larger than 1 and align with our baseline results.

Next, we shift our focus from the rate of ADHD prescriptions to the rate of ADHD diagnoses. To identify ADHD diagnoses, we use the Read Codes, a clinical terminology system used in general practices across the UK. We define ADHD using the same codes as Bushe et al. (2015).[33] Our dependent variable $ADHDdiagnosis_{i,a}$ ($ADHDdiagnosis_{i,g}$) takes value 1 if at least one relevant ADHD Read Code was recorded in the child's i medical history or the child received a prescription for ADHD at any point in time up to age a (grade g). In Figure A.9, we plot the share of children with an ADHD diagnosis by age group. Using this dependent variable, we re-estimate our baseline model and plot the proportional increase in ADHD diagnoses rate caused by being an early starter from age 5 to 15 and grades 0 to 10 in Figure 15. The results are similar in magnitude and significance to those in Figure 6.[34]

Finally, we also explore whether these results differ across the different birth cohorts included in our study. As shown in Figure A.10, there has been an increase in the prescription rates across cohorts. Such increase has also been observed in other developed countries, such as the US, and has raised concerns about whether children are being correctly diagnosed and treated for ADHD (Setlik et al., 2009). Nevertheless, when considering the effect of being an early starter on ADHD prescription by age and separately for three different adjacent birth cohorts (children born in 2002-2004, 2005-2007, and 2008-2010), we do not find any statistically significant differences across cohorts (see Figure A.11).

9 Conclusions and Discussion

Children who start school at a younger age are more likely to be diagnosed with ADHD and receive prescriptions for ADHD. Given the high prevalence of ADHD among children and adoles-

[33] See Bushe et al. 2015 for a complete list of codes.

[34] As mentioned in Section 6.2, we interpret the negative but insignificant coefficients for reception class and grade 1 as a confirmation of our intuition that these negative coefficients were driven by medications for ADHD typically not being recommended for children under the age of 5.

cents, as well as its long-term impact on educational and labor market outcomes, it is crucial to understand the underlying mechanisms behind this association. Specifically, it is important to investigate (1) what mechanisms explain the effect of starting school early on the probability of being diagnosed with and prescribed drugs for ADHD and (2) why this effect does not diminish with age as it does for the impact on a child's socio-emotional and cognitive skills, as found in previous research (e.g., Elder and Lubotsky, 2009; Robertson, 2011; Crawford et al., 2013).

Shedding light on these questions is essential for developing effective interventions to improve children's health and education. Our empirical answers can be summarized in the following two key findings. First, the effect of starting school early on ADHD prescriptions from age 8 onward is mainly explained by a relative age effect. Second, the long-term impact of early school entry on ADHD, observed from age 9 onward, is primarily explained by the persistence of prescriptions made between ages 5 and 8.

Our findings are aligned with previous evidence in two dimensions. First, we documented similar patterns in the increased rate of ADHD prescriptions for early starters as those in previous studies. Second, our results reinforce the hypothesis raised in Evans et al. (2010), Elder (2010), and Furzer et al. (2022) that the relative age difference between early and late starters is the main mechanism causing a gap in the ADHD prescription rate in the long-term.

The relative age effect is likely to be driven by an amplified perception of ADHD symptoms in early relative to late starters, which is caused by a peer-comparison bias. Therefore, a gap in ADHD between early and late starters implies, in the best scenario, more marginal diagnoses for early starters (Persson et al., 2021) and, in the worst scenario, an over-diagnosis of early starters or an under-diagnosis of late starters (see, e.g., Furzer et al., 2022). Over-diagnoses can affect negatively schooling performance (e.g., Currie et al., 2014); under-diagnoses may worsen children's future outcomes (e.g., Chorniy and Kitashima, 2016) and lead to negative spillover effects school mates (e.g., Aizer, 2008) and younger family members (e.g., Breining, 2014; Persson et al., 2021); while marginal diagnoses can lead to prescriptions with modest or even negative effects on health and with long-term economic costs (e.g., Persson et al., 2021). In conclusion, a gap in ADHD rates between early and late starters suggests that there is a misallocation of medical resources.

Based on our findings, we offer some key recommendations to parents and policymakers to reduce the gap in ADHD. First, a cautious approach is necessary when delaying children's school entry (red-shirting). This is because it may have varying effects on different subgroups of children. The effect of being an early starter may harm subgroups of children who have a high risk of ADHD over-diagnosis, e.g., high-SES children, leading to a potential increase in such risk of over-diagnosis or to marginal diagnosis with no or little benefit. Conversely, starting early may be beneficial for under-diagnosed groups, such as girls.

Second, increasing awareness of ADHD symptoms and of the effect of relative age among parents and teachers and improving early referrals and diagnosis between ages 5 and 8 could help mitigate biases and lead to reduced long-term prescription gaps after age 8. Improvements in teachers' assessment of ADHD could be achieved by training teachers to recognize children with ADHD behavioral issues and other learning disabilities. Furzer et al. (2022) provide some empirical evidence that teachers with training in learning disabilities report ADHD symptoms with no discontinuity at the cutoff date for school entry. Another possibility is to modify the ADHD diagnostic process to give less weight to teachers' reports when the child is a very early starter or a very late starter observed in the first years of primary school and with mild symptoms. Persson et al. (2021) provide a similar suggestion to reduce the propagation of marginal or misdiagnoses caused by "hereditary tagging" – i.e., by the fact that other cases of ADHD in the family can lead to an increase in ADHD treatment even for children who have very mild symptoms.

Lastly, another potential way to address the negative consequences of relative age bias is to sort children into separate classes based on the month and year of birth to ensure they have classmates of similar ages. This could be achieved by changing how children are sorted into classes within each cohort of students or by having multiple school entry points throughout the academic year.

References

Aizer, A. (2008). Peer effects and human capital accumulation: The externalities of ADD. Technical report, National Bureau of Economic Research.

Alalouf, M., S. Miller, and L. R. Wherry (2019). What difference does a diagnosis make? evidence from marginal patients. Technical report, National Bureau of Economic Research, Forthcoming at American Journal of Health Economics.

Aveyard, P., M. Gao, N. Lindson, J. Hartmann-Boyce, P. Watkinson, D. Young, C. A. Coupland, P. San Tan, A. K. Clift, D. Harrison, et al. (2021). Association between pre-existing respiratory disease and its treatment, and severe covid-19: a population cohort study. *The Lancet Respiratory Medicine 9*(8), 909–923.

Baker, R., D. Kendrick, L. J. Tata, and E. Orton (2017). Association between maternal depression and anxiety episodes and rates of childhood injuries: A cohort study from England. *Injury Prevention 23*(6), 396–402.

Bedard, K. and E. Dhuey (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics 121*(4), 1437–1472.

Biederman, J., E. Mick, S. V. Faraone, E. Braaten, A. Doyle, T. Spencer, T. E. Wilens, E. Frazier, and M. A. Johnson (2002). Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. *American Journal of Psychiatry 159*(1), 36–42.

Björkenstam, E., C. Björkenstam, B. Jablonska, and K. Kosidou (2018). Cumulative exposure to childhood adversity, and treated attention deficit/hyperactivity disorder: A cohort study of 543,650 adolescents and young adults in Sweden. *Psychological Medicine 48*(3), 498–507.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2011). Too young to leave the nest? The effects of school starting age. *The Review of Economics and Statistics 93*(2), 455–467.

Blanden, J., E. Del Bono, S. McNally, and B. Rabe (2016). Universal pre-school education: The case of public funding with private provision. *The Economic Journal 126*(592), 682–723.

Bos, M., A. Hertzberg, and A. Liberman (2023). The effects of diagnosing a young adult with a mental illness: Evidence from randomly assigned doctors.

Breining, S. N. (2014). The presence of adhd: Spillovers between siblings. *Economics Letters 124*(3), 469–473.

Brewer, N. T., T. Salz, and S. E. Lillie (2007). Systematic review: the long-term effects of false-positive mammograms. *Annals of internal medicine 146*(7), 502–510.

Buckles, K. S. and D. M. Hungerman (2013). Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics 95*(3), 711–724.

Bushe, C., B. Wilson, F. Televantou, M. Belger, and L. Watson (2015). Understanding the treatment of attention deficit hyperactivity disorder in newly diagnosed adult patients in general practice: A UK database study. *Pragmatic and Observational Research 6*, 1.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal 17*(2), 372–404.

Cascade, E., A. H. Kalali, and S. B. Wigal (2010). Real-world data on: Attention deficit hyperactivity disorder medication side effects. *Psychiatry (Edgmont) 7*(4), 13.

Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association 115*(531), 1449–1455.

Chen, K., N. Fortin, and S. Phipps (2015). Young in class: Implications for inattentive/hyperactive behaviour of canadian boys and girls. *Canadian Journal of Economics/Revue canadienne d'économique 48*(5), 1601–1634.

Chen, M.-H., W.-H. Lan, Y.-M. Bai, K.-L. Huang, T.-P. Su, S.-J. Tsai, C.-T. Li, W.-C. Lin, W.-H. Chang, T.-L. Pan, et al. (2016). Influence of relative age on diagnosis and treatment of attention-deficit hyperactivity disorder in Taiwanese children. *Journal of Pediatrics 172*, 162–167.

Chorniy, A. and L. Kitashima (2016). Sex, drugs, and ADHD: The effects of ADHD pharmacological treatment on teens' risky behaviors. *Labour Economics 43*, 87–105.

Cornelissen, T. and C. Dustmann (2019). Early school exposure, test scores, and noncognitive outcomes. *American Economic Journal: Economic Policy 11*(2), 35–63.

Coupland, C., S. Harcourt, Y. Vinogradova, G. Smith, C. Joseph, M. Pringle, and J. Hippisley-Cox (2007). Inequalities in uptake of influenza vaccine by deprivation and risk group: Time trends analysis. *Vaccine 25*(42), 7363–7371.

Crawford, C., L. Dearden, and E. Greaves (2013). The impact of age within academic year on adult outcomes. Technical report, IFS Working Papers.

Crawford, C., L. Dearden, and E. Greaves (2014). The drivers of month-of-birth differences in children's cognitive and non-cognitive skills. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 177*(4), 829–860.

Cuddy, E. and J. Currie (2020). Rules vs. discretion: Treatment of mental illness in us adolescents. Technical report, National Bureau of Economic Research, forthcoming Journal of Political Economy.

Currie, J. and M. Stabile (2006). Child mental health and human capital accumulation: The case of ADHD. *Journal of Health Economics 25*(6), 1094–1118.

Currie, J., M. Stabile, and L. Jones (2014). Do stimulant medications improve educational and behavioral outcomes for children with ADHD? *Journal of Health Economics 37*, 58–69.

Currie, J., M. Stabile, P. Manivong, and L. L. Roos (2010). Child health and young adult outcomes. *Journal of Human resources 45*(3), 517–548.

Currie, J. and E. Zwiers (2023). Medication of postpartum depression and maternal outcomes: Evidence from geographic variation in dutch prescribing. *Journal of Human Resources*.

Dalsgaard, S., M. K. Humlum, H. S. Nielsen, and M. Simonsen (2012). Relative standards in ADHD diagnoses: The role of specialist behavior. *Economics Letters 117*(3), 663–665.

Dalsgaard, S., M. K. Humlum, H. S. Nielsen, and M. Simonsen (2014). Common Danish standards in prescribing medication for children and adolescents with ADHD. *European Child & Adolescent Psychiatry 23*, 841–844.

Dalsgaard, S., H. S. Nielsen, and M. Simonsen (2014). Consequences of ADHD medication use for children's outcomes. *Journal of Health Economics 37*, 137–151.

Dhuey, E., D. Figlio, K. Karbownik, and J. Roth (2019). School starting age and cognitive development. *Journal of Policy Analysis and Management 38*(3), 538–578.

Ding, W., S. F. Lehrer, J. N. Rosenquist, and J. Audrain-McGovern (2009). The impact of poor health on academic performance: New evidence using genetic markers. *Journal of Health Economics 28*(3), 578–597.

Einav, L., A. Finkelstein, T. Oostrom, A. Ostriker, and H. Williams (2020). Screening and selection: The case of mammograms. *American Economic Review 110*(12), 3836–3870.

Elder, T. and Y. Zhou (2021). The black-white gap in noncognitive skills among elementary school children. *American Economic Journal: Applied Economics 13*(1), 105–32.

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. and D. H. Lubotsky (2009). Kindergarten entrance age and children's achievement: Impacts of state policies, family background, and peers. *Journal of Human Resources 44*(3), 641–683.

Emma Degroote, M.-C. B. and M. V. Houtte (2022). Suspicion of adhd by teachers in relation to their perception of students cognitive capacities: do cognitively strong students escape verdict? *International Journal of Inclusive Education 0*(0), 1–15.

Evans, W. N., M. S. Morrill, and S. T. Parente (2010). Measuring inappropriate medical diagnosis and treatment in survey data: The case of ADHD among school-age children. *Journal of Health Economics 29*(5), 657–673.

Faraone, S. V., J. Sergeant, C. Gillberg, and J. Biederman (2003). The worldwide prevalence of ADHD: Is it an American condition? *World Psychiatry 2*(2), 104.

Fleming, M., A. Bandyopadhyay, J. S. McLay, D. Clark, A. King, D. F. Mackay, R. A. Lyons, K. Sayal, S. Brophy, and J. P. Pell (2022). Age within schoolyear and attention-deficit hyperactivity disorder in Scotland and Wales. *BMC Public Health 22*(1), 1–9.

Fletcher, J. and B. Wolfe (2008). Child mental health and human capital accumulation: The case of ADHD revisited. *Journal of Health Economics 27*(3), 794–800.

Fletcher, J. and B. Wolfe (2009). Long-term consequences of childhood ADHD on criminal activities. *The Journal of Mental Health Policy and Economics 12*(3), 119.

Fletcher, J. M. (2014). The effects of childhood ADHD on adult labor market outcomes. *Health Economics 23*(2), 159–181.

Froehlich, T. E., B. P. Lanphear, J. N. Epstein, W. J. Barbaresi, S. K. Katusic, and R. S. Kahn (2007). Prevalence, recognition, and treatment of attention-deficit/hyperactivity disorder in a national sample of US children. *Archives of Pediatrics & Adolescent Medicine 161*(9), 857–864.

Furzer, J., E. Dhuey, and A. Laporte (2022). ADHD misdiagnosis: Causes and mitigators. *Health Economics 31*(9), 1926–1953.

Gao, M., C. Piernas, N. M. Astbury, J. Hippisley-Cox, S. O'Rahilly, P. Aveyard, and S. A. Jebb (2021). Associations between body-mass index and COVID-19 severity in 6.9 million people in England: a prospective, community-based, cohort study. *The Lancet Diabetes & Endocrinology 9*(6), 350–359.

Gould, M. S., B. T. Walsh, J. L. Munfakh, M. Kleinman, N. Duan, M. Olfson, L. Greenhill, and T. Cooper (2009). Sudden death and use of stimulant medications in youths. *American Journal of Psychiatry 166*(9), 992–1001.

Hartman, C. A., N. Rommelse, C. L. van der Klugt, R. B. Wanders, and M. E. Timmerman (2019). Stress exposure and the course of ADHD from childhood to young adulthood: Comorbid severe emotion dysregulation or mood and anxiety problems. *Journal of Clinical Medicine 8*(11), 1824.

Hippisley-Cox, J. and C. Coupland (2010). Unintended effects of statins in men and women in england and wales: population based cohort study using the qresearch database. *BMJ 340*, c2197.

Hippisley-Cox, J., C. Coupland, and P. Brindle (2014). The performance of seven qprediction risk scores in an independent external sample of patients from general practice: a validation study. *BMJ open 4*(8), e005809.

Hippisley-Cox, J., Y. Vinogradova, C. Coupland, and M. Pringle (2005). Comparison of key practice characteristics between general practices in England and Wales and general practices in the QRESEARCH data.

Humphreys, K. L., E. L. Watts, E. L. Dennis, L. S. King, P. M. Thompson, and I. H. Gotlib (2019). Stressful life events, adhd symptoms, and brain structure in early adolescence. *Journal of Abnormal Child Psychology 47*, 421–432.

Jepsen, J. R. M. and M. Michel (2006). ADHD and the symptom dimensions inattention, impulsivity, and hyperactivity: A review of aetiological twin studies from 1996 to 2004. *Nordic Psychology 58*(2), 108–135.

Kiessling, L. and J. Norris (2023). The long-run effects of peers on mental health. *The Economic Journal 133*(649), 281–322.

Krabbe, E., E. Thoutenhoofd, M. Conradi, S. Pijl, and L. Batstra (2014). Birth month as predictor of adhd medication use in dutch school classes. *European Journal of Special Needs Education 29*(4), 571–578.

Layton, T. J., M. L. Barnett, T. R. Hicks, and A. B. Jena (2018). Attention deficit–hyperactivity disorder and month of school enrollment. *New England Journal of Medicine 379*(22), 2122–2130.

Livingstone, L. T., W. L. Coventry, R. P. Corley, E. G. Willcutt, S. Samuelsson, R. K. Olson, and B. Byrne (2016). Does the environment have an enduring effect on ADHD? A longitudinal study of monozygotic twin differences in children. *Journal of Abnormal Child Psychology 44*(8), 1487–1501.

Mannuzza, S. and R. G. Klein (2000). Long-term prognosis in attention-deficit/hyperactivity disorder. *Child and Adolescent Psychiatric Clinics of North America 9*(3), 711–726.

McEwan, P. J. and J. S. Shapiro (2008). The benefits of delayed primary school enrollment: Discontinuity estimates using exact birth dates. *Journal of Human Resources 43*(1), 1–29.

Morrow, R. L., E. J. Garland, J. M. Wright, M. Maclure, S. Taylor, and C. R. Dormuth (2012). Influence of relative age on diagnosis and treatment of attention-deficit/hyperactivity disorder in children. *Canadian Medical Association Journal 184*(7), 755–762.

Nafilyan, V., B. Humberstone, N. Mehta, I. Diamond, C. Coupland, L. Lorenzi, P. Pawelek, R. Schofield, J. Morgan, P. Brown, et al. (2021). An external validation of the qcovid risk prediction algorithm for risk of mortality from covid-19 in adults: a national validation cohort study in england. *The Lancet Digital Health 3*(7), e425–e433.

NICE (2018). Attention deficit hyperactivity disorder: Diagnosis and management.

Öster, C., M. Ramklint, J. Meyer, and J. Isaksson (2020). How do adolescents with ADHD perceive and experience stress? An interview study. *Nordic Journal of Psychiatry 74*(2), 123–130.

Persson, P., X. Qiu, and M. Rossin-Slater (2021). Family spillover effects of marginal diagnoses: The case of ADHD. NBER Working Paper Series 28334, National Bureau of Economic Research.

Pottegård, A., J. Hallas, and H. Zoëga (2014). Children's relative age in class and use of medication for adhd: a Danish Nationwide Study. *Journal of Child Psychology and Psychiatry 55*(11), 1244–1250.

Riglin, L., S. Collishaw, A. K. Thapar, S. Dalsgaard, K. Langley, G. D. Smith, E. Stergiakouli, B. Maughan, M. C. ODonovan, and A. Thapar (2016). Association of genetic risk variants with attention-deficit/hyperactivity disorder trajectories in the general population. *JAMA Psychiatry 73*(12), 1285–1292.

Robertson, E. (2011). The effects of quarter of birth on academic outcomes at the elementary school level. *Economics of Education Review 30*(2), 300–311.

Rockhill, C., I. Kodish, C. DiBattisto, M. Macias, C. Varley, and S. Ryan (2010). Anxiety disorders in children and adolescents. *Current problems in pediatric and adolescent health care 40*(4), 66–99.

Root, A., J. P. Brown, H. J. Forbes, K. Bhaskaran, J. Hayes, L. Smeeth, and I. J. Douglas (2019). Association of relative age in the school year with diagnosis of intellectual disability, attention-deficit/hyperactivity disorder, and depression. *JAMA Pediatrics 173*(11), 1068–1075.

Saigal, S., J. Pinelli, L. Hoult, M. M. Kim, and M. Boyle (2003). Psychopathology and social competencies of adolescents who were extremely low birth weight. *Pediatrics 111*(5), 969–975.

Sarzosa, M. and S. Urzúa (2021). Bullying among adolescents: The role of skills. *Quantitative Economics 12*(3), 945–980.

Scahill, L. and M. Schwab-Stone (2000). Epidemiology of ADHD in school-age children. *Child and Adolescent Psychiatric Clinics of North America 9*(3), 541–555.

Schnorrbusch, C., G. A. Fabiano, A. M. Aloe, and R. C. Toro Rodriguez (2020). Attention deficit hyperactivity disorder and relative age: A meta-analysis. *School psychology review 49*(1), 2–19.

Schwandt, H. and A. Wuppermann (2016). The youngest get the pill: ADHD misdiagnosis in Germany, its regional correlates and international comparison. *Labour Economics 43*, 72–86.

Setlik, J., G. R. Bond, and M. Ho (2009). Adolescent prescription ADHD medication abuse is rising along with prescriptions for these medications. *Pediatrics 124*(3), 875–880.

Shigeoka, H. (2015). School entry cutoff date and the timing of births. NBER Working Paper Series 21402, National Bureau of Economic Research.

Thomas, R., S. Sanders, J. Doust, E. Beller, and P. Glasziou (2015). Prevalence of attention-deficit/hyperactivity disorder: A systematic review and meta-analysis. *Pediatrics 135*(4), e994–e1001.

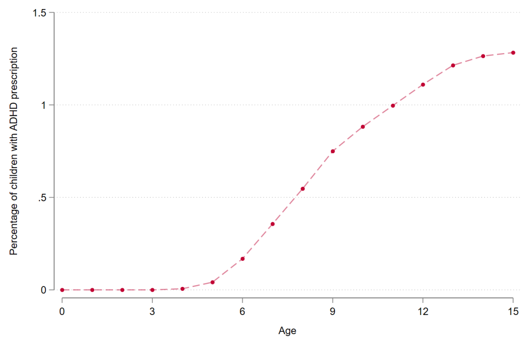
Townsend, P., P. Phillimore, and A. Beattie (1988). *Health and deprivation: Inequality and the North.* Routledge.

Whitely, M., M. Raven, S. Timimi, J. Jureidini, J. Phillimore, J. Leo, J. Moncrieff, and P. Landman (2018). Attention deficit hyperactivity disorder late birthdate effect common in both high and low prescribing international jurisdictions: Systematic review. *Journal of Child Psychology and Psychiatry 60*(4), 380–391.

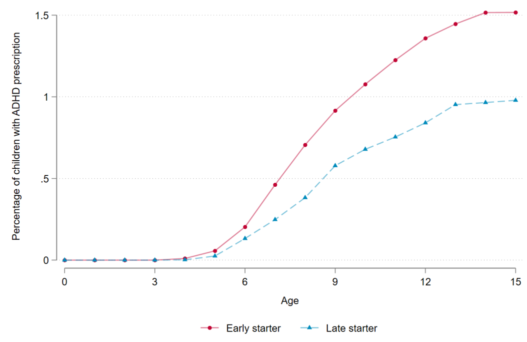
Zoëga, H., U. A. Valdimarsdóttir, and S. Hernández-Díaz (2012). Age, academic performance, and stimulant prescribing for ADHD: a nationwide cohort study. *Pediatrics 130*(6), 1012–1018.

Figures

Figure 1: Rate of ADHD Prescriptions by Age



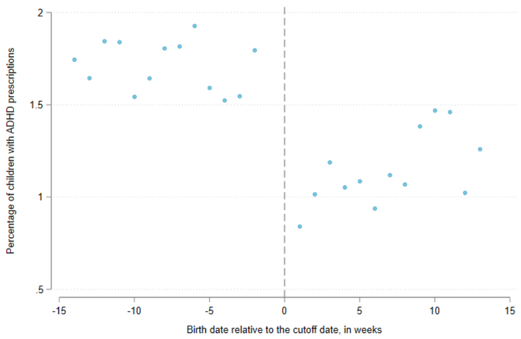
(a) Pooled sample



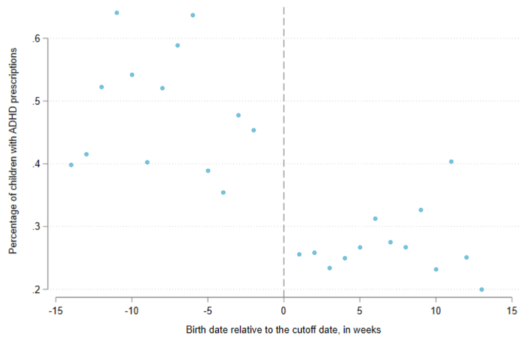
(b) Early vs late starters

Notes: Each marker captures the percentage of children receiving an ADHD prescription in the corresponding age period. Panel (a) considers the full sample of children born between July and October, while Panel (b) considers early starters (born in July and August) and late starters (born in September and October) separately.

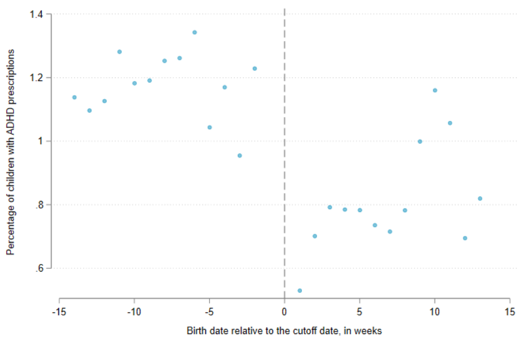
Figure 2: Prevalence of ADHD Prescriptions by Week of Birth



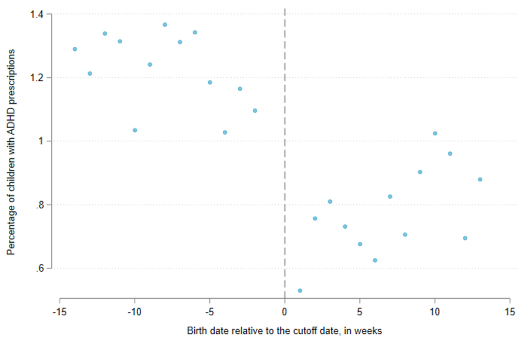
(a) All Ages (5 to 15)



(b) Ages 5 to 7



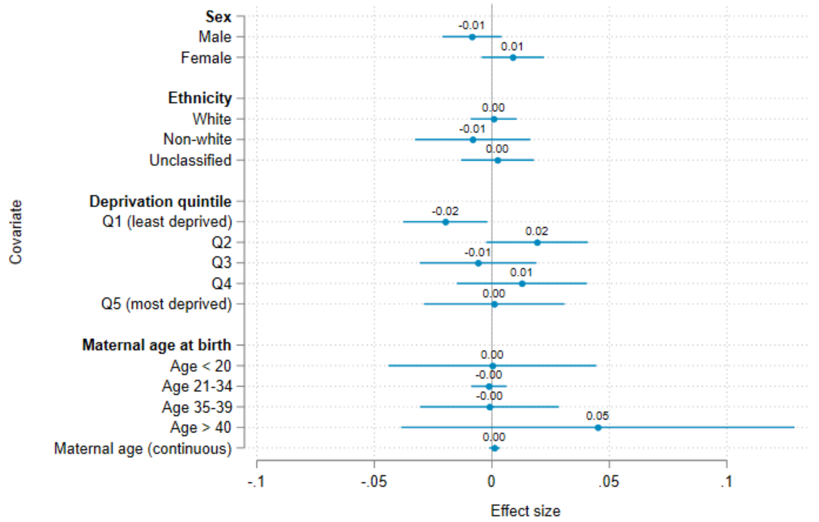
(c) Ages 8 to 10



(d) Ages 11 to 15

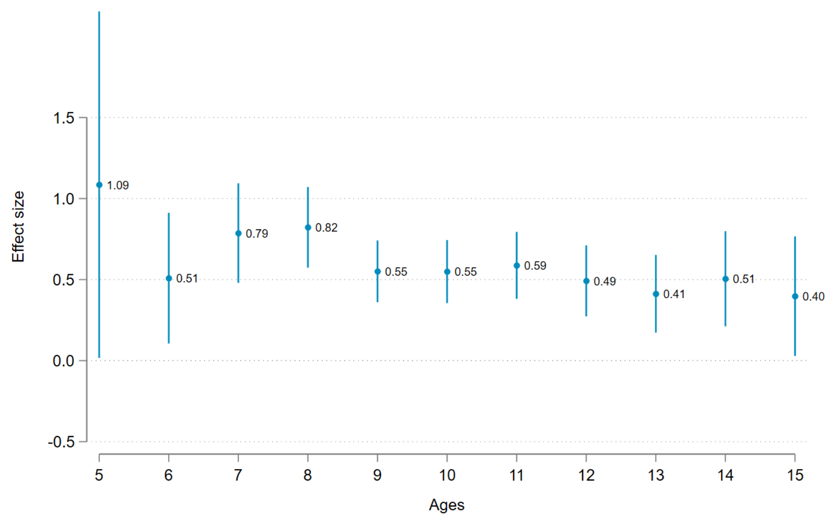
Notes: Each marker captures the percentage of children receiving a prescription for ADHD by birth date in weeks relative to September 1. Panel (a) consider children at any age between 5 and 15, while panels (b) to (d) focus on different age groups, 5-7, 8-10 and 11-15. The vertical dashed line represents September 1 – the cut-off point for school entry.

Figure 3: Balance of Observable Characteristics After Controlling for General Practice and Year Fixed Effects



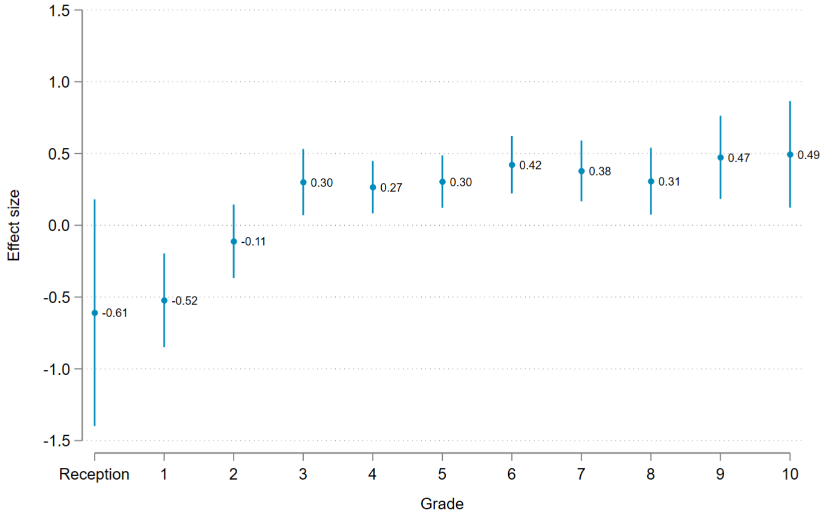
Notes: Each covariate is regressed on the dummy for early starters while controlling for general practice and year-fixed effects. The figure plots the estimated effect of the early starters dummy and the corresponding 95% confidence intervals.

Figure 4: Effect of Early Start of School on ADHD Prescriptions by Age



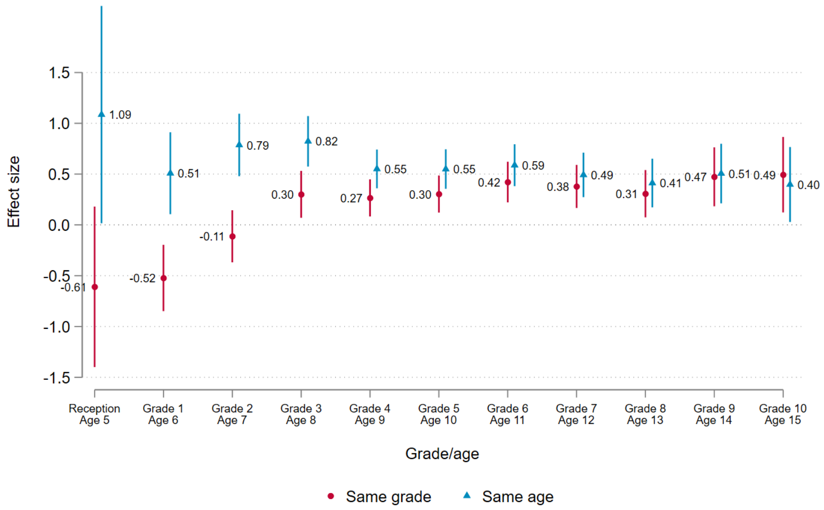
Notes: The figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 15. The effects are expressed as a proportional increase in the prescription rate for early starters relative to late starters, $r_a = \gamma_a / Pr(ADHD_{i,a} = 1 | Early_i = 0)$. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include GP and year of birth fixed effects.

Figure 5: Effect of Early Start of School on ADHD Prescriptions By Grade



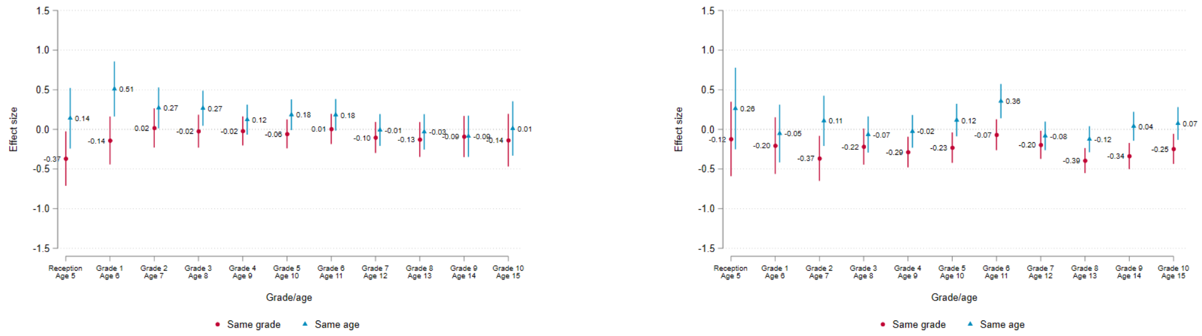
Notes: The figure plots point estimates and 95% confidence intervals of separate regressions, one for each grade from reception class to grade 10. The effects are expressed as a proportional increase in the prescription rate for early starters relative to late starters, $r_g = \gamma_g / Pr(ADHD_{i,g} = 1 | Early_i = 0)$. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects.

Figure 6: Effect of Early Start of School on ADHD Prescriptions by Age and Grade



Notes: The figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 15 and grade from reception class to grade 10. The effects are expressed as a proportional increase in the prescription rate for early starters relative to late starters, $r_t = \gamma_t / Pr(ADHD_{i,t} = 1 | Early_i = 0)$ with $t = a$ for age and g for grade. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects.

Figure 7: Effect of Early Start of School on Anxiety Prescriptions and Diagnoses by Age and Grade

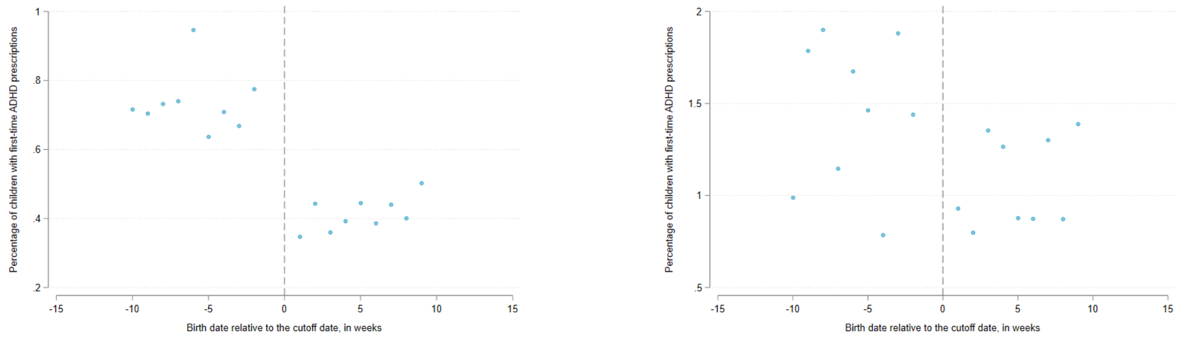


(a) Prescriptions for Anxiety

(b) Diagnoses of Anxiety

Notes: The figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 15 and grade from reception class to grade 10. The effects are expressed as a proportional increase in the prescription (panel (a)) and diagnosis (panel (b)) rate for early starters relative to late starters, $r_t = \gamma_t / Pr(Anxiety_{i,t} = 1 | Early_i = 0)$ with $t = a$ for age and g for grade. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects.

Figure 8: Incidence of ADHD Prescriptions by Week of Birth

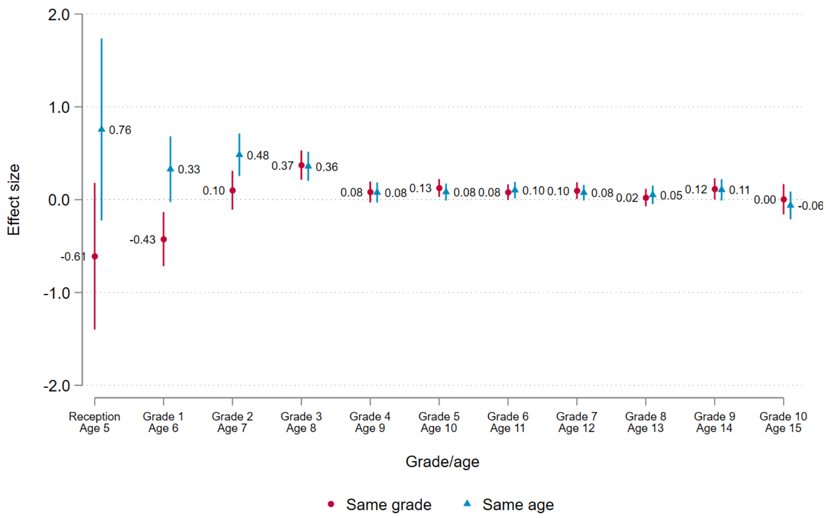


(a) Ages 5 to 8

(b) Ages 9 to 15

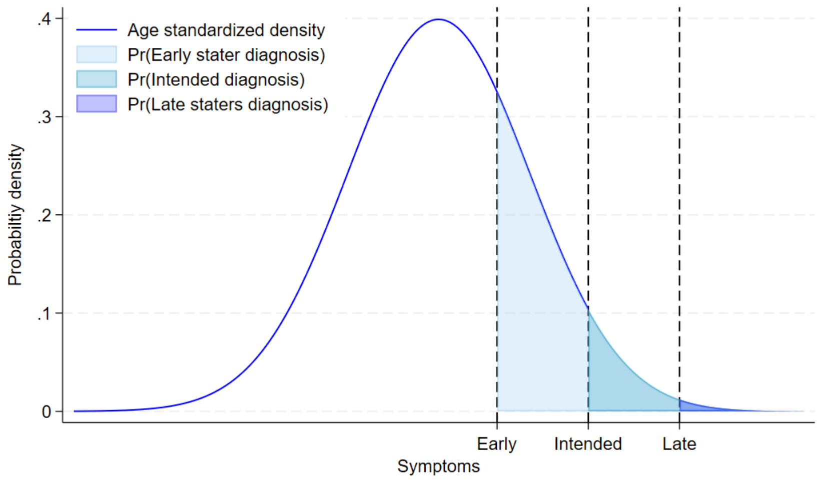
Notes: Each marker captures the percentage of children receiving a first-time prescription for ADHD by birth date in weeks relative to September 1. Panel (a) and (b) consider prescriptions for children aged 5-8 and 9-15, respectively. The vertical dashed line represents September 1 – the cut-off point for school entry.

Figure 9: Effect of Early Start of School on ADHD First-Time Prescriptions by Age and Grade



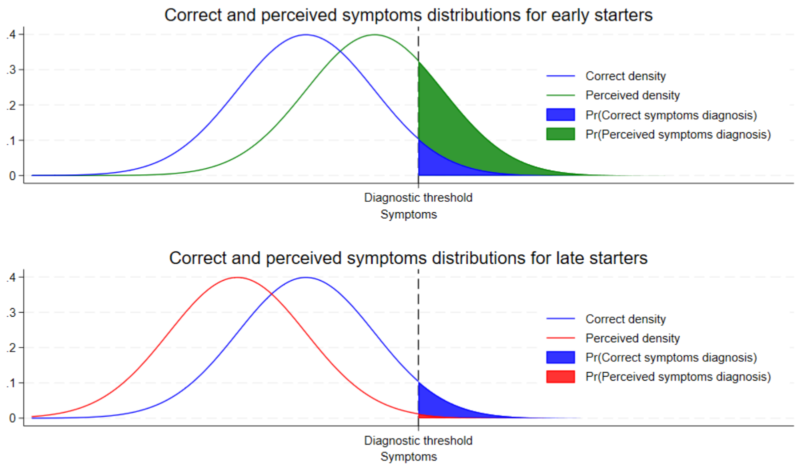
Notes: The figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 15 and grade from reception class to grade 10. The effects measure the part of the proportional increase in ADHD prescriptions explained by differences in first-time prescriptions between early and late starters (see 9). In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects.

Figure 10: Stylized Representation of the Distribution of Age-standardized ADHD Symptoms with Diagnostic Thresholds



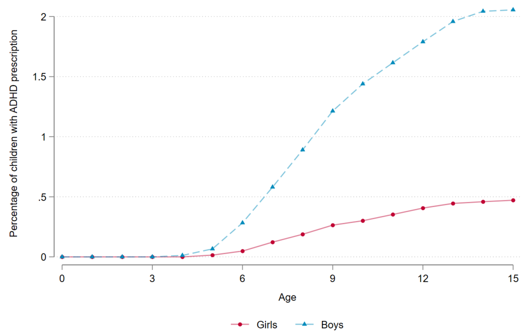
Notes: The vertical dashed lines represent the early-starters, the intended (age-standardized), and late-starters diagnostic thresholds potentially adopted by specialists.

Figure 11: Stylized Representation of the Distributions of Perceived and Age-standardized (Correct) ADHD Symptoms for Early and Late Starters

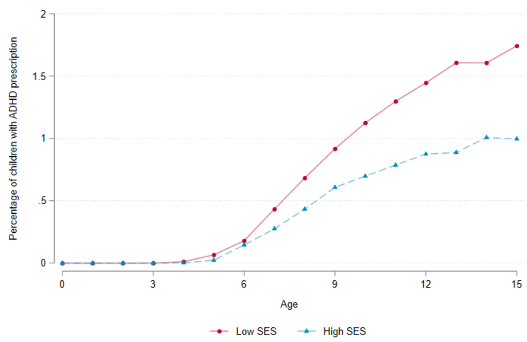


Notes: The vertical dashed lines represent the diagnostic thresholds teachers use.

Figure 12: Rate of ADHD Prescriptions by Age: Heterogeneity by Gender and Socioeconomic Background



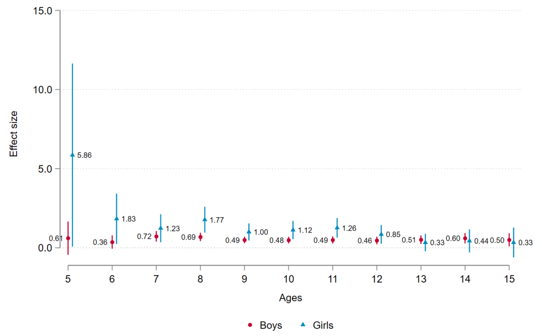
(a) By gender



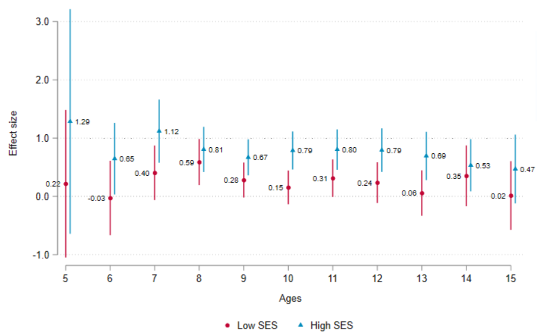
(b) By socioeconomic area background

Notes: Each marker captures the percentage of children receiving an ADHD prescription in the corresponding age period. Low (high) SES is defined as living in deprived (non-deprived) areas where the Townsend deprivation score is in the top (bottom) two quintiles.

Figure 13: Effect of Early Start of School on ADHD Prescriptions by Age: Heterogeneity by Gender and Socioeconomic Background



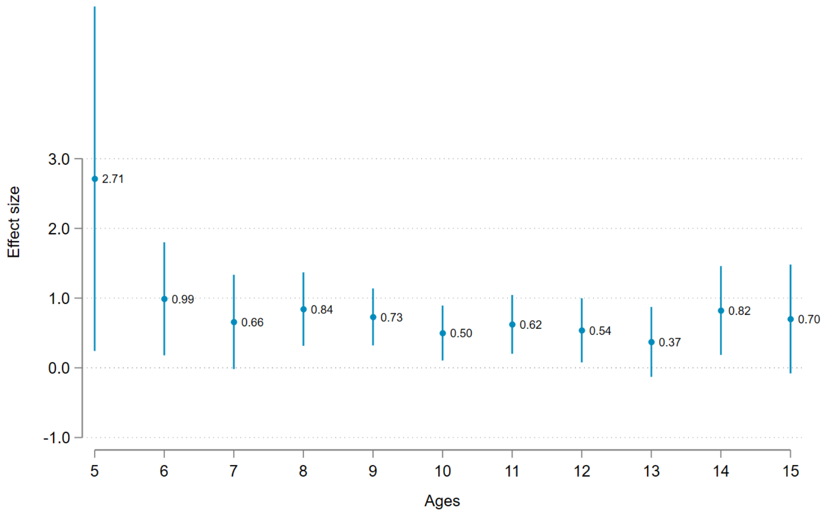
(a) By gender



(b) By socioeconomic area background

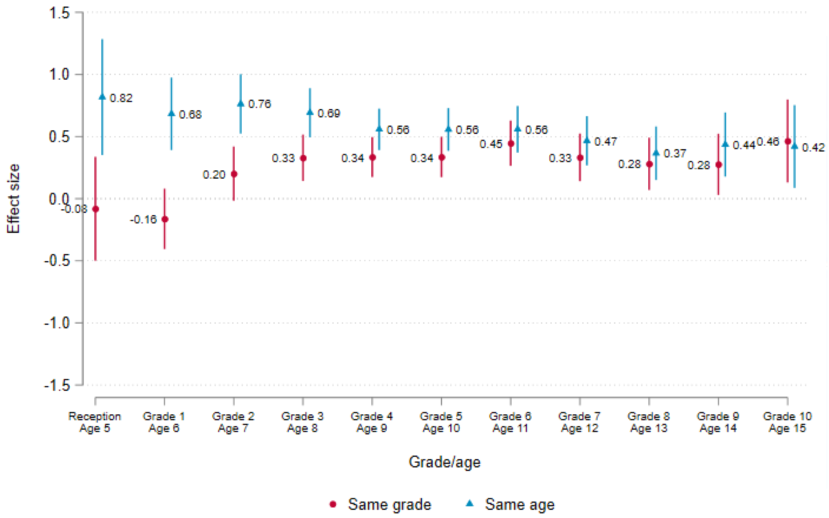
Notes: the figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 15. Low (high) SES is defined as living in deprived (non-deprived) areas where the Townsend deprivation score is in the top (bottom) two quintiles. The effects are expressed as a proportional increase in the prescription rate for early starters relative to late starters, $r_a = \gamma_a / Pr(ADHD_{i,a} = 1 | Early_i = 0)$. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects.

Figure 14: Effect of Early Start of School on ADHD Prescriptions by Age, Using a Regression Discontinuity Design



Notes: The figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 15. The effects are expressed as a proportional increase in the prescription rate for early starters relative to late starters, $r_a = \tilde{\gamma}_a / Pr(ADHD_{i,a} = 1 | Early_i = 0)$. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects.

Figure 15: Effect of Early Start of School on ADHD Diagnoses by Age and Grade



Notes: The figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 15 and grade from reception class to grade 10. The effects are expressed as a proportional increase in the diagnosis rate for early starters relative to late starters, $r_t = \gamma_t / Pr(ADHD_{i,t} = 1 | Early_i = 0)$ with $t = a$ for age and g for grade. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects.

Tables

Table 1: Number of Observations by Age

(1) Age	(2) Observations
5	96,698
6	96,698
7	96,698
8	96,698
9	96,698
10	80,805
11	66,434
12	52,599
13	39,290
14	27,446
15	17,384

Table 2: Summary Statistics, By School Starting Age

	(1) Full sample	(2) Early starters	(3) Late starters	(4) Difference
Sex				
Male	51.1	50.9	51.3	-0.4
Female	48.9	49.1	48.7	0.4
Ethnicity				
White	54.6	54.8	54.3	0.6*
Non-white	14.7	14.9	14.6	0.3
Unclassified	30.7	30.3	31.1	-0.9***
Deprivation index in quintiles				
Q1 (least deprived)	28.4	27.8	28.9	-1.1***
Q2	24.9	25.0	24.8	0.1
Q3	20.2	20.2	20.2	0.0
Q4	16.1	16.3	15.8	0.5**
Q5 (most deprived)	10.4	10.7	10.2	0.5**
Maternal age at birth				
$\leq$20	7.5	7.4	7.6	-0.2
21-34	74.4	74.3	74.6	-0.3
35-39	15.6	15.7	15.4	0.3
$\geq$40	2.5	2.6	2.4	0.2**
ADHD rates in percentage points				
Average ADHD diagnosis	0.5	0.6	0.4	0.2***
Average ADHD prescription	0.4	0.5	0.3	0.2***
N	96,698	49,157	47,541	96,698

Notes: All variables are dummies, and we report the average expressed in percentages for the full sample in column 1, for early starters in column 2, for late starters in column 4, and the difference in the average between early and late starters in column 4. * $p<0.10$, ** $p<0.05$, *** $p<0.001$. The averages of the dummies for ADHD diagnosis and prescription are again expressed in percentage and computed by pooling together children observed from 5 to 15 years old.

Table 3: Testing Differences Between Same Age and Same Grade Estimates on the Effect Early Start of School on ADHD Prescriptions

Age/grade	(1) Same age	(2) Same grade	(3) p-value
Age 5/Reception	1.086	-0.609	0.058
	(0.55)	(0.4)	
Age 6/Grade 1	0.509	-0.522	0.037
	(0.21)	(0.17)	
Age 7/Grade 2	0.787	-0.112	0.036
	(0.16)	(0.13)	
Age 8/Grade 3	0.823	0.301	0.052
	(0.13)	(0.12)	
Age 9/Grade 4	0.551	0.266	0.077
	(0.1)	(0.09)	
Age 10/Grade 5	0.549	0.304	0.097
	(0.1)	(0.09)	
Age 11/Grade 6	0.588	0.422	0.141
	(0.11)	(0.1)	
Age 12/Grade 7	0.492	0.379	0.226
	(0.11)	(0.11)	
Age 13/Grade 8	0.412	0.307	0.253
	(0.12)	(0.12)	
Age 14/Grade 9	0.505	0.473	0.602
	(0.15)	(0.15)	
Age 15/Grade 10	0.398	0.494	0.379
	(0.19)	(0.19)	

Notes: The table reports the point estimates and standard errors of separate regression, one for each age from 5 to 15 (column 1) and grade from reception class to grade 10 (column 2). The estimates and standard errors are expressed as a proportional increase in the diagnosis rate for early starters relative to late starters, $r_t = \gamma_t / Pr(ADHD_{i,t} = 1 | Early_i = 0)$ with $t = a$ for age and g for grade. These estimates are the same as the ones reported in Figure 6. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects. In column 3, we report the p-values of a test of differences in the results of columns 1 and 2, with the null hypothesis assuming no difference between these estimates.

Table 4: Testing Differences Between Same Age and Same Grade Estimates on the Effect Early Start of School on Anxiety Prescriptions

Age/grade	(1) Same age	(2) Same grade	(3) p-value
Age 5/Reception class	0.141	-0.368	0.077
	0.2	0.18	
Age 6/Grade 1	0.51	-0.14	0.055
	0.18	0.15	
Age 7/Grade 2	0.271	0.018	0.114
	0.13	0.13	
Age 8/Grade 3	0.267	-0.022	0.087
	0.11	0.11	
Age 9/Grade 4	0.123	-0.02	0.155
	0.1	0.09	
Age 10/Grade 5	0.184	-0.057	0.094
	0.1	0.09	
Age 11/Grade 6	0.184	0.006	0.131
	0.1	0.1	
Age 12/Grade 7	-0.007	-0.102	0.242
	0.1	0.1	
Age 13/Grade 8	-0.031	-0.127	0.268
	0.11	0.11	
Age 14/Grade 9	-0.087	-0.09	0.985
	0.13	0.13	
Age 15/Grade 10	0.011	-0.137	0.236
	0.18	0.17	

Notes: The table reports the point estimates and standard errors of separate regression, one for each age from 5 to 15 (column 1) and grade from reception class to grade 10 (column 2). The estimates and standard errors are expressed as a proportional increase in the diagnosis rate for early starters relative to late starters, $r_t = \gamma_t / Pr(Anxiety_{i,t} = 1 | Early_i = 0)$ with $t = a$ for age and g for grade. These estimates are the same as the ones reported in Figure 7a. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects. In column 3, we report the p-values of a test of differences in the results of columns 1 and 2, with the null hypothesis assuming no difference between these estimates.

Table 5: Testing Differences Between Same Age and Same Grade Estimates on the Effect Early Start of School on Anxiety Diagnoses

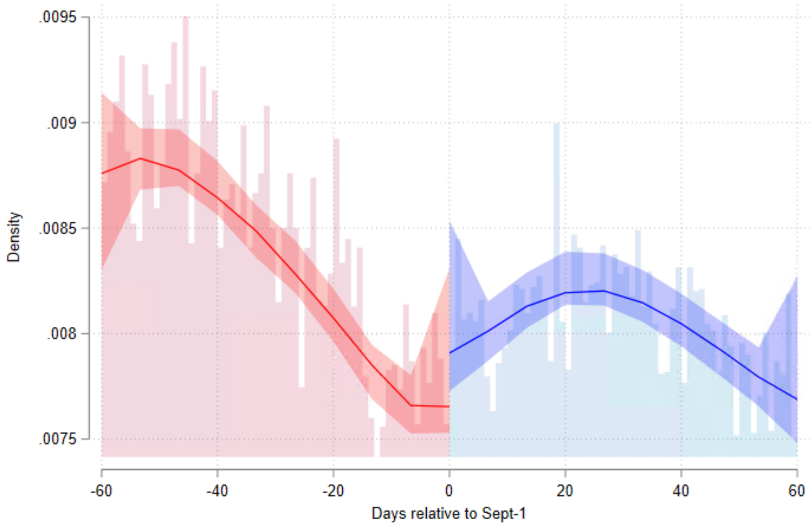
Age/grade	(1) Same age	(2) Same grade	(3) p-value
Age 5/Reception class	0.264	-0.121	0.113
	0.26	0.24	
Age 6/Grade 1	-0.052	-0.205	0.215
	0.19	0.18	
Age 7/Grade 2	0.108	-0.366	0.059
	0.16	0.14	
Age 8/Grade 3	-0.065	-0.218	0.14
	0.12	0.12	
Age 9/Grade 4	-0.025	-0.286	0.065
	0.11	0.1	
Age 10/Grade 5	0.117	-0.23	0.05
	0.1	0.1	
Age 11/Grade 6	0.357	-0.068	0.04
	0.11	0.1	
Age 12/Grade 7	-0.082	-0.195	0.154
	0.09	0.09	
Age 13/Grade 8	-0.124	-0.394	0.06
	0.08	0.08	
Age 14/Grade 9	0.04	-0.336	0.043
	0.09	0.08	
Age 15/Grade 10	0.074	-0.246	0.054
	0.11	0.1	

Notes: The table reports the point estimates and standard errors of separate regression, one for each age from 5 to 15 (column 1) and grade from reception class to grade 10 (column 2). The estimates and standard errors are expressed as a proportional increase in the diagnosis rate for early starters relative to late starters, $r_t = \gamma_t / Pr(Anxiety_{i,t} = 1 | Early_i = 0)$ with $t = a$ for age and g for grade. These estimates are the same as the ones reported in Figure 7b. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects. In column 3, we report the p-values of a test of differences in the results of columns 1 and 2, with the null hypothesis assuming no difference between these estimates.

Appendix

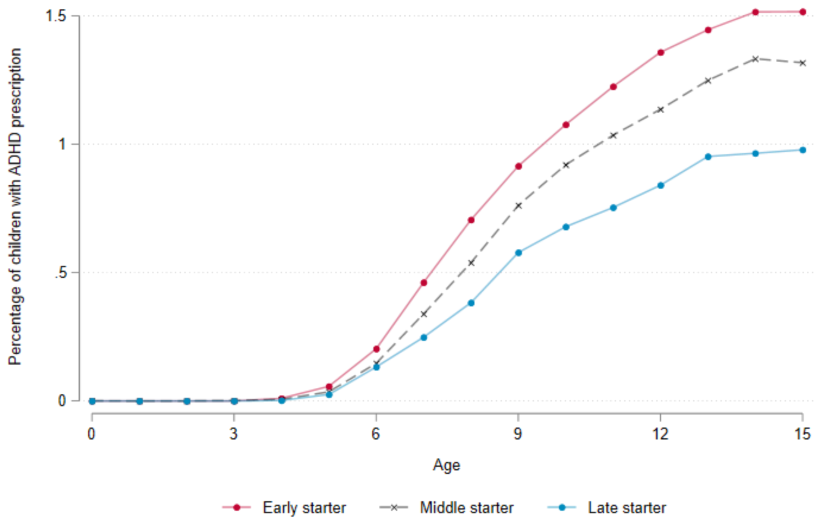
Figures

Figure A.1: Manipulation Testing: Estimated Density of the Running Variable



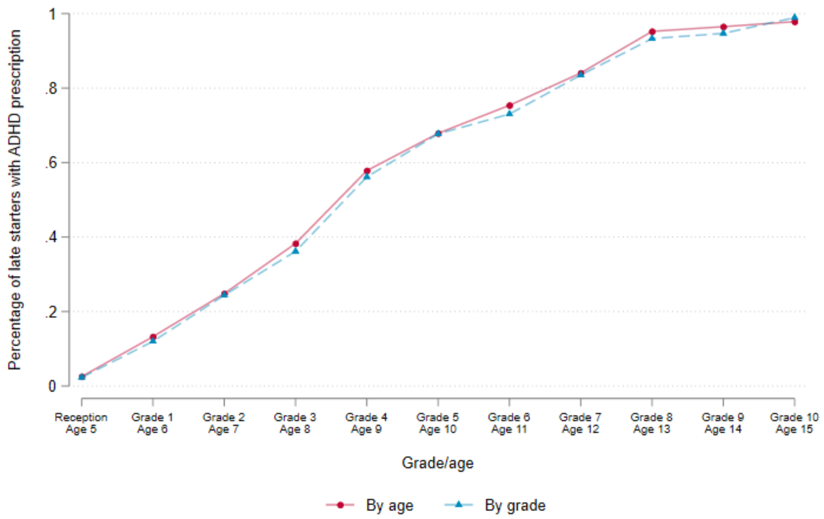
Notes: The horizontal axis depicts the date of birth relative to the cut-off point at school entry of September 1. The solid lines depict the local polynomial density estimate (blue and red on each side of the September 1 cutoff). The shaded areas capture the robust bias-corrected confidence intervals. The histogram of the running variable, the children's birth date in days centered around the cutoff for school entry, is shown in the background.

Figure A.2: Rate of ADHD Prescriptions by Age



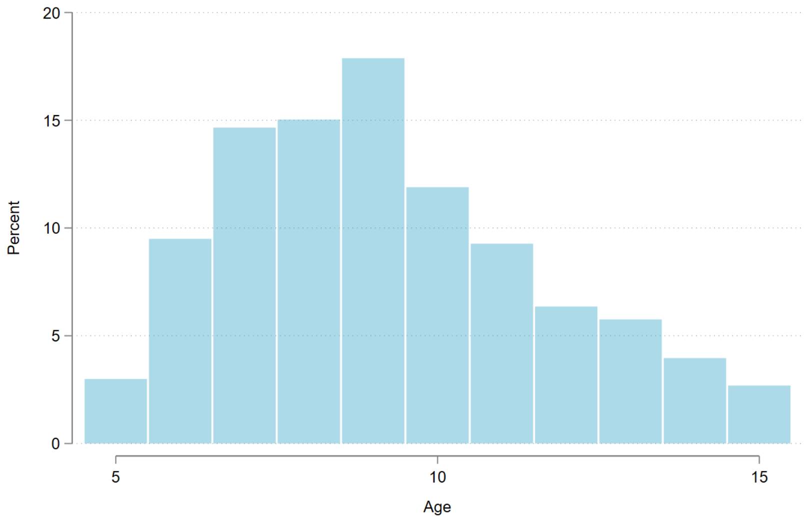
Notes: Each marker captures the percentage of children receiving an ADHD prescription in the corresponding age period. The figure complements Figure 1b, which considers early starters (born in July and August) and late starters (born in September and October) separately and includes middle starters (born between November and June. As reported in Table 2, our main sample of early and late starters consists of $N = 96,698$ individuals (49,157 early starters and 47,541 late starters). Additionally, the figure above includes $N = 196,938$ middle starters, who are not part of our main analyses.

Figure A.3: Rate of ADHD Prescriptions for Late Starters by Age and Grade



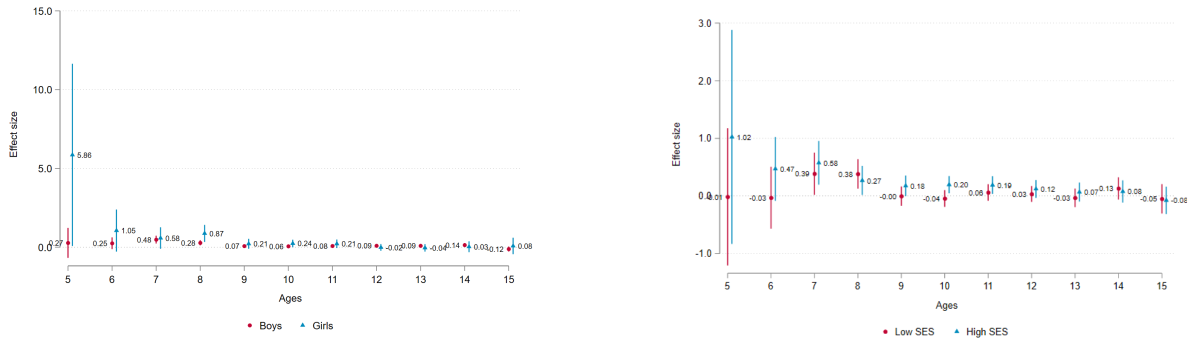
Notes: Each marker captures the percentage of late starters (i.e., born in September and October) receiving an ADHD prescription in the corresponding age (from ages 5 to 15) and grade period (from reception class to grade 10).

Figure A.4: Histogram of the Age at Which Children Receive Their First ADHD Prescription



Notes: To construct this figure, we consider a sub-sample of children who received an ADHD prescription between ages 5 and 15. We plot the distribution of the age at which these children receive their first prescription.

Figure A.5: Effect of Early Start of School on ADHD First-Time Prescriptions by Age: Heterogeneity by Gender and Socioeconomic Background

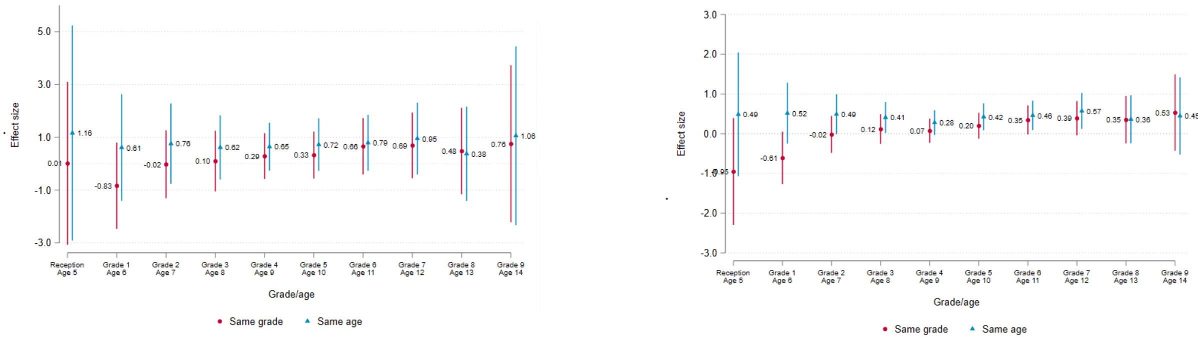


(a) By gender

(b) By socioeconomic area background

Notes: the figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 15. Low (high) SES is defined as living in deprived (non-deprived) areas where the Townsend deprivation score is in the top (bottom) two quintiles. The effects measure the part of the proportional increase in ADHD prescriptions explained by differences in first-time prescriptions between early and late starters (see 9). In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects.

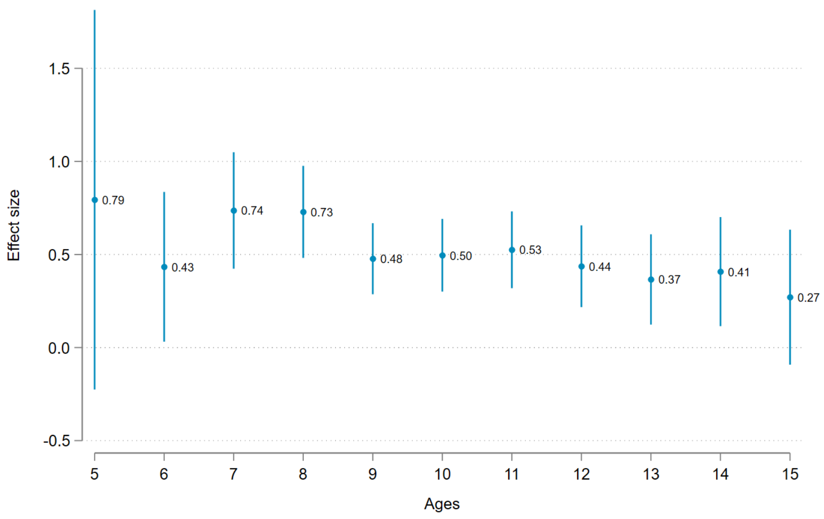
Figure A.6: Effect of Early Start of School on ADHD Prescriptions by Age, Including Maternal Fixed Effects



(a) Bandwidth = 60 days **(b)** Bandwidth = 120 days

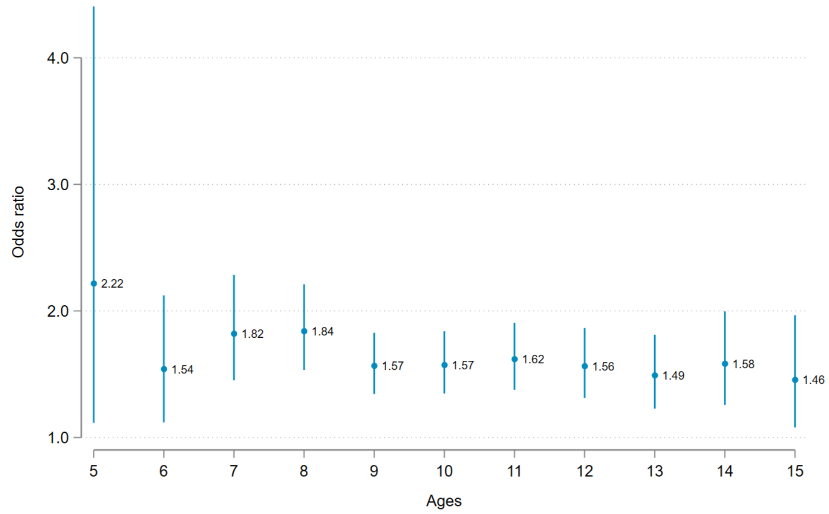
Notes: The figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 15 and grade from reception class to grade 10. The effects are expressed as a proportional increase in the prescription rate for early starters relative to late starters, $r_t = \gamma_t / Pr(ADHD_{i,t} = 1 | Early_i = 0)$ with $t = a$ for age and g for grade. In each regression, we control for sex and include mother and year of birth fixed effects. Panel (a) uses all sibling combinations in a sample of children born within 60 days of the September 1 cutoff date. In Panel (b), we expand this bandwidth to 120 days at either side of the September 1 cutoff point. The effective sample sizes are reported in Table A.1.

Figure A.7: Effect of Early Start of School on ADHD Prescriptions by Age, Excluding Children Born Around September 1



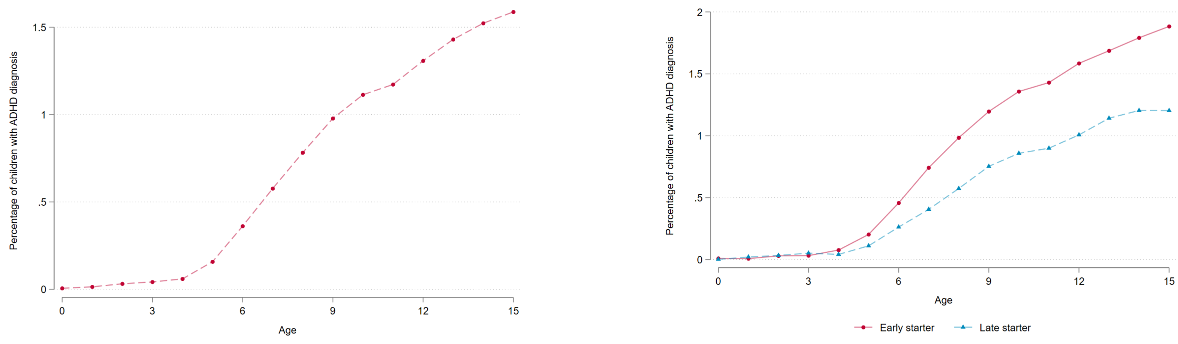
Notes: The figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 15. The effects are expressed as a proportional increase in the prescription rate for early starters relative to late starters, $r_a = \gamma_a / Pr(ADHD_{i,a} = 1 | Early_i = 0)$. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects. Compared to the results in Figure 4, we exclude children born within +/− 3 days of the September 1 threshold.

Figure A.8: Effect of Early Start of School on ADHD Prescriptions by Age, Using a Logistic Model



Notes: The figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 15. The effects are expressed as odds ratios. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects.

Figure A.9: Average Rate of ADHD Diagnoses by Age

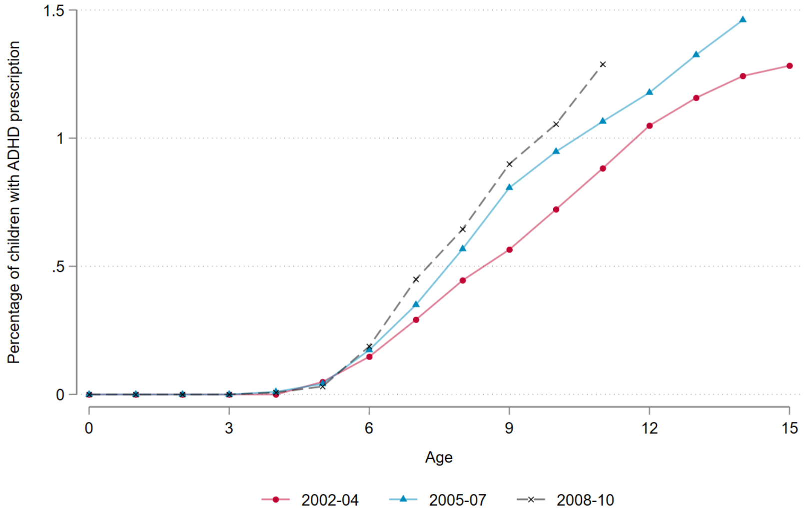


(a) Pooled sample

(b) By school starting age

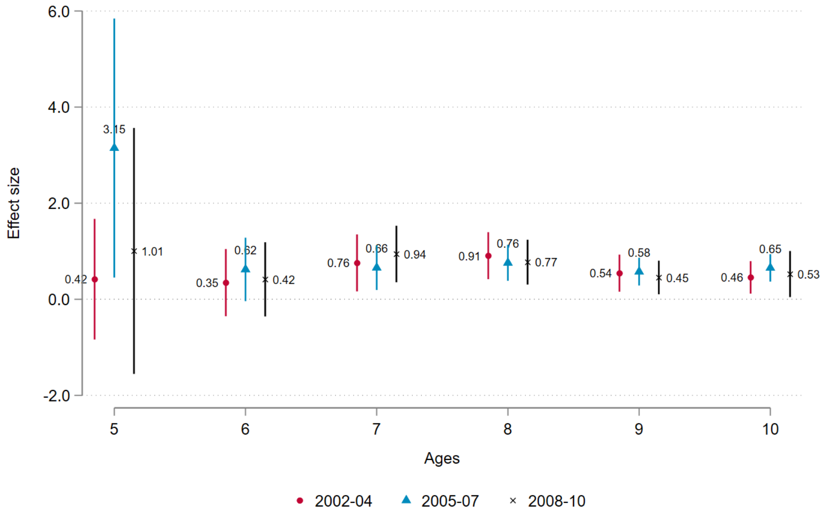
Notes: Each marker captures the percentage of children diagnosed with ADHD in the corresponding age period.

Figure A.10: Rate of ADHD Prescriptions by Cohort



Notes: Each marker captures the percentage of children receiving an ADHD prescription in the corresponding age period. Since we only observe children until 2020, children born in later cohorts (in blue and black) cannot be tracked until age 15.

Figure A.11: Effect of Early Start of School on ADHD Prescriptions by Age and Cohort



Notes: The figure plots point estimates and 95% confidence intervals of separate regressions, one for each age from 5 to 10 and separately for three different birth cohorts. The effects are expressed as a proportional increase in the prescription rate for early starters relative to late starters, $r_a = \gamma_a / Pr(ADHD_{i,a} = 1 | Early_i = 0)$. In each regression, we control for sex, ethnicity, postal code SES, and maternal age at birth, and include general practice and year of birth fixed effects.

Tables

Table A.1: Number of Observations by Age When Including Mother Fixed Effects

(1) Age	(2) Observations with bandwidth = 60	(3) Observations with bandwidth = 120
5	14,163	53,410
6	14,163	53,410
7	14,163	53,410
8	14,163	53,410
9	14,163	53,410
10	10,802	41,194
11	7,825	30,432
12	5,239	20,642
13	3,061	12,434
14	1,429	6,119
15	401	1,886