Antidepressant Treatment in Childhood*

Sonia Bhalotra
University of Warwick

N. Meltem Daysal
University of Copenhagen

Mircea Trandafir
Rockwool Foundation Research Unit

July 2026

Abstract

Mental health disorders tend to emerge in childhood, with half starting by age 14. This makes early intervention important, but treatment rates are low, and antidepressant treatment for children remains controversial since an FDA warning in 2004 that highlighted adverse effects. Linking individuals across Danish administrative registers, we provide some of the first evidence of impacts of antidepressant treatment in childhood on objectively measured mental health indicators and economic outcomes over time, and the first attempt to investigate under- vs over-treatment. Leveraging conditional random assignment of patients to psychiatrists with different prescribing tendencies, we find that treatment during ages 8-15 improves test scores at age 16, particularly in Math, increases enrollment in post-compulsory education at age 18, and that it leads to higher employment and earnings and lower welfare dependence at ages 25–30. We demonstrate, on average, a reduction in suicide attempts, self-harm, and hospital visits following AD initiation. The gains to treatment are, in general, larger for low SES children, but they are less likely to be treated. Using a marginal treatment effects framework and Math scores as the focal outcome, we show positive returns to treatment among the untreated. Policy simulations confirm that expanding treatment among low SES children (and boys) generates substantial net benefits, consistent with under-treatment in these groups. Our findings underscore the potential of early mental health treatment to improve longer term economic outcomes and reducing inequality.

Keywords: Antidepressants, mental health, education, test scores, human capital, Denmark, physician leniency, marginal treatment effects

JEL Classifications: I11, I12, I18, J13

*Bhalotra: Department of Economics, University of Warwick, CAGE, IFS, CEPR, IEA, IZA (email: sonia.bhalotra@warwick.ac.uk); Daysal: Department of Economics, University of Copenhagen, CEBI, CESifo, IZA (email: meltem.daysal@econ.ku.dk); Trandafir: Rockwool Foundation Research Unit and IZA (email: mt@rff.dk). We are grateful to Thomas Cornelissen for many insightful discussions, and to Angelina Nazarova for outstanding research assistance. We received helpful comments from Richard Blundell, Joe Doyle, David Jaeger, James Fenske, Eric French, Torben Heien Nielsen, Martin Karlsson, Therese Nilsson, Niels Skipper and participants at several seminars and workshops. The research was conducted with financial support from the Danish Foundation for Independent Research under grant 4182-00231 and the ESRC-funded CAGE Centre at Warwick (under grant ES/Z504701/1) where Bhalotra and Daysal are co-investigators. The activities of CEBI are financed by the Danish National Research Foundation (Grant DNRF134).

1 Introduction

Global estimates indicate that about 15% of children aged 10–19 suffer from a mental health illness (MHI).[1] Most MHI emerge in childhood, with half starting by age 14 and two-thirds by age 24 (Kessler et al., 2005). If left untreated, they tend to persist (Kessler and Bromet, 2013), highlighting the need for early intervention. Yet, treatment rates are low: about half of those in need do not receive care (Moitra et al., 2022; Evans-Lacko et al., 2018). Pharmacological treatment for children, in particular, has been controversial since an FDA warning in 2004 that highlighted suicidal ideation as a potential side effect of antidepressant use (Boaden et al., 2020). We examine the effects of antidepressant (AD) treatment during childhood and adolescence on short, medium and longer term economic outcomes. We provide some of the first evidence of impacts of AD treatment in childhood on economic outcomes over time, and objectively measured mental health indicators, and the first attempt to investigate under- vs over-treatment with AD.

We address two key challenges in identifying the impact of AD treatment during childhood on later outcomes. First, we leverage longitudinal, individual-level data from multiple Danish registers to track the onset of mental health conditions, treatment decisions, and the progression of multiple health, cognitive, and economic outcomes over time. Our analysis focuses on children diagnosed with a psychiatric condition, defined as those referred by their general practitioner to a child psychiatrist at any point between the ages of 8 and 15. We investigate antidepressant (SSRI) use, which is predominantly prescribed in the treatment of depression and anxiety (Jack et al., 2020; Schröder et al., 2017). These conditions are linked to cognitive and behavioral impairments, including reduced concentration, memory distortions, and altered economic preferences, beliefs, and marginal utility (Ridley et al., 2020; de Quidt and Haushofer, 2016). Given this evidence, our primary outcome is performance in national academic tests administered at age 16. We track our sample cohorts to examine the cumulative effects of AD treatment in childhood on educational enrollment at age 18, and on education, employment, earnings, and welfare dependence at ages 25 to 30. We also report indicators of psychiatric and somatic health in the three months following the initiation of AD treatment, offering objective evidence of treatment effectiveness.

Second, we address the potential bias arising from AD use being correlated with unobserved factors that also affect outcomes. We implement a 2SLS strategy that relies on quasi-random variation in prescribing rates across child psychiatrists.[2] In particular, we use the leave-one-out prescribing propensity of the child's first psychiatrist as an instrument for AD use. Given that virtually all children in our sample are initially treated with psychotherapy, our estimates identify the additional

[1] WHO factsheet October 2024, available here

[2] In Denmark, while general practitioners may prescribe AD to adults, children are typically referred to psychiatrists for this. Of all children age 8-15, 6.8% see a psychiatrist.

impact of AD treatment.

There is substantial variation across psychiatrists in their AD prescribing propensity, and the instrument strongly predicts AD use: moving from a psychiatrist at the 25th percentile of the prescribing propensity distribution to one at the 75th percentile increases the likelihood of AD use by 7.9 percentage points (52% of the mean). We show that the prescribing tendency is uncorrelated with a comprehensive set of observables related to the outcomes. We also present evidence supporting the validity of the exclusion restriction and the monotonicity assumption. Our 2SLS estimates identify the local average treatment effect (LATE) for compliers—children who would not have received AD treatment had they seen a psychiatrist with a lower prescribing rate. Compliers are marginal cases, a relevant subpopulation for policy, and the group for whom concerns about over- or under-prescribing may be most relevant.

Our main findings are as follows. We estimate that AD use by age 15 is associated with large increases in test scores at age 16, with more precisely estimated effects for Math than for Danish. To examine the impact of outliers, we use the Kaplan and Sun (2017) estimator to show that treatment effects are fairly uniform across the test score distribution. We conjecture that these sizable impacts reflect early treatment disrupting the mutually reinforcing dynamic between poor mental health and poor school performance, the effects of which accumulate over time. Such impacts likely include improvements in attention, motivation, and skill accumulation.[3] The estimated effects completely close the baseline mean gap in test scores associated with having a diagnosed mental health condition.

We examine heterogeneity in treatment effects by gender and socioeconomic status (SES), defining children whose mothers have higher education as high SES. We find that AD treatment has similar-sized impacts on the tests scores of girls and boys,[4] and much larger impacts for low-SES than for high SES children. We can reject that the SES gradient is explained by low SES children being more likely to see psychiatrists with lower prescribing rates. It is consistent with mothers with higher education being better able to shield their children's school performance from the effects of mental illness. Notably, higher returns to treatment among low SES children contrast with their lower rates of receiving treatment.[5] We explore this further in the policy simulations discussed below.

We examine heterogeneity in treatment effects by unobservables that determine selection into AD use by estimating marginal treatment effects (MTE). Our findings indicate positive selection into treatment (sorting on gains to treatment), though this is weaker among low SES children, con-

[3]One dynamic path is that poor mental health undermines school performance, and poor school performance, in turn, worsens mental health. Another is state dependence or self-productivity in each domain: untreated mental health issues tend to persist (Kessler and Bromet, 2013), and because learning is cumulative, children who fall behind academically often struggle to catch up (Heckman, 2007).

[4]The implied population-level impact is larger for girls because more girls than boys receive AD, and girls exhibit higher rates of depression and anxiety starting around age 12 (pubertal years).

[5]Children from low SES families are more likely to experience mental health issues, less likely to have a psychiatric visit, and, conditional on having a visit, less likely to receive AD.

sistent with information frictions in this group. For Math scores, the MTE is almost always positive and the average treatment effect among children who do not use AD (ATU) is positive. If the policy objective were to improve performance in age-16 Math tests, or if improvements in mental health were proxied by gains in Math scores, these results would suggest that children in our sample are under-treated with AD. However, for Danish scores, the MTE is negative for some children, and the average effect among untreated children is negative, albeit small. These contrasting patterns complicate drawing unqualified policy conclusions.

We therefore use the MTE framework to conduct policy simulations. We simulate a policy that makes psychiatrists "identity blind." Specifically, we increase the prescribing rate for low SES children to equal that for high SES children and, subsequently, we increase the rate for boys to equal that for girls. In both cases we find that children pushed into treatment tend to have high test score returns to treatment in both Math and Danish, which is consistent with under-treatment in these groups. Allowing that under-treatment of a group of children might occur alongside over-treatment of another group, we also conduct the reverse simulation: bringing the prescribing rate of high SES children and of girls down to that of low SES children and boys, respectively. We find in both cases that this harms more children than it benefits. Next, in light of public debates concerning "rules vs discretion" in medical guidelines, we simulate a reduction in discretion among psychiatrists by compressing the distribution of prescribing rates. First, we push up the bottom 25% of the prescribing propensity distribution and we find positive test score effects in both Math and Danish for virtually all of the children affected. Then we push down the top 25% and we find that this hurts much more children than it benefits. This, again, is evidence consistent with under-treatment being a better characterization of prescribing behaviour in our sample than over-treatment.

In the last part of the paper, we examine the effects of AD use on a broader set of outcomes. Amid continued debate over medicating mental health conditions in children, we investigate the health effects of AD use. AD use in children has long faced medical and regulatory scrutiny, especially following concerns in the early 2000s about increased risks of suicidal thoughts and behaviors. In response, the U.S. FDA issued a "black box" warning in 2004, and the European Medicines Agency followed with similar warnings in 2005. However, later research has questioned the analyses underlying these actions and argued that the risks of untreated mental health conditions may be greater than those associated with treatment (Fornaro et al., 2019; Cipriani et al., 2018; Lu et al., 2014; Rihmer, 2007). Using more objective administrative data and larger samples than most existing studies, we find that AD treatment is associated with reductions in attempted suicide and self-harm, as well as an overall decline in hospital visits, including ER visits. This pattern is consistent with treatment improving mental health—a mechanism that likely underpins the identified impacts on economic outcomes.[6]

[6]Our findings are consistent with evidence that adverse effects have not been identified in the more recent classes of SSRI that are prescribed in Denmark (Boaden et al., 2020). We cannot rule out that a small fraction of AD users

We then turn to the effects on post-compulsory educational attainment and labor market performance. We find that AD treatment in childhood significantly increases the likelihood of being enrolled in education at age 18 across gender and SES groups, indicating impacts beyond test scores. Disaggregated results show that AD treatment raises academic track enrollment among girls and high-SES children and vocational track enrollment among boys and low-SES children. Following our cohorts to age 25-30, we find that those treated with AD in childhood are more likely to be employed, have higher labor and total income, and are less likely to receive welfare or disability benefits, with larger gains among low SES individuals and men. These results highlight the potential of improved mental health treatment in childhood to enhance longer term educational and economic trajectories.

We subject our primary outcome to several tests. We mitigate concerns over the exclusion restriction by controlling for observed characteristics and behaviours of psychiatrists. In particular, we show that our estimates of AD treatment effects on test scores are not driven by psychotherapy (which has a positive direct effect) or by other mental health medications (which have a negative direct effect). However, one might still worry that high prescribers are higher quality psychiatrists.[7] In fact we see no reason why this should be the case. Moreover, there are two reasons that we might imagine that high prescribers are in fact lower quality psychiatrists. First, if the consequences of under-treatment are more severe than the consequences of over-treatment and if low quality providers are more uncertain, they will tend to over-treat (Chan et al., 2022). Second, if lower quality psychiatrists feel less capable of providing psychotherapy or other forms of non-pharmaceutical support, they may be more likely to resort to drug treatment. Nevertheless, we bound the IV estimates to allow for failure of the exclusion restriction, using outside evidence to suggest that it is small relative to the primary effect (Conley et al., 2012).

We conduct numerous other tests on the instrument, examining potential endogenous mobility across psychiatrists, sensitivity to clinic size, and the impact of excluding outliers in prescribing behavior. To assess the external validity of the LATE estimates, we estimate the share of compliers. At 33%, this share is fairly large. We also find that the compliers are broadly similar to the analysis sample in their demographic and SES characteristics. We also show that our LATE estimates are robust to potentially endogenous test-taking, to excluding children who took tests during the pandemic, to controlling for school or individual fixed effects, to relaxing sample restrictions, and to modeling treatment intensity. Finally, we test the robustness of the MTE analysis, including checks of pairwise monotonicity (Frandsen et al., 2023), sensitivity to functional form and first-stage propensity score estimation (Cornelissen et al., 2018), and to including higher order polynomials of a covariate

experience side effects, including increased suicidality.

[7] Define quality as an unobservable trait of the psychiatrist that is predictive of the primary outcome, school performance of the child, underlying which is improved mental health.

index (Devereux, 2022). While the main analysis focuses on the 2SLS and MTE estimates, we also provide reduced form estimates.

Related literature and contributions We now delineate our contributions relative to the literature. Our main contribution is to provide causal evidence of antidepressant treatment of school-age children on their future educational and economic outcomes. There is a growing literature on the economics of child mental health, most of which examines ADHD and uses sibling fixed effects to address omitted variable bias (see, e.g., Currie et al., 2010; Currie and Stabile, 2009; Fletcher and Wolfe, 2008; Currie and Stabile, 2006). Our work is more closely related to research on the effectiveness of medical treatments.[8] Among studies that investigate AD treatment, the available evidence primarily looks at contemporaneous impacts on treating *adults* on their labor market outcomes[9], and the results are mixed.[10] Our results are consistent with larger returns to early life intervention, in line with dynamic feedback loops that we elaborate in the next section. Although childhood and adolescence is a critical period for the emergence of depression and anxiety, there is limited causal evidence on the dynamic returns to mental health intervention in this period (Hendren and Sprung-Keyser, 2020; Almond et al., 2018).[11]

Second, we provide evidence on the effects of AD treatment using objective measures of health. Clinical trials typically report short-term impacts of AD use on self-reported symptoms (Currie, 2024; Cipriani et al., 2018). Using objective health measures is important given ongoing contro-

[8] As noted by Currie (2024), the strongest evidence to date again concerns ADHD treatments (Chorniy, 2025; Currie et al., 2014; Dalsgaard et al., 2014). Biasi et al. (Forthcoming) focuses on bipolar disorder, a more extreme condition, for which SSRI are not the first line of treatment. They leverage the adoption of lithium as a treatment, comparing differences in outcomes between people with bipolar disorder and their healthy siblings, for cohorts aged 20 before and after the treatment became available. For a recent review of interventions that improve child mental health, see Currie (2025).

[9] We are aware of two studies of AD use and test scores. Busch et al. (2014) show that the sharp aggregate drop in AD prescribing following the 2004 FDA warning coincided with a decline in grade point averages. Cowan and Hao (2021) however show that Medicaid expansions led to increased use of AD, with no improvements in academic outcomes. These studies do not look at mental health or future economic outcomes and the marginal patient in their analysis is not comparable to ours.

[10] Bütikofer et al. (2020) estimate the reduced-form effects of the FDA's 2007 expanded black box warning on antidepressants, finding no effects on young adults but reduced employment among women aged 35–49. Shapiro (2022) leverages variation driven by the borders of television markets and finds that exposure to direct-to-consumer advertising of AD significantly decreases missed days of work. Laird and Nielsen (2016) rely on the fact that the prescribing rate of primary care physicians may change following a residential move, which affects the likelihood of being prescribed AD, to estimate no significant effects in Denmark of adult AD use on labor market outcomes. Currie and Zwiers (Forthcoming) use Dutch data and instrument for AD use in the postpartum period with the propensity of the local practitioner to prescribe AD to women 46–65 years old. They find that AD use leads to lower employment rates in the first year after birth, with no other significant labor market effects. In a developing country context, Angelucci and Bennett (2024) find no impact of AD treatment for adults on labor market outcomes following a randomized controlled trial set in India, though Lund et al. (2024) report impacts on employment in developing countries. In their review of the literature, Ridley et al. (2020) highlight the paucity of evidence of impacts of depression treatment on earnings.

[11] A considerable body of evidence suggests large economic returns to intervening on early childhood health (Bhalotra and Venkataramani, 2015; Almond and Mazumder, 2011).

versy around AD treatment for children, especially concerns about (self-reported) side effects such as increased suicidal ideation in a small fraction of users (Boaden et al., 2020; Fornaro et al., 2019). Our findings offer new, more robust evidence that challenges these concerns about suicidality.[12]

Third, we provide insights into the debate over whether AD are over- or under-prescribed to children. Childhood and adolescence are critical periods of neurobiological development when trajectories can shift quickly in either direction (Dahl et al., 2018). Our finding that low-SES children in our sample are under-prescribed despite higher returns to treatment highlights mental illness as an important source of inequality in educational and economic outcomes.

Finally, our work is also related to the broader literature documenting how physician discretion affects health (Cuddy and Currie, Forthcoming; Cutler et al., 2019; Finkelstein et al., 2016). Treatment decisions in mental health are especially uncertain, as symptoms are inherently subjective and assessed by both patients and clinicians (Marquardt, 2021; Currie and MacLeod, 2020; Berndt et al., 2015). We show that child psychiatrists' discretion has substantial impacts on children's long-term economic outcomes, and our simulations indicate that tightening medical guidelines could improve these outcomes.

It is estimated that, in 2021, mental health disorders affected 13.9% of the world's population and accounted for 17.2% of total years of life lost on account of disability. Depression alone is the leading cause of ill health and disability, and among the most significant risk factors for teen suicide (Galaif et al., 2007). In this paper, we draw attention to the early origins of mental health illness in adulthood, and provide evidence of persistent impacts of treatment in childhood. Our results highlight the potential for treatment to alter economic trajectories and improve social mobility.

2 Background

2.1 Conceptual Framework

In this section we discuss the size of the burden of mental health illnesses in childhood and adolescence, evidence from neuroscience on brain changes at this life stage, evidence from medicine and psychology on the efficacy of antidepressant use, and reasons why we expect that mental health recovery may influence economic outcomes.

Childhood mental health Mental health disorders are the leading cause of childhood disability in high-income countries (Barican et al., 2022). Across the world, 15% of children aged 10 to 19 experience a mental health disorder (World Health Organization, 2022). Neuroscientific research has

[12] The closest related study is Ludwig et al. (2009), who use cross-country variation in AD sales to estimate that an increase of one pill per capita reduces suicides by 5%.

identified adolescence as a critical period of brain development: gray matter in the cerebral cortex (which enables individuals to control emotions) steadily declines while white matter (connecting different regions of the brain) increases (Gogtay et al., 2004; Giedd et al., 1999). These neurodevelopmental changes are thought to heighten emotional sensitivity and, in particular, vulnerability to negative emotional stimuli. Pre-clinical studies, for instance, demonstrate that chronic stress alters both the structure and functioning of the hypothalamic-pituitary-adrenal axis (Isgor et al., 2004), with downstream effects on cardiovascular health, immune function (e.g., inflammatory responses) and, critically for our focus, learning and memory (Stephens and Wand, 2012).

Effectiveness of antidepressants A comprehensive review of 522 clinical trials involving 116,477 adults with moderate to severe depression concludes that all antidepressants evaluated were more effective than placebo in reducing symptoms, with odds ratios ranging from 2.13 to 1.37 (Cipriani et al., 2018). A recent review focusing on children and adolescents highlights variation in efficacy across different antidepressant drugs (Boaden et al., 2020). The review refers to the continuing debate surrounding prescription of AD to young people. In relation to this, it reports insufficient data on tolerability (commonly measured by treatment discontinuation), but it provides some evidence that, among SSRIs, paroxetine may increase suicidality, whereas sertraline may reduce it.

However, several limitations in the existing evidence base should be noted. First, clinical trials typically rely on self-reported symptom reduction, measured using psychometric tools such as the PHQ-9. These measures are inherently subjective, influenced by placebo effects, and may themselves be endogenous to treatment. Second, most trials are based on relatively small sample sizes. Third, outcomes are generally assessed over short durations (usually no more than eight weeks) and thus provide limited insight into longer-term effectiveness. Fourth, the overall quality of the evidence is mixed.[13] Administrative data offer a valuable complement to clinical trials by enabling the evaluation of treatment effects using objective outcomes measured in large, representative populations over time. As we describe in Section 3, we will use register data to assess impacts of AD treatment on objective markers of health.

Dynamic effects of antidepressants on academic achievement Mental health disorders are associated with symptoms such as fatigue, disrupted sleep, and pessimism, all of which can impair concentration, distort memory, and reduce productivity. Beyond these cognitive and behavioral impairments, poor mental health can also distort economic preferences, beliefs, and marginal utility (Ridley et al., 2020; de Quidt and Haushofer, 2016). Moreover, there is state dependence in mental

[13] In their adult-focused review, Cipriani et al. (2018) rated 82% of the 522 trials as having a moderate to high risk of bias. They also noted that 78% of studies were funded by pharmaceutical companies, although they found no systematic differences in response or attrition rates based on funding source. In contrast, the review focused on children judged the included studies on children to be of moderate to high quality (Boaden et al., 2020).

health, with conditions tending to persist if left untreated. This may be reinforced by dynamic complementarities between mental health and economic outcomes as they evolve over the lifecourse. If treatment is effective in improving mental health, it has the potential to break these negative feedback loops. If left unaddressed, their cumulative longer term impacts on educational and labour market outcomes can be large.

2.2 Diagnosis and treatment of childhood mental health disorders

Denmark has universally subsidised health care, provided in a stepped system (Mathiesen et al., 2016). A child with a mental health condition will initially consult their general practitioner (GP) who can provide advice and prescribe psychotherapy but GPs tend not to prescribe antidepressant medication for children. Following guidelines of the Danish Health and Medicines Authority, GPs refer children at risk of a mental health disorder to secondary care. It is recommended that GPs refer mild cases to private practice child psychiatrists and more severe cases to hospitals with psychiatric departments. While direct costs are unlikely to deter mental health care utilization, patients face long waiting times due to a limited supply of providers. During 2008–2019, the mean (median) wait time for children after a GP referral was 41 (30) days.[14] In 2022, the Danish Health and Medicine Authority released a 10-year action plan to strengthen the provision of mental health services citing lack of availability and limited treatment offers among the biggest challenges to be addressed (Danish Health Authority, 2022).

For children who reach secondary care with depression, anxiety, or OCD, national guidelines recommend non-pharmacological interventions as the first-line of treatment (Sørensen et al., 2020). When pharmacological intervention is needed, the guidelines follow international protocols suggesting selective serotonin reuptake inhibitors (SSRI) as the first-line pharmacologic intervention (Dwyer and Bloch, 2019).[15] Pharmaceutical treatments are never recommended on their own but rather as a supplement to boost non-pharmacological interventions (Sørensen et al., 2020). Antidepressants are subsidised under the National Health Insurance Scheme, a 3 month supply costing around 70 DKK (10 USD).

2.3 Education system

Education in Denmark is compulsory at the primary level, which starts from the year the child turns six until they complete 9th grade. If their primary school offers it, children have the option of extending primary school by enrolling in a 10th grade- this is attended by roughly half of all children

[14] Available here, last accessed in June 2025.

[15] SSRIs are the most popular antidepressants due to their relatively few side effects, and they are the only class of antidepressants licensed for pediatric use (Pottegård et al., 2014). Over 96% of antidepressant prescriptions for children in our sample are SSRI, for which reason, we use the terms interchangeably.

completing 9th grade. Further education is voluntary and can be academic (three year high-school) or vocational (four-year trade school). Primary school exit examinations, performance in which is a primary outcome of our analysis are compulsory. All municipalities implement psycho-social guidance programs aimed at aiding parents and health care professionals in detecting student mental health issues, schools do not do referrals to GPs or offer mental health treatment.

3 Data

Using unique personal identifiers, we link individual-level records across several population-level administrative data sets from Denmark. Appendix Table 1 provides definitions of the variables used in the analysis, and their sources. A brief discussion is provided here.

Outcomes Our primary outcome is subject-specific test scores from 9th grade qualifying exams. We focus on grades in the written tests for Danish (reading) and mathematics, two subjects for which written tests are both mandatory for all students and graded blind by a school teacher and an external evaluator (who can overrule the teacher in case of grade discrepancies). These exams are high stakes as enrollment in further education (high school or vocational school) depends on their results.[16] In robustness checks, we examine longitudinal data including test scores from grades 2–8 (see Appendix C.1). We standardize all test scores within subject and test year.

We additionally examine enrollment beyond compulsory education (9 years during our sample period), at age 18, distinguishing the academic and vocational track. For the earlier birth cohorts in our sample, we study labor market outcomes and welfare dependency at ages 25–30. We also examine health outcomes. We define an indicator for suicide attempts and self-harm within 3 months of antidepressant initiation, as well as indicators for hospital contacts including ER, inpatient and outpatient visits in that window.

Antidepressant use The prescription register provides the exact date when the prescription was filled, the personal identifier of the person for whom the prescription was written, the identifier of the prescribing medical practice (the private clinic or the hospital department), the Anatomical Therapeutic Chemical (ATC) Classification of the medication, and the amount included in the prescription measured in defined daily doses (DDD). Importantly, we observe all filled AD prescriptions in Denmark, including those issued by providers outside the national public insurance network, starting from 1995. Using these data we define an indicator that takes on the value 1 if a child has ever filled a prescription for selective serotonin reuptake inhibitors (SSRI, defined as all

[16] Individual exceptions are made but there are no general exemptions for mental health conditions (see the law on mandatory final assessment, Folkeskoleloven Act No. 313 of 19 April 2006).

drugs within ATC-group N06AB) by age 15, typically the year before the exit exams in grade 9 (age 16).[17]

Covariates The administrative data link children to both parents, irrespective of whether the parents live together. We have a rich set of child and parent characteristics. For children we use indicators for gender, birth order, birth year, year of first psychiatric contact (which we refer to as year of diagnosis), municipality of residence in the year of diagnosis, and number of siblings. For each parent we control for age, marital/cohabitation status, employment status, income decile in the gender-specific full population, highest education degree, and their own SSRI use. Since we look at children age 8 onwards, we define this and all time-varying parental characteristics (other than age) at their value when the child is 6 years old. [18] We define household socioeconomic status (SES) as a binary variable which indicates high SES if the mother has higher education (some college and above) and low SES otherwise.

Treatments other than antidepressants We investigate the two other treatments that are of substantive relevance in our setting: non-SSRI psychotropic medications (defined as all drugs within the ATC class "N" except for N06AB) and psychotherapy. We utilize two indicators of use of therapy. The first indicates whether the child received any counseling services from GPs or psychologists during the year prior to their first psychiatrist visit. The second indicator captures psychotherapy treatment following the child's first contact with the child psychiatrist, and this includes therapy provided by psychiatrists, GPs, and psychologists. The information comes from medical claims submitted by providers in the national public insurance network, but therapy can be sought at clinics outside this network. There are no systematic records of this, so the measures of therapy will understate the amount used.

Psychiatrist characteristics In investigating the exclusion restriction, we use a register that records the characteristics of the psychiatrist. This includes their age, experience, gender, marital status, and whether they practice independently or with another psychiatrist in the clinic. As many as 80% of clinics have one psychiatrist. For multi-provider clinics, the characteristics represent the average age and experience across psychiatrists, and indicators for whether all psychiatrists are female or married.

Analysis sample We begin with the universe of 839,523 native children born between 1991–2005 who are observed every year between the ages of 6 and 16. We restrict our analysis to native children,

[17] A recent meta analysis of AD trials identifies paroxetine as an SSRI with elevated suicidality risk. It is notable that fewer than 1% of SSRI prescriptions are for paroxetine, 1.9% in the analysis sample, and 0.6% in the hospital sample.

[18] Parental data are missing for a small number of observations. In these cases, we assign the mode of each variable to the missing value.

defined as children whose parents are both born in Denmark, in order to reduce heterogeneity in attitudes toward mental health disorders and their treatment (Alonso et al., 2008). Because the prescription drug data start in 1995 and the outcomes of interest are realized at age 16 or later, there is a trade-off between the age at which we can start following children and the number of cohorts we can study. We choose to follow children starting from age 6 because that is the age when children are required to start school and teachers, school nurses, and school counselors may help identify children with mental health disorders.

We impose the following restrictions to arrive at our analysis sample. First, we focus on children who do not have a psychiatric visit at ages 6–7, but who do have such a visit between the ages of 8–15. This is because, in Denmark, antidepressants cannot be prescribed to children under the age of 8. As discussed in the previous section, the medical guidelines suggest that GPs refer children with suspected mental health disorders to psychiatrists. Close to 7% of children age 8-15 are referred to a child psychiatrist. We further restrict the sample to children whose first psychiatric visit is at a private clinic. Consultations at private clinics are fully reimbursed under Denmark's universal health coverage system. We focus on private clinics because the vast majority (80%) of these clinics have one psychiatrist, in contrast to psychiatric hospital departments that have several. In the clinic sample we can identify the psychiatrist and we can directly associate prescriptions with them through the claims register. We also have data on their characteristics. As a result, our instrument (discussed in the next section) is stronger in the clinic sample, and we are also in a stronger position to investigate the exclusion restriction.[19] However, since many more children are seen at hospitals, merging the hospital sample with the clinic sample significantly increases power. Therefore, in Section 5.3, we show that our results are robust to expanding the sample to include children whose first psychiatrist consultation is at a hospital. This is what we might expect if the hospital sample includes fewer marginal cases since, as we explain in the next section, our estimates are for marginal cases, defined as cases over which psychiatrists differ in whether they treat with AD.

To limit noise in our measure of specialist prescribing tendency, we require that the initial clinic treats a minimum of 50 children aged 8–15 during the year of the first consultation (which results in exclusion of 311 children). This leaves us with 7,211 children, approximately 78% of whom take the exit exams at the end of 9th grade. In Section 5.3, we investigate robustness of our results to different thresholds.

[19]Hospitals often have many psychiatrists in a department, and we have no information on the number, or their characteristics. The Danish registers simply come with this limitation for public sector providers. Hospitals experienced a process of consolidation over the analysis period, which resulted in even larger departments, and some churn. We can only associate prescriptions with the department and not with the individual psychiatrist. As a result, the instrument (which is detailed in the next section) is constructed as the average prescribing tendency at the department level, and it is weaker than the instrument in the clinic sample.

Descriptive statistics Descriptives for the analysis sample are in Table 1. Comparing children who have had no contact with a psychiatrist with children who see a child psychiatrist at a clinic (our primary analysis sample) (columns 1 and 2, Panel A), we observe a mental health penalty on test scores of 0.376 for Math and 0.273 for Danish. Within the analysis sample of children who see a psychiatrist, those on antidepressants (SSRI) exhibit better school performance than those who are not (columns 3 and 4). This holds across the distribution, see Appendix Figure A1. Our analysis will seek to identify the causal component of this descriptive difference.

Panel B of Table 1 reports means for all treatments that the children are on. Fewer than 10% of children receive more than one drug type at any time. Children in our sample who are on SSRI are also more likely to be on non-SSRI medications for mental health and to have been on therapy after their first contact with the psychiatrist.[20] In Section 5.2, we investigate these correlations, and argue that they do not bias our estimates of impacts of SSRI. The table shows that median days between the first contact with a psychiatrist and psychotherapy is 28 days, and children treated with AD are also more likely to have received psychological counseling *before* their first psychiatric contact. Median time to SSRI prescription is 106.5 days, consistent with therapy being the first line of treatment. For children using non-SSRI drugs, median time to medication is shorter, at 23 days, consistent with these including medications for more extreme conditions such as psychosis.[21]

Panels C and E of Table 1 shows the age at first diagnosis is 11.64. Comparing columns 1 and 2, boys are over-represented among children referred to a psychiatrist, as are children whose mothers and fathers are less likely to be employed, are not married, and were more likely to have used SSRI before their child's referral. Differences in age, whether the parent has a college education, and income are small. Comparing columns 3 and 4, conditional on referral, children on SSRI are more likely to be girls, to have a mother and father with a college education (40% vs 34%), who is employed, and who used SSRI when the child was age 6-7. Difference by age and income are small. In our analysis, we attempt to address the issues of selection into SSRI use.

Panel D describes the characteristics of child psychiatrists in the clinic sample. The average propensity to prescribe AD in the full sample (i.e. unconditional on test taking) is 10.5%, and there is a 3 percentage point difference between the prescribing tendencies of the psychiatrists seen by children who are vs are not on SSRI. Close to 80% of all private clinics have just one psychiatrist, most of the rest have two. 41% of psychiatrists are men, mean age 57, mean experience is 28 years, 66% are married, and on average, they treat 170 children each year. In Section 5.2, we investigate

[20] The table shows that 71% of children in the sample receive psychotherapy on average. We only observe therapy provided by in-network providers. Survey evidence suggests that the true rate of therapy use is higher (Sørensen et al., 2020; Buhl Sørensen et al., 2003), likely due to some children receiving care from out-of-network providers. Over 90% of therapy is provided by the child psychiatrist.

[21] Use of red-flag medications, for children, violating clinical guidelines is rare in Denmark, in contrast to the USA (Cuddy and Currie, 2020). When used, duration is brief, for instance, median benzodiazepine use among children is 6–9 days.

whether these observable characteristics of psychiatrists are correlated with AD use, and show that our estimates are robust to controlling for them.

In Appendix Table A2 we describe differences between children whose first contact is with a clinic vs a hospital; as discussed we merge the hospital sample into the clinic sample (our primary analysis sample) as a robustness check, and to investigate outcomes for which the analysis sample is underpowered. Children in the clinic analysis sample are more likely to be prescribed SSRI (15.2% vs. 9.2%) and other mental health medications (46.2% vs. 42.2%) than those whose first contact is at a psychiatric hospital.

4 Empirical strategy

We are interested in estimating the impact of AD treatment on child outcomes in a sample of children referred to a psychiatrist in a setting in which the first line of treatment is psychotherapy. Our estimates will thus capture the impact of adding AD treatment to psychotherapy. The main estimating equation is given by:

$$Y_i^a = \beta_0 + \beta_1 AD_i + \delta \mathbf{X}_i + \epsilon_i, \tag{1}$$

where Y_i^a is an outcome for child i measured at age a, AD_i is an indicator for antidepressant use by age 15, and $\mathbf{X}_i$ is a vector of baseline child and family characteristics including fixed effects for year of birth, year of diagnosis, and municipality of residence in the year of diagnosis.[22,23] The standard errors are clustered at the psychiatrist and year-of-diagnosis level. We will also display results with SE clustered at the municipality and year level (Abadie et al., 2023).

It is plausible that AD use may be correlated with unobservable characteristics of the child or their parents, such as conscientiousness, that also affect their outcomes. Therefore, OLS estimates of β_1 may be biased by selection into AD use. In contrast to the common setting where sorting is motivated by an outcome gain, in our setting, this is unclear. While parents of children with higher outcome potential (e.g. more educated parents whose children have higher test score potential) may seek out higher quality psychiatrists, there is no reason to believe that higher quality psychiatrists are more likely to prescribe AD- or, equivalently, that parents want AD. To address this we implement a two stage least squares (2SLS) strategy that exploits variation in AD prescribing propensities across child psychiatrists.[24] In particular, we construct a leave-one-out instrument that is defined as the

[22] The fixed effects allow us to purge any local-level associations between child mental health and school performance. They will remove time-invariant attitudes toward mental health treatments, information and GP traits including GP referral tendencies).

[23] Our identifying assumption is similar to conditional random assignment based on, for instance, court, date, and/or day of the week assumed in some of the judge leniency literature (e.g., Agan et al., 2023; Dobbie et al., 2018). Instrument validity is discussed and tested in more detail in Section 5.2.

[24] Our data include the identifier of the clinic issuing the prescription rather than the identifier of the individual

AD prescribing rate of the first psychiatrist seen by the index child, defined over *other* children of the same age (8–15), seen in the same year. We allow the instrument to vary at the psychiatrist-year level to accommodate learning, or differential responses to changes in national guidelines (Mueller-Smith, 2015). The instrument for child i can be written as:

$$PP_i = \frac{\widetilde{AD}_{j(i)t(i)} - AD_i}{\widetilde{N}_{j(i)t(i)} - 1}, \tag{2}$$

where $\widetilde{AD}_{j(i)t(i)}$ is the total number of children prescribed AD by the psychiatrist $j(i)$ who treats child i in year $t(i)$, and $\widetilde{N}_{j(i)t(i)}$ is the total number of children treated by the psychiatrist in that year.[25] We define the instrument based on the first psychiatrist that the child sees to minimize concerns over endogenous mobility across psychiatrists, so that $t(i)$ is the year of the first psychiatric visit.[26]

The premise for the instrument is that the prescribing behavior of the psychiatrist for similar-age children predicts the probability that the index child will use AD. The idea derives from the larger judge leniency literature (e.g., Agan et al., 2023; Bhuller et al., 2020; Dobbie et al., 2018; Aizer and Doyle, 2015; Dahl et al., 2014; Maestas et al., 2013; Doyle, 2008, 2007; Kling, 2006). A number of recent papers exploit variation in physician prescribing tendencies (Currie and Zwiers, Forthcoming; Costa-Ramón et al., 2023; Eichmeyer and Zhang, 2023, 2022; Cuddy and Currie, 2020; Laird and Nielsen, 2016; Dalsgaard et al., 2014), but we are unaware of research leveraging variation in psychiatrist prescribing tendencies. This is a case in which we may expect significant idiosyncratic variation, given the subjectivity inherent in the assessment of depression and anxiety, the absence of definitive clinical tests, and the consequent discretion in decisions pertaining to the introduction of pharmacotherapy.

In order for the 2SLS method to yield consistent estimates, the instrument must satisfy three conditions. First, the relevance condition requires that the psychiatrist's prescribing tendency be a sufficiently strong predictor of AD use to mitigate the finite-sample bias inherent in 2SLS. Second, the exogeneity condition assumes that assignment to psychiatrists is conditionally random, i.e., chil-

specialist. For that reason, we define the instrument at the psychiatric clinic level. As documented in Table 1, 80% of children in our sample are treated in clinics with a single provider. 18% of children are seen in clinics with two psychiatrists, and 2% in clinics with three psychiatrists. We show that our results are robust to restricting the sample to solo-provider clinics.

[25] So as to increase the predictive power of the instrument, we define the instrument on all children seen at the same psychiatric clinic in the year of each child's first consultation. This includes out-of-sample information on children diagnosed before age 8, those diagnosed prior to our sample period, in-sample children diagnosed in previous years, and immigrant children. Therefore, our instrument is similar in spirit but is not equivalent to a leave-one-out psychiatrist fixed effect.

[26] In the analysis sample, 6.3% of children switch psychiatric clinics. We find no evidence that this mobility is correlated with the instrument i.e., the prescribing tendency of the first psychiatrist they see.

dren are not systematically matched to providers based on their prescribing behavior. Third, the exclusion restriction requires that psychiatrists influence child outcomes solely through their antidepressant prescribing decisions. Relevance can be empirically assessed by examining the strength of the first stage. While exogeneity and exclusion are fundamentally untestable, we discuss their plausibility at length in Section 5.2 and present supporting evidence.

When treatment effects are heterogeneous, the instrument must satisfy an additional condition, the monotonicity condition, to ensure that the 2SLS estimand can be interpreted as a local average treatment effect (LATE). In our setting, a sufficient condition is average monotonicity, which requires that children who receive an AD prescription from a high prescribing psychiatrist would also receive AD, on average, from a low prescribing psychiatrist. Recent evidence from the judge leniency literature suggests that violations of average monotonicity introduce minimal bias (Sigstad, Forthcoming), but we nevertheless provide evidence in Section 5.2 that monotonicity is likely to hold in our context.

The 2SLS model identifies a LATE for children whose receipt of AD treatment was influenced by the prescribing tendency of the first psychiatrist they consulted- children on AD who may not have been on AD had they seen a psychiatrist with a lower prescribing rate. These marginal patients are of particular interest when considering the question of AD being over- or under-prescribed, or when evaluating the extent of discretion granted to psychiatrists under existing medical guidelines. So as to illuminate the external validity of our results, we compute the proportion of compliers, as well as their average characteristics.

Distributional effects and heterogeneity by observable and unobservable characteristics We complement the baseline results with quantile IV regression estimates that reveal how AD treatment shifts the distribution of test scores. We then explore treatment heterogeneity along both observable and unobservable dimensions, which is relevant for policy targeting. Among observables, we focus on child gender and household socio-economic status. To capture heterogeneity conditional on observables, we apply the marginal treatment effects (MTE) framework developed by Heckman and Vytlacil (2007, 2005).

The MTE represents the average effect of AD use for children at the margin of treatment, defined by the distribution of unobserved resistance to treatment (see Appendix E for details). Estimation of the MTE relies on the same assumptions as the LATE plus a stricter version of the monotonicity assumption and an assumption that observed and unobserved heterogeneity enter the outcome equation additively. We provide evidence on their plausibility in Section 5.5. The MTE framework allows the unobserved gain from treatment to be correlated with unobserved characteristics that affect selection into treatment. The sign of this correlation is meaningful for consideration of the benefits of treatment for the infra-marginal *vs* the marginal patient, and thus relevant to questions of

expanding or retracting treatment. The MTE framework gives us tools to tackle the otherwise difficult question of over- vs under-treatment. Aggregating the MTE with appropriate weights, we can estimate average treatment effects on the treated (ATT) and the untreated (ATUT). It also provides an approach to policy simulations, which we utilize to vary the prescribing behavior of psychiatrists. We estimate policy relevant treatment effects (PRTEs) resulting from policy experiments that change the treatment status of children of low vs high SES, and of boys vs girls, relative to the status quo. This allows us to determine which, if any groups of children are harmed by or benefit from expansions or reductions in access to AD. We similarly estimate PRTE to simulate a tightening of medical guidelines ("rules vs discretion") proxied by compressing the prescribing rate distribution, and to simulate policies that eliminate either low or high prescribing behaviour.

5 Results

5.1 Effects of antidepressant treatment on school performance

First stage and relevance The first stage relationship presented in panel A of Table 2 shows that the psychiatrist's prescribing tendency is a strong predictor of AD use. The coefficient is 0.928 and the F-statistic is 82.8. Challenging the rule of thumb that the F-statistic be at least 10, Lee et al. (2022) show that weak instrument bias can remain substantial if the first-stage F-statistic is below 104.7. To account for this, we follow Andrews et al. (2019) and Sun (2018) and report Anderson–Rubin confidence intervals.[27]

The histogram in Figure 1 shows the distribution of the instrument conditional on fixed effects for year of diagnosis, year of birth and municipality of residence. There is substantial variation in AD prescribing tendencies, with (unconditional) prescribing rates at the psychiatrist-year level ranging from zero to about 34%. On the premise that assignment of children to psychiatrists is conditionally random (discussed in the next section), this is a measure of the extent of psychiatrist discretion. Mental health treatment is an area where we may expect idiosyncratic variation given the inherent subjectivity in how patients experience and report their symptoms. In addition, although medical guidelines exist, the absence of definitive clinical tests means that psychiatrists must rely on their judgment in interpreting the information, introducing further subjectivity.

Overlaid on the histogram is the fitted curve from a local linear regression of antidepressant use on the instrument.[28] We see that the probability of using AD is monotonically increasing in the

[27] The number of observations in Table 2 is smaller than the number of children with a mental health contact noted in Table 1 because not all children take the 9th grade exams. The distribution of the instrument in the sample of test-takers is very similar to the distribution in the full sample of children with a psychiatric visit but the mean AD use is slightly lower at 14.7%. We show in Section 5.3 that our main findings are not driven by selective test-taking.

[28] As a result of conditioning on the fixed effects, the position (intercept) of the curve is not meaningful, but the slope

prescribing tendency of the child psychiatrist and close to linear for most of the distribution. An interquartile shift in the distribution (which involves moving up from a prescribing rate of 6.2% to a rate of 14.1%) increases the probability of AD use by 7.9 percentage points, which is about 52% of the mean prescription rate of 15.2% among children who see a psychiatrist.

Impact of antidepressant treatment on test scores Panel B of Table 2 presents the reduced form estimates which show how much psychiatrist discretion impacts test scores. We find that an interquartile shift in the prescribing propensity of the psychiatrist raises Math scores by 0.038 standard deviation (henceforth SD) and Danish scores by 0.029 SD. IV (2SLS) estimates of AD treatment on test scores shown in panel C indicate that AD use in school-going years is associated with a 0.52 SD increase in Math scores and a 0.40 SD increase for Danish. For Math, the Anderson–Rubin confidence interval is entirely positive. For Danish the confidence interval is largely positive, but it strays into the negative territory. Larger and more precise impacts on Math scores are also evident in response to use of an anti-anxiety app (Cavatorta et al., 2021), and in response to cannabis prohibition (Marie and Zölitz, 2017), broadly consistent with the mechanisms by which mental health impacts test scores (Section 2.1). That said, the point estimates for Math and Danish are not statistically significantly different from one another.

In Appendix Figure A2, we present results from an instrumental variables quantile regression to estimate the impact of AD use at each decile of the test score distribution of children in our analysis sample (Kaplan, 2022; Kaplan and Sun, 2017).[29] We find that the estimated effects of AD are fairly uniform across the distribution, suggesting that the average effects are not driven by outliers.

Performance on test scores may be regarded as a measure of the impact of AD treatment on functionality, an objective "hard" outcome that plausibly reflects improvements in mental health. If AD treatment places children on a higher trajectory of mental health accumulation (see Section 6.1 for evidence in line with this), this can lead to persistent gains in cognitive performance at later ages. This would be consistent with dynamic complementarities between mental health and learning, as well as self-productivity in each.[30]

Characterizing compliers The 2SLS results represent a local average treatment effect for the subpopulation of compliers: children whose AD use is impacted by their first psychiatrist's AD pre-

is.

[29] Recall that our sample consists of children who have seen a psychiatrist. Their median Math and Danish test scores in our sample fall around the 35th and 39th percentile of the respective distributions for children without any mental health contact.

[30] For example, a child who begins AD treatment may show immediate improvements in school performance relative to a similarly diagnosed peer who does not receive treatment. These initial gains can trigger reinforcing effects: enhanced academic performance fosters further learning due to the self-productivity of cognitive skills, while academic success may, in turn, improve mental health through increased self-confidence and stronger peer relationships, demonstrating the cross-productivity between mental health and educational outcomes.

scribing tendency. The size and composition of the complier population is important when considering the relevance of the 2SLS results. Following Chyn et al. (Forthcoming) , we multiply the first stage coefficient by the difference between the maximum and minimum AD-prescribing propensity to obtain the fraction of compliers.[31] The results are presented in Appendix Table B1. The share of compliers in the full sample is 31.3%. It is higher for girls (38.6% vs 27.4% among boys), and for children from higher SES families (40.6% vs 25.6% in low SES). These are large fractions of the population, consistent with substantial discretion and variation in AD prescribing behaviour. To obtain the average characteristics of compliers, we interact each characteristic with the treatment and regress that on the instrument. Aside from being more likely to be girls, high-SES, and to have parents with a history of AD use, compliers are similar to the full analysis sample in terms of average characteristics. This suggests that the LATE estimates have considerable external validity.

Benchmarking the estimated effects of AD use The descriptive statistics in Table 1 show that children who have had a mental health contact during their school years, relative to children who have not, score 0.376 (0.273) SD lower on the 9th grade math (Danish) tests. Our results imply that the (cumulative) AD treatment effect wipes out the baseline (control group) penalty in test scores associated with a mental health diagnosis. The effects are large, but plausible in view of self-productivity and dynamic complementarities in the production of mental health (Section 2.1). The effect sizes are comparable to effects found in other studies evaluating medical interventions during childhood (Daysal et al., 2022; Bharadwaj et al., 2013).[32]

5.2 Instrument validity

In this section we discuss the validity of the the exogeneity condition, the exclusion restriction, and the monotonicity assumption.[33]

5.2.1 Conditional independence

For our instrument to be valid, conditional on year of birth, year of diagnosis and municipality of residence fixed effects, children should not be systematically matched to providers based on their prescribing behavior.[34] We assess this assumption in two ways. First, in the right panel ofAppendix

[31] There is a slight difference in the estimating samples for the Math and Danish test scores. We characterize compliers using the Math sample; results are similar using the Danish sample.

[32] Daysal et al. (2022) and Bharadwaj et al. (2013) focus on very low birth-weight children, and find that access to medical treatments at birth increases their math test scores by 0.32 SD in Denmark and 0.22–0.48 SD in Norway.

[33] The checks implemented in this section are based on the estimating sample for 9th grade math test score. The results are virtually the same if we use the slightly different sample of test-takers for the Danish test.

[34] This assumption is similar to the conditional random assignment based on, for instance, court, date, and/or day of the week assumed in some of the judge leniency literature (e.g., Agan et al., 2023; Dobbie et al., 2018).

Figure B1 we provide a test of balance on observable characteristics using a rich set of mother, father and child characteristics. The left panel demonstrates that there is some sorting into AD use on observable characteristics – girls are more likely to use AD by age 15, and children are more likely to use AD if their mothers are married, (weakly) more educated, have themselves used AD, or if the child received counseling before their first psychiatric visit. If children with these characteristics are also more likely to perform well in school, then any association between AD use and school performance will be contaminated by selection into AD use. The right panel shows that the instrument is balanced on these characteristics. In addition, we find no evidence that the observable characteristics jointly predict the psychiatrist's propensities to prescribe AD (the F-statistic is 1.5). We summarize this finding in Appendix Figure B2 where we overlay a local linear regression of predicted AD use (using a linear regression including the displayed child and family observables) on psychiatrist tendencies to prescribe AD. This confirms that predicted AD use is not correlated with the instrument. In a related check, we follow Bhuller et al. (2020) and assess the stability of the first-stage estimate to the inclusion of covariates. If the instrument is in fact orthogonal to patient characteristics, then the first-stage coefficient should be stable to the choice of controls. The results in Appendix Table B2 show that this is indeed the case. These checks increase our confidence that the assignment of children to the first psychiatrist they consult is as good as random.

In our setting, we are not surprised that we see no evidence of sorting into high AD prescribers. While it is plausible, for example, that more educated parents — whose children tend to have higher academic potential — seek out higher-quality psychiatrists, there is no reason to believe that higher-quality psychiatrists are more likely to prescribe AD.

In the next section, we investigate further the question of sorting into psychiatrist on quality-indicators, and the question of whether proxies for quality are correlated with the instrument.[35]

5.2.2 Exclusion restriction

Quasi-random allocation of patients to psychiatrists is sufficient to interpret the reduced-form effect of prescribing tendency on outcomes as causal. However, for the 2SLS estimates to reflect the causal impact of AD, the exclusion restriction must hold. This requires that a psychiatrist's tendency to prescribe AD influences child outcomes solely through the AD treatment. We investigate this as follows. **Pairwise monotonicity**. Our *first strategy* is to show that pairwise monotonicity holds,

see 5.5. This is equivalent to testing whether psychiatrists have a common ordering (or ranking) of

[35] Unlike the judge leniency setup, where a defendant always prefers a more lenient judge, patients do not universally prefer AD treatment. Depending on individual symptom dynamics and the effectiveness of therapy, medication may or may not be the optimal course of treatment. Conversations with a couple of psychiatrists suggest that, conditional on municipality of residence and year of diagnosis, children are typically treated by the psychiatrist with the shortest waiting time operating within a reasonable distance from the child's residence.

patients according to their suitability for AD prescription, with differences in observed prescribing rates reflecting differences in the threshold at which they prescribe. In this case, psychiatrists vary only in the probability that they prescribe. This adds plausibility to the exclusion restriction; see, for instance, the discussion in (Frandsen et al., 2023).

Characteristics and and broader prescribing behaviours of psychiatrists. However, the exclusion restriction could still fail if psychiatrists directly impact future outcomes like test scores through a channel other than their tendency to prescribe AD. There are two natural candidates here: (a) psychiatrist characteristics, which include age, experience, gender, marital status, and whether they practice alone or with other psychiatrists in the clinic, and (b) behaviors, which refers to their prescribing behavior for treatments other than AD. The two empirically relevant treatments for mental health are psychotherapy and other N-class drugs. *The second strategy* we adopt adjusts for these potential direct impacts in three steps. We start with checking for balance on these observables, then we control for characteristics and prescribing propensities and, finally, we replace the prescribing propensities with actual treatment rates instrumented with the prescribing propensity.

The left panel of Appendix Figure B3 focuses on *actual AD use* and shows that it is not correlated with psychiatrist characteristics, but that it is positively correlated with use of other mental health medications and psychotherapy.[36] The right panel shows associations with *AD prescribing leniency*, our instrument. With the exception of physician gender, the instrument is not correlated with CP characteristics. Moreover, these observable characteristics explain little of the variation in the instrument: their inclusion increases the adjusted-R^2 from 31% to 35%. While the AD instrument is uncorrelated with an indicator for psychotherapy, it is positively correlated with the use of non-SSRI drugs.

To account for this, we follow Bhuller et al. (2020) in examining this in two steps. First, we add psychiatrist characteristics and prescribing propensities for alternative treatments as controls in our baseline model. Column 2 of Appendix Table B4 shows that the IV estimates for both Math and Danish are similar to the baseline estimates and statistically significant at 10% based on the Anderson–Rubin confidence intervals. In a second step, in column 3 of Appendix Table B4, we replace the prescribing tendencies for therapy and non-SSRI drugs with indicators for the focal child being prescribed these other treatments, instrumenting these individual indicators with their psychiatrist's prescribing tendency for that treatment. Appendix Table B3 shows that the first-stage regression for AD treatment is robust to adding controls for the other treatments (column 2), and that the first stage for therapy and non-SSRI drugs is strong (columns 3, 4).[37] The augmented IV results are in

[36] This refers to therapy after their first contact with a psychiatrist.

[37] Examining the off-diagonal coefficients (capturing the effect of the psychiatrist's tendency to prescribe one treatment on the likelihood of a child receiving another) we find some evidence of cross-treatment influence for pharmaceutical treatments. However, these effects are considerably smaller than the direct effects shown on the principal diagonal.

Appendix Table B4. We continue to see a positive significant impact of AD treatment on test scores. We now also observe that psychotherapy has an independent positive impact on test scores, but that non-SSRI N-class drugs (which include stimulants) reduce test scores.[38] These findings align with medical studies demonstrating the efficacy of both therapy and antidepressants (Boaden et al., 2020; Cipriani et al., 2018). The negative impact of non-SSRI drugs, which include ADHD medications, is consistent with Currie et al. (2014), who find adverse effects of ADHD medication on the educational outcomes of marginal patients.

Unobservable quality of psychiatrists. The remaining concern is that the AD prescribing tendency proxies for an unobservable dimension of psychiatrist behaviour that we label "quality". Unobserved quality could include, for instance, the psychiatrist's ability to gain the cooperation and trust of the child, or their diagnostic skill that determines where on the continuum of "treatability" of the patient the prescribing threshold is set. These unobservables would cause failure of the exclusion restriction if they were correlated with the instrument and directly impacted test scores. There is no reason to expect that high prescribers are more or less likely to make a child feel comfortable and positive. As regards diagnostic skill, we might imagine that high prescribers have weaker skills. See, for instance, (Chan et al., 2022), who investigates diagnosis of pneumonia through radiography, arguing that as long as the risks associated with failing to treat a patient who needs treatment are larger than the risks associated with treating a patient who does not need treatment[39], physicians with weaker diagnostic skills will be more likely to treat. For these reasons, we regard it as unlikely that our instrument is positively correlated with psychiatrist quality. We nevertheless now discuss *a third strategy* that we use to investigate this.

We adopt the "plausibly exogenous" framework of Conley et al. (2012) to gauge how large a direct effect of the instrument (AD prescribing propensity) on the outcome (test scores) would have to be to wipe out our baseline estimate. We allow the instrument to affect test scores directly by an amount γ in addition to its effect through AD use. Let $\gamma = \rho\, \gamma_q$, where ρ is the correlation between AD prescribing propensity and psychiatrist quality, and γ_q is the effect of psychiatrist quality on test scores. We benchmark γ_q using estimates from a literature on school-based social and emotional learning (SEL). In the more credible studies in the Corcoran et al. (2018) meta-analysis (randomized designs, larger samples, and more recent cohorts), SEL programs raise math scores by 0.18–0.23 SD. To be conservative, we set γ_q to 0.23 SD. It follows that the plausibly exogenous confidence interval will exclude zero as long as γ is below about 0.1 SD, which requires ρ to be below roughly 0.4. In other words, for our estimates to be driven by direct impacts of high prescribers on test scores, the

[38] The Lewis and Mertens (Forthcoming) statistic reported in Appendix Table B4 indicates that the three instruments are not jointly strong. We therefore provide Anderson-Rubin confidence intervals obtained through the projection method for each coefficient on the three treatments. Given that the instruments are jointly weak, the confidence intervals are wide, and we focus on the fact that they are entirely positive rather than on the point estimates.

[39] This is plausible in our setting given that treatment effects are expected to dominate side effects.

direct effect would have to be almost as large as the impact of AD.[40] This gives us some confidence that the estimated effect of AD on test scores, and math in particular, is likely to remain statistically different from zero even under a violation of the exclusion restriction.

5.2.3 Extensivity

Our treatment variable is a binary indicator reflecting whether a child used AD by age 15. Our IV strategy makes the implicit assumption that a high prescribing psychiatrist increases the probability of using AD but not the duration of treatment for children who are on AD. A violation of this assumption, where the instrument also affects treatment intensity rather than just initiation, would constitute a specific form of exclusion restriction failure. We implement two checks to assess the extensivity assumption, in line with Norris et al. (2021). These rely on data on AD volume measured as the total number of defined daily doses (DDDs) filled by age 15.

First, we discretize the positive range of AD volume into 20 ventiles. We then run 20 regressions of having a treatment duration smaller than that ventile on the psychiatrist prescribing tendency and baseline control variables. Under extensivity, the estimated coefficients on the instrument should increase monotonically. Appendix Figure B4 shows that this is indeed the case. Second, we replace in our baseline specification the AD indicator with AD volume but still use the psychiatrist tendency to prescribe AD as the instrument. The IV estimates from this model represent a convex combination of extensive and intensive effects even if the extensivity assumption does not hold (Angrist and Imbens, 1995). The results reported in Appendix Table B5 are consistent with those using the binary AD indicator: increasing AD volume by 450 DDDs (the median AD duration) increases Math by approximately 0.235 SD and Danish by 0.177.[41]

5.2.4 Monotonicity

We follow the literature and implement two indirect tests. First, we show in Appendix Table B6 that the first-stage estimate is positive and statistically significant in different subgroups defined by child gender, household socio-economic status, quartiles of the predicted test score distribution, and characteristics of the first psychiatrist consulted (Norris, 2022; Norris et al., 2021). Second, we

[40] One way of assessing this is to note that correlations of the instrument with observable psychiatrist characteristics shown in Appendix Figure B3 are smaller.

[41] We also estimated two other specifications similar to our baseline. In the first, we use AD volume as the endogenous variable and the psychiatrist propensity to prescribe higher volumes of AD as the instrument. The results, shown in column 3 of Appendix Table B5, are very similar to the previous column and indicate that 450 additional DDDs of AD increase Math by 0.284 SD and Danish by 0.290 SD. The second specification adds the psychiatrist tendency to prescribe AD (our instrument) as a control. Intuitively, identification in this model relies on a comparison between psychiatrists who treat a similar proportion of children with AD, but differ in the average duration of treatment they prescribe. Column 4 of Appendix Table B5 shows that, although AD volume still has a positive effect on test scores, the first stage is essentially flat. This suggests that the effect of AD volume on test scores work through the psychiatrist tendency to prescribe AD and not volume.

estimate the first-stage in different subgroups using reverse (or cross-) sample instruments (Frandsen et al., 2023). For example, when we examine the first-stage in the sample of girls, we define the psychiatrist's prescribing tendency based on the sample of boys. We again confirm that the first-stage estimates are positive and statistically significant, indicating that psychiatrists who are lenient prescribers for a given subgroup are also lenient prescribers for other groups (see Appendix Table B7).

5.3 Robustness checks

We assess the robustness of our main results across a range of sample restrictions and model specifications. The variations considered are summarized in Figure 2, which reports the IV point estimates alongside Andersen–Rubin confidence intervals. In all cases, the estimates remain stable in both sign and statistical significance. Below, we provide the motivation for each robustness check and discuss the corresponding results.[42]

5.3.1 Robustness checks related to the instrument

Volume of the psychiatric clinic Our analysis sample includes children treated in clinics with at least 50 patients during the first year of contact, and the instrument is constructed at the clinic-year level. Following best practice outlined in Chyn et al. (Forthcoming), in Appendix Table C1, we investigate the sensitivity of the estimates to alternative thresholds. Across all specifications, the IV coefficients remain stable and statistically robust. Increasing the threshold to 75 or 100 patients yields slightly larger coefficients for Math, while lowering it to 25 results in a smaller coefficient for Math and a larger coefficient for Danish, both of which are statistically significant. To better isolate prescribing behavior, we restrict the sample to the 80% of clinics with a single provider, and the resulting estimates closely align with our main findings. As an additional robustness check, we change the clustering of the standard errors from the clinic-year level to the clinic level, as is standard in the judge fixed effects literature, and to the municipality and year of diagnosis level, which is the level of random allocation in our setup (Chyn et al., Forthcoming; Abadie et al., 2023). Our inference remains unchanged.

Outliers We next probe the importance of outliers by excluding children treated by the top-3 psychiatrists, ranked by their average propensity to prescribe antidepressants. The IV coefficients increase slightly but remain statistically indistinguishable from the main estimates. Given the relatively small number of psychiatric clinics, one potential concern is that models including municipality fixed effects may rely on limited identifying variation. To address this, we replace fixed effects for

[42] Appendix C reports the corresponding regressions.

the 98 municipalities that children reside in with fixed effects for Denmark's five health care regions. The estimates are again similar to the main results.

5.3.2 Robustness checks related to the analysis sample

Endogenous test taking While a mental health diagnosis does not automatically exempt students from taking exit exams, exceptions can be granted on a case-by-case basis. A potential concern is that students on AD who miss the exam may be those who would have performed poorly, introducing selection bias. To address this, Appendix Table C2 examines the impact of AD treatment on test-taking behavior. Panel A analyzes the likelihood of deferring the exam, while Panel B focuses on the likelihood of never taking it. We find no statistically significant evidence of endogenous test-taking. However, some coefficients are economically meaningful in magnitude, prompting further investigation.

To assess the potential impact of selection, we estimate bounds on the baseline treatment effect. Because students do not take any high-stakes standardized tests prior to age 16, we cannot identify the likely position in the test score distribution of those who did not take the exit exam.[43] We address this by re-estimating the baseline model, adding back non-test-takers, sequentially assigning the AD-treated a test score percentile (from the distribution of untreated students) ranging from the 10th to the 90th percentile. The resulting estimates, shown in Appendix Figure C1, are consistently positive and not statistically distinguishable from the baseline (indicated by the horizontal red line). This pattern reduces the concern that our main findings are driven by selective test-taking among AD-treated students.

Removing tests taken during the pandemic Given that the COVID-19 pandemic increased rates of mental health disorders among children and led to school closures, including those who took age-16 exams during this period may bias our estimates. Excluding them produces estimate similar to the baseline (Appendix Table C3).

Including children first seen in hospitals As discussed in Section 3, the main analysis excludes children whose initial mental health contact occurred at a hospital psychiatric department. Appendix Table C3 assesses the robustness of our findings to including these children. The F-statistic decreases from 82 to 47.[44] The Math coefficient is slightly lower (0.46 vs. 0.52) but estimated with greater precision, while the Danish coefficient is now close to zero.[45]

[43] National tests administered in grades 2–8 are low-stakes and limited to public schools.

[44] Appendix Figure C2 documents the balance check on the instrument in this sample.

[45] This greater sensitivity of Danish scores to changing sample composition is possibly explained by greater heterogeneity in treatment effects by characteristics (see Appendix Table A2).

5.3.3 Robustness checks related to AD treatment

Endogenous mobility A primary threat to our identification strategy is the possibility that children seeking to improve their 9th-grade test performance selectively initiate AD treatment by targeting high-prescribing psychiatrists. To alleviate this concern, we define the instrument using the prescribing propensity of the first-seen psychiatrist. In our analysis sample, 6.3% of children move across psychiatric clinics. Appendix Table C4 shows that subsequent psychiatrist changes are uncorrelated with prescribing rates.

AD use in the year of taking the test We next examine whether our results may be influenced by children who are using AD at the time of the test. This includes two distinct groups. The first consists of children who begin AD treatment before age 15 but continue their medication through the exam period. The second group includes children who initiate AD treatment after age 16; under our current treatment definition (AD use by age 15) these children are classified as part of the control group. We show in Appendix Table C5 that our results become larger when we drop the first group of children. Further dropping the children in the second group does not impact the coefficient estimates. These results are consistent with early interventions having a larger impact on test scores due to dynamic complimentarities.

School effects Schools are a significant part of a child's environment and an extensive literature in economics documents the importance of peer effects (Sacerdote, 2011). We re-estimate our baseline model including school fixed-effects. The estimated effect for Math is very similar to the baseline and remains statistically significant at 10% based on the Andersen-Rubin confidence intervals (see Appendix Table C5).

5.3.4 Robustness test using longitudinal data on children

While our instrument is balanced across observable characteristics, concerns may remain that it does not fully eliminate unobserved differences between treated and untreated children. An alternative approach leverages longitudinal data, which allows us to control for time-invariant individual characteristics through child fixed effects. For a subsample of children, we observe repeated scores from low-stakes, nationwide standardized tests in Math and Danish administered in grades 2 through 8. Using this panel data and a modified version of our instrument, we estimate the effect of AD use on test performance while controlling for individual fixed effects (see Appendix C.1 for details). Due to the limited number of children with repeated test scores, Appendix Table C6 reports results both for our main analysis sample and for the broader sample of children with any psychiatric contact. The findings are broadly consistent with our baseline estimates: following AD initiation, test scores increase by 0.472 SD in Math and 0.318 SD in Danish.

5.4 Heterogeneity in treatment effects by observable characteristics

Child gender Medical studies indicate that girls have higher rates of AD use (Wesselhoeft et al., 2020). This is also true in our analysis sample. Conditional on having a psychiatric contact, girls are more likely to have used AD than boys (19.2% versus 12.5%). There are no statistically significant differences by child gender in the average impact of AD use on test scores, see columns 2 and 3 of Table 3. However, there is a tendency for girls to show larger improvements than boys in the upper end of the test score distribution across both subjects (Appendix Figure D2), echoing the findings of Bhalotra et al. (2022), Cavatorta et al. (2021), and Marie and Zölitz (2017). The gender-specific first-stage relationships are in the top Panel of Appendix Figure D1.

Household SES We next split the sample by household SES, defining high SES families as those in which the mother has at least some college education. So as to confirm that defining SES on mother's education is meaningful, columns 3 and 4 of Appendix Table D1 compare the characteristics of children from low- and high SES backgrounds in our analysis sample. Children whose mothers have less than a college education are much less likely to have highly educated fathers, both parents have lower earnings and employment rates, they are less likely to be married, and also more likely to contribute boys to the analysis sample.

Turning to treatments, low and high SES children are similar in age at the time of their first psychiatric contact, and in the median days between diagnosis and initiation of therapy. However, low SES children are less likely to use antidepressants (14% vs. 17.3%), and they start AD treatment slightly later. Low SES children are also more likely to use stimulants (35.9% vs. 30.6%) and antipsychotics (7.8% vs. 6.9%).[46]

Table 3 presents IV estimates of our baseline model by household SES (columns 4 and 5).[47] We find that AD treatment has similar effects on Danish test scores for the two groups, but the impact on Math scores is substantially larger among low SES children. The Andersen–Rubin confidence interval for the low SES group is entirely positive, despite a weaker first-stage. This gradient indicates that higher education among mothers acts as a partial substitute for antidepressant treatment, with more educated mothers possibly being better equipped to buffer their children's academic performance from the adverse effects of mental health conditions. This aligns with a medical literature showing that the effectiveness of AD treatment depends on the patient's broader environment (Elwadhi and Cohen, 2020). The spirit of our finding is reminiscent of the finding of Cornelissen et al. (2018) that immigrant children, who would benefit most from childcare in Germany, are less likely

[46] Currie et al. (2024) report similar differences by SES in the pharmaceutical treatment of mental health conditions in Canada.

[47] The bottom panel of Appendix Figure D1 shows the first-stage. Due to smaller sample sizes, the instrument is weaker in both subgroups, with F-statistics of 37 for the low SES group and 56 for the high SES group.

to take it up.

Appendix Figure D3 illustrates how AD treatment shifts the distribution of test scores by SES. Among low SES children, Math scores increase most clearly at the lower and upper ends of the distribution, while the effect is relatively uniform for high SES children. For Danish, the results are not precise but scores tend to rise in the lower half of the distribution for low SES children and in the upper half for high SES children.

A potential concern with the observed SES gradient is that it may be driven by positive selection into the analysis sample, specifically, that low SES children with higher potential gains from treatment may be more likely to seek care in clinics rather than hospital psychiatric departments. To investigate this, we re-estimate our model using an expanded sample that includes children first seen in hospitals. We find that the gradient is in fact steeper in the expanded sample (Appendix Table D2), so this is unlikely to be the case. The observed SES gradient is also not explained by differential sorting. Average prescribing rates of the psychiatrists seen by low vs high SES children are similar (10.8% for low SES vs. 10% for high SES children), and the psychiatrist's prescribing tendency is uncorrelated with household SES, see Appendix Figure B1. It is also unlikely to be driven by differences by SES in use of therapy or non-SSRI drugs- we showed in Appendix Table B4 that the estimated impacts of AD were robust to controlling for these other treatments, instrumented by their own prescribing rates.

Overall, we see limited heterogeneity in the impact of AD treatment by gender, and stark heterogeneity by whether or not the mother has higher education. In the next section, we estimate marginal treatment effects, thus allowing for variation in treatment effects conditional on observables.

5.5 Marginal Treatment Effects

Since our instrument is continuous, we can estimate marginal treatment effects (MTE), which represent the derivative of expected test scores with respect to the predicted probability of AD use. Compared to the LATE estimates discussed thus far, the MTE approach provides a richer characterization of treatment effect heterogeneity driven by unobserved factors influencing take-up. By aggregating MTEs with appropriate weights, we can recover not only the LATE, but also the average impact of antidepressant use on children who use AD (the average treatment effect on the treated, ATT) and on those who are diagnosed but do not use AD (the average treatment effect on the untreated, ATUT). We provide a detailed discussion of the MTE framework and its estimation in Appendix E.

A particular advantage of the MTE framework in our setting is that it illuminates the policy question of whether children in our sample are being over- or under-treated with antidepressants. The LATE is policy relevant in its own right as it reflects the effect of AD use for those children who

are on the margin of treatment under current prescribing practices. However, the LATE cannot inform us about the consequences of any changes in prescribing behavior because the compliers change as the policy (guidelines) change. The MTE provides us with a framework within which we can examine the consequences of policy experiments that expand or reduce treatment with AD, both along the distribution of unobserved resistance to treatment and on average.

Appendix Figure E1 shows the distributions of the predicted probability of AD use (propensity score) for treated and untreated children, with vertical lines marking the upper and lower bounds of common support over which the MTE can be estimated.[48] Figure 3 plots the MTE curves, assumed to follow a second-order polynomial in the propensity score, and the corresponding 95% confidence intervals. The curves show the effect of using AD for the compliers along the distribution of resistance to treatment (the negative of the prescribing propensity), v, which is the residual in the first stage equation. Intuitively, at every percentile of v, the MTE is the treatment effect for children who would be "pushed into treatment" if treated by a marginally more lenient psychiatrist.

The MTE curves are both negatively sloped, indicating positive selection on gains to treatment conditional on observables. This is consistent with patients having private information on the degree to which taking AD will benefit them.[49] The height of the curve indicates the range of values of v for which the MTE is positive. For Math, the MTE are positive for almost the entire distribution of common support. For Danish, the MTE turns negative around the middle of the common support range. However, the magnitude of the effect of AD also depends on the distribution of compliers (or of treated children) along the distribution of v. The dots in Appendix Figure E4 indicate the fraction of compliers at each percentile of v and show that the majority of compliers come from the range where the MTE are large and positive in both Math and Danish, which explains why both LATEs are large and positive. Note also that the LATE obtained by aggregating the MTE is similar to the LATE obtained via 2SLS, which gives us further confidence in the validity of the MTE framework.

The estimated ATT is large and positive for both Math and Danish, suggesting benefits to all treated children not just the compliers, consistent with positive selection into treatment on unobservables. The ATUT is negative and relatively small in Danish, but positive and large in Math, indicating that children who are not treated with AD would have benefited from them in terms of performance in the Math exam. Thus, if improvements in Math test scores were the objective in deciding on AD treatment, then AD are being under-prescribed. In any case, the positive ATT and our finding that the LATE, which captures marginal cases, is positive suggest that AD are probably not being over-prescribed.[50]

[48] In order to avoid biases due to small number of observations, we trim 1% from both tails of the distributions.

[49] For instance, conditional upon the prescribing tendency of their psychiatrist, people who know they will adhere to treatment protocols and/or adopt complementary lifestyle strategies may be more likely to take up an AD prescription.

[50] There is no such thing as an optimal prescribing rate: patients subjectively present their symptoms to a practitioner, and the practitioner subjectively parses the information with reference to their medical training and national guidelines.

Robustness checks on the MTE In addition to the conditions required for LATE identification, the MTE require pairwise monotonicity, a stronger version of the monotonicity condition. In our setup, this condition requires that all psychiatrists have the same ranking of children in terms of whether they should get AD, but that they differ on the threshold they use for prescribing AD. Appendix Table E1 compares the LATE and the p-value from the Frandsen et al. (2023) joint test of pairwise monotonicity and exclusion restriction between our analysis sample and the subsample with at least 20 observations per psychiatrist-year.[51] The estimated 2SLS effects of AD use are very similar between the two samples, especially for Math. Importantly, the Frandsen et al. (2023) test fails to reject the pairwise monotonicity assumption in both samples for Math and in the full sample for Danish. We cautiously interpret this as evidence that the MTE are identified, at least for Math.

We conduct several additional checks, described in detail in Appendix E.3. First, given that both the slope and the height of the MTE curve are consequential, we consider alternative functional forms for the MTE. Appendix Figure E7 shows that functional form has limited bearing on both attributes of the MTE curve. Second, the MTE depends on the correct estimation of the propensity score in the first stage. Following Cornelissen et al. (2018), Appendix Figure E8 shows that a semiparametrically estimated propensity score has a similar shape to the baseline propensity score, with a correlation coefficient close to 0.90. Third, in the spirit of Devereux (2022), we control for higher order polynomials in a covariate index because the MTE may be sensitive to the set of covariates included. The MTE curves plotted in Appendix Figure E3 are virtually identical to the baseline MTE curves, especially for Math.[52] All these checks reinforce our confidence in the estimation of the MTE curve, and in particular the MTE curve for Math.

Interacted MTE—differences by gender and SES The results in Section 5.4 indicate different effects of AD use, on average, by gender and SES. To investigate the MTE by these characteristics, we model it as a second-degree polynomial in the propensity score interacted with an indicator for the characteristic (see Appendix E.4 for details). Appendix Figure E5 plots the MTE curves by gender. The shapes of the curves are slightly different, but the properties (slope and where they cross the horizontal axis) are roughly similar, as are the ATE, ATT and ATUT.

Our statements on over- or under-prescribing are based on the estimated benefits to children in terms of academic performance. However, the benefits of AD use likely extend beyond just academic achievement given that test scores in the 9th grade exam, particularly in Math, are strongly correlated with labor market performance in adulthood. We present evidence on this, and also on health effects, in Section 6.

[51] The Frandsen et al. (2023) test requires a "large enough" number of observations per examiner-cell to provide sufficient power in finite samples. This condition is not fully satisfied in our analysis sample because the instrument varies at the psychiatrist-year level. Column 2 in Appendix Table E1 restricts the data to psychiatrist-year cells with at least 20 children.

[52] Note that we cannot superimpose the MTE curves because the distribution of v, which determines the horizontal axis, changes every time we change the first stage specification. We cannot compare the LATE for the same reason. However, we can compare the ATE, ATT and ATUT between the two sets of results, and they are very similar.

The MTE by SES are in Appendix Figure E6. While the two curves are similar for Danish scores, they are different for Math in slope and height. The MTE curvey for low-SES curve is flatter, indicating more limited selection into treatment on unobservables, while the high-SES curve mimics the average curve, indicating positive selection on gains to treatment. Low-SES children have positive MTE over the entire relevant part of the distribution of v, while the MTE for high SES children turns negative about halfway. The ATU is also different between the two groups and suggests that the Math test scores of untreated low-SES children would benefit from treatment with AD. This motivates the policy experiments we conduct in the next section.

5.6 Policy Experiments

As discussed, the MTE framework can be leveraged to throw some light on over- or under-treatment with antidepressants. The LATE cannot inform us about the consequences of changes in prescribing behavior because the compliers change as policy guidelines change. We estimate policy-relevant marginal treatment effect (PRMTE) curves, which show impacts of new (simulated) policies for the children induced into treatment by the policy (relative to the *status quo*) at each point in the distribution of unobserved resistance to treatment. Using weights equal to the share of children induced into treatment, we can aggregate the PRMTE into the policy-relevant treatment effect (PRTE), which gives the average effect per net person shifted into treatment (see Appendix E.5). We conduct two policy simulations motivated by our results: making psychiatrists "identity-blind" (specifically, to the SES or the gender of the patient), and tightening medical guidelines to reduce physician discretion. These policies affect the treatment status of all children in the same direction, which means that the PRTE represents the ATT among affected children.

5.7 Identity blinding

In this section we discuss closing the SES and gender gaps in AD treatment by manipulating the propensity score in both directions. First, given our previous finding of a positive ATUT among low SES children, we increase the AD treatment rate of low-SES children to equal that of high-SES children. If there is under-treatment of low-SES children, we should see positive PRMTE and PRTE. Figure 4(a) shows that this policy leads to an increase of 1.6 (Math) or 1.3 (Danish) percentage points in the total share of children who receive treatment, corresponding to approximately 10% of all children receiving AD. The PRMTE curve for Math is positive over the entire relevant range and lies above the MTE curve. This means that the low-SES children who are switched into treatment under our policy see larger test score gains than children (of low and high-SES) who were already treated under the *status quo*. The PRMTE weights, which represent the share among all children affected by the policy at each point in the distribution of unobserved resistance to treatment, are

almost monotonically increasing with the MTE, indicating that most of the affected children have relatively high returns to treatment. As a result, the estimated PRTE is 1.023 SD, higher than the baseline ATT of 0.782 SD.

The PRMTE for Danish test scores in Figure 4(b) is slightly below the baseline MTE and is positive only until roughly the middle of the relevant range of unobserved resistance to treatment. However, the PRTE weights show again that the vast majority of children treated with AD because of the policy come from the range with positive returns to treatment and that relatively few children (18% of the affected children) are harmed by the policy. As a result, the PRTE is smaller than the baseline ATT of 0.906 SD, but still positive and large at 0.633 SD. In conclusion, all children affected by this policy do better in Math, and most but not all do better in Danish, suggesting under-treatment among the low-SES children in our sample.

Even if low-SES children are under-treated with AD, it may be that high-SES children are over-treated. To investigate this, we reduce the high-SES treatment rate to equal the rate for low-SES children. If there was over-treatment among high-SES children, then this simulation should show negative PRMTE and PRTE (i.e., the children who are denied treatment under the policy were being harmed by their use of AD under the *status quo*). Our policy reduces the share of all children receiving AD by 0.8–1.0 percentage points. The estimated PRMTE plotted in Figures 4(c) and (d) both become negative around the middle of the relevant range of unobserved resistance to treatment. This is consistent with some children who are switched away from treatment being harmed by using AD. However, the PRMTE weights show that most of the switchers (83% in both Math and Danish) come from among those with relatively high returns to treatment, resulting in relatively large and positive PRTEs of 0.482 SD (Math) and 0.674 SD (Danish), albeit smaller than the baseline ATTs. Overall, these results suggest that reducing the treatment rate among high SES children harms more children than it benefits. There is thus scant evidence of over-treatment.

We repeat the exercise for the gender gap in AD treatment. The results, shown in Figure 5, are similar both when we simulate an increase in the use of AD among boys to the level of girls (3.8 percentage points increase in affected children) and when we reduce the prescribing rate among girls (2.7 percentage point reduction). In all four Figures, the PRMTE curve is virtually identical to the baseline MTE curve, indicating that the children affected by the policy have similar gains from treatment as the children treated under the *status quo*. As in the case of SES, the vast majority of affected children come from the range of unobserved resistance to treatment where the returns to treatment are positive, resulting in relatively large and positive values of the PRTE. As in the previous experiment, the conclusion is that expanding treatment with AD among boys tends to generate benefits across the board, while reducing the use of AD among girls harms more children than it benefits.

5.8 Rules vs discretion

As discussed and shown in Figure 1, there is discretion in the decision to prescribe antidepressants. One lever that policy makers have to reduce discretion is to strengthen guidelines. We proxy this with a compression of the distribution of prescribing rates. Motivated by the preceding results indicating under-treatment, at least for low-SES children, we first simulate the impact of expanding AD treatment by imposing a floor on the prescribing rate- we shift prescribing rates under the 25th percentile to the 25th percentile of the distribution under the *status quo*). This increases the number of children treated by about 0.5 percentage points. Although the PRMTE curves for both Math and Danish become negative a bit above the midpoint of the relevant range, the PRTE weights show that almost all of the children pushed into treatment by this policy come from the range with positive PRMTE (95% for both Math and Danish). As a result, the PRTE in both Math and Danish is high and comparable to the ATT under the *status quo*.

Next, we consider the reverse scenario: we proxy a concern with over-treatment with a cap that pushes all rates above the 75th percentile of the prevailing distribution down to the 75th percentile. This policy reduces the number of treated children by about 1.5 percentage points. In both Math and Danish, the PRMTE curve lies above the MTE curve, indicating that the children from whom treatment would be withheld by the cap benefit more from treatment than the children to whom treatment is still available (see Appendix Figures E9(c) and E9(d)). Although the children shifted out of treatment tend to be more evenly distributed along the range of unobserved resistance to treatment, the resulting PRTE is positive and relatively large for both Math and Danish. Overall, restricting access to AD treatment benefits at best a small fraction of children, harming a larger fraction (specifically, 93% in Math and 59% in Danish).

Taken together, our policy experiments indicate that the evidence emerging from test score outcomes is inconsistent with over-prescribing and consistent with under-prescribing, especially among low-SES children and boys.

6 Additional Outcomes

In this section, we explore the effects of AD use on a broader set of outcomes. Given the ongoing debate surrounding the pharmaceutical treatment of mental health in children, we begin by examining the short-term health impacts of AD use. We then turn to its effects on long-term educational attainment, labor market performance and welfare dependency. To improve statistical power, these analyses are conducted using the full sample of children with a psychiatric contact between ages 8

and 15, including those first seen in hospital settings.[53]

6.1 Effects of antidepressant use on (mental) health

The use of AD in children has long been subject to medical and regulatory scrutiny. Concerns over increased risks of suicidal thoughts and behaviors gained prominence in the early 2000s. In response to these concerns, the U.S. Food and Drug Administration (FDA) issued a "black box" warning in 2004, advising caution in prescribing SSRI to pediatric patients. The European Medicines Agency followed in 2005 with similar warnings across the EU, emphasizing the potential mental health risks of AD use in young populations. However, subsequent medical research has highlighted methodological flaws in the meta-analysis that informed these regulatory actions and argued that the risks associated with untreated mental health conditions may outweigh those linked to treatment (Cipriani et al., 2018; Lu et al., 2014; Rihmer, 2007).

In Table 4, we examine the health effects of AD use, looking at the likelihood of suicide attempts or self-harm (indicators of mental health), as well as at any inpatient or outpatient hospital contact within three months of treatment initiation. For children not treated with AD, we assign a placebo initiation date based on the average age of AD initiation among same-gender peers diagnosed in the same year. On average, we find no evidence of increased adverse health outcomes following AD initiation. Across all children, we see significant reductions in suicide attempts and self-harm, and in hospital visits, driven by outpatient visits. There are unambiguous benefits for girls on all of these outcomes and, additionally, on emergency room (ER) visits. Among boys, the average effects are smaller, and the Andersen–Rubin confidence intervals are wide, making it hard to rule out adverse effects. When we stratify by socioeconomic status, we observe a significant reduction in suicide attempts and self-harm among low-SES children, and a decline in hospital outpatient and ER visits among high-SES children. These findings suggest overall improvements in mental health that plausibly contribute to the documented test score gains.

Our finding that AD treatment for children referred to psychiatrists improves indicators of mental health and overall health aligns with medical evidence on antidepressant efficacy discussed in Section 2.1. Our estimates offer three key advantages. First, we track individuals for three months following treatment initiation, whereas most clinical trials have shorter follow-up periods. Second, our outcomes are based on objective measures of contact with the medical system, as captured in administrative data. In contrast, medical trials typically assess effectiveness through changes in self-reported symptoms. Not only is self-reporting inherently subjective and variable across individuals, but it may also be endogenous—if AD treatment improves well-being, individuals may become better at recognizing and articulating their symptoms, potentially biasing trial-based estimates of

[53] Statistical power is limited for health outcomes due to their rarity, and for long-run outcomes because only children from earlier birth cohorts (born between 1991–1996) can be observed.

effectiveness downward. Third, clinical trials tend to be constrained to smaller samples.

6.2 Effects of antidepressant use on longer run economic outcomes

We now look at the cumulative impacts of AD treatment on human capital accumulation, employment, earnings and welfare dependency.

Enrollment beyond compulsory schooling In Table 5, we examine the effect of AD use on the likelihood of being enrolled in education at age 18. The estimate in column 1 of Panel A indicates a statistically significant increase of 42.8%. This effect is similar in magnitude and precision across gender and socioeconomic status groups (see row 1). The fact that we observe significant effects on higher education for high-SES children even though we did not observe significant increases in test scores for this group suggests that AD treatment may influence educational attainment not only through cognitive performance, but also by enhancing intrinsic motivation.

Disaggregating the results by educational track reveals meaningful heterogeneity by gender and SES. AD treatment increases the likelihood of enrollment in the academic track among girls and high-SES children, and in the vocational track among boys and low-SES children. The effects are statistically significant for all groups except high SES children, for whom the magnitude is meaningful but significance is on the margin.[54] Overall, our results suggest that improvements in mental health can enhance educational progression at a critical transition point.

Labor market outcomes and welfare receipt Table 6 shows estimates for these outcomes measured at ages 25–30 for individuals born during 1991–1996. Our results indicate that young adults who received AD treatment in childhood are significantly more likely to be employed, less likely to receive welfare or disability insurance benefits, and they report higher labor income—reflected also in higher total income. All of these long-run economic benefits are evident for the low-SES group. In the high-SES group, we observe a significant increase in total income but we interpret the estimates for this group with caution since the first-stage (shown in the penultimate row of the Table) is weak. Turning to gender differences, we find significant gains in labor income for both men and women, with larger absolute gains for men. However, increases in employment and reductions in welfare reliance are statistically significant only for men.

Overall, in line with our discussion of mutually reinforcing dynamic complementarities between mental health and human capital and labour market outcomes in Section 2.1, our findings suggest that childhood AD treatment can avert a vicious cycle with in which poor mental health inhibits economic outcomes and, in turn, weak economic outcomes feedback into poor mental health both

[54] The differences in the effects on educational track choice may reflect gender-based comparative advantage in cognitive versus physical skill-intensive paths (Bhalotra and Venkataramani, 2013; Rosenzweig and Zhang, 2013).

directly and through intermediate outcomes such as marital stability and crime. This is particularly so among low-SES children who experience higher rates of mental illness, and for whom antidepressant treatment yields large improvements in mental health, test scores, labor market outcomes and welfare dependency.

7 Conclusions

The burden of mental health disorders is rising to hitherto unseen proportions, challenging capacity in health provision, limiting labor market participation and triggering debates about disability benefits reform.[55] We draw attention to the fact that mental health disorders tend to emerge in childhood and, in many cases, to persist if left untreated. Against the backdrop of some controversy over the effectiveness of AD treatment in general, and concerns over side effects for children in particular, we estimate the impact of antidepressant use among children referred to child psychiatrists in Denmark, in particular, the impacts of adding antidepressants, given that therapy is the first line of treatment. We find that antidepressant treatment in childhood is effective in improving both health and economic outcomes, over the medium and longer term. Our estimates suggest a strong economic payoff to treating depression and anxiety, and particularly so for children from less advantaged family backgrounds. Policy simulations suggest that expanding treatment to low SES children would benefit them.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2023. "When Should You Adjust Standard Errors for Clustering?" *Quarterly Journal of Economics* 138 (1): 1–35.

Agan, Amanda, Jennifer L Doleac, and Anna Harvey. 2023. "Misdemeanor Prosecution." *Quarterly Journal of Economics* 138 (3): 1453–1505.

Aizer, Anna, and Joseph J. Doyle. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *Quarterly Journal of Economics* 130 (2): 759–803.

Almond, Douglas, Janet Currie, and Valentina Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature* 56 (4): 1360–1446.

Almond, Douglas, and Bhashkar Mazumder. 2011. "Health Capital and the Prenatal Environment: The Effect of Ramadan Observance during Pregnancy." *American Economic Journal: Applied Economics* 3 (4): 56–85.

Alonso, J., A. Buron, R. Bruffaerts, Y. He, J. Posada-Villa, J-P. Lepine, M. C. Angermeyer, et al. 2008. "Association of Perceived Stigma and Mood and Anxiety Disorders: Results from the World Mental Health Surveys." *Acta Psychiatrica Scandinavica* 118 (4): 305–314.

[55] About half of all claimants for disability benefits claim on the grounds of mental health.

Andresen, Martin Eckhoff. 2018. "Exploring Marginal Treatment Effects: Flexible Estimation Using Stata." *Stata Journal* 18 (1): 118–158.

Andrews, Isaiah, James H. Stock, and Liyang Sun. 2019. "Weak Instruments in Instrumental Variables Regression: Theory and Practice." *Annual Review of Economics* 11 (1): 727–753.

Angelucci, Manuela, and Daniel Bennett. 2024. "The Economic Impact of Depression Treatment in India: Evidence from Community-Based Provision of Pharmacotherapy." *American Economic Review* 114 (1): 169–198.

Angrist, Joshua D., and Guido W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association* 90 (430): 431–442.

Barican, Jenny Lou, Donna Yung, Christine Schwartz, Yufei Zheng, Katholiki Georgiades, and Charlotte Waddell. 2022. "Prevalence of Childhood Mental Disorders in High-Income Countries: A Systematic Review and Meta-Analysis to Inform Policymaking." *Evidence Based Mental Health* 25 (1).

Berndt, Ernst R., Robert S. Gibbons, Anton Kolotilin, and Anna Levine Taub. 2015. "The Heterogeneity of Concentrated Prescribing Behavior: Theory and Evidence from Antipsychotics." *Journal of Health Economics* 40:26–39.

Beuchert, Louise, and Anne Nandrup. 2018. "The Danish National Tests at a Glance." *Nationaløkonomisk Tidsskrift* 1.

Bhalotra, Sonia, Martin Karlsson, Therese Nilsson, and Nina Schwarz. 2022. "Infant Health, Cognitive Performance, and Earnings: Evidence from Inception of the Welfare State in Sweden." *Review of Economics and Statistics* 104 (6): 1138–1156.

Bhalotra, Sonia R., and Atheendar Venkataramani. 2013. "Cognitive Development and Infectious Disease: Gender Differences in Investments and Outcomes." IZA Discussion Paper No. 7833.

———. 2015. "Shadows of the Captain of the Men of Death: Early Life Health Interventions, Human Capital Investments, and Institutions."

Bharadwaj, Prashant, Katrine Vellesen Løken, and Christopher Neilson. 2013. "Early Life Health Interventions and Academic Achievement." *American Economic Review* 103 (5): 1862–1891.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. 2020. "Incarceration, Recidivism, and Employment." *Journal of Political Economy* 128 (4): 1269–1324.

Biasi, Barbara, Michael S. Dahl, and Petra Moser. Forthcoming. "Career Effects of Mental Health." *Journal of Political Economy Microeconomics*.

Boaden, Katharine, Anneka Tomlinson, Samuele Cortese, and Andrea Cipriani. 2020. "Antidepressants in Children and Adolescents: Meta-review of Efficacy, Tolerability and Suicidality in Acute Treatment." *Frontiers in Psychiatry* 11.

Buhl Sørensen, Christine, Ea Bøhm Jepsen, Per Hove Thomsen, and Søren Dalsgaard. 2003. "Indications for and Use of Antidepressants in Child and Adolescent Psychiatry." *European Child & Adolescent Psychiatry* 12 (3): 114–121.

Busch, Susan H., Ezra Golberstein, and Ellen Meara. 2014. "The FDA and ABCs: Unintended Consequences of Antidepressant Warnings on Human Capital." *Journal of Human Resources* 49 (3): 540–571.

Bütikofer, Aline, Christopher J. Cronin, and Meghan M. Skira. 2020. "Employment Effects of Healthcare Policy: Evidence from the 2007 FDA Black Box Warning on Antidepressants." *Journal of Health Economics* 73:102348.

Cavatorta, Elisa, Simona Grassi, and Mark Lambiris. 2021. "Digital Antianxiety Treatment and Cognitive Performance: An Experimental Study." *European Economic Review* 132:103636.

Chan, David C, Matthew Gentzkow, and Chuan Yu. 2022. "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." *Quarterly Journal of Economics* 137 (2): 729–783.

Chorniy, Anna. 2025. "AddItAll: The Effects of ADHD Treatment on Children's Well-Being."

Chyn, Eric, Brigham Frandsen, and Emily C. Leslie. Forthcoming. "Examiner and Judge Designs in Economics: A Practitioner's Guide." *Journal of Economic Literature*.

Cipriani, Andrea, Toshi A. Furukawa, Georgia Salanti, Anna Chaimani, Lauren Z. Atkinson, Yusuke Ogawa, Stefan Leucht, et al. 2018. "Comparative Efficacy and Acceptability of 21 Antidepressant Drugs for the Acute Treatment of Adults with Major Depressive Disorder: A Systematic Review and Network Meta-Analysis." *Lancet* 391 (10128): 1357–1366.

Conley, Timothy G, Christian B Hansen, and Peter E Rossi. 2012. "Plausibly exogenous." *Review of Economics and Statistics* 94 (1): 260–272.

Corcoran, Roisin P., Alan C. K. Cheung, Elizabeth Kim, and Chen Xie. 2018. "Effective Universal School-Based Social and Emotional Learning Programs for Improving Academic Achievement: A Systematic Review and Meta-Analysis of 50 Years of Research." *Educational Research Review* 25:56–72.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2018. "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance." *Journal of Political Economy* 126 (6): 2356–2409.

Costa-Ramón, Ana, N. Meltem Daysal, and Ana Rodriguez-Gonzalez. 2023. "The Oral Contraceptive Pill and Adolescents' Mental Health." CEPR Discussion Paper no. 18269.

Cowan, Benjamin W., and Zhuang Hao. 2021. "Medicaid Expansion and the Mental Health of College Students." *Health Economics* 30 (6): 1306–1327.

Cuddy, Emily, and Janet Currie. 2020. "Treatment of Mental Illness in American Adolescents Varies Widely within and across Areas." *Proceedings of the National Academy of Sciences* 117 (39): 24039–24046.

———. Forthcoming. "Rules vs. Discretion: Treatment of Mental Illness in U.S. Adolescents." *Journal of Political Economy*.

Currie, Janet. 2024. "The Economics of Child Mental Health: Introducing the Causes and Consequences of Child Mental Health Special Issue." *Journal of Human Resources* 59 (S): S1–S13.

———. 2025. "Investing in Children to Address the Child Mental Health Crisis."

Currie, Janet, Paul Kurdyak, and Jonathan Zhang. 2024. "Socioeconomic Status and Access to Mental Health Care: The Case of Psychiatric Medications for Children in Ontario Canada." *Journal of Health Economics* 93:102841.

Currie, Janet, and W. Bentley MacLeod. 2020. "Understanding Doctor Decision Making: The Case of Depression Treatment." *Econometrica* 88 (3): 847–878.

Currie, Janet, and Mark Stabile. 2006. "Child Mental Health and Human Capital Accumulation: The Case of ADHD." *Journal of Health Economics* 25 (6): 1094–1118.

———. 2009. "Mental Health in Childhood and Human Capital." Edited by Jonathan Gruber. In *The Problems of Disadvantaged Youth: An Economic Perspective*, 115–148. University of Chicago Press.

Currie, Janet, Mark Stabile, and Lauren Jones. 2014. "Do Stimulant Medications Improve Educational and Behavioral Outcomes for Children with ADHD?" *Journal of Health Economics* 37:58–69.

———. 2010. "Child Health and Young Adult Outcomes." *Journal of Human Resources* 45 (3): 517–548.

Currie, Janet, and Esmée Zwiers. Forthcoming. "Medication of Postpartum Depression and Maternal Outcomes: Evidence from Geographic Variation in Dutch Prescribing." *Journal of Human Resources*.

Cutler, David, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg. 2019. "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending." *American Economic Journal: Economic Policy* 11 (1): 192–221.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad. 2014. "Family Welfare Cultures." *Quarterly Journal of Economics* 129 (4): 1711–1752.

Dahl, Ronald E., Nicholas B. Allen, Linda Wilbrecht, and Ahna Ballonoff Suleiman. 2018. "Importance of Investing in Adolescence from a Developmental Science Perspective." *Nature* 554 (7693): 441–450.

Dalsgaard, Søren, Helena Skyt Nielsen, and Marianne Simonsen. 2014. "Consequences of ADHD Medication Use for Children's Outcomes." *Journal of Health Economics* 37:137–151.

Danish Health Authority. 2022. *Strengthening Mental Health Care. Recommendations for a 10-Year Action Plan in Denmark*. Copenhagen: Danish Health Authority.

Daysal, N. Meltem, Marianne Simonsen, Mircea Trandafir, and Sanni Breining. 2022. "Spillover Effects of Early-Life Medical Interventions." *Review of Economics and Statistics* 104 (1): 1–16.

De Quidt, Jonathan, and Johannes Haushofer. 2016. "Depression for Economists." NBER Working Paper no. 22973.

Devereux, Paul J. 2022. "Fragility of the Marginal Treatment Effect." University College Dublin Working Paper WP22/04.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108 (2): 201–240.

Doyle, Joseph J. 2007. "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review* 97 (5): 1583–1610.

———. 2008. "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care." *Journal of Political Economy* 116 (4): 746–770.

Dwyer, Jennifer B., and Michael H. Bloch. 2019. "Antidepressants for Pediatric Patients." *Current Psychiatry* 18 (9): 26–42F.

Eichmeyer, Sarah, and Jonathan Zhang. 2022. "Pathways into Opioid Addiction: Evidence from Practice Variation in Emergency Departments." *American Economic Journal: Applied Economics* 14 (4): 271–300.

———. 2023. "Primary Care Providers' Influence on Opioid Use and Its Adverse Consequences." *Journal of Public Economics* 217:104784.

Elwadhi, Deeksha, and Alex Cohen. 2020. "Social Inequalities in Antidepressant Treatment Outcomes: A Systematic Review." *Social Psychiatry and Psychiatric Epidemiology* 55, no. 10 (10): 1241–1259.

Evans-Lacko, S., S. Aguilar-Gaxiola, A. Al-Hamzawi, J. Alonso, C. Benjet, R. Bruffaerts, W. T. Chiu, et al. 2018. "Socio-Economic Variations in the Mental Health Treatment Gap for People with Anxiety, Mood, and Substance Use Disorders: Results from the WHO World Mental Health (WMH) Surveys." *Psychological Medicine* 48 (9): 1560–1571.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics* 131 (4): 1681–1726.

Fletcher, Jason, and Barbara Wolfe. 2008. "Child Mental Health and Human Capital Accumulation: The Case of ADHD Revisited." *Journal of Health Economics* 27 (3): 794–800.

Fornaro, Michele, Annalisa Anastasia, Alessandro Valchera, Alessandro Carano, Laura Orsolini, Federica Vellante, Gabriella Rapini, et al. 2019. "The FDA "Black Box" Warning on Antidepressant Suicide Risk in Young Adults: More Harm than Benefits?" *Frontiers in Psychiatry* 10:294.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie. 2023. "Judging Judge Fixed Effects." *American Economic Review* 113 (1): 253–277.

Galaif, Elisha R, Steve Sussman, Michael D Newcomb, and Thomas F Locke. 2007. "Suicidality, Depression, and Alcohol Use among Adolescents: A Review of Empirical Findings." *International Journal of Adolescent Medicine and Health* 19 (1): 27–36.

Giedd, Jay N., Jonathan Blumenthal, Neal O. Jeffries, F. X. Castellanos, Hong Liu, Alex Zijdenbos, Tomáš Paus, Alan C. Evans, and Judith L. Rapoport. 1999. "Brain Development during Childhood and Adolescence: A Longitudinal MRI Study." *Nature Neuroscience* 2 (10): 861–863.

Gogtay, Nitin, Jay N. Giedd, Leslie Lusk, Kiralee M. Hayashi, Deanna Greenstein, A. Catherine Vaituzis, Tom F. Nugent, et al. 2004. "Dynamic Mapping of Human Cortical Development during Childhood through Early Adulthood." *Proceedings of the National Academy of Sciences* 101 (21): 8174–8179.

Heckman, James J. 2007. "The Economics, Technology, and Neuroscience of Human Capability Formation." *Proceedings of the National Academy of Sciences* 104 (33): 13250–13255.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Review of Economics and Statistics* 88 (3): 389–432.

Heckman, James J., and Edward Vytlacil. 2001. "Policy-Relevant Treatment Effects." *American Economic Review* 91 (2): 107–111.

———. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica* 73 (3): 669–738.

———. 2007. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments." Edited by James J. Heckman and Edward E. Leamer. In *Handbook of Econometrics*, Chapter 71, 4875–5143. Vol. 6. Elsevier.

Hendren, Nathaniel, and Ben Sprung-Keyser. 2020. "A Unified Welfare Analysis of Government Policies." *Quarterly Journal of Economics* 135 (3): 1209–1318.

Isgor, Ceylan, Mohamed Kabbaj, Huda Akil, and Stanley J. Watson. 2004. "Delayed Effects of Chronic Variable Stress during Peripubertal-Juvenile Period on Hippocampal Morphology and on Cognitive and Stress Axis Functions in Rats." *Hippocampus* 14 (5): 636–648.

Jack, Ruth H., Rebecca M. Joseph, Carol Coupland, Debbie Butler, Chris Hollis, Richard Morriss, Roger David Knaggs, Andrea Cipriani, Samuele Cortese, and Julia Hippisley-Cox. 2020. "Secondary Care Specialist Visits Made by Children and Young People Prescribed Antidepressants in Primary Care: A Descriptive Study Using the QResearch Database." *BMC Medicine* 18 (1): 93.

Kaplan, David M. 2022. "Smoothed Instrumental Variables Quantile Regression." *Stata Journal* 22 (2): 379–403.

Kaplan, David M., and Yixiao Sun. 2017. "Smoothed Estimating Equations for Instrumental Variables Quantile Regression." *Econometric Theory* 33 (1): 105–157.

Kessler, Ronald C., Patricia Berglund, Olga Demler, Robert Jin, Kathleen R. Merikangas, and Ellen E. Walters. 2005. "Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication." *Archives of General Psychiatry* 62 (6): 593.

Kessler, Ronald C., and Evelyn J. Bromet. 2013. "The Epidemiology of Depression across Cultures." *Annual Review of Public Health* 34:119–138.

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review* 96 (3): 863–876.

Laird, Jessica, and Torben Heien Nielsen. 2016. "The Effects of Physician Prescribing Behaviors on Prescription Drug Use and Labor Supply: Evidence from Movers in Denmark."

Lee, David S., Justin McCrary, Marcelo J. Moreira, and Jack Porter. 2022. "Valid T-Ratio Inference for IV." *American Economic Review* 112 (10): 3260–3290.

Lewis, Daniel J., and Karel Mertens. Forthcoming. "A Robust Test for Weak Instruments for 2SLS with Multiple Endogenous Regressors." *Review of Economic Studies*.

Lu, Christine Y., Fang Zhang, Matthew D. Lakoma, Jeanne M. Madden, Donna Rusinak, Robert B. Penfold, Gregory Simon, et al. 2014. "Changes in Antidepressant Use by Young People and Suicidal Behavior after FDA Warnings and Media Coverage: Quasi-Experimental Study." *BMJ* 348:g3596.

Ludwig, Jens, Dave E. Marcotte, and Karen Norberg. 2009. "Anti-Depressants and Suicide." *Journal of Health Economics* 28 (3): 659–676.

Lund, Crick, Kate Orkin, Marc Witte, John H. Walker, Thandi Davies, Johannes Haushofer, Sarah Murray, et al. 2024. "The Effects of Mental Health Interventions on Labor Market Outcomes in Low- and Middle-Income Countries." NBER Working Paper no. 32423.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review* 103 (5): 1797–1829.

Marie, Olivier, and Ulf Zölitz. 2017. ""High" Achievers? Cannabis Access and Academic Performance." *Review of Economic Studies* 84 (3): 1210–1237.

Marquardt, Kelli. 2021. "Identifying Physician Practice Style for Mental Health Conditions."

Mathiesen, Pernille, Sabine Frølich Maarbjerg, Kirsten Lykke, and Thomas Balslev. 2016. "The Child Health System in Denmark: Current Problems and Successes." *Journal of Pediatrics* 177:S60–S62.

Moitra, Modhurima, Damian Santomauro, Pamela Y. Collins, Theo Vos, Harvey Whiteford, Shekhar Saxena, and Alize J. Ferrari. 2022. "The Global Gap in Treatment Coverage for Major Depressive Disorder in 84 Countries from 2000–2019: A Systematic Review and Bayesian Meta-Regression Analysis." *PLOS Medicine* 19 (2): e1003901.

Montiel Olea, José Luis, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics* 31 (3): 358–369.

Mueller-Smith, Michael. 2015. "The Criminal and Labor Market Impacts of Incarceration."

Norris, Samuel. 2022. "Measuring Examiner Consistency and Skill: Evidence from Refugee Decisions."

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver. 2021. "The Effects of Parental and Sibling Incarceration: Evidence from Ohio." *American Economic Review* 111 (9): 2926–2963.

Pottegård, Anton, Helga Zoëga, Jesper Hallas, and Per Damkier. 2014. "Use of SSRIs among Danish Children: A Nationwide Study." *European Child & Adolescent Psychiatry* 23 (12): 1211–1218.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel. 2020. "Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms." *Science* 370 (6522): eaay0214.

Rihmer, Zoltán. 2007. "Suicide Risk in Mood Disorders." *Current Opinion in Psychiatry* 20 (1): 17.

Rosenzweig, Mark R., and Junsen Zhang. 2013. "Economic Growth, Comparative Advantage, and Gender Differences in Schooling Outcomes: Evidence from the Birthweight Differences of Chinese Twins." *Journal of Development Economics* 104:245–260.

Sacerdote, Bruce. 2011. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" Edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. In *Handbook of the Economics of Education*, 249–277. Vol. 3. Elsevier.

Schröder, Carsten, Michael Dörks, Bianca Kollhorst, Tilo Blenk, Ralf W. Dittmann, Edeltraut Garbe, and Oliver Riedel. 2017. "Outpatient Antidepressant Drug Use in Children and Adolescents in Germany between 2004 and 2011." *Pharmacoepidemiology and Drug Safety* 26 (2): 170–179.

Shapiro, Bradley T. 2022. "Promoting Wellness or Waste? Evidence from Antidepressant Advertising." *American Economic Journal: Microeconomics* 14 (2): 439–477.

Sigstad, Henrik. Forthcoming. "Monotonicity among Judges: Evidence from Judicial Panels and Consequences for Judge IV Designs." *American Economic Review*.

Sørensen, Johanne Østerby, Annette Rasmussen, Troels Roesbjerg, and Anne Katrine Pagsberg. 2020. "Clinician Compliance to Recommendations Regarding the Risk of Suicidality with Selective Serotonin Reuptake Inhibitors in the Treatment of Children and Adolescents." *European Child & Adolescent Psychiatry* 29 (5): 707–718.

Stephens, Mary Ann C., and Gary Wand. 2012. "Stress and the HPA Axis." *Alcohol Research : Current Reviews* 34 (4): 468–483.

Sun, Liyang. 2018. "Implementing Valid Two-Step Identification-Robust Confidence Sets for Linear Instrumental-Variables Models." *Stata Journal* 18 (4): 803–825.

Wesselhoeft, R., P. B. Jensen, A. Talati, J. Reutfors, K. Furu, K. Strandberg-Larsen, P. Damkier, A. Pottegård, and M. Bliddal. 2020. "Trends in Antidepressant Use among Children and Adolescents: A Scandinavian Drug Utilization Study." *Acta Psychiatrica Scandinavica* 141 (1): 34–42.

World Health Organization. 2022. *World Mental Health Report: Transforming Mental Health for All*. Geneva: World Health Organization.

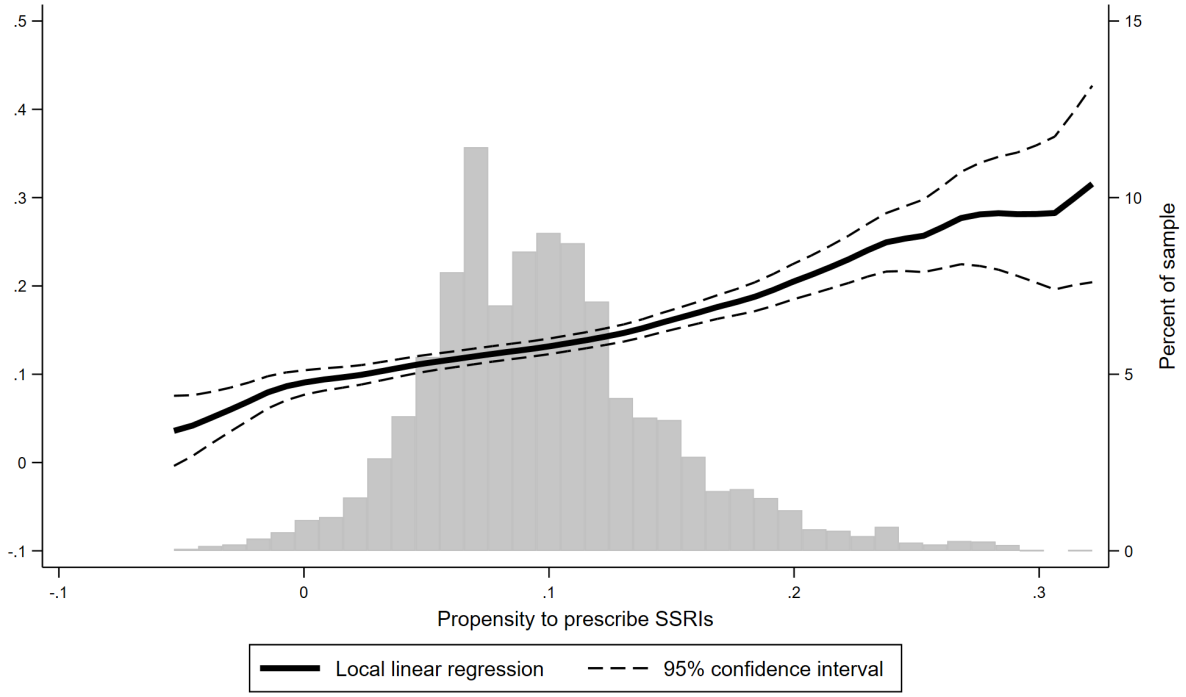


Figure 1: Distribution of specialist tendency to prescribe ADs and the probability of receiving ADs by age 15

Notes: Both the propensity to prescribe AD and the probability of filling a prescription by age 15 are net of year-of-diagnosis and municipality fixed effects. The gray bars represent the histogram of the adjusted tendency to prescribe antidepressants. The solid line plots the local polynomial regression of the probability of receiving antidepressants by age 15 on the tendency to prescribe antidepressants of the treating specialist, while the dashed lines indicate the corresponding 95% confidence interval.

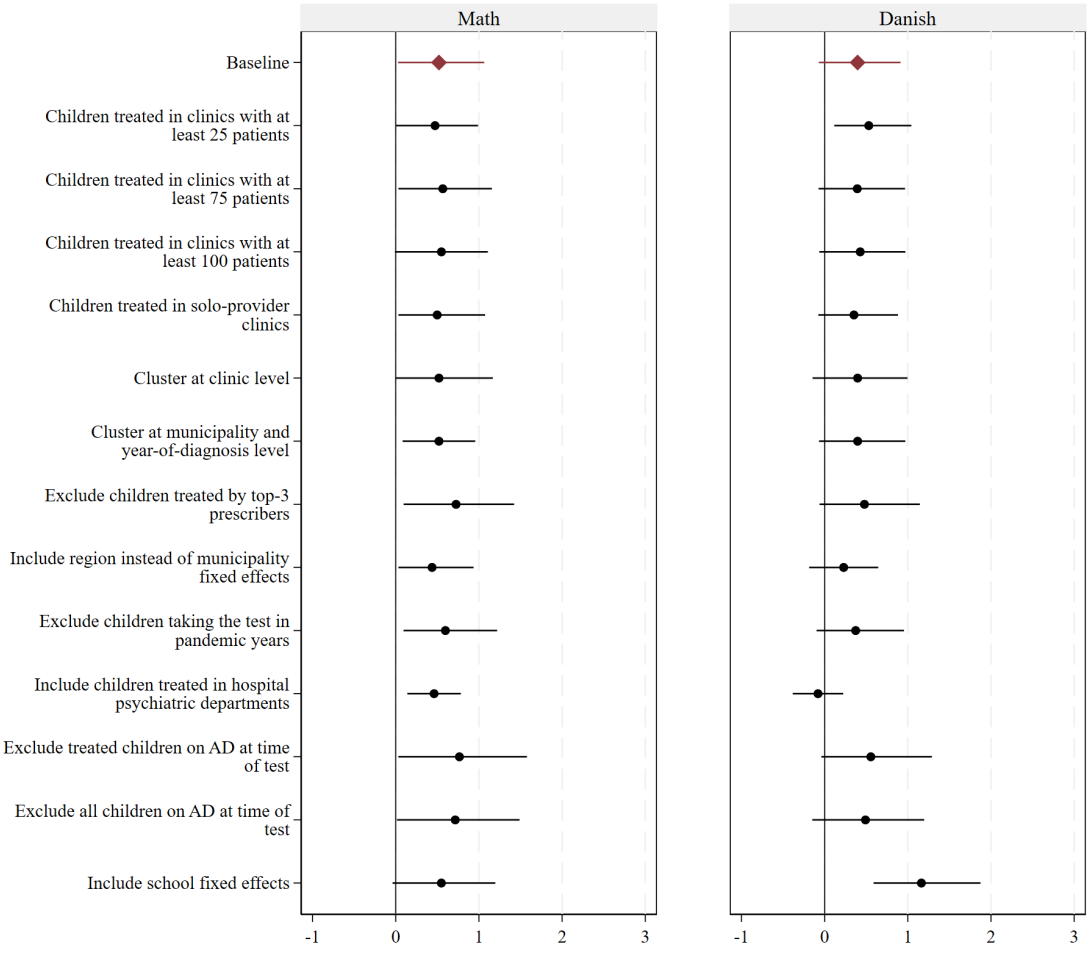
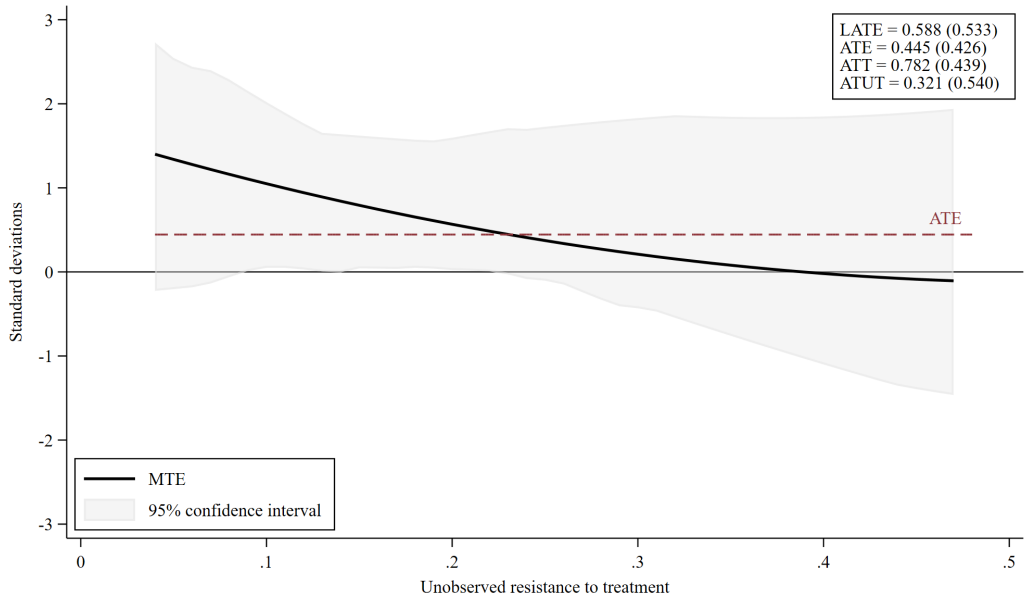
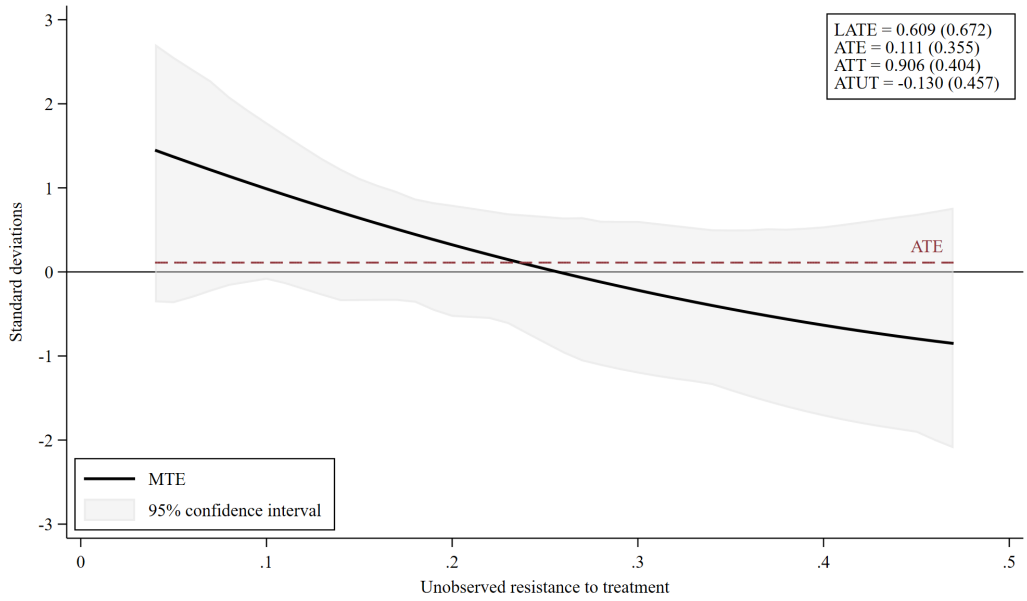


Figure 2: Robustness of the effect of AD use on test scores

Notes: Each panel plots the estimated effect of AD use and corresponding Anderson-Rubin 95% confidence intervals from a separate 2SLS regression of the outcome listed in the panel heading, in the sample or with the specification indicated on the vertical axis. Unless otherwise mentioned, all specifications include the full set of control variables described in the notes to Table 2 and the standard errors are clustered at the clinic-year of diagnosis level.



(a) Math

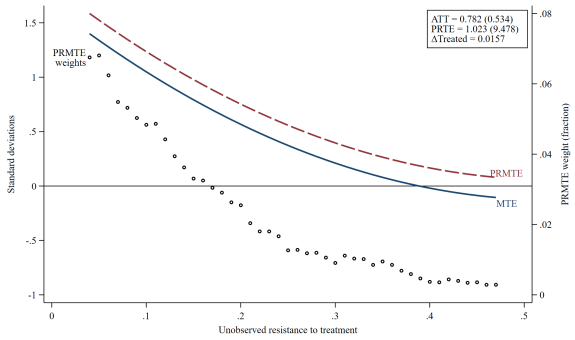


(b) Danish

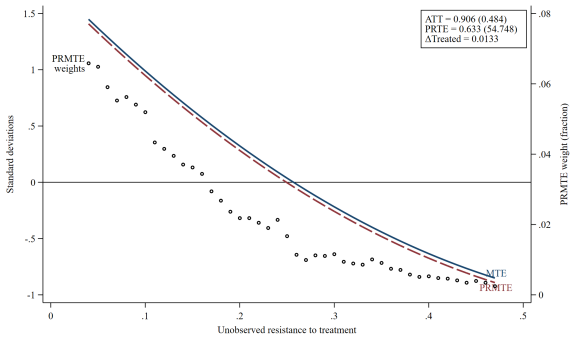
Figure 3: Marginal treatment effects

Notes: The solid line in each panel plots the marginal treatment effects (MTE) curve for the standardized test score indicated in the panel and the shaded area represent the 95% confidence interval obtained from 100 replications of a Bayesian bootstrap clustered at the clinic-year of diagnosis level. The horizontal dashed line shows the average treatment effect calculated by averaging the MTEs over the relevant range. The MTE is estimated via local instrumental variables from a specification including the full set of controls listed in the notes to Table 2. The MTE curve is assumed to be a polynomial of second degree in unobserved resistance to treatment and is estimated over the range of common support trimmed by 1% at both tails.

I. Increase the prescribing rate for low SES children to the level of high SES children

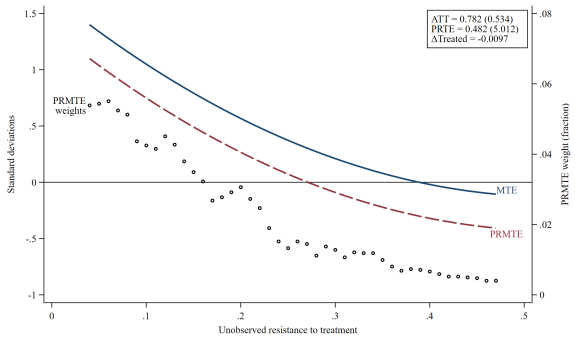


(a) Math

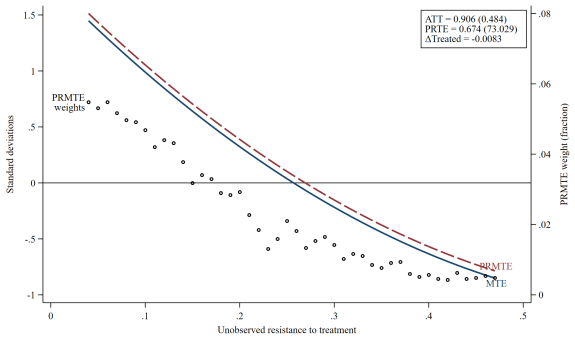


(b) Danish

II. Reduce the prescribing rate for high SES children to the level of low SES children



(c) Math

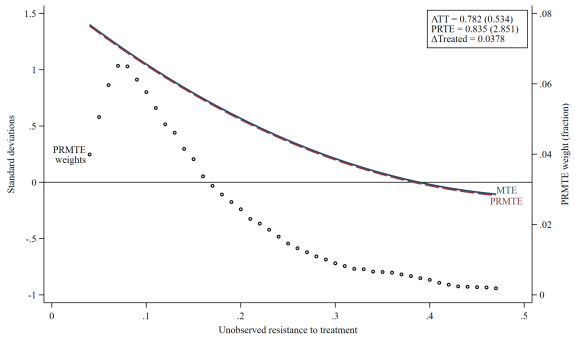


(d) Danish

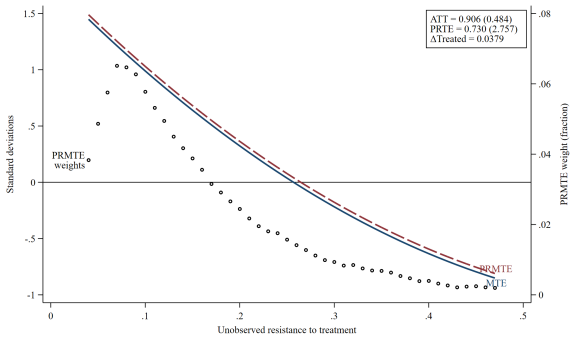
Figure 4: Policy experiments

Notes: The two lines in each panel plot the marginal treatment effects (MTE) curves under the *status quo* (solid line) and for the children whose treatment status is changed by the new policy (dashed line), for the standardized test score indicated in the panel. The MTE curves are estimated as detailed in the notes to Figure 3. The dots show the fraction of the children affected by the policy at each level of unobserved resistance to treatment. Each panel reports the average treatment effect on the treated (ATT) under the *status quo*, the average treatment effect among the children pushed into or out of treatment with AD by the policy (PRTE), and the fraction of children affected by the policy (ΔTreated).

I. Increase the prescribing rate for boys to the level of girls

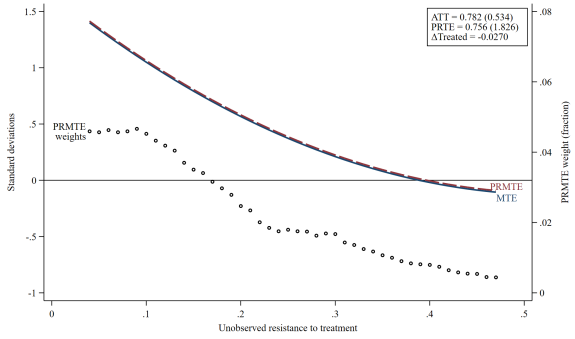


(a) Math

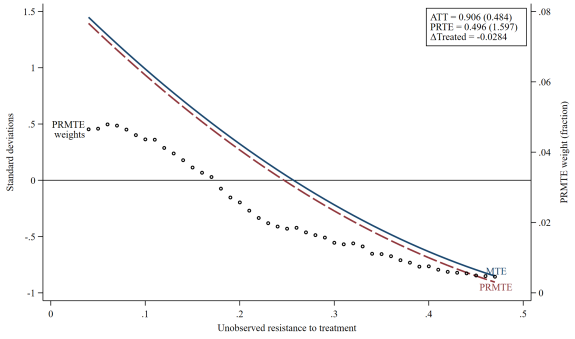


(b) Danish

II. Reduce the prescribing rate for girls to the level of boys



(c) Math



(d) Danish

Figure 5: Policy experiments

Notes: The two lines in each panel plot the marginal treatment effects (MTE) curves under the *status quo* (solid line) and for the children whose treatment status is changed by the new policy (dashed line), for the standardized test score indicated in the panel. The MTE curves are estimated as detailed in the notes to Figure 3. The dots show the fraction of the children affected by the policy at each level of unobserved resistance to treatment. Each panel reports the average treatment effect on the treated (ATT) under the *status quo*, the average treatment effect among the children pushed into or out of treatment with AD by the policy (PRTE), and the fraction of children affected by the policy (ΔTreated).

Table 1: Characteristics of the analysis sample

	No mental health contact (1)	Analysis sample All (2)	No SSRI (3)	SSRI (4)	*p*-value (5)
A. Outcomes					
Standardized test score, math	0.093	−0.283	−0.305	−0.156	0.000
	(0.968)	(1.016)	(1.021)	(0.982)	
Standardized test score, Danish	0.074	−0.199	−0.241	0.042	0.000
	(0.983)	(1.025)	(1.019)	(1.026)	
B. Treatments					
Therapy within 1 year before first MH visit	—	0.116	0.094	0.243	0.000
Therapy after first MH visit, by age 15	—	0.713	0.687	0.858	0.000
Pharmaceutical treatments by age 15:					
SSRI	0.001	0.152	0.000	1.000	—
Tricyclic antidepressant	0.001	0.002	0.002	0.003	0.866
Benzodiazepine	0.006	0.015	0.012	0.032	0.000
Stimulant	0.005	0.341	0.351	0.284	0.000
Antipsychotic	0.001	0.075	0.052	0.203	0.000
Any non-SSRI N-drug	0.062	0.462	0.446	0.552	0.000
Age when first SSRI was filled	13.481	12.674	—	12.674	
	(1.980)	(1.778)		(1.778)	
Median days between first MH visit and:					
First SSRI prescription	—	106.500	—	106.500	
First N-class non-SSRI prescription	—	−23.000	—	−23.000	
First therapy	—	28.000	28.000	28.000	0.140
Median days b/w first therapy and first SSRI	—	125.000	—	125.000	
C. Child characteristics					
Boy	0.504	0.596	0.615	0.489	0.000
Birth order:					
1	0.453	0.490	0.496	0.460	0.031
2	0.374	0.364	0.361	0.382	0.201
3+	0.174	0.145	0.143	0.158	0.210
Number of siblings:					
0	0.128	0.187	0.189	0.173	0.189
1	0.561	0.560	0.560	0.563	0.860
2	0.223	0.178	0.174	0.195	0.121
3+	0.087	0.076	0.077	0.070	0.440
Parents are married	0.810	0.680	0.665	0.758	0.000
High SES	0.365	0.352	0.344	0.401	0.000
Age at first diagnosis	—	11.637	11.583	11.942	0.000
		(2.069)	(2.079)	(1.984)	

Table 1 (cont.): Characteristics of the analysis sample

	No mental health contact (1)	Analysis sample All (2)	No SSRI (3)	SSRI (4)	*p*-value (5)
D. Psychiatrist characteristics					
Male	—	0.413	0.413	0.412	0.960
Age (years)	—	57.075	57.110	56.884	0.168
		(5.232)	(5.283)	(4.930)	
Experience (years)	—	27.934	27.893	28.164	0.118
		(5.612)	(5.682)	(5.205)	
Married	—	0.657	0.658	0.655	0.840
Solo practice	—	0.793	0.792	0.803	0.392
Number of patients treated during year	—	169.781	170.157	167.685	0.340
		(79.220)	(79.271)	(78.937)	
Propensity to prescribe SSRIs	—	0.105	0.101	0.130	0.000
		(0.061)	(0.059)	(0.066)	
E. Mother's characteristics at child age 6					
Age	35.879	35.708	35.598	36.321	0.000
	(4.629)	(5.066)	(5.116)	(4.731)	
Some college or more	0.365	0.352	0.344	0.401	0.000
Employed	0.852	0.780	0.774	0.816	0.001
Annual gross total income	333.004	326.692	325.940	330.897	0.262
	(189.383)	(150.718)	(154.009)	(130.803)	
Used SSRI between child ages 6–7	0.086	0.167	0.161	0.201	0.002
F. Father's characteristics at child age 6					
Age	38.247	38.122	38.025	38.667	0.000
	(5.405)	(5.732)	(5.755)	(5.573)	
Some college or more	0.296	0.280	0.274	0.315	0.007
Employed	0.932	0.887	0.881	0.916	0.000
Annual gross total income	450.972	427.379	425.405	438.405	0.151
	(561.056)	(331.281)	(342.182)	(262.040)	
Used SSRI between child ages 6–7	0.050	0.088	0.086	0.097	0.270
Observations	770,520	7,211	6,116	1,095	

Notes: Column 1 describes the sample of all native children born between 1991–2005, who are observed every year between the ages of 6 and 16, and who have no contact with a psychiatrist between ages 6–15. Columns 2 describes our analysis sample of all native children born between 1991–2005, who are observed every year between the ages of 6 and 16, who have no contact with a psychiatrist between ages 6–7, and whose first psychiatric visit between ages 8–15 is with a child psychiatrist in a private clinic. Columns 3 and 4 split the analysis sample by whether the child filled an AD prescription by age 15 (column 4) or not (column 3). The *p*-value in Column 5 corresponds to a test of equality of the average characteristics of children treated and not treated with AD. Means and standard deviations in brackets, unless otherwise indicated. Number of siblings refers to the number of siblings in the household when the child is aged 6. Children of mothers with short cycle higher education or above are classified as high SES. Parents with short cycle higher education or above are classified as high SES. All parental characteristics, including marital status, are measured when the child is aged 6. All monetary variables are in 2015 DKK. See Appendix Table A1 for details on how each variable is constructed.

Table 2: Effect of AD use in childhood on 9th grade test scores

	Standardized test scores	
	Math	Danish
	(1)	(2)
A. First stage (dependent variable: Ever filled an SSRI prescription by age 15)		
Propensity to prescribe SSRIs	0.928***	0.920***
	(0.102)	(0.101)
Effect of interquartile change in instrument	0.074	0.073
Mean of outcome	0.147	0.147
First-stage F statistic	82.8	82.2
Observations	5,495	5,383
B. Reduced form		
Propensity to prescribe SSRIs	0.482**	0.365
	(0.240)	(0.224)
Effect of interquartile change in instrument	0.038	0.029
Mean of outcome	−0.283	−0.199
Observations	5,495	5,383
C. Two-stage least squares		
Ever filled an SSRI prescription by age 15	0.520**	0.397*
	(0.266)	(0.253)
	[0.028, 1.063]	[−0.070, 0.913]
Mean of outcome	−0.305	−0.241
Observations	5,495	5,383

Notes: Analysis sample as described in the notes to Table 1. Each cell presents the results from a separate regression of the outcome listed in the column for the specification listed in the row. All specification include: indicators for child gender, birth order, family size; indicators for mother's and father's highest education degree attained, employment status, income deciles, and marriage status at child age 6; indicators for mother's and father's use of AD during child age 6–7; mother's and father's age; and fixed effects for year of birth of the child, year of diagnosis (first psychiatric visit), and municipality of residence during the year of diagnosis. The mean of the outcome is calculated among children not using AD in Panels A and D, and among all children in Panels B and C. In addition to the coefficient estimate, Panels B and C list the effect on the outcome of an increase in the value of the instrument from the 25th to the 75th percentile in the full analysis sample, corresponding to approximately 7.9 percentage points. Panel D reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance (based on Anderson-Rubin confidence intervals in Panel C) at the 1%, 5%, and 10% levels, respectively.

Table 3: Effect of AD use in childhood on 9th grade test scores in subpopulations

	All	Gender		Socioeconomic status	
		Girls	Boys	Low SES	High SES
	(1)	(2)	(3)	(4)	(5)
A. 2SLS estimates					
Standardized test score, math	0.520**	0.541*	0.575	0.720**	0.236
	(0.266)	(0.316)	(0.411)	(0.428)	(0.336)
	[0.028, 1.063]	[−0.031, 1.173]	[−0.175, 1.325]	[0.020, 1.669]	[−0.371, 0.843]
First-stage F statistic	82.8	52.9	53.0	37.3	55.6
Mean of outcome	−0.305	−0.326	−0.292	−0.518	0.050
Observations	5,495	2,310	3,185	3,391	2,104
Standardized test score, Danish	0.397*	0.392	0.465	0.333	0.408
	(0.253)	(0.332)	(0.392)	(0.393)	(0.344)
	[−0.070, 0.913]	[−0.209, 0.992]	[−0.250, 1.255]	[−0.386, 1.128]	[−0.277, 1.028]
First-stage F statistic	82.2	53.0	48.2	35.3	54.6
Mean of outcome	−0.241	0.128	−0.500	−0.431	0.074
Observations	5,383	2,326	3,057	3,307	2,076
B. Reduced form estimates					
Standardized test score, math	0.482**	0.618*	0.468	0.546*	0.284
	(0.240)	(0.358)	(0.336)	(0.301)	(0.404)
Effect of interquartile change in instrument	0.038	0.049	0.037	0.043	0.023
Mean of outcome	−0.283	−0.294	−0.276	−0.488	0.048
Observations	5,495	2,310	3,185	3,391	2,104
Standardized test score, Danish	0.365	0.442	0.376	0.251	0.482
	(0.224)	(0.377)	(0.307)	(0.290)	(0.418)
Effect of interquartile change in instrument	0.029	0.035	0.030	0.020	0.038
Mean of outcome	−0.199	0.157	−0.469	−0.394	0.113
Observations	5,383	2,326	3,057	3,307	2,076

Notes: Analysis sample as described in the notes to Table 1. Each cell presents the results from a separate regression of the outcome listed in the row for the specification listed in the panel title, estimated in the sample indicated in the column. Children of mothers with short cycle higher education or above are classified as high SES. All specifications include the full set of controls listed in the notes to Table 2. The mean of the outcome is calculated among all children in Panel A and among children not using AD in Panel B. In addition to the coefficient estimate, Panel B lists the effect on the outcome of an increase in the value of the instrument from the 25th to the 75th percentile in the full analysis sample, corresponding to approximately 7.9 percentage points. Panel A reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance (based on Anderson-Rubin confidence intervals in Panel A) at the 1%, 5%, and 10% levels, respectively.

Table 4: Side effects of AD use on hospital visits within 3 months of AD initiation, all children diagnosed with a mental health disorder

	All	Gender		Socioeconomic status	
		Girls	Boys	Low SES	High SES
	(1)	(2)	(3)	(4)	(5)
Suicide attempt or self-harm	−0.029***	−0.052***	0.005	−0.037***	−0.015
	(0.012)	(0.021)	(0.005)	(0.015)	(0.012)
	[−0.053, −0.010]	[−0.097, −0.016]	[−0.004, 0.014]	[−0.068, −0.012]	[−0.039, 0.005]
Mean of outcome	0.003	0.008	0.000	0.004	0.002
Any hospital visit	−0.217***	−0.303***	−0.106	−0.142	−0.365***
	(0.091)	(0.119)	(0.114)	(0.105)	(0.135)
	[−0.405, −0.047]	[−0.550, −0.103]	[−0.320, 0.109]	[−0.360, 0.055]	[−0.671, −0.138]
Mean of outcome	0.197	0.239	0.170	0.204	0.179
Outpatient visit	−0.189**	−0.216**	−0.129	−0.137	−0.280***
	(0.078)	(0.101)	(0.106)	(0.092)	(0.122)
	[−0.336, −0.042]	[−0.424, −0.047]	[−0.348, 0.070]	[−0.309, 0.036]	[−0.531, −0.077]
Mean of outcome	0.156	0.192	0.133	0.160	0.146
ER visit	−0.065	−0.188**	0.054	−0.045	−0.118*
	(0.059)	(0.087)	(0.064)	(0.073)	(0.071)
	[−0.187, 0.034]	[−0.368, −0.042]	[−0.066, 0.174]	[−0.195, 0.091]	[−0.264, 0.001]
Mean of outcome	0.054	0.065	0.046	0.058	0.041
Inpatient visit	−0.015	−0.053	0.040	−0.001	−0.040
	(0.035)	(0.051)	(0.041)	(0.043)	(0.051)
	[−0.080, 0.051]	[−0.159, 0.043]	[−0.037, 0.118]	[−0.081, 0.080]	[−0.144, 0.054]
Mean of outcome	0.030	0.046	0.020	0.031	0.028
First-stage F statistic	129.9	84.3	97.3	108.2	57.5
Observations	52,360	21,658	30,702	37,636	14,724

Notes: Sample of all native children born between 1991–2005, who are observed every year between the ages of 6 and 16, who have no contact with a psychiatrist between ages 6–7, and whose first psychiatric visit between ages 8–15 is with a child psychiatrist in a private clinic of with a hospital psychiatric department. Each cell presents the results from a separate regression of the outcome listed in the row estimated in the sample indicated in the column. Children not using SSRI are assigned an initiation date based on the average age at AD initiation of children of the same gender diagnosed in the same year. Children of mothers with short cycle higher education or above are classified as high SES. All specification include the set of controls described in the notes of Table 2. The mean of the outcome is calculated among children not using SSRI. In addition to the coefficient estimate, each cell reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

Table 5: Effects of AD use on school enrollment at age 18, sample of all children diagnosed with a mental health disorder at age 8-15

	All	Gender		Socioeconomic status	
		Girls	Boys	Low SES	High SES
	(1)	(2)	(3)	(4)	(5)
Education beyond compulsory schooling	0.278***	0.297***	0.294*	0.257**	0.302**
	(0.112)	(0.130)	(0.157)	(0.136)	(0.153)
	[0.090, 0.509]	[0.079, 0.566]	[0.000, 0.618]	[0.029, 0.540]	[0.047, 0.617]
Mean of outcome	0.648	0.653	0.644	0.614	0.748
Academic track	0.091	0.242**	−0.022	0.027	0.238
	(0.105)	(0.123)	(0.141)	(0.121)	(0.157)
	[−0.107, 0.288]	[0.036, 0.497]	[−0.286, 0.242]	[−0.199, 0.254]	[−0.055, 0.562]
Mean of outcome	0.432	0.499	0.386	0.376	0.598
Vocational track	0.187**	0.055	0.316**	0.230**	0.064
	(0.084)	(0.078)	(0.135)	(0.109)	(0.116)
	[0.029, 0.361]	[−0.092, 0.217]	[0.063, 0.595]	[0.026, 0.455]	[−0.152, 0.302]
Mean of outcome	0.216	0.154	0.258	0.238	0.150
First-stage F statistic	109.8	74.2	78.7	91.7	48.4
Observations	42,894	18,189	24,705	31,579	11,315

Notes: Sample of all native children born between 1991–2005, who are observed every year between the ages of 6 and 16, who have no contact with a psychiatrist between ages 6–7, and whose first psychiatric visit between ages 8–15 is with a child psychiatrist in a private clinic of with a hospital psychiatric department. Each cell presents the results from a separate regression of the outcome listed in the row estimated in the sample indicated in the column. Children not using SSRI are assigned an initiation date based on the average age at AD initiation of children of the same gender diagnosed in the same year. Children of mothers with short cycle higher education or above are classified as high SES. All specification include the set of controls described in the notes of Table 2. The mean of the outcome is calculated among children not using SSRI. In addition to the coefficient estimate, each cell reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

Table 6: Effect of AD use on labor market outcomes at ages 25–30, sample of all children diagnosed with a mental health disorder at age 8–15

	All	Gender		Socioeconomic status	
		Girls	Boys	Low SES	High SES
	(1)	(2)	(3)	(4)	(5)
Employed	0.472***	0.133	0.752***	0.554***	0.039
	(0.154)	(0.164)	(0.255)	(0.170)	(0.340)
	[0.214, 0.822]	[−0.174, 0.504]	[0.324, 1.382]	[0.269, 0.941]	[−0.662, 1.208]
Mean of outcome	0.469	0.430	0.494	0.474	0.448
On welfare or disability benefits	−0.312***	−0.089	−0.542***	−0.405***	0.133
	(0.138)	(0.158)	(0.222)	(0.154)	(0.313)
	[−0.599, −0.080]	[−0.416, 0.207]	[−1.089, −0.171]	[−0.756, −0.146]	[−0.698, 0.963]
Mean of outcome	0.252	0.253	0.251	0.273	0.172
Total income	183.480***	73.882	259.472***	167.860***	261.666*
	(59.548)	(54.394)	(99.366)	(57.659)	(202.803)
	[83.423, 318.850]	[−17.330, 197.287]	[92.686, 504.746]	[71.013, 298.888]	[−37.524, 1358.690]
Mean of outcome	153.065	129.660	168.331	152.345	155.727
Labor income	158.510***	86.826*	198.741***	162.360***	182.383
	(53.280)	(55.281)	(85.346)	(56.729)	(154.351)
	[68.985, 279.633]	[−5.873, 212.242]	[55.487, 409.408]	[67.076, 291.274]	[−45.327, 1017.320]
Mean of outcome	148.881	128.268	162.327	148.765	149.311
First-stage F statistic	59.6	33.2	32.0	53.3	7.7
Observations	59,263	24,804	34,459	45,908	13,355

Notes: Sample of all native children born between 1991–2005, who are observed every year between the ages of 6 and 16, who have no contact with a psychiatrist between ages 6–7, and whose first psychiatric visit between ages 8–15 is with a child psychiatrist in a private clinic of with a hospital psychiatric department. Each cell presents the results from a separate regression of the outcome listed in the row estimated in the sample indicated in the column. Children of mothers with short cycle higher education or above are classified as high SES. All specification include the set of controls described in the notes of Table 2. The mean of the outcome is calculated among children not using SSRI. In addition to the coefficient estimate, each cell reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the person level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

Antidepressant Treatment in Childhood

Online Appendix

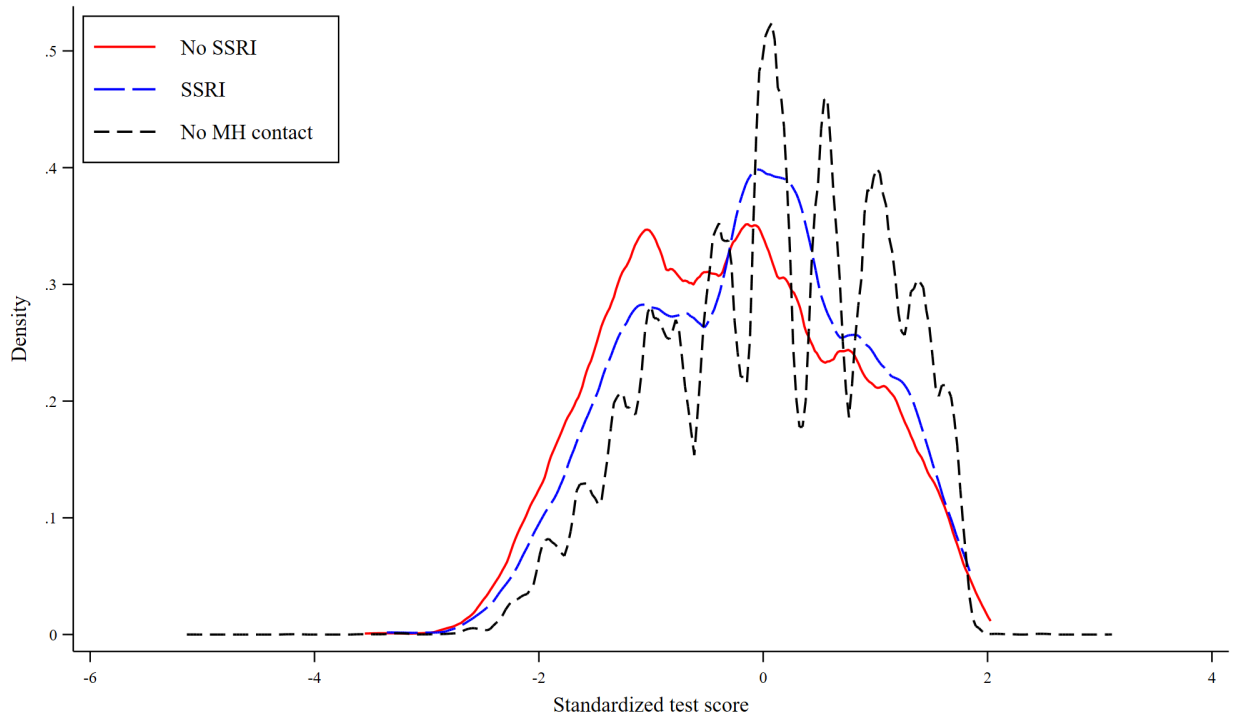
Sonia Bhalotra
University of Warwick

N. Meltem Daysal
University of Copenhagen, CEBI, CESifo, and IZA

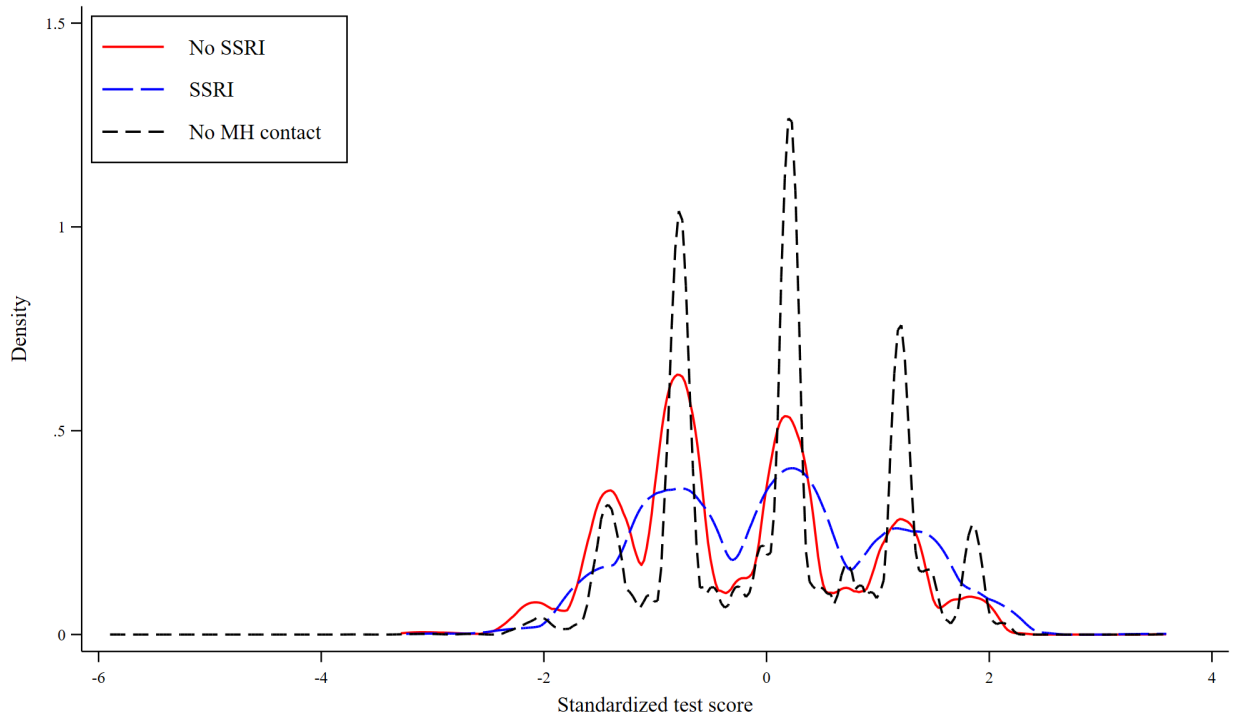
Mircea Trandafir
Rockwool Foundation Research Unit and IZA

Appendix A

Additional Descriptive Statistics



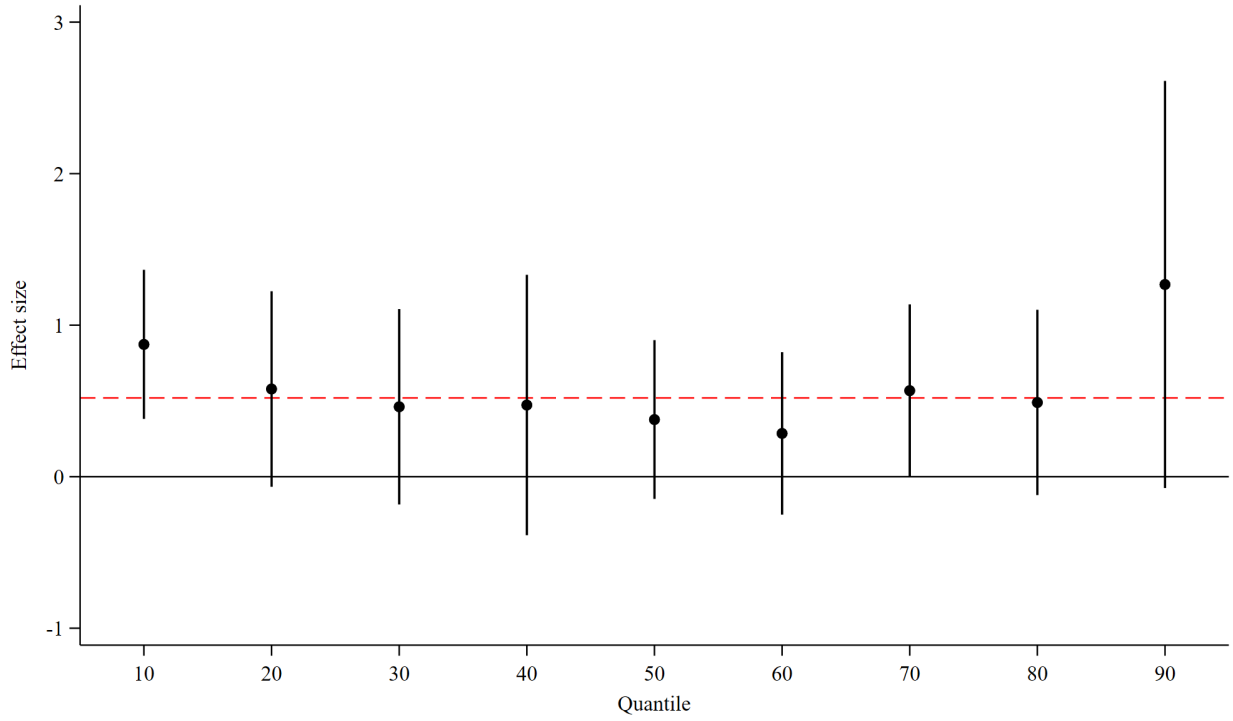
(a) Math



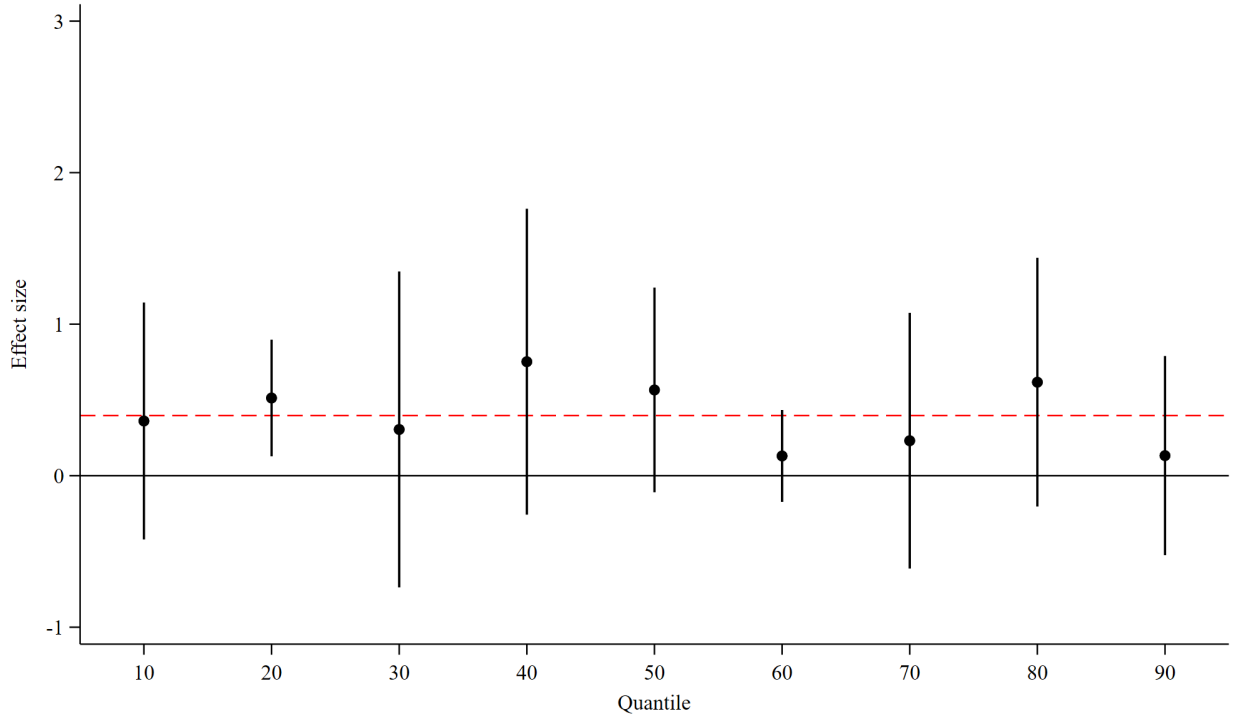
(b) Danish

Appendix Figure A1: Distribution of test scores by AD use in the analysis sample

Notes: Each panel plots the smoothed distribution of the standardized test score indicated in the panel, separately for children who did not use antidepressants by age 15 (solid line) and for children who did (long-dashed line), as well as for children with no psychiatric contact between the ages of 6–15 (short-dashed line).



(a) Math



(b) Danish

Appendix Figure A2: Distribution of test scores by AD use in the analysis sample

Notes: In each panel, each point and line segment represent the estimated effect and corresponding 95% confidence interval from an instrumental variables quantile regression estimated at the quantile indicated on the horizontal axis for the outcome indicated in the panel. Each regression includes the full set of controls listed in Table 2 and is estimated using smoothed estimating equations (Kaplan, 2022; Kaplan and Sun, 2017). Standard errors obtained from 100 bootstrap replications. Some confidence intervals are truncated to improve the legibility of the figures.

Appendix Table A1: Data sources and definitions

Variable	Definition	Years available	Register
A. Outcomes			
9th grade test scores	Grades in the written tests for Danish (reading) and mathematics, taken at the end of the 9th grade, standardized within subject and test year.	2001–2021	Academic Achievement Register (UDFK)
Longitudinal test scores (2nd–8th grade)	Scores in the national standardized tests for Danish and mathematics administered in grades 2–8, standardized at the grade and academic year level	2010–2022	DNT registers from STIL
Short-term health outcomes	inpatient visit; outpatient visit; emergency room visit; any visit related to a suicide atempt (ICD-10 codes T39, T40, T42.3, T42.4, T42.7, T43, T50.9, T58, X40-X42, X44, X46, X47, Y10-Y12, Y16, Y17).	1977–2018	National Patient Register (LPR_ADM)
Enrollment beyond compulsory education	Indicators for enrollment at age 18 in any educational program beyond 9th grade (academic high school or vocational school), and separate indicators for enrollment in an academic high school or in a vocational school	1980–2021	Education register (UDDA)
Labor market outcomes	Total annual income in thousands DKK, annual labor income (salary and self-employment income) in thousands DKK, indicators for being employed, for being enrolled in higher education, and for receiving welfare or disability benefits. Monetary variables are deflated to 2015 DKK.	1980–2021	Income register (IND)
B. Medical treatments after diagnosis			
AD use	Separate indicators for filling a prescription (ATC code in brackets) by age 15 for: selective serotonin reuptake inhibitors, SSRI (N06AB).	1995–2021	National Prescription Register (LMDB)
Non-SSRI mental health medication use	Separate indicators for filling a prescription (ATC code in brackets) by age 15 for: tricyclic antidepressant (N06AA); benzodiazepines (N05BA); stimulant (N06BA); antipsychotics (N05A); any ATC-class-N drug, not SSRI (N* excluding N06AB).	1995–2021	National Prescription Register (LMDB)
Psychotherapy after diagnosis	Indicator for a visit (after the first visit to a psychiatrist and by age 15) for: counseling services provided by general practitioners, GPs (SPECIALE codes 802147–802149, 804003, 804021–804027, 804050, 804063, 804106, 804116, 806100, 806101); psychologists (SPECIALE codes starting with 63); or therapy provided by child psychiatrists (SPECIALE codes 260131, 260132, 260133, 260134)	1990–2021	National health insurance service register (SYSI, SSSY)

Appendix Table A1 (cont.): Data sources and definitions

Variable	Definition	Years available	Register
C. Control variables, measured in the calendar year when the child turns 6 years old			
Year of diagnosis	Indicators for the year of the first visit at a privately-practicing child psychiatrist (SPECIALE codes starting with 26), defined as the first contact after turning 8 for children with no contact between the ages of 6–8	1990–2021	National health insurance service register (SYSI, SSSY)
Psychotherapy before diagnosis	Indicator for a visit (within 1 year before the first visit to a psychiatrist) for: counseling services provided by general practitioners, GPs (SPECIALE codes 802147–802149, 804003, 804021–804027, 804050, 804063, 804106, 804116, 806100, 806101); or psychologists (SPECIALE codes starting with 63)	1990–2021	National health insurance service register (SYSI, SSSY)
Child demographics	Indicators for gender, birth order, year of birth, number of siblings, municipality of residence	1986–2021	Population register (BEF, BEF_ADR)
Parent demographics	Age of each parent, indicator for whether they are married	1986–2021	Population register (BEF)
Parental education	For each parent, indicators for the highest degree obtained (at the 1-digit level)	1980–2021	Education register (UDDA)
Parental labor market outcomes	For each parent, indicators for the income decile within the gender-specific full population , as well as indicators for whether they are employed	1980–2021	Income register (IND) Register-based Labour Force Statistics (RAS)
Parental use of antidepressants	For each parent, indicators for filling an SSRI (ATC: N06AB) prescription at any point when the child was 6–8 years old	1995–2021	National Prescription Register (LMDB)
D. Psychiatrist characteristics			
Psychiatrist characteristics	Indicators for: the clinic having only male psychiatrists for the entire year; all the psychiatrists in the clinic being married; the clinic having only one psychiatrist for the entire year. Averages over all the psychiatrists employed in the clinic of: age; experience (number of years since graduation)	1995–2021	Yderregister (1 and 2) from Sundhedsdatastyrelsen, linked with BEF and UDDA

Appendix Table A2: Characteristics of the full sample of children by type of specialist of first mental health visit

	First contact with mental health specialist between ages 8–15			*p*-value (1)–(3)	*p*-value (2)–(3)
	None	Hospital	Private clinic		
	(1)	(2)	(3)	(4)	(5)
A. Outcomes					
Standardized test score, math	0.093	−0.400	−0.283	0.000	0.000
	(0.968)	(1.022)	(1.016)		
Standardized test score, Danish	0.074	−0.281	−0.199	0.000	0.000
	(0.983)	(1.045)	(1.025)		
B. Treatments					
Therapy within 1 year before first MH visit	—	0.104	0.116	0.000	0.002
Therapy after first MH visit, by age 15	—	—	0.713	0.000	0.000
Pharmaceutical treatments by age 15:					
SSRI	0.001	0.092	0.152	0.000	0.000
Tricyclic antidepressant	0.001	0.003	0.002	0.024	0.906
Benzodiazepine	0.006	0.023	0.015	0.000	0.000
Stimulant	0.005	0.263	0.341	0.000	0.000
Antipsychotic	0.001	0.074	0.075	0.000	0.878
Any non-SSRI N-drug	0.062	0.422	0.462	0.000	0.000
Age when first SSRI was filled	13.481	13.161	12.674	0.000	0.000
	(1.980)	(1.624)	(1.778)		
Median days between first MH visit and:					
First SSRI prescription	—	154.000	106.500	—	0.000
First non-SSRI prescription	—	−24.000	−23.000	—	0.166
C. Child characteristics					
Boy	0.504	0.574	0.596	0.000	0.000
Birth order:					
1	0.453	0.483	0.490	0.000	0.281
2	0.374	0.358	0.364	0.110	0.280
3+	0.174	0.159	0.145	0.000	0.003
Number of siblings:					
0	0.128	0.178	0.187	0.000	0.093
1	0.561	0.537	0.560	0.882	0.000
2	0.223	0.199	0.178	0.000	0.000
3+	0.087	0.085	0.076	0.000	0.005
Parents are married	0.810	0.660	0.680	0.000	0.001
High SES	0.365	0.278	0.352	0.021	0.000
Age at first diagnosis	—	11.899	11.637	—	0.000
		(2.082)	(2.069)		

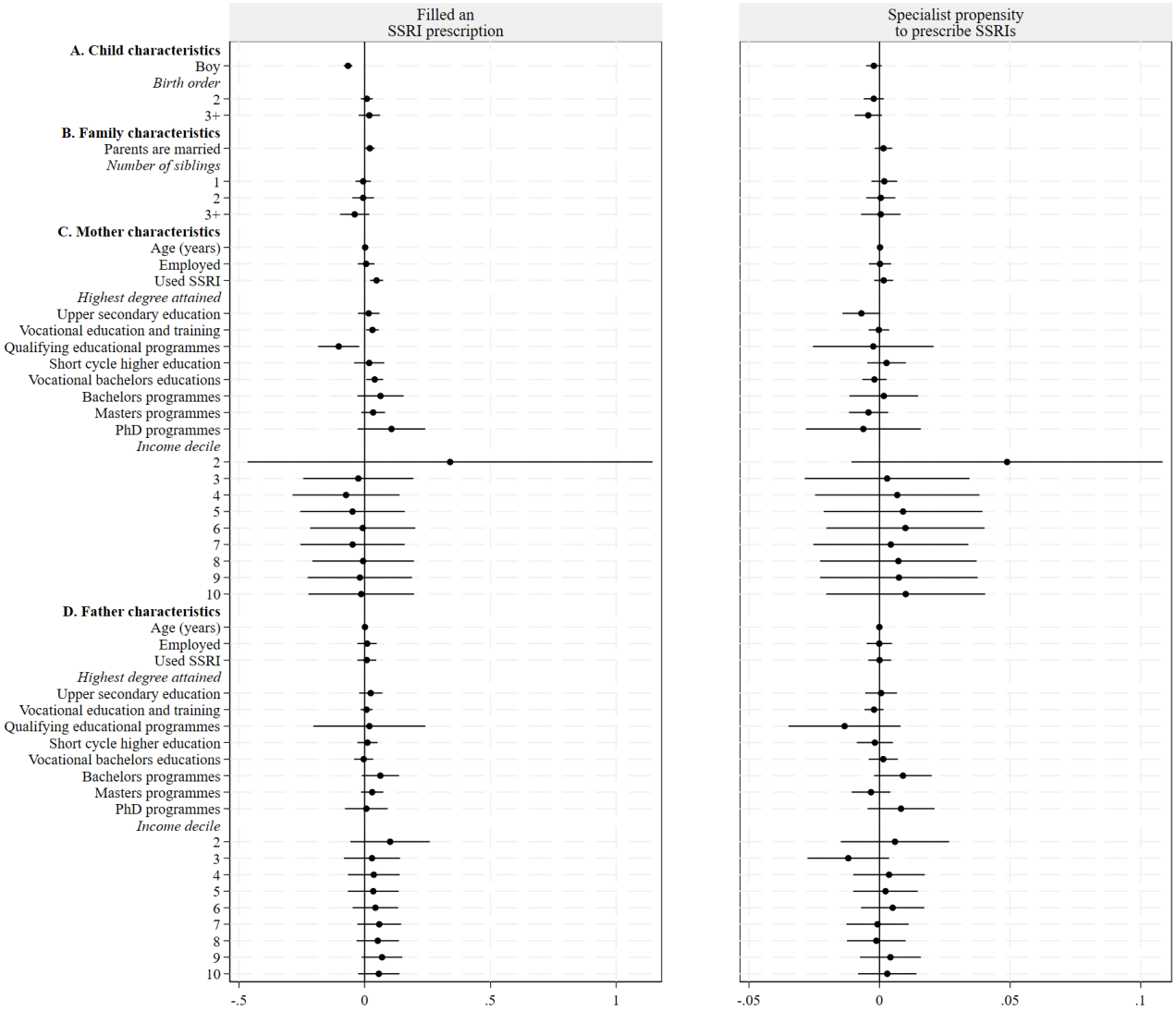
Appendix Table A2 (cont.): Characteristics of the full sample of children by type of specialist of first mental health visit

	First contact with mental health specialist between ages 8–15			*p*-value (1)–(3)	*p*-value (2)–(3)
	None	Hospital	Private clinic		
	(1)	(2)	(3)	(4)	(5)
D. Psychiatrist characteristics					
Female	—	—	0.413		
Age (years)	—	—	57.075		
			(5.232)		
Experience (years)	—	—	27.934		
			(5.612)		
Married	—	—	0.657		
Solo practice	—	—	0.793		
Number of patients treated during year	—	2,977.441	169.781		0.000
		(2,043.735)	(79.220)		
Propensity to prescribe SSRIs	—	0.060	0.105		0.000
		(0.031)	(0.061)		
E. Mother's characteristics at child age 6					
Age	35.879	35.211	35.708	0.004	0.000
	(4.629)	(5.092)	(5.066)		
Some college or more	0.365	0.278	0.352	0.021	0.000
Employed	0.852	0.748	0.780	0.000	0.000
Annual gross total income	333.004	309.579	326.692	0.000	0.000
	(189.383)	(125.057)	(150.718)		
Used SSRI between child ages 6–7	0.086	0.178	0.167	0.000	0.016
F. Father's characteristics at child age 6					
Age	38.247	37.799	38.122	0.065	0.000
	(5.405)	(5.755)	(5.732)		
Some college or more	0.296	0.212	0.280	0.003	0.000
Employed	0.932	0.871	0.887	0.000	0.000
Annual gross total income	450.972	396.071	427.379	0.000	0.000
	(561.056)	(353.161)	(331.281)		
Used SSRI between child ages 6–7	0.050	0.094	0.088	0.000	0.087
Observations	770,520	49,411	7,211		

Notes: Sample of all native children born between 1991–2005, who are observed every year between the ages of 6 and 16, and who have no contact with a psychiatrist (either in a private clinic or hospital department) between ages 6–7. Column (1) includes children with no psychiatric visit by age 15, Column (2) those children whose first mental health-related visit between the ages of 8–15 was at a hospital psychiatric department, and Column (3) the children whose first psychiatric visit was in a private child psychiatry clinic. Columns (4) and (5) present the *p*-values for the test of equality of means between Columns (1) and (3), and between Columns (2) and (3), respectively. See the notes to Table 1 for details on the variables listed.

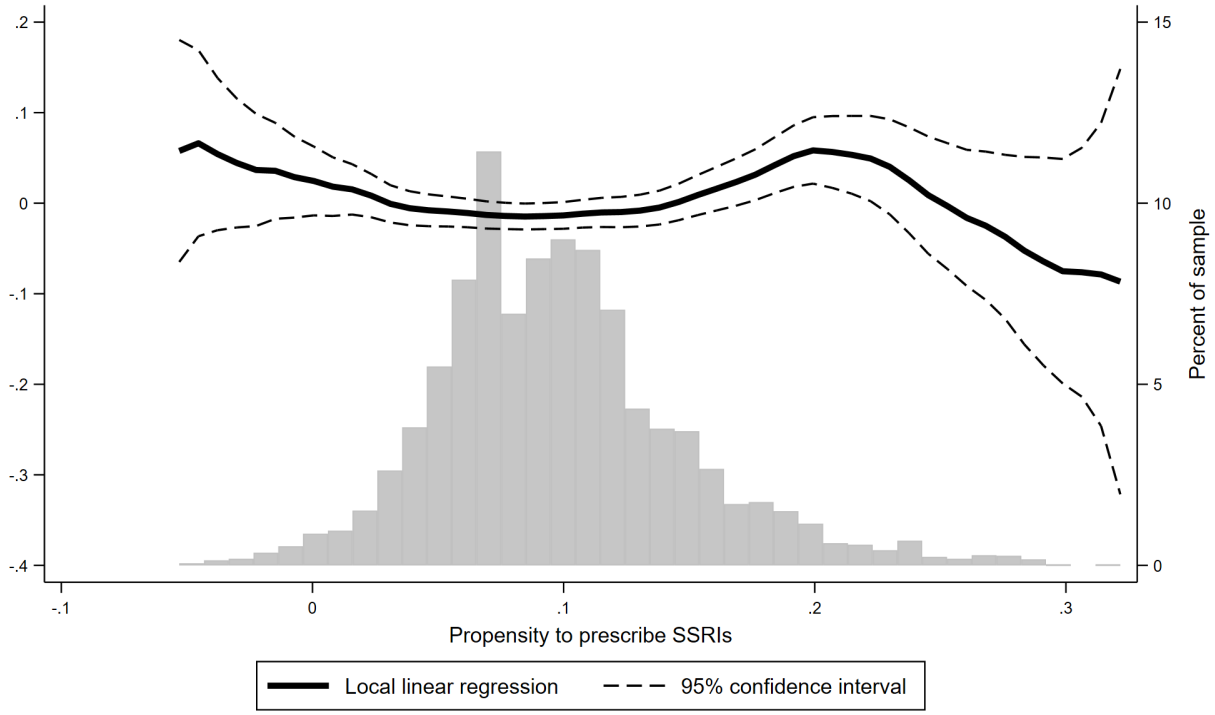
Appendix B

Instrument Validity and Generalizability of LATE

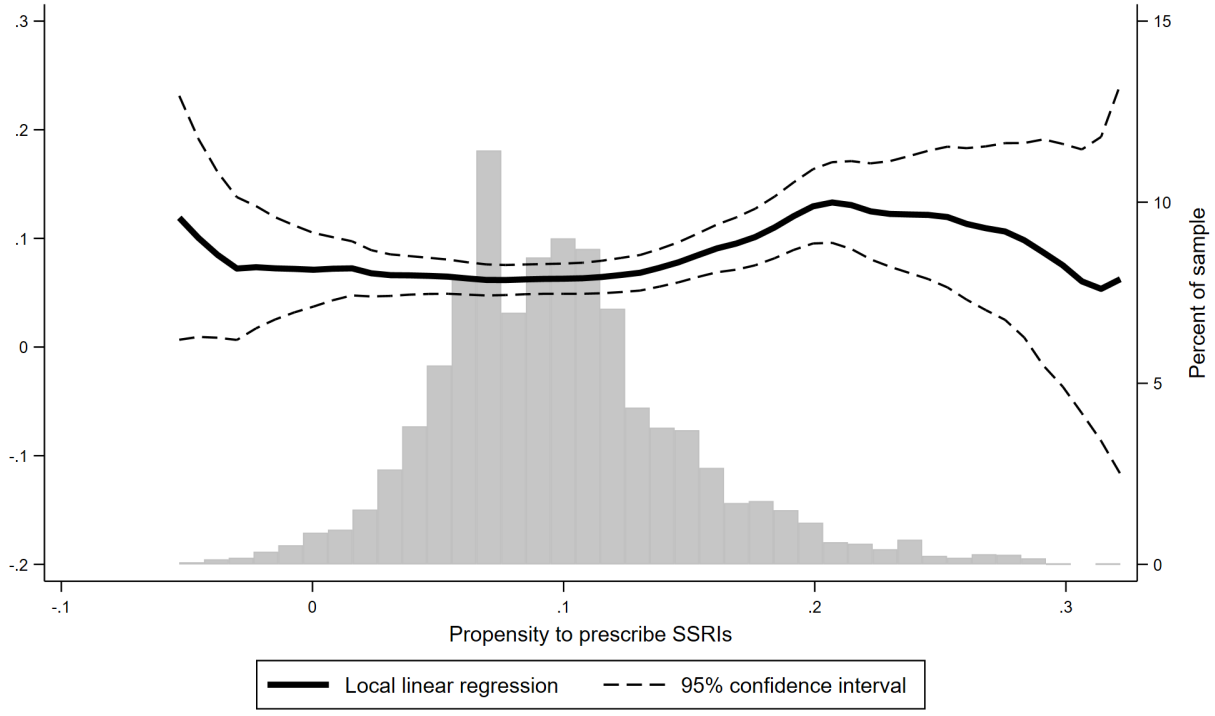


Appendix Figure B1: Effects of child, parent, family, and clinic characteristics, on the probability of filling an AD prescription and on the specialist propensity to prescribe AD

Notes: Each panel plots the coefficients and corresponding 95% confidence intervals from a regression of the outcome listed in the panel heading on all the variables listed, as well as indicators for year of birth of the child, year of diagnosis, and municipality of residence. The regressions are estimated in the sample of children who took the Math test (the results from the sample of children who took the Danish test are virtually identical). The unlisted variables and the variables in Panels A–D constitute the set of baseline controls.



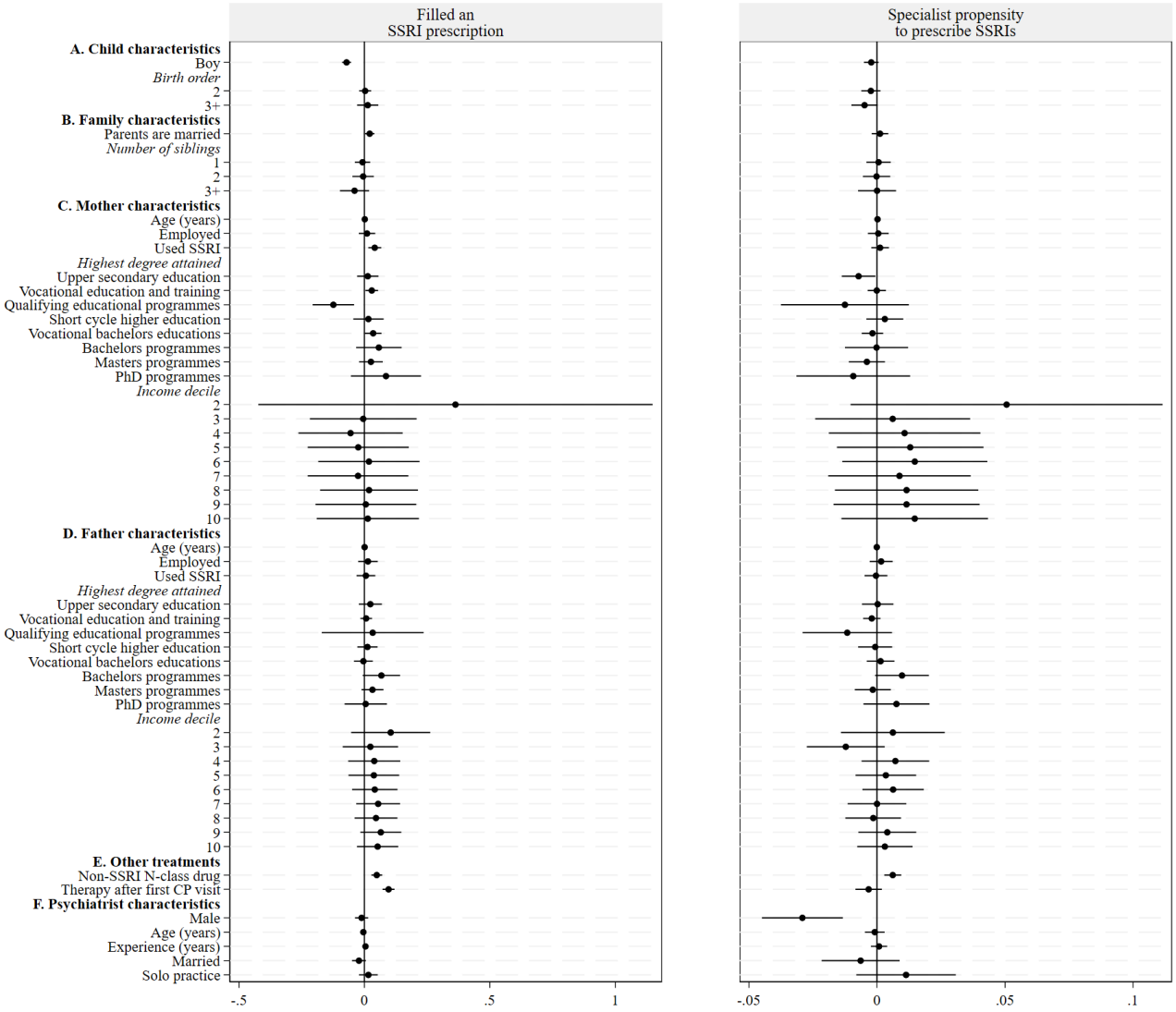
(a) Math



(b) Danish

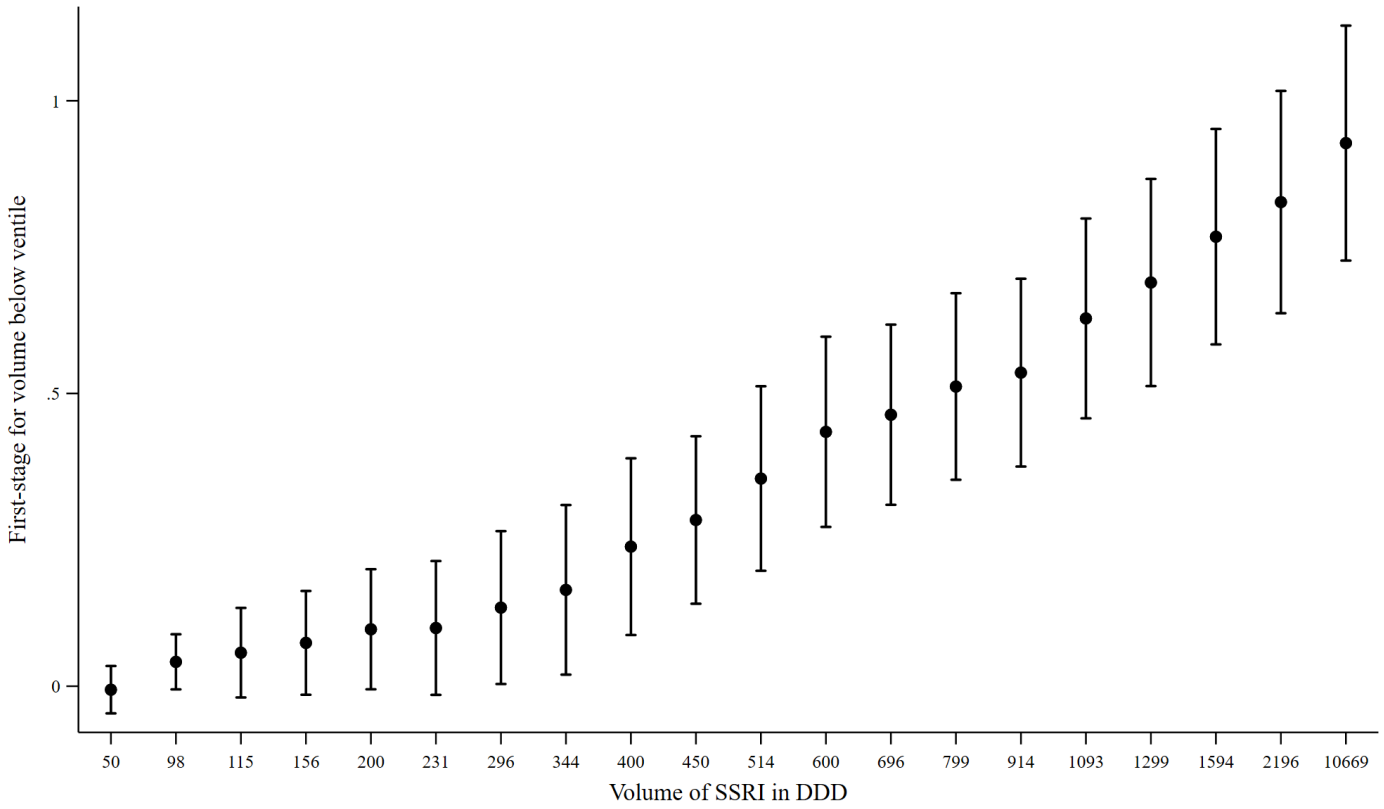
Appendix Figure B2: Distribution of specialist tendency to prescribe ADs and predicted test scores

Notes: Both the propensity to prescribe AD and the predicted test scores are net of year-of-diagnosis and municipality fixed effects. Test scores are predicted using the full set of control variables described in the notes to Table 2. The gray bars represent the histogram of the adjusted tendency to prescribe antidepressants. The solid line plots the local polynomial regression of the predicted outcome indicated in the figure caption on the tendency to prescribe antidepressants of the treating specialist, while the dashed lines indicate the corresponding 95% confidence interval.



Appendix Figure B3: Effects of child, parent, family, and psychiatrist treatment choices and characteristics, on the probability of filling an AD prescription and on the specialist propensity to prescribe AD

Notes: Each panel plots the coefficients and corresponding 95% confidence intervals from a regression of the outcome listed in the panel heading on all the variables listed, as well as indicators for year of birth of the child, year of diagnosis, and municipality of residence. The regressions are estimated in the sample of children who took the Math test (the results from the sample of children who took the Danish test are virtually identical). The unlisted variables and the variables in Panels A–D constitute the set of baseline controls.



Appendix Figure B4: Effect of psychiatrist leniency on the probability of children filling prescriptions with cumulated total volume of DDD less than given volume

Notes: Each dot and line segment represent the coefficient and corresponding 95% confidence interval from a regression of the probability that the total volume of AD prescribed by age 15 is lower than the ventile of the distribution of volume of AD indicated on the horizontal axis. Each regression controls for the set of controls listed in the notes to Table 2. Regressions estimated in the sample of children who took the Math test (the results in the sample of children who took the Danish test are virtually identical).

Appendix Table B1: Average characteristics for the analysis sample and for compliers

	Panel A. Share of compliers				
	All (1)	Girls (2)	Boys (3)	Low SES (4)	High SES (5)
Share of compliers	0.313	0.386	0.274	0.256	0.406

	Panel B. Characteristics of compliers relative to the full analysis sample		
	All (1)	Compliers (2)	Ratio (3)
A. Child characteristics			
Boy	0.580	0.453	0.782
First born	0.495	0.540	1.090
High SES	0.383	0.456	1.190
Parents are married	0.700	0.812	1.160
B. Mother's characteristics			
Age	35.902	37.196	1.036
Employed	0.812	0.818	1.008
Annual income (thousands)	336.917	342.241	1.016
Used SSRI between child ages 6-8	0.166	0.205	1.236
C. Father's characteristics			
Age	38.233	38.968	1.019
Employed	0.906	0.908	1.002
Annual income (thousands)	445.526	469.055	1.053
Used SSRI between child ages 6-8	0.089	0.112	1.252

Notes: Analysis sample as described in the notes to Table 1. Sample of children taking the math test (the estimates for the sample of children taking the Danish test are virtually identical and available upon request). Panel A lists the share of compliers in the sample indicated in the column, calculated as the product between the first-stage coefficient in that sample multiplied by the range of the instrument. Panel B lists the means of the variables listed in the row among the analysis sample (Column 1) and among compliers (Column 2). The mean of a characteristic among compliers is the estimated coefficient on the instrument from a regression of the interaction between the characteristic and the indicator for filling an AD prescription by age 15 on the instrument and the full set of controls listed in the notes to Table 2. Column 3 lists the ratio between the mean among compliers (Column 2) and the mean in the full analysis sample (Column 1).

Appendix Table B2: First-Stage Estimates of AD Use on CP Leniency

	Baseline	Excluding parental controls	Excluding parental and child controls
	(1)	(2)	(3)
Specialist tendency to prescribe AD	0.928***	0.947***	0.962***
	(0.102)	(0.105)	(0.102)
Effect of interquartile change in instrument	0.074	0.075	0.076
Mean of outcome	0.147	0.147	0.147
F statistic	82.8	81.6	88.9
Observations	5,495	5,495	5,495

Notes: Analysis sample as described in the notes to Table 1. Sample of children taking the math test (the estimates for the sample of children taking the Danish test are virtually identical and available upon request). Each cell presents the results from a separate regression of the indicator for filling an AD prescription by age 15 in the specification listed in the column. All specifications include fixed effects for the year of diagnosis and municipality of residence during the year of diagnosis. The specification in Column (1) includes the full set of controls listed in the notes to Table 2. The specification in Column (2) excludes the indicators for family size; the indicators for mother's and father's education degree, employment status, income deciles, and marriage status at child age 6; the indicators for mother's and father's use of SSRI during child age 6–7; and mother's and father's age. In addition, the specification in Column (3) excludes the indicators for child gender and birth order, as well the fixed effects for year of birth. The mean of the outcome is calculated among all children in the estimation sample. In addition to the coefficient estimate, each cell lists the effect of an increase in the value of the instrument from the 25th to the 75th percentile in the full analysis sample, corresponding to approximately 7.9 percentage points. Each cell also includes the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table B3: First-stage estimates of effect of SSRI use on test scores, taking into account other treatments and psychiatrist characteristics

	Baseline	Including other prescribing tendencies and psychiatrist characteristics		
Dependent variable	AD use	AD use	Non-SSRI N-drugs	Therapy
	(1)	(2)	(3)	(4)
Psychiatrist tendency to treat with:				
— AD	0.928***	0.907***	0.355***	−0.132
	(0.102)	(0.101)	(0.123)	(0.123)
— Non-SSRI N-class medication		0.108***	0.681***	0.066
		(0.038)	(0.051)	(0.042)
— Therapy		−0.080	−0.045	0.250***
		(0.073)	(0.115)	(0.091)
Mean of outcome	0.147	0.147	0.418	0.711
Observations	5,495	5,495	5,495	5,495
Psychiatrist characteristics	No	Yes	Yes	Yes

Notes: Analysis sample as described in the notes to Table 1. Sample of children taking the math test (the estimates for the sample of children taking the Danish test are virtually identical and available upon request). Each column presents the results from a separate regression of an indicator for the use of the treatment listed in the column on the set of controls listed in the notes to Table 2, the full set of psychiatrist characteristics (Columns 2–4), and the specialist tendencies to provide treatments as indicated in the rows. Column 1 repeats our baseline results and includes only the psychiatrist tendency to prescribe AD. Columns 2–4 include all psychiatrist characteristics as well as separate variables for the psychiatrist tendency to prescribe AD, to prescribe non-SSRI N-class medication, and to provide therapy. Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table B4: 2SLS estimates of effect of SSRI use on test scores, taking into account other treatments and psychiatrist characteristics

	Baseline	Include psych. leniency and chars.	Incl. other treatments and psych. chars.
	(1)	(2)	(3)
A. Standardized test score, math			
Treatment received:			
— AD	0.520**	0.405	0.613***
	(0.266)	(0.277)	(0.398)
	[0.028, 1.063]	[−0.106, 0.970]	[0.613, 2.147]
— Non-SSRI N-class medication			−0.273***
			(0.211)
			[−1.085, −0.273]
— Therapy			0.702***
			(0.896)
			[0.702, 5.877]
Lewis-Mertens statistic			7.1
Lewis-Mertens 95% critical value			21.0
Mean of outcome	−0.305	−0.305	−0.305
Observations	5,495	5,495	5,495
B. Standardized test score, Danish			
Treatment received:			
— AD	0.397*	0.402*	0.859***
	(0.253)	(0.287)	(0.625)
	[−0.070, 0.913]	[−0.072, 0.987]	[0.859, 4.465]
— Non-SSRI N-class medication			−0.494***
			(0.325)
			[−2.993, −0.494]
— Therapy			2.205***
			(1.566)
			[2.205, +∞]
Lewis-Mertens statistic			6.5
Lewis-Mertens 95% critical value			21.4
Mean of outcome	−0.241	−0.241	−0.241
Observations	5,383	5,383	5,383
Psychiatrist characteristics	No	Yes	Yes

Notes: Analysis sample as described in the notes to Table 1. Each column presents the results from a separate regression of the outcome indicated in the panel using the specification indicated in the column. All specifications include the full set of controls listed in Table 2. In addition, Column 2 controls for the full set of psychiatrist characteristics (Columns 2–4), as well as separate variables for the psychiatrist tendency to prescribe non-SSRI N-class medication and to provide therapy. Column 3 replaces the two psychiatrist tendency variables with two indicator variables, one for the use of non-SSRI N-drug medication and one for the use of therapy, instrumented for with the two psychiatrist tendency variables. Each panel reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the Lewis and Mertens (Forthcoming) statistic for instrument strength and the corresponding 95% critical value. The mean of the outcome is calculated among children not using AD. Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

Appendix Table B5: 2SLS estimates of the effects of AD dosage on test scores

	Endogenous variable			
	Indicator for SSRI use	Volume of SSRI used (hundred DDDs)		
	(1)	(2)	(3)	(4)
Standardized test score, math	0.520*	0.052*	0.053*	0.063
	(0.266)	(0.027)	(0.028)	(0.180)
	[0.028, 1.063]	[0.007, 0.108]	[0.006, 0.116]	—
First-stage F statistic	82.8	60.1	34.2	1.0
Mean of outcome	−0.305	−0.284	−0.284	−0.284
Effect of median volume	—	0.235	0.239	0.284
Observations	5,495	5,495	5,495	5,495
Standardized test score, Danish	0.397	0.039	0.064**	0.393
	(0.253)	(0.025)	(0.028)	(0.455)
	[−0.070, 0.913]	[−0.007, 0.090]	[0.018, 0.128]	—
First-stage F statistic	82.2	60.9	33.3	0.8
Mean of outcome	−0.241	−0.200	−0.200	−0.200
Effect of median volume	—	0.177	0.290	1.770
Observations	5,383	5,383	5,383	5,383
Instrument = PP for	AD	AD	Volume	Volume
Control for PP for AD	No	No	No	Yes

Notes: Analysis sample as described in the notes to Table 1. Each cell presents the results from a separate regression of the outcome listed in the row for the specification listed in the column. Column 1 repeats our baseline results and reports the coefficient on the indicator for AD use by age 15. The specifications in columns 2–4 replace the indicator for AD use with the total volume of AD (in hundreds of daily-defined doses) filled by age 15. All specifications include the full set of controls listed in the notes to Table 2. In addition, the specification in Column 4 includes the psychiatrist tendency to prescribe AD. The specifications in Columns 1 and 2 instrument for the endogenous variable (the indicator for AD use in Column 1, the volume of AD in Column 2) with the psychiatrist tendency to prescribe AD. The specifications in Columns 3 and 4 instrument for the endogenous variable (the volume of AD) with the psychiatrist tendency to prescribe volume of AD, defined as the average volume prescribed by the psychiatrist to the other children age 8–15 treated by the psychiatrist in the year of diagnosis of the child. Each panel reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). In addition, Columns 2–4 report the estimated effect of the median volume of AD on the outcome. Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

Appendix Table B6: First-Stage Estimates of AD Use on CP Leniency by Subgroup

	A. Gender		B. Socioeconomic status		C. Quartile of predicted outcome			
	Girls	Boys	Low SES	High SES	First	Second	Third	Fourth
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Specialist tendency to prescribe AD	1.143***	0.813***	0.758***	1.204***	0.673***	0.664***	0.750***	1.408***
	(0.157)	(0.112)	(0.124)	(0.162)	(0.190)	(0.177)	(0.183)	(0.178)
Effect of interquartile change in instrument	0.091	0.064	0.060	0.095	0.053	0.053	0.059	0.112
First-stage F statistic	52.9	53.0	37.3	55.6	12.5	14.0	16.8	62.5
Observations	2,310	3,185	3,391	2,104	1,374	1,374	1,374	1,373

	D. Characteristics of the physicians in the child psychiatry clinic of first mental health contact							
	Age (years)		Experience (years)		Gender		Number of physicians	
	Below median	Above median	Below median	Above median	All female	Not all female	Solo practice	Multiple physicians
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Specialist tendency to prescribe AD	1.024***	0.945***	1.071***	0.634***	0.789***	0.927***	0.930***	0.631**
	(0.126)	(0.161)	(0.128)	(0.148)	(0.119)	(0.186)	(0.110)	(0.284)
Effect of interquartile change in instrument	0.081	0.075	0.085	0.050	0.063	0.074	0.074	0.050
First-stage F statistic	65.7	34.6	70.2	18.2	43.8	24.8	70.9	4.9
Observations	2,736	2,759	3,054	2,441	2,821	2,674	4,395	1,100

Notes: Analysis sample as described in the notes to Table 1. Sample of children taking the math test (the estimates for the sample of children taking the Danish test are virtually identical and available upon request). Each cell presents the results from a separate regression of the indicator for filling an AD prescription by age 15 in the sample of children with the characteristic listed in the column (Panels A–C) or treated by psychiatrists with the characteristic listed in the column (Panel D). Children of mothers with short cycle higher education or above are classified as high SES. The median age of psychiatrists is 57.375 years, and the median experience is 29 years. All specifications include the full set of controls listed in the notes to Table 2. The mean of the outcome is calculated among all children. In addition to the coefficient estimate, each Panel lists the effect of an increase in the value of the instrument from the 25th to the 75th percentile in the full analysis sample, corresponding to approximately 7.9 percentage points. Each panel also reports the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table B7: First-Stage Estimates of AD Use on CP Leniency Using Cross-Sample Instrument

	Gender		Socioeconomic status	
	Girls	Boys	Low SES	High SES
	(1)	(2)	(3)	(4)
Specialist tendency to prescribe AD	1.383***	0.513***	0.829***	1.050***
	(0.199)	(0.075)	(0.132)	(0.115)
Effect of interquartile change in instrument	0.110	0.041	0.066	0.083
First-stage F statistic	48.4	46.8	39.6	82.8
Observations	2,310	3,185	3,391	2,104

Notes: Analysis sample as described in the notes to Table 1. Sample of children taking the math test (the estimates for the sample of children taking the Danish test are virtually identical and available upon request). Each cell presents the results from a separate regression of the indicator for filling an AD prescription by age 15 in the sample of children with the characteristic listed in the column. Children of mothers with short cycle higher education or above are classified as high SES. The instrument in Column (1) is defined as the fraction of 8–15 year old boys treated in the year of diagnosis by the same psychiatrist who fill an AD prescription in that year. The instruments in Columns (2), (3), and (4) are constructed in a similar manner, using only girls, high-SES, and low-SES children, respectively. All specifications include the full set of controls listed in the notes to Table 2. The mean of the outcome is calculated among all children. In addition to the coefficient estimate, each Panel lists the effect of an increase in the value of the instrument from the 25th to the 75th percentile in the full analysis sample, corresponding to approximately 7.9 percentage points. Each panel also reports the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Appendix C

Robustness Checks

C.1 National tests

Starting from 2010, students enrolled in Danish public schools are required to take a number of national standardized tests. These tests are computerized adaptive tests, in which the performance of the student in the previous questions determines the difficulty of the next question.[1] They are usually administered between mid-January and April and are scored electronically, without input from their teachers (who also do not prepare the questions). The tests are low stakes: they are mainly used to inform teachers and parents on the performance of students. Although they are mandatory and more than 90% of public school students take them, there is no mandated punishment for truancy and individual exemptions may be granted if teachers and parents agree that the student would not be able to perform in an informative way.

In our paper, we use the data from the reading tests in Danish, administered in grades 2, 4, 6, and 8, and from the tests in Mathematics, administered in grades 3, 6, and 8. Each test calculates scores across several performance areas (e.g., for Danish, these are language comprehension, decoding, and reading comprehension). We follow standard practice in the literature and we standardized each score at the cohort level, take the simple average across the different domains, and restandardize at the cohort level to obtain one score per subject (Beuchert and Nandrup, 2018). We also use the exact date at which each test was administered in order to determine if it occurred before or after an AD prescription was filled.

Empirical strategy To estimate the effects of AD use on test score, we implement a modified version of our baseline strategy in Section 4. The main estimating equation is:

$$Y_{igt} = \beta_0 + \beta_1 AD_{it} + \delta \mathbf{X}_{it} + \alpha_i + \mu_g + \nu_t + \epsilon_{igt},$$

where Y_{igt} is a test score for child i in grade g during year t, AD_{igt} is an indicator for the child having filled an AD prescription before the date of the test, X_{it} is a set of time-varying characteristics (an indicator for whether the child has skipped the previous test, and an indicator for whether the previous test and the current test were administered using different algorithms), and α_i, μ_g, and ν_t are individual, calendar year, and grade fixed effects.

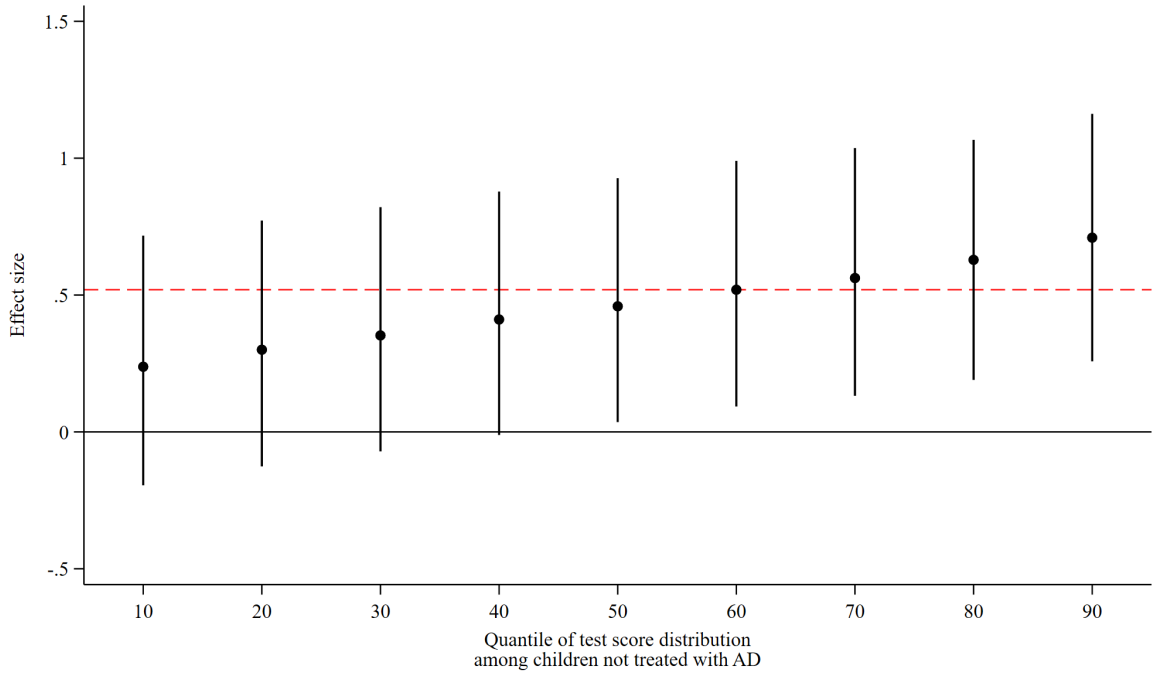
As before, it is possible that the use of AD is correlated with some unobservable (time-varying) individual characteristics that also influence school performance. Therefore, we implement a panel 2SLS estimation strategy in which the first stage equation is:

$$AD_{igt} = \beta_0^{FS} + \beta_1^{FS} \mathbb{1}(t \geq V_i) PP_i + \delta^{FS} \mathbf{X}_{it} + \alpha_i^{FS} + \mu_g^{FS} + \nu_t^{FS} + u_{igt},$$

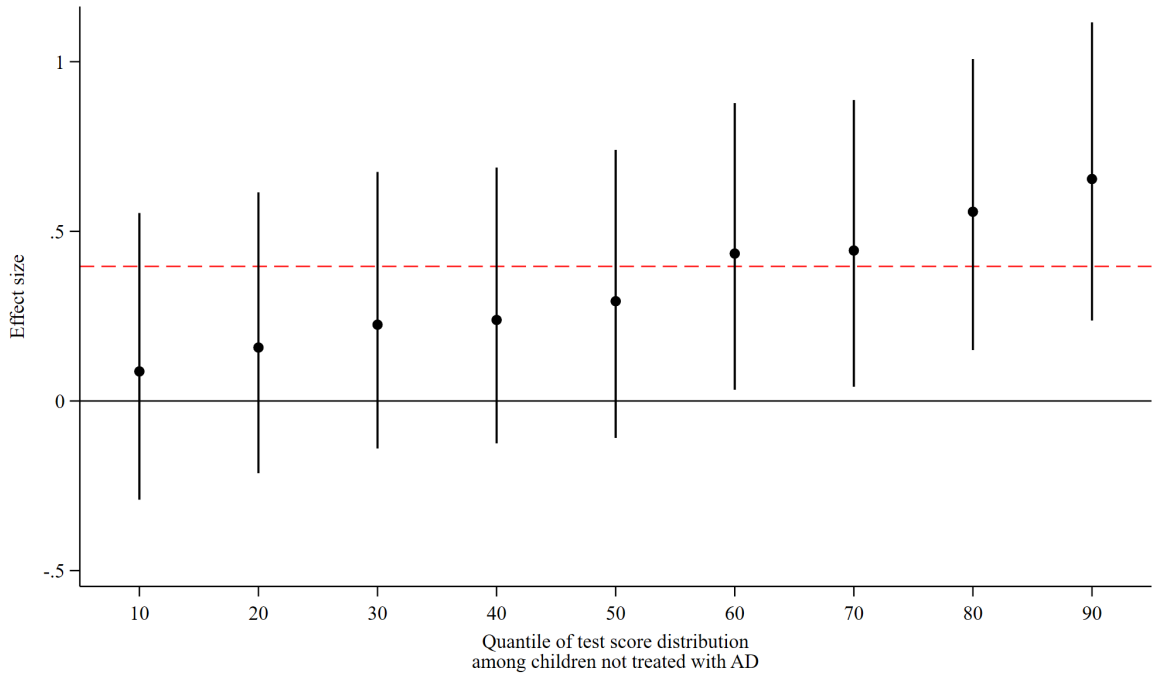
where $\mathbb{1}(t \geq V_i)$ indicates whether the date of the test t is after the date V_i of the child's first mental health visit, and PP_i is our instrument as defined in Equation (2):

$$PP_i = \frac{\sum_{j \neq i} AD_j \mathbb{1}(\text{psychiatrist}_j = \text{psychiatrist}_i)}{\sum_{j \neq i} \mathbb{1}(\text{psychiatrist}_j = \text{psychiatrist}_i)}.$$

[1] The algorithm changed in 2015, changing the difficulty of the first question and of the second and third questions following a correct or wrong response.



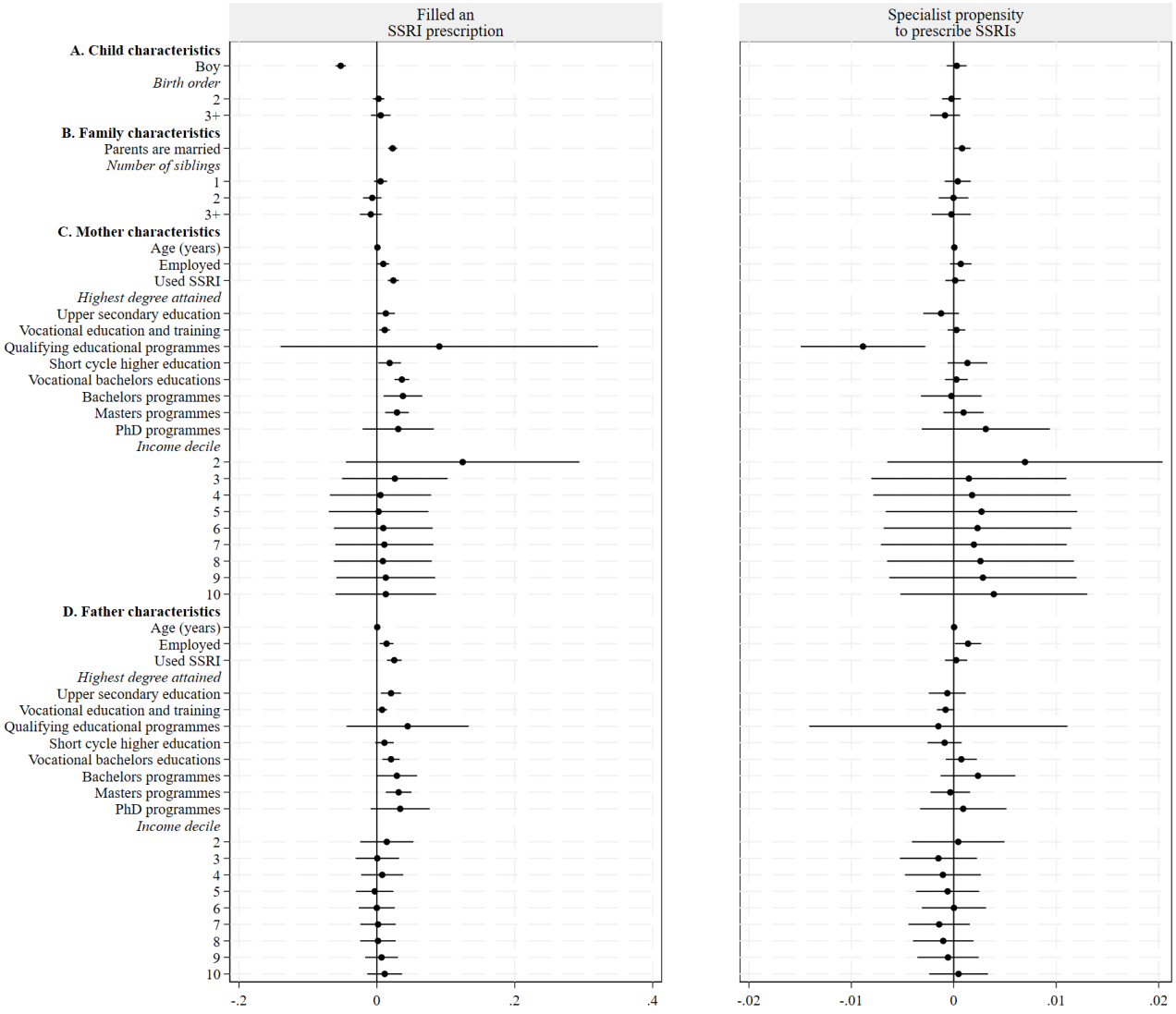
(a) Math



(b) Danish

Appendix Figure C1: Bounding the effect of using AD on test scores

Notes: In each panel, each point and line segment represent the estimate and corresponding 95% Anderson-Rubin confidence interval from an instrumental variables regression similar to our baseline regression in Equation (1). In these regressions, children treated with AD and who did not take the test are assigned the percentile indicated on the horizontal axis from the distribution among children not treated with AD, for the outcome indicated in the panel. Each regression includes the full set of controls listed in Table 2. Standard errors clustered at the specialist and year level.



Appendix Figure C2: Effects of child, parent, family, and clinic characteristics, on the probability of filling an SSRI prescription and on the specialist propensity to prescribe SSRIs in the full MH sample

Notes: Each panel plots the coefficients and corresponding 95% confidence intervals from a regression of the outcome listed in the panel heading on all the variables listed, as well as indicators for year of birth of the child, year of diagnosis, and municipality of residence. The regressions are estimated in the sample of children who took the Math test (the results from the sample of children who took the Danish test are virtually identical and available upon request). The unlisted variables and the variables in Panels A–D constitute the set of baseline controls.

Appendix Table C1: Robustness Checks Related to the Instrument

	Math (1)	Danish (2)
Baseline	0.520**	0.397*
	(0.266)	(0.253)
	[0.028, 1.063]	[−0.070, 0.913]
First-stage F statistic	82.786	82.197
Mean of outcome	−0.305	−0.241
Observations	5,495	5,383
Children treated in clinics with at least 25 patients	0.472**	0.531**
	(0.253)	(0.251)
	[0.004, 0.990]	[0.116, 1.043]
First-stage F statistic	84.345	84.632
Mean of outcome	−0.304	−0.238
Observations	5,713	5,601
Children treated in clinics with at least 75 patients	0.566**	0.392
	(0.289)	(0.282)
	[0.032, 1.156]	[−0.073, 0.967]
First-stage F statistic	71.873	70.446
Mean of outcome	−0.306	−0.245
Observations	5,188	5,079
Children treated in clinics with at least 100 patients	0.549*	0.428*
	(0.303)	(0.267)
	[−0.009, 1.107]	[−0.064, 0.971]
First-stage F statistic	66.367	65.185
Mean of outcome	−0.319	−0.271
Observations	4,377	4,276
Children treated in solo-provider clinics	0.498**	0.353
	(0.283)	(0.261)
	[0.032, 1.074]	[−0.076, 0.882]
First-stage F statistic	70.912	71.041
Mean of outcome	−0.305	−0.231
Observations	4,395	4,318

Appendix Table C1 (cont.): Robustness Checks Related to the Instrument

	Math (1)	Danish (2)
Clustering at the clinic level	0.520*	0.397
	(0.322)	(0.298)
	[−0.006, 1.168]	[−0.146, 0.996]
First-stage F statistic	50.686	54.078
Mean of outcome	−0.305	−0.241
Observations	5,495	5,383
Clustering at municipality and year-of-diagnosis level	0.520**	0.397*
	(0.239)	(0.284)
	[0.083, 0.956]	[−0.068, 0.971]
First-stage F statistic	70.1	68.3
Mean of outcome	−0.305	−0.241
Observations	5,495	5,383
Excluding children treated by top-3 prescribers	0.725**	0.479*
	(0.343)	(0.327)
	[0.094, 1.423]	[−0.061, 1.145]
First-stage F statistic	61.213	61.554
Mean of outcome	−0.314	−0.254
Observations	4,795	4,700
Including region instead of municipality fixed effects	0.437**	0.229
	(0.242)	(0.223)
	[0.032, 0.936]	[−0.186, 0.644]
First-stage F statistic	89.338	89.395
Mean of outcome	−0.305	−0.241
Observations	5,495	5,383

Notes: Analysis sample as described in the notes to Table 1. Each cell presents the results from a separate regression of the outcome listed in the column for the specification listed in the row. All specifications include the full set of controls listed in the notes to Table 2. The number of patients in clinics refers to the number of 8–15 year old children treated in the year of diagnosis of the child. Solo-provider clinics are clinics operated by a single psychiatrist throughout the entire year of diagnosis. Psychiatrists are ranked based on their average prescribing tendency over the entire sample period. The mean of the outcome is calculated among children not using AD. In addition to the coefficient estimate, each panel reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level unless mentioned otherwise. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

Appendix Table C2: Effect of AD use on the probability of delaying or forgoing taking the 9th grade tests

	All	Gender		Socioeconomic status	
		Girls	Boys	Low SES	High SES
	(1)	(2)	(3)	(4)	(5)
A. Took the test after 9th grade					
Math	−0.023	0.042	−0.053	0.040	−0.097
	(0.052)	(0.067)	(0.078)	(0.068)	(0.068)
	[−0.119, 0.073]	[−0.079, 0.176]	[−0.197, 0.106]	[−0.086, 0.179]	[−0.234, 0.027]
First-stage F statistic	84.8	53.7	65.0	52.3	55.5
Mean of outcome	0.042	0.045	0.040	0.046	0.033
Observations	7,211	2,914	4,297	4,671	2,540
Danish	−0.004	0.024	−0.001	0.064	−0.097
	(0.061)	(0.058)	(0.101)	(0.088)	(0.069)
	[−0.117, 0.109]	[−0.082, 0.141]	[−0.186, 0.204]	[−0.097, 0.242]	[−0.235, 0.028]
First-stage F statistic	84.8	53.7	65.0	52.3	55.5
Mean of outcome	0.050	0.044	0.053	0.056	0.039
Observations	7,211	2,914	4,297	4,671	2,540
B. Never took the test					
Math	0.039	0.052	0.038	0.015	0.061
	(0.096)	(0.114)	(0.142)	(0.131)	(0.129)
	[−0.139, 0.217]	[−0.178, 0.260]	[−0.223, 0.299]	[−0.251, 0.256]	[−0.173, 0.295]
First-stage F statistic	84.8	53.7	65.0	52.3	55.5
Mean of outcome	0.192	0.155	0.214	0.225	0.128
Observations	7,211	2,914	4,297	4,671	2,540
Danish	0.088	0.094	0.092	0.108	0.029
	(0.103)	(0.126)	(0.143)	(0.147)	(0.126)
	[−0.102, 0.279]	[−0.135, 0.323]	[−0.171, 0.354]	[−0.163, 0.378]	[−0.199, 0.258]
First-stage F statistic	84.8	53.7	65.0	52.3	55.5
Mean of outcome	0.200	0.152	0.229	0.235	0.132
Observations	7,211	2,914	4,297	4,671	2,540

Notes: Analysis sample as described in the notes to Table 1. Each cell presents the results from a separate regression of the outcome listed in the row for the specification listed in the panel title, estimated in the sample indicated in the column. Children of mothers with short cycle higher education or above are classified as high SES. All specification include: indicators for child gender, birth order, family size; indicators for mother's and father's education degree, employment status, income deciles, and marriage status at child age 6; indicators for mother's and father's use of SSRI during child age 6–7; mother's and father's age; and fixed effects for year of birth, year of diagnosis, and municipality of residence during the year of diagnosis. The mean of the outcome is calculated among children not using SSRI. In addition to the coefficient estimate, each cell reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

Appendix Table C3: Robustness Checks Related to the Sample

	Math (1)	Danish (2)
Baseline	0.520**	0.397*
	(0.266)	(0.253)
	[0.028, 1.063]	[−0.070, 0.913]
First-stage F statistic	82.786	82.197
Mean of outcome	−0.3	−0.2
Observations	5,495	5,383
Excluding children taking the test in pandemic years	0.597**	0.373
	(0.306)	(0.285)
	[0.093, 1.219]	[−0.097, 0.954]
First-stage F statistic	58.285	57.499
Mean of outcome	−0.3	−0.2
Observations	4,596	4,502
Including children treated in hospital psychiatric departments	0.461***	−0.079
	(0.202)	(0.184)
	[0.125, 0.876]	[−0.458, 0.263]
First-stage F statistic	46.724	46.628
Mean of outcome	−0.4	−0.3
Observations	39,975	38,675

Notes: Analysis sample as described in the notes to Table 1. Each cell presents the results from a separate regression of the outcome listed in the column for the specification listed in the row. All specifications include the full set of controls listed in the notes to Table 2. The number of patients in clinics refers to the number of 8–15 year old children treated in the year of diagnosis of the child. Solo-provider clinics are clinics operated by a single psychiatrist throughout the entire year of diagnosis. Psychiatrists are ranked based on their average prescribing tendency over the entire sample period. The mean of the outcome is calculated among children not using AD. In addition to the coefficient estimate, each panel reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

Appendix Table C4: Effect of psychiatrist tendency to prescribe AD on mobility across psychiatric clinics

	All	Gender		Socioeconomic status	
		Girls	Boys	Low SES	High SES
	(1)	(2)	(3)	(4)	(5)
Changed between private clinics	−0.061	−0.040	−0.099	−0.038	−0.086
	(0.100)	(0.121)	(0.127)	(0.096)	(0.151)
Effect of interquartile change in instrument	−0.005	−0.003	−0.008	−0.003	−0.007
Mean of outcome	0.063	0.063	0.064	0.062	0.065
Observations	5,495	2,310	3,185	3,391	2,104

Notes: Analysis sample as described in the notes to Table 1. Sample of children taking the math test (the estimates for the sample of children taking the Danish test are virtually identical and available upon request). Each cell presents the results from a separate reduced-form regression of the outcome listed in the row on the specialist tendency to prescribe AD in the sample indicated in the column. All specifications include the full set of controls listed in the notes to Table 2. Children of mothers with short cycle higher education or above are classified as high SES. The mean of the outcome is calculated among children not using AD. In addition to the coefficient estimate, each panel also lists the effect on the outcome of an increase in the value of the instrument from the 25th to the 75th percentile in the full analysis sample, corresponding to approximately 7.9 percentage points. Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table C5: Robustness Checks Related to AD treatment

	Math (1)	Danish (2)
Baseline	0.520**	0.397*
	(0.266)	(0.253)
	[0.028, 1.063]	[−0.070, 0.913]
First-stage F statistic	82.786	82.197
Mean of outcome	−0.3	−0.2
Observations	5,495	5,383
Excluding treated children on AD at time of test	0.765**	0.556*
	(0.398)	(0.360)
	[0.031, 1.577]	[−0.039, 1.290]
First-stage F statistic	48.7	50.1
Mean of outcome	−0.305	−0.241
Observations	5,180	5,070
Excluding all children on AD at time of test	0.715**	0.491
	(0.379)	(0.346)
	[0.015, 1.489]	[−0.148, 1.197]
First-stage F statistic	49.4	50.4
Mean of outcome	−0.308	−0.248
Observations	5,097	4,986
Including school fixed effects	0.549*	1.163***
	(0.360)	(0.395)
	[−0.038, 1.197]	[0.588, 1.874]
First-stage F statistic	54.6	49.4
Mean of outcome	−0.245	−0.168
Observations	4,819	4,800

Notes: Analysis sample as described in the notes to Table 1. Each cell presents the results from a separate regression of the outcome listed in the column for the specification listed in the row. All specifications include the full set of controls listed in the notes to Table 2. We consider a child as using AD at the time of the test if the number of daily defined doses from the last prescription filled before the test is larger than the number of days between the date of the last prescription filled and the date of the test. School fixed effects are based on the school in which the student was enrolled at the time of the test. The mean of the outcome is calculated among children not using AD. In addition to the coefficient estimate, each panel reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

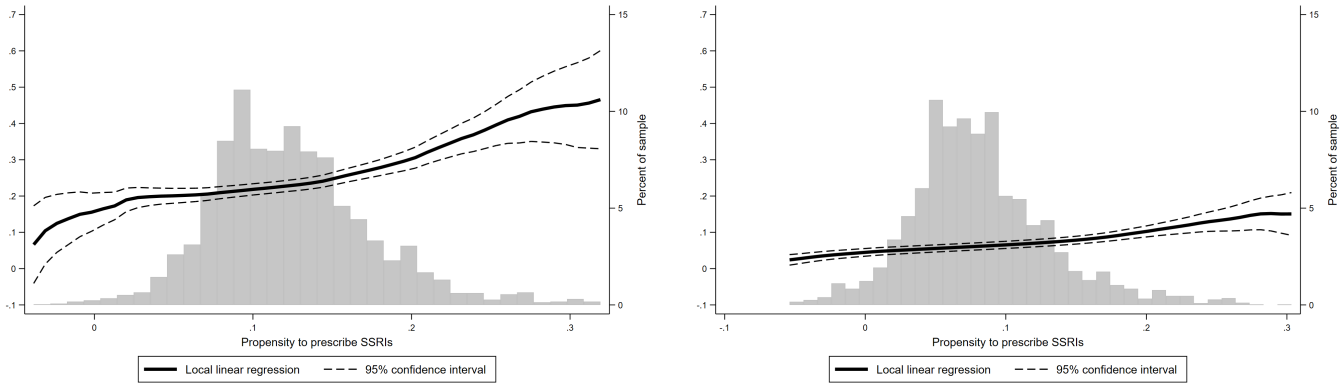
Appendix Table C6: Effect of AD use on 2nd–8th grade test scores

	Analysis sample (1)	Full MH sample (2)
Standardized test score, math	0.385	0.472
	(1.267)	(0.969)
	[−3.491, +∞]	[−1.196, 2.510]
First-stage F statistic	8.0	14.6
Mean of outcome	−0.250	−0.319
Observations	1,416	11,619
Children	690	5,717
Standardized test score, Danish	−0.030	0.318
	(0.299)	(0.206)
	[−0.609, 0.549]	[−0.079, 0.716]
First-stage F statistic	50.9	174.2
Mean of outcome	−0.186	−0.227
Observations	5,789	52,787
Children	1,745	16,223

Notes: The analysis sample is described in the notes to Table 1, while the full MH sample is the sample of all children whose first mental health-related visit between the ages of 8–15 was either at a hospital psychiatric department or with a privately-practicing psychiatrist (see the notes to Table 4). In both cases, the sample is restricted to children who took the test at least once before diagnosis (first mental health-related visit) and at least once after diagnosis. Each cell presents the results from a separate regression of the outcome listed in the row in the sample listed in the column. All specification include: indicators for the calendar year when the test is taken, for the grade in which the child is, for whether the student skipped a test, and for the scale on which the test was graded (pre- or post-2014); and child fixed effects. The main independent variable is an indicator for filling an AD prescription before the date of the test. The instrument is the interaction between the specialist tendency to prescribe AD as defined in Equation (2) and an indicator for the test being taken after the first mental health visit. The mean of the outcome is calculated among children not using AD. In addition to the coefficient estimate, each panel reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

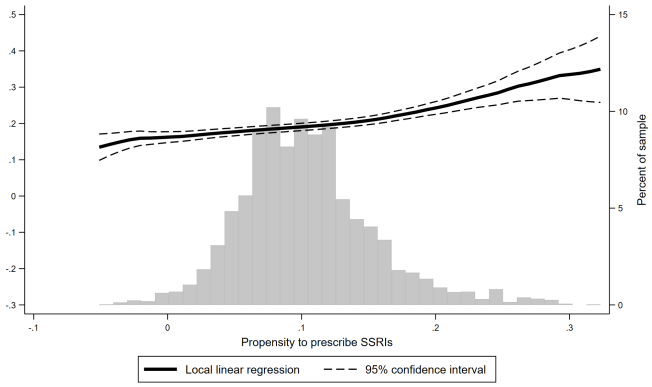
Appendix D

Treatment Heterogeneity by Observable Characteristics

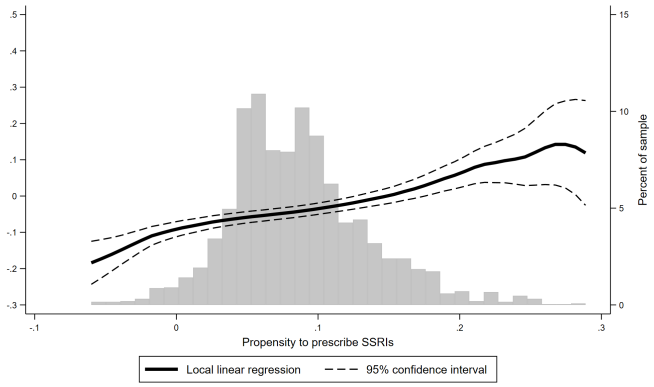


(a) Girls

(b) Boys



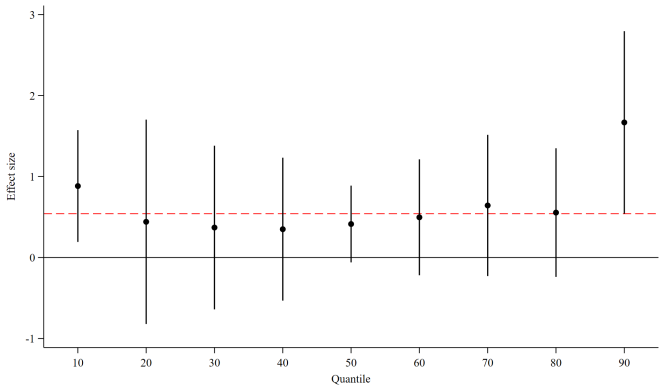
(c) Low SES



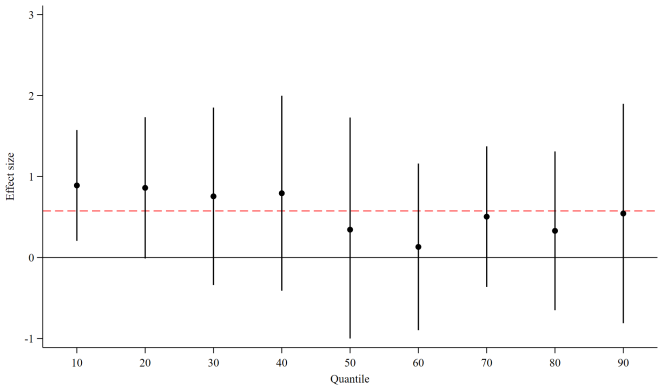
(d) High SES

Appendix Figure D1: Distribution of psychiatrist tendency to prescribe ADs by gender and SES

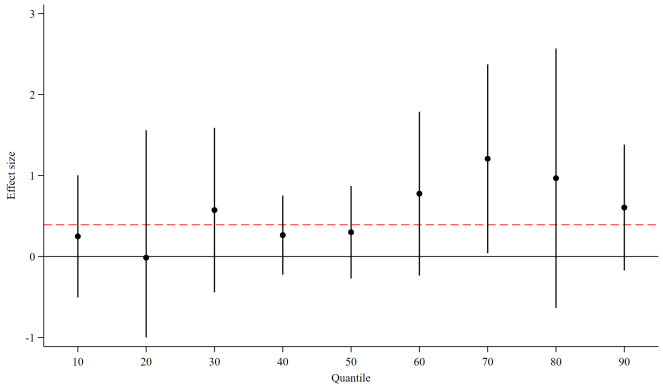
Notes: Both the propensity to prescribe AD and the probability of filling a prescription by age 15 are net of year-of-diagnosis and municipality fixed effects. Hence, the location of the fitted curves is not meaningful but the slopes are. The gray bars represent the histogram of the adjusted tendency to prescribe antidepressants. The solid line plots the local polynomial regression of the probability of receiving antidepressants by age 15 on the tendency to prescribe antidepressants of the treating specialist within the group indicated in the panel, while the dashed lines indicate the corresponding 95% confidence interval.



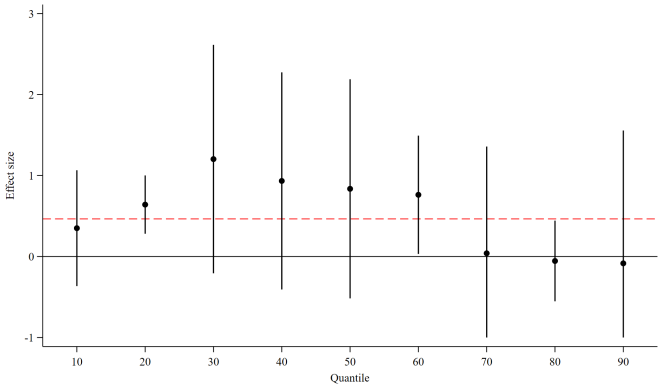
(a) Girls, math



(b) Boys, math



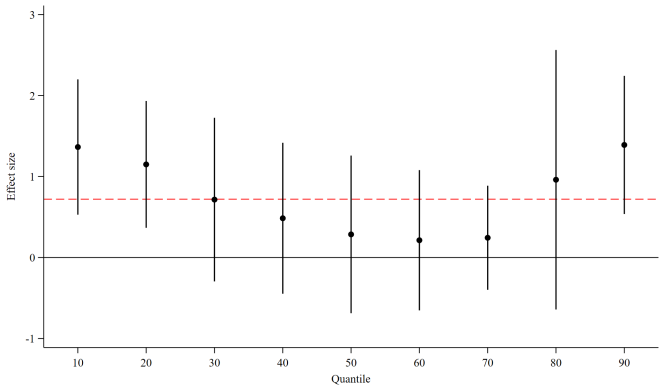
(c) Girls, Danish



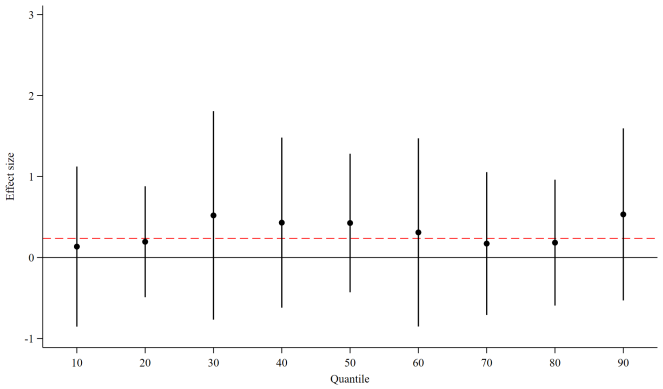
(d) Boys, Danish

Appendix Figure D2: Effects of AD use along the distribution of test scores by gender

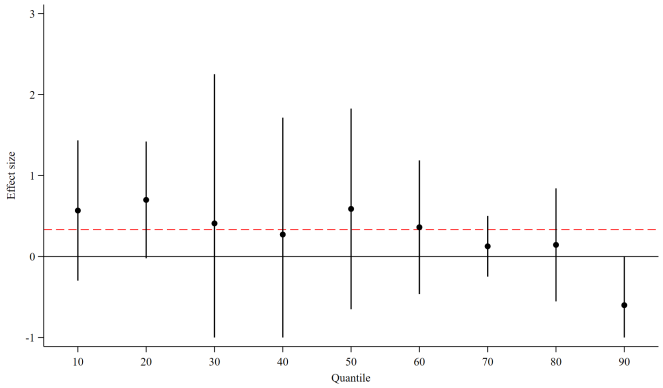
Notes: In each panel, each point and line segment represent the estimated effect and corresponding 95% confidence interval from an instrumental variables quantile regression estimated at the quantile indicated on the horizontal axis for the outcome and subsample indicated in the panel. Each regression includes the full set of controls listed in Table 2 and is estimated using smoothed estimating equations (Kaplan, 2022; Kaplan and Sun, 2017). Standard errors obtained from 100 bootstrap replications. Some confidence intervals are truncated to improve the legibility of the figures.



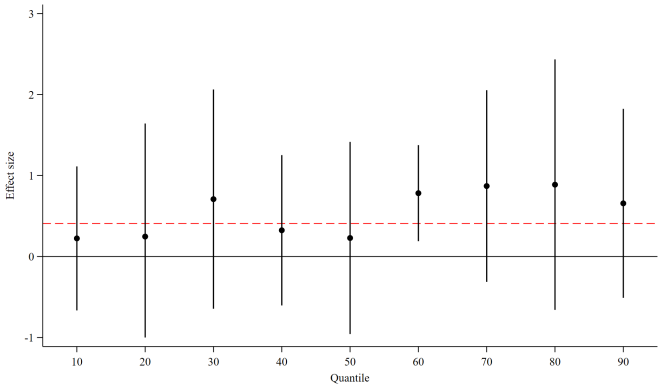
(a) Low SES, math



(b) High SES, math



(c) Low SES, Danish



(d) High SES, Danish

Appendix Figure D3: Effects of AD use along the distribution of test scores by SES

Notes: In each panel, each point and line segment represent the estimated effect and corresponding 95% confidence interval from an instrumental variables quantile regression estimated at the quantile indicated on the horizontal axis for the outcome and subsample indicated in the panel. Each regression includes the full set of controls listed in Table 2 and is estimated using smoothed estimating equations (Kaplan, 2022; Kaplan and Sun, 2017). Standard errors obtained from 100 bootstrap replications. Some confidence intervals are truncated to improve the legibility of the figures.

Appendix Table D1: Characteristics of the analysis sample and of the sample of children with no mental health contact by SES

	No mental health contact		Analysis sample					
			All		No SSRI		SSRI	
	Low SES (1)	High SES (2)	Low SES (3)	High SES (4)	Low SES (5)	High SES (6)	Low SES (7)	High SES (8)
A. Outcomes								
Standardized test score, math	−0.116	0.440	−0.488	0.047	−0.518	0.047	−0.302	0.043
	(0.965)	(0.868)	(0.985)	(0.979)	(0.986)	(0.979)	(0.959)	(0.981)
Standardized test score, Danish	−0.097	0.358	−0.394	0.111	−0.432	0.070	−0.164	0.326
	(0.967)	(0.944)	(0.978)	(1.022)	(0.972)	(1.019)	(0.985)	(1.014)
B. Treatments								
Therapy within 1 year before first MH visit	—	—	0.112	0.125	0.089	0.103	0.252	0.230
Therapy after first MH visit, by age 15	—	—	0.730	0.682	0.705	0.653	0.881	0.822
Pharmaceutical treatments by age 15:								
SSRI	0.001	0.001	0.140	0.173	0.000	0.000	1.000	1.000
tricyclic antidepressant	0.001	0.001	0.002	0.003	0.002	0.003	0.003	0.002
benzodiazepine	0.006	0.005	0.015	0.017	0.013	0.012	0.026	0.041
stimulant	0.006	0.003	0.359	0.306	0.370	0.313	0.291	0.273
antipsychotic	0.001	0.001	0.078	0.069	0.058	0.041	0.204	0.200
Any non-SSRI N-drug	0.064	0.057	0.469	0.451	0.458	0.425	0.535	0.576
Age when first SSRI was filled	13.517	13.416	12.673	12.675	—	—	12.673	12.675
	(1.968)	(2.002)	(1.768)	(1.795)			(1.768)	(1.795)
Median days between first CP visit and:								
first SSRI prescription	—	—	111.000	103.000	—	—	111.000	103.000
first non-SSRI prescription	—	—	−18.000	−25.500	—	—	−18.000	−25.500
First therapy	—	—	28.000	28.000	28.000	28.000	28.000	28.000
Median days b/w first therapy and first SSRI	—	—	139.000	112.000	—	—	139.000	112.000
C. Child characteristics								
Boy	0.502	0.508	0.613	0.565	0.631	0.584	0.502	0.472
Birth order:								
1	0.448	0.461	0.491	0.489	0.494	0.498	0.468	0.449
2	0.376	0.370	0.363	0.367	0.359	0.366	0.392	0.367
3+	0.176	0.169	0.146	0.144	0.147	0.136	0.140	0.185
Number of siblings:								
0	0.143	0.103	0.201	0.160	0.202	0.164	0.194	0.141
1	0.555	0.571	0.550	0.579	0.547	0.584	0.569	0.554
2	0.211	0.244	0.171	0.189	0.171	0.181	0.175	0.223
3+	0.091	0.081	0.078	0.072	0.080	0.070	0.062	0.082
Parents are married	0.771	0.879	0.617	0.795	0.602	0.786	0.706	0.836
Age at first diagnosis	—	—	11.612	11.684	11.560	11.626	11.931	11.958
			(2.081)	(2.046)	(2.093)	(2.052)	(1.976)	(1.997)

Appendix Table D1 (cont.): Characteristics of the analysis sample and of the sample of children with no mental health contact by SES

	No mental health contact		Analysis sample					
			All		No SSRI		SSRI	
	Low SES (1)	High SES (2)	Low SES (3)	High SES (4)	Low SES (5)	High SES (6)	Low SES (7)	High SES (8)
D. Psychiatrist characteristics								
Female	—	—	0.426	0.389	0.424	0.391	0.434	0.378
Age (years)	—	—	57.149	56.940	57.185	56.966	56.931	56.814
			(5.318)	(5.068)	(5.345)	(5.162)	(5.146)	(4.594)
Experience (years)	—	—	27.911	27.976	27.877	27.923	28.120	28.229
			(5.599)	(5.638)	(5.637)	(5.766)	(5.356)	(4.977)
Married	—	—	0.652	0.667	0.655	0.663	0.636	0.683
Solo practice	—	—	0.803	0.774	0.800	0.775	0.825	0.770
Number of patients treated by specialist during year	—	—	171.912	165.863	172.055	166.529	171.040	162.672
			(80.110)	(77.419)	(80.170)	(77.414)	(79.799)	(77.451)
Propensity to prescribe SSRIs	—	—	0.108	0.100	0.104	0.095	0.133	0.127
			(0.061)	(0.060)	(0.059)	(0.057)	(0.067)	(0.065)
E. Mother's characteristics at child age 6								
Age	35.067	37.290	34.578	37.785	34.483	37.728	35.159	38.057
	(4.727)	(4.088)	(5.062)	(4.368)	(5.123)	(4.375)	(4.634)	(4.330)
Some college or more	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
Employed	0.804	0.937	0.710	0.909	0.706	0.902	0.735	0.938
Annual gross total income ('000 2015 DKK)	297.880	394.026	290.227	393.751	290.036	394.551	291.394	389.926
	(146.857)	(234.182)	(112.906)	(184.701)	(115.211)	(191.105)	(97.686)	(150.449)
Used SSRI between child ages 6–7	0.098	0.065	0.177	0.147	0.173	0.136	0.203	0.198
F. Father's characteristics at child age 6								
Age	37.599	39.374	37.228	39.768	37.140	39.716	37.765	40.014
	(5.459)	(5.118)	(5.679)	(5.459)	(5.692)	(5.491)	(5.571)	(5.302)
Some college or more	0.152	0.547	0.135	0.548	0.131	0.548	0.160	0.547
Employed	0.916	0.959	0.860	0.935	0.855	0.932	0.893	0.950
Annual gross total income ('000 2015 DKK)	409.058	523.789	380.862	512.921	378.946	514.187	392.591	506.864
	(513.712)	(628.356)	(249.319)	(431.295)	(256.496)	(450.761)	(199.623)	(322.602)
Used SSRI between child ages 6–7	0.054	0.044	0.086	0.091	0.084	0.091	0.099	0.093
Observations	489,031	281,489	4,671	2,540	4,015	2,101	656	439

Notes: Columns 1 and 2 describe the sample of all native children born between 1991–2005, who are observed every year between the ages of 6 and 16, and who have no contact with a psychiatrist between ages 6–15. Columns 3–8 describe our analysis sample of all native children born between 1991–2005, who are observed every year between the ages of 6 and 16, who have no contact with a psychiatrist between ages 6–7, and whose first psychiatric visit between ages 8–15 is with a child psychiatrist in a private clinic. Children of mothers with short cycle higher education or above are classified as high SES. Number of siblings refers to the total number of siblings in the household when the child is aged 6. All parental characteristics, including marital status, are measured when the child is aged 6. All monetary variables are in 2015 DKK. See Appendix Table A1 for details on how each variable is constructed.

Appendix Table D2: Effects of AD use in the analysis sample and in the full MH sample, by SES

	Low SES		High SES	
	Analysis sample (1)	Full MH sample (2)	Analysis sample (3)	Full MH sample (4)
Standardized test score, math	0.720**	0.787***	0.236	−0.176
	(0.428)	(0.219)	(0.336)	(0.296)
	[0.020, 1.669]	[0.419, 1.241]	[−0.371, 0.843]	[−0.788, 0.377]
First-stage F statistic	37.3	98.2	55.6	53.6
Mean of outcome	−0.518	−0.585	0.050	−0.021
Observations	3,391	27,027	2,104	12,948
Standardized test score, Danish	0.333	−0.061	0.408	−0.066
	(0.393)	(0.189)	(0.344)	(0.309)
	[−0.386, 1.128]	[−0.414, 0.293]	[−0.277, 1.028]	[−0.704, 0.511]
First-stage F statistic	35.3	91.4	54.6	59.4
Mean of outcome	−0.431	−0.454	0.074	0.009
Observations	3,307	26,047	2,076	12,628

Notes: The analysis sample is described in the notes to Table 1, while the full MH sample is the sample of all children whose first mental health-related visit between the ages of 8–15 was either at a hospital psychiatric department or with a privately-practicing psychiatrist (see the notes to Table 4). Each cell presents the results from a separate regression of the outcome listed in the row in the sample listed in the column. All specifications include the full set of controls listed in the notes to Table 2. The mean of the outcome is calculated among children not using AD. In addition to the coefficient estimate, each panel reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the robust first-stage F-statistic, which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

Appendix Table D3: Effect of SSRI use on test scores, taking into account other treatments and psychiatrist characteristics, by SES

	Baseline		Taking into account other treatments and psychiatrist characteristics	
	Low SES (1)	High SES (2)	Low SES (3)	High SES (4)
A. 2SLS, Standardized test score, math				
Treatment received:				
— AD	0.720**	0.236	1.319	0.050
	(0.428)	(0.336)	(0.908)	(0.371)
	[0.020, 1.669]	[−0.371, 0.843]	[−∞, −2.145] ∪ [1.319, +∞]	[−0.646, 1.443]
— Non-SSRI N-class medication			−0.624	0.005
			(0.377)	(0.322)
			[−∞, −0.624] ∪ [0.813, +∞]	[−1.206, 0.610]
— Therapy			1.065	0.893
			(1.524)	(1.031)
			[−∞, −7.649] ∪ [1.065, +∞]	[−1.043, 8.637]
Lewis-Mertens statistic			4.3	5.2
Lewis-Mertens 95% critical value			20.5	20.1
Mean of outcome	−0.518	0.050	−0.518	0.050
Observations	3,391	2,104	3,391	2,104
B. 2SLS, Standardized test score, Danish				
Treatment received:				
— AD	0.333	0.408	1.919	0.081
	(0.393)	(0.344)	(1.606)	(0.537)
	[−0.386, 1.128]	[−0.277, 1.028]	[−∞, +∞]	[−0.927, 2.095]
— Non-SSRI N-class medication			−0.981	−0.013*
			(0.622)	(0.431)
			[−∞, +∞]	[−∞, −0.013]
— Therapy			3.025	1.963**
			(2.988)	(1.419)
			[−∞, −8.360] ∪ [3.025, +∞]	[1.963, +∞]
Lewis-Mertens statistic			3.7	4.9
Lewis-Mertens 95% critical value			20.7	19.5
Mean of outcome	−0.431	0.074	−0.431	0.074
Observations	3,307	2,076	3,307	2,076
Psychiatrist characteristics	No	No	Yes	Yes

Notes: Analysis sample as described in the notes to Table 1. Each column presents the results from a separate regression of the outcome indicated in the panel using the specification and in the subsample indicated in the column. All specifications include the full set of controls listed in Table 2. In addition, Columns 3 and 4 include the full set of psychiatrist characteristics, as well as two indicator variables, one for the use of non-SSRI N-drug medication and one for the use of therapy, instrumented for with the corresponding psychiatrist tendencies. Each panel reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the Lewis and Mertens (Forthcoming) statistic for instrument strength and the corresponding 95% critical value. The mean of the outcome is calculated among children not using AD. Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.

Appendix E

Marginal Treatment Effects

E.1 General setup

Suppose the potential outcomes of child i when treated and when not treated, and the observed outcome, respectively, are given by:

$$\begin{aligned}
Y_i^1 &= \mu^1(X_i) + u_i^1, \\
Y_i^0 &= \mu^0(X_i) + u_i^0, \\
Y_i &= Y_i^1 \cdot AD_i + Y_i^0 \cdot (1 - AD_i) = \mu^0(X_i) + AD_i\left[\mu^1(X_i) - \mu^0(X_i) + u_i^1 - u_i^0\right] + u_i^0 \\
&= \mu^0(X_i) + AD_i \cdot \Delta_i + u_i^0,
\end{aligned}$$

where X_i is a set of observable characteristics of the child and of the parents, and Δ_i is the treatment effect of AD on child i.

Suppose also that the decision to take AD treatment (the first stage) depends on the entire set of instruments Z_i, which includes the observable characteristics X and the propensity to prescribe AD of the specialist PP, and on an unobserved "resistance to treatment" V:

$$AD_i = \mathbb{1}\left[\mu(Z_i) - V_i \geq 0\right].$$

The condition in the equation above can be rewritten as $F_v[\mu(X_i, PP_i)] \geq F_v(V_i)$, where $F_v(\cdot)$ is the cumulative distribution function of V. The left-hand side of this inequality is the probability of receiving AD as a function of the observable characteristics X and the instrument PP, which is just the propensity score $P(Z_i)$ derived from the first-stage regression:

$$P(z_i) = Pr(AD_i = 1 | Z_i = z_i) = \mu(X_i, PP_i) + \epsilon_i.$$

The right-hand side of the inequality represents the quantiles of the distribution of V, $U_{Di} = F_v(V_i)$. Therefore, child i will be treated with AD if the propensity score exceeds their place (quantile) in the distribution of unobserved resistance to treatment.

The marginal treatment effect (MTE) is given by the treatment effect at a particular point in the distribution of unobserved resistance to treatment:

$$MTE(x, u_D) = E[Y_i^1 - Y_i^0 | X_i = x, U_{Di} = u_D] = E[\Delta_i | X_i = x, V_i = v].$$

The marginal treatment effect can then be used to derive other parameters of interest, such as the average treatment effect ATE, the average treatment effect on the treated ATT and on the untreated $ATUT$, as well as the IV

estimator from a 2SLS approach:

$$
\begin{aligned}
ATE &= E(Y^1 - Y^0 | X = x) = \int_0^1 MTE(x, u_D) du_D, \\
ATT &= E(Y^1 - Y^0 | X = x, AD = 1) = \int_0^1 MTE(x, u_D) \omega_{ATT}(x, u_D) du_D, \\
ATUT &= E(Y^1 - Y^0 | X = x, AD = 0) = \int_0^1 MTE(x, u_D) \omega_{ATUT}(x, u_D) du_D, \\
LATE &= \int_0^1 MTE(x, u_D) \omega_{LATE}(x, u_D) du_D,
\end{aligned}
\tag{E.1}
$$

where the ω's are parameter-specific weights (see, e.g., Heckman et al., 2006, for details and formulas). It can be shown that for two given values of the instrument z and z' such that $AD(z) = 1$ and $AD(z') = 0$, the local average treatment effect is:

$$
\begin{aligned}
LATE(z, z') &= E[Y_i^1 - Y_i^0 | X_i = x, AD_i(z) = 1, AD_i(z') = 0] \\
&= E[Y_i^1 - Y_i^0 | X_i = x, u'_D < U_{iD} < u_D].
\end{aligned}
$$

This implies that, as $u'_D \to u_D$, the $LATE$ converges to the MTE.

E.2 Estimation

We assume linearity in both the potential outcomes equations ($\mu_i^1(X_i) = X_i\beta^1$ and $\mu_i^0(X_i) = X_i\beta^0$) and in the first stage ($\mu(X_i, PP_i) = X_i\delta_X + PP_i\delta$). We also assume additive separability between observed and unobserved heterogeneity:

$$
\begin{aligned}
E[Y_i^1 | X_i = x, U_{Di} = u] &= x\beta^1 + E[u_i^1 | X_i = x, U_{Di} = u], \\
E[Y_i^0 | X_i = x, U_{Di} = u] &= x\beta^0 + E[u_i^0 | X_i = x, U_{Di} = u].
\end{aligned}
\tag{E.2}
$$

Then the model outlined in the previous section is a Heckman selection model, which can be estimated via joint maximum likelihood by assuming that the error terms follow a normal distribution. In this *normal parametric* case, the MTE is given by:

$$
MTE(x, u_D) = E[Y_i^1 - Y_i^0 | X_i = x, U_{Di} = u_D] = x(\beta^1 - \beta^0) + (\rho^1 - \rho^0)\Phi^{-1}(u_D),
$$

where ρ^1 and ρ^0 are the correlation coefficients between U_i^1 and V_i, and U_i^0 and V_i, respectively. This approach restricts the shape of the MTE curve to be monotonic, given by $\Phi^{-1}(\cdot)$. The sign of the difference $(\rho^0 - \rho^1)$ (the reverse of the term above) gives the direction of the selection on unobserved gains.

The normality assumption is very restrictive, so the usual estimation approach (and the one we follow in this paper) is the method of *local instrumental variables (local IV)*, in which incremental changes in the propensity score serve as local instruments that identify the MTE at each value of V. In other words, at each value of V there are some persons on the margin for taking AD treatment and the MTE is the difference in the potential outcomes with and without the treatment for these people. Heckman and Vytlacil (2005) show that in this case, the MTE can be

estimated as the change in observed outcomes induced by a marginal change in the propensity score:

$$MTE(x, p) = \frac{\partial E[Y|P(Z) = p]}{\partial p}.$$

When the instrument has enough variation so that the propensity score has full common support within cells defined by X_i, then the MTE curve can be estimated nonparametrically within each cell. In general, however, the propensity score will not have full common support. In this case, we need to impose the parametric assumptions in Equation (E.2) and the full independence assumption $(X_i, Z_i) \perp\!\!\!\perp (U_i^0, U_i^1, V_i)$, which implies that the observable characteristics X can affect the intercept of the MTE curve but not its shape (which is given by the way in which U_i^1 and U_i^0 depend on V_i).

Under these assumptions, the MTE can be rewritten as:

$$MTE(x, u) = x(\beta_1 - \beta_0) + E\left[u_i^1 - u_i^0 | X_i = x, U_{Di} = u\right].$$

Given that the potential outcomes are assumed to be linear in X_i and that the MTE has constant shape across X_i, the outcome equation becomes:

$$E[Y_i|X_i = x, P(Z_i) = p] = X_i\beta^0 + X_i(\beta^1 - \beta^0)p + K(p),$$

where $K(\cdot)$ is a nonlinear function in the propensity score. The coefficients on the interaction term $X_i p$ show how the observable characteristics affect the treatment effect and the intercept of the MTE curve. The last term, $K(p)$, gives the shape of the MTE curve and, importantly, is estimated across all values of X_i instead of within cells of X_i as in the nonparametric approach. Therefore, this requires unconditional full common support, a much weaker assumption than full common support within the cells defined by X_i.

Finally, taking the derivative of the outcome equation with respect to the propensity score yields the MTE:

$$MTE(x, U_{Di} = p) = x(\beta_1 - \beta_0) + \frac{\partial K(p)}{\partial p}.$$

The form of the estimator above depends on the functional form chosen for $K(\cdot)$. Two common approaches are the *semiparametric* form proposed by Heckman et al. (2006), or a *parametric* approach in which $K(p)$ is a polynomial in p.

Our preferred specification is a parametric one with a third-degree polynomial in p. This implies that the MTE is a second-degree polynomial in p:

$$MTE(x, U_{Di} = p) = x(\beta_1 - \beta_0) + \left(\pi_1 + \pi_2 p + \pi_3 p^2\right). \tag{E.3}$$

We conduct the estimation using the `mtefe` Stata package provided by Andresen (2018). We should note at this point that the estimation procedure discretizes the distribution of V (i.e., it uses a grid of values for U_{Di}), meaning that the weights ω in Equation (E.1) represent the fraction of children with that level of unobserved resistance to treatment among all the children contributing to the calculation of the specific treatment effect.

We should also note that the more continuous the distribution of the instrument, the easier it is to estimate

the MTE. In many applications of the judge fixed effect framework, continuity boils down to a few discrete values because typically the instrument is constructed at the judge level and there are only two possible values of the instrument per judge. The distribution of our instrument is shown in Figure 1. The 411 unique values range from 0 to 0.3375, with a median difference between two consecutive values of 0.0008, or less than 1% of the range. Therefore, the distribution of our instrument appears to be reasonably continuous.

E.3 Robustness of MTE curves

We conduct three types of specification checks. First, we compare the estimated curves from the parametric normal, semiparametric, and parametric approaches with different polynomials in p. If our baseline model is correctly specified, then we should not find large differences between the MTE curves estimated under the three different estimation strategies.

Second, a misspecified first stage can lead to bias in the MTE curve. Following Cornelissen et al. (2018), we compare the propensity score estimated using our baseline approach to a propensity score estimated semiparametrically as follows. We first classify the instrument into equal-size bins of width 0.05. Next, we predict the probability of receiving AD from a linear regression on all the control variables included in our baseline specification and we classify this predicted probability into equally-sized bins of size 0.05. We obtain the semiparametric propensity score by regressing the AD indicator on the full set of interactions between the two sets of bins constructed above. Finally, we plot the two propensity scores against the instrument and we calculate the correlation coefficient between them. A high correlation would suggest that our baseline first stage is a sufficiently flexible approximation of the propensity score.

Third, Devereux (2022) notes that the estimated MTE curves are sensitive to the set of controls included. Of particular concern are correlations between the propensity score and higher-order terms in the controls included in the MTE specification Equation (E.3). In the spirit of the recommendations in Devereux (2022), we construct two indices of the controls: a linear index obtained as the predicted values from an OLS regression of the probability of AD use on all the controls, and a non-linear index obtained in a similar fashion from a probit regression. We then augment the MTE specification to include second-order terms in each of the two indices. If the baseline MTE is biased because the estimated propensity score is correlated with these terms, then we should see a difference between the baseline MTE curves and the MTE curves obtained through this strategy.

E.4 Interacted MTEs

In principle, there are no theoretical restrictions on the function $K(\cdot)$. This is particularly useful for our case because we are interested in letting the shape of the MTE curve vary across groups defined by gender or SES. In particular, we modify the parametric approach by including group-specific polynomials in the propensity score:

$$K(p) = \sum_{m=0}^{M} \left(\pi_m + \sum_{j=2}^{J} G_i^j \theta_m^j \right) p^m,$$

where J is the number of groups, G_i^j an indicator for person i belonging to group j, and M is the degree of the polynomial in p. In our case, both the classification by gender and that by SES consist of only two groups so the

functional form above simplifies to:

$$K(p) = \sum_{m=0}^{M} (\pi_m + G_i\theta_m)\, p^m,$$

where G_i is an indicator for child i being a boy or high-SES. As a result, the shape of the MTE is allowed to differ between the two groups:

$$\begin{aligned} MTE(x, U_{Di} = p) &= x(\beta_1 - \beta_0) + \sum_{m=1}^{M} (\pi_m + G_i\theta_m)\, p^{m-1} \\ &= x(\beta_1 - \beta_0) + \left(\pi_1 + \pi_2 p + \pi_3 p^2\right) + G_i \left(\theta_1 + \theta_2 p + \theta_3 p^2\right) \end{aligned}$$

We estimate the MTE above using a modified version of the `mtefe` Stata package provided by Andresen (2018). This modified package, available from the authors, allows the shape of the MTE curve to differ across two or more different groups. As before, we can calculate the usual treatment effects by averaging the MTE with appropriate weights. Both in the estimation of the MTE and of the treatment effects, we use the average values of the covariates over the entire sample so that the difference between the two sets of estimates (MTE and treatment effects) is driven entirely by differences in the shape of the polynomial in p.

E.5 Policy relevant treatment effects

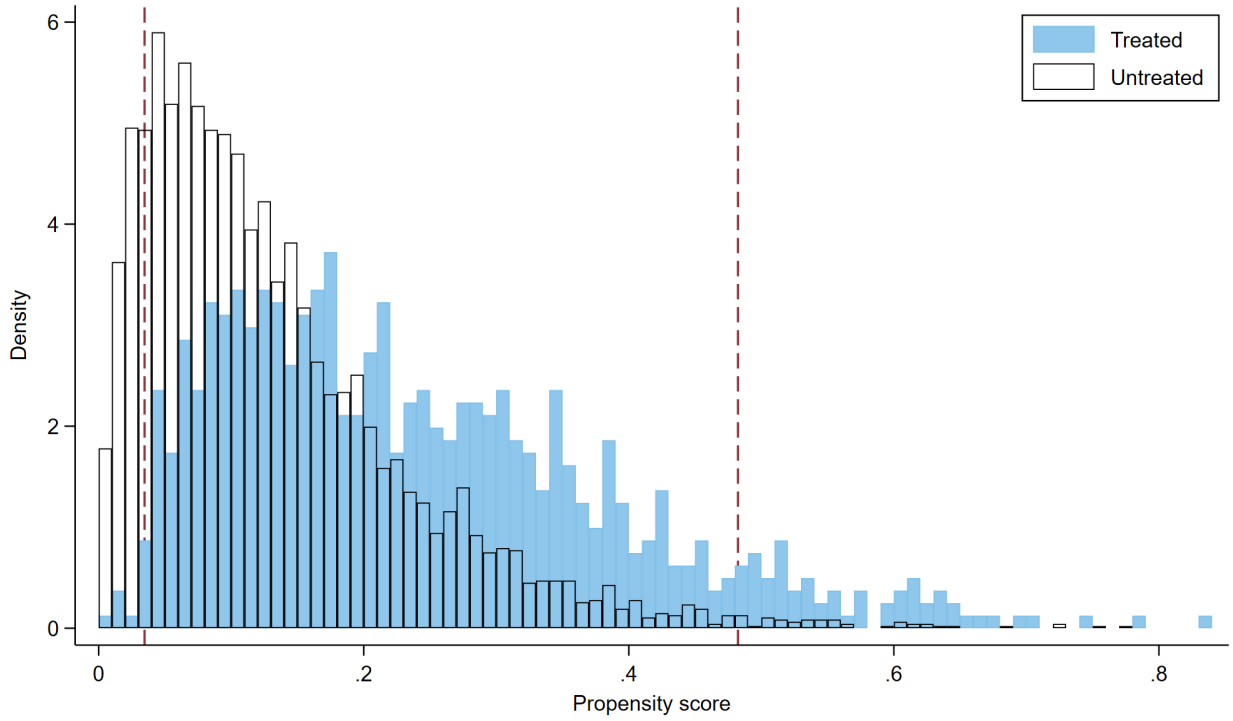
The MTE framework allows us to conduct policy experiments and estimate the effects of specific policies that change the *status quo* (Heckman and Vytlacil, 2005, 2001). Consider a policy that changes the probability of receiving the treatment (the propensity score) for some children, but not the potential outcomes from treatment (Y_i^0 and Y_i^1) or the unobserved selection into treatment (V_i). This policy does not change the benefits from treatment (the MTEs), but rather who gets the treatment. The effect of the policy, or the policy-relevant treatment effect $PRTE$, can be defined as the average change in outcomes between the new policy and the status quo divided by the change in the share of treated children:

$$PRTE = \frac{E[Y_i|\text{new policy}] - E[Y_i|\text{status quo}]}{E[AD_i|\text{new policy}] - E[AD_i|\text{status quo}]}.$$

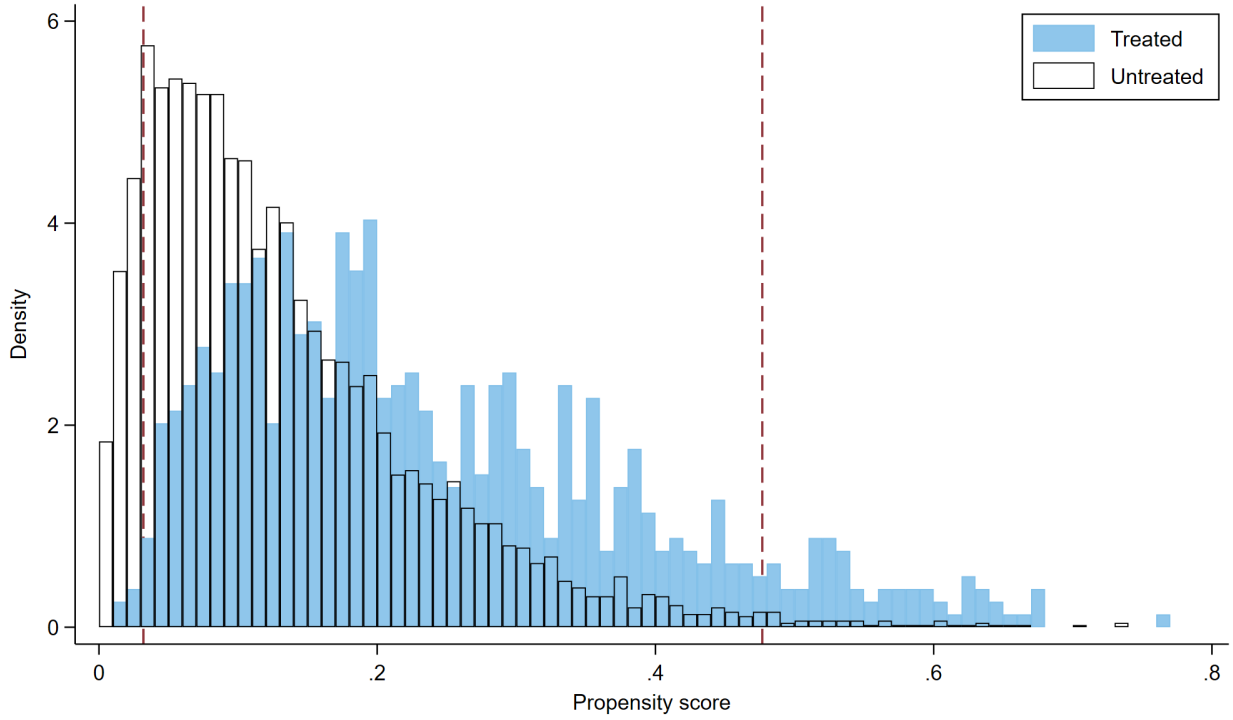
The $PRTE$ is the average change in outcomes per net person shifted into treatment. This is more difficult to interpret when the policy shifts some people into and some out of treatment. However, the policies studied in this paper change the treatment status of all affected children in the same direction. In this case, the $PRTE$ becomes the average treatment effect (on the treated) per child affected by the policy.

In practice, we first define a new policy that changes the treatment status of some children. Next, we estimate the MTEs for the affected children (those whose treatment status would change if this policy were implemented) at each level of unobserved resistance to treatment, which yield the *policy-relevant marginal treatment effect (PRMTE)* curve. As in Equation (E.1), the calculation of the $PRTE$ involves integrating weighted PRMTEs, and the corresponding discretized weights represent the share of children at each level of unobserved resistance to treatment among all the children affected by the policy. Using these weights we can calculate the $PRTE$ and the total fraction of children affected by the policy.

There are several ways in which we assess the policies we consider. First, at each level of unobserved resistance to treatment, we can compare the returns to treatment for children affected by the policy relative to children treated under the status quo by comparing the PRMTE curve to the MTE curve. If the PRMTE curve lies above the MTE curve, then the children affected by the policy have higher returns from treatment than the children already treated, and vice-versa. Second, the PRMTE weights indicate what fraction of the affected children are at each point in the distribution of returns to treatment. We can then use these weights to calculate the fraction of children negatively affected by the policy. For example, if the policy involves shifting children into treatment, then the sum of the PRMTE weights corresponding to the range of negative returns to AD treatment yields the fraction of children affected by the policy who are harmed by it. Finally, we compare the PRMTE to the ATT under the status quo to assess how the average treatment effect for the children affected by the policy compares to the average treatment effect for the children treated under the status quo.



(a) Math

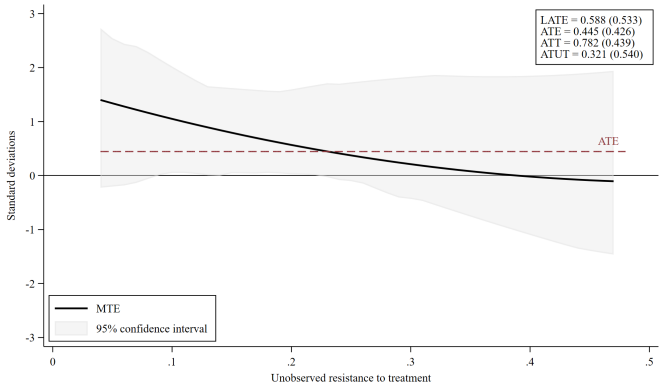


(b) Danish

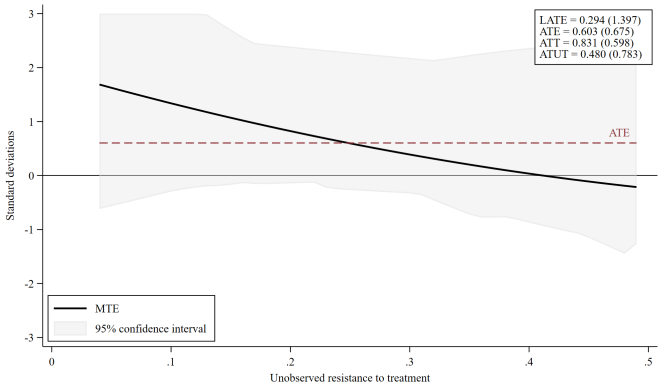
Appendix Figure E1: Common support for the propensity score used to estimate the marginal treatment effects

Notes: Each panel plots the distribution of the propensity score used to estimate the marginal treatment effects, separately for treated (filled bars) and untreated (hollow bars) children. The dashed vertical lines represent the upper and lower bounds of common support after 1% trimming of both tails. The propensity score is predicted from a probit regression of the indicator for AD use by age 15 on the instrument and the full set of controls listed in Table 2.

I. Math

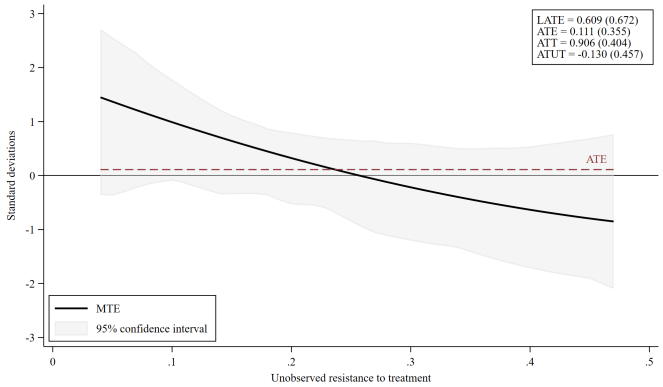


(a) Full analysis sample

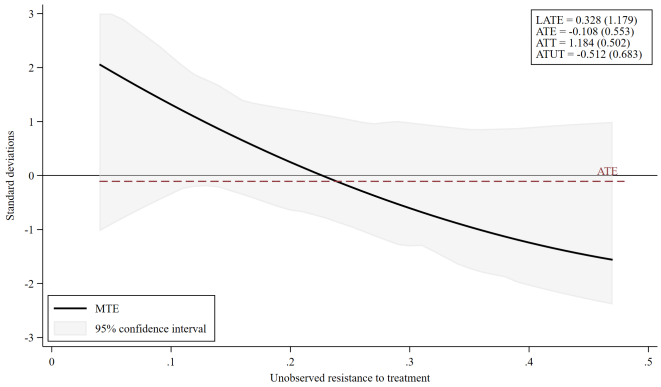


(b) Psychiatrists with more than 20 patients

II. Danish



(c) Full analysis sample

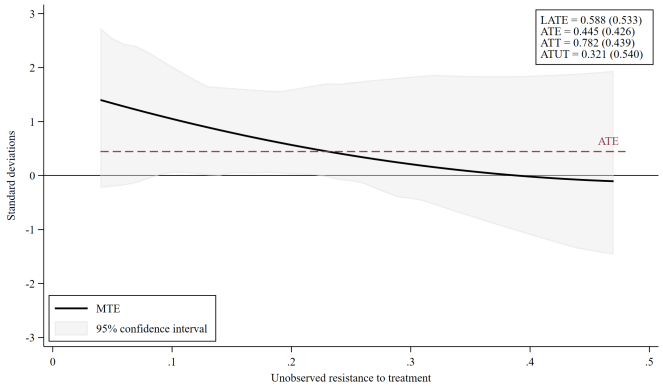


(d) Psychiatrists with more than 20 patients

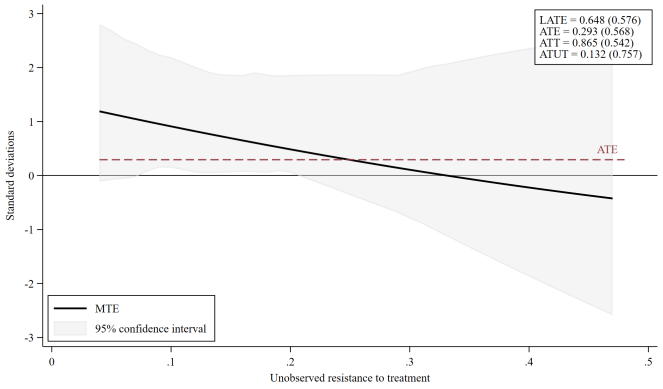
Appendix Figure E2: MTE in the full analysis sample and in the sample of children treated by psychiatrists with at least 20 newly diagnosed patients

Notes: The solid line in each panel plots the marginal treatment effects (MTE) curve for the standardized test score indicated in the heading estimated in the sample in the panel, and the shaded areas represent the 95% confidence interval obtained from 100 replications of a Bayesian bootstrap clustered at the clinic-year level. The horizontal dashed line shows the average treatment effect calculated by averaging the MTEs over the relevant range. The MTE is estimated via local instrumental variables from a specification including all the control variables described in the notes to Table 2. The MTE curve is assumed to be a polynomial of second degree in unobserved resistance to treatment and is estimated over the range of common support trimmed by 1% at both tails. The sample in Figures (b) and (d) includes only the children in the analysis sample whose first psychiatrist was also the first psychiatrist for at least 19 other children in the analysis sample in that same year. The confidence interval in Figures (b) and (d) is trimmed for legibility.

I. Math

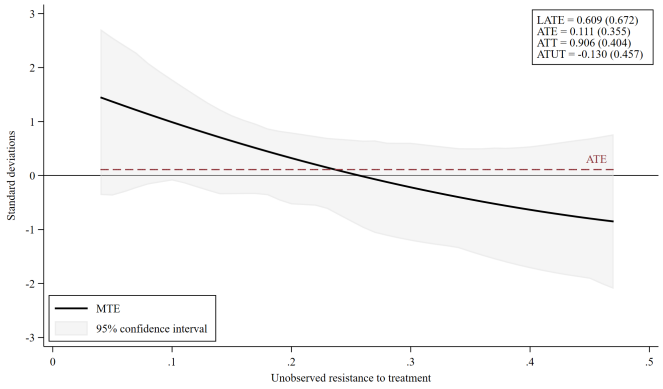


(a) Baseline

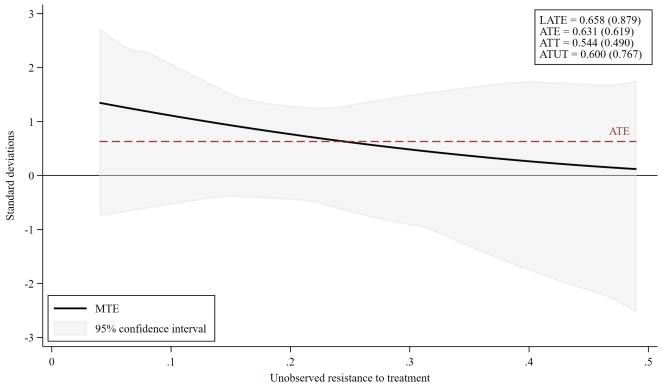


(b) Including covariate indices

II. Danish



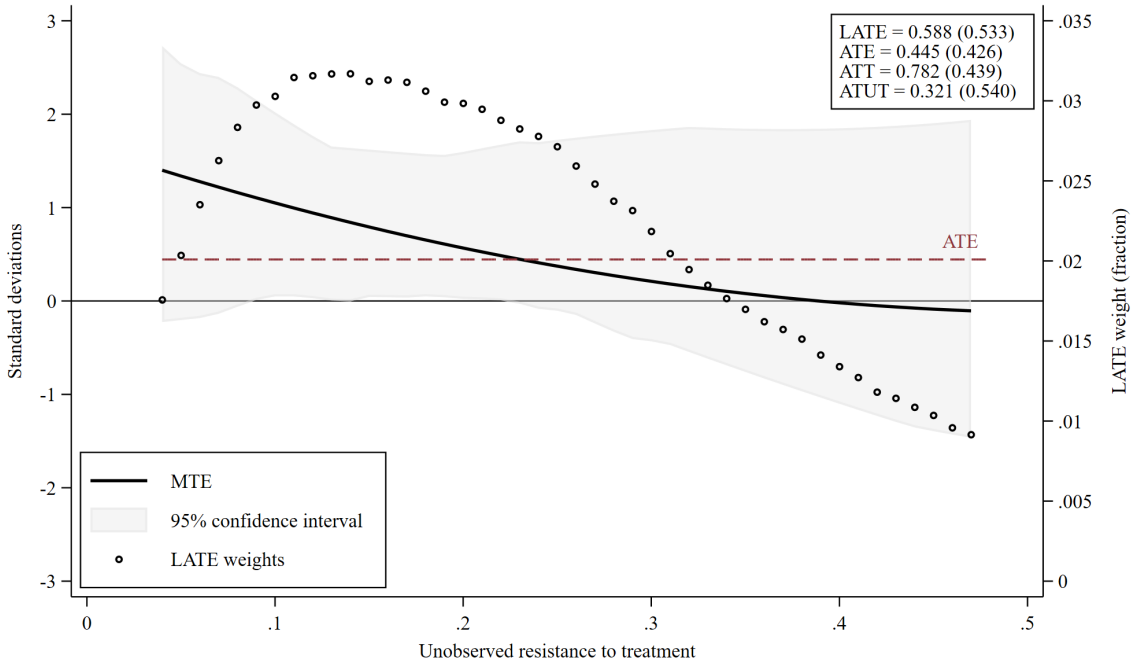
(c) Baseline



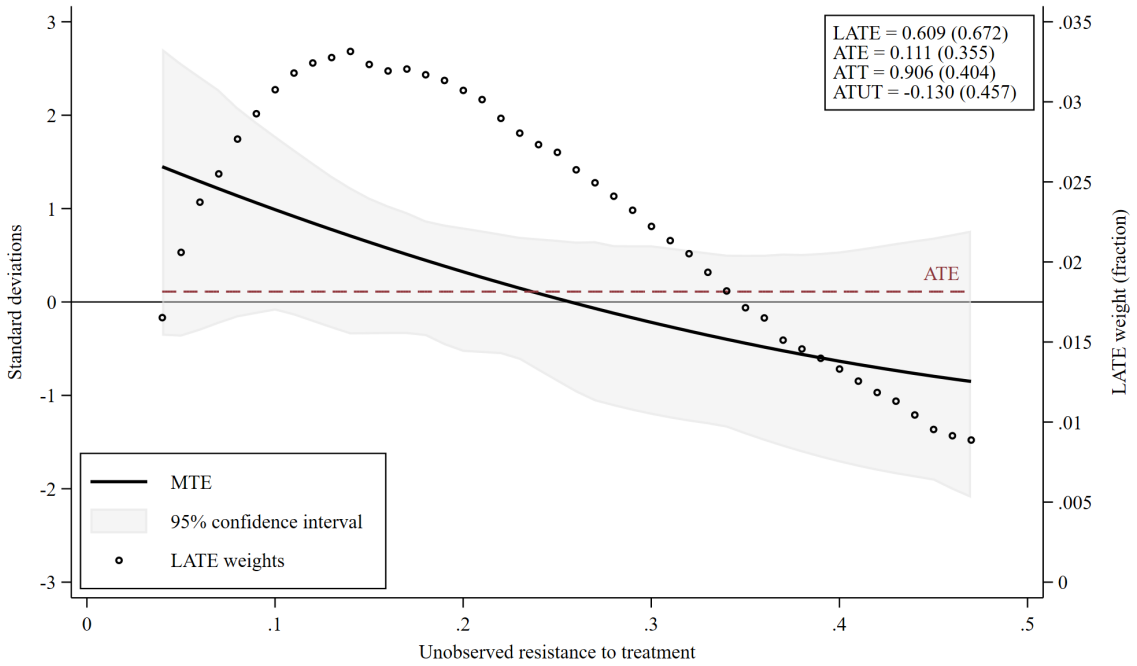
(d) Including covariate indices

Appendix Figure E3: Comparison of MTE estimated using the baseline specification and the specification including the covariate indices

Notes: The solid line in each panel plots the marginal treatment effects (MTE) curve for the standardized test score indicated in the heading estimated using the specification in the panel, and the shaded areas represent the 95% confidence interval obtained from 100 replications of a Bayesian bootstrap clustered at the clinic-year level. The horizontal dashed line shows the average treatment effect calculated by averaging the MTEs over the relevant range. The MTE is estimated via local instrumental variables from a specification including all the control variables described in the notes to Table 2. In addition, the specifications in Figures (b) and (d) include the squared values of a linear index in the full set of control variables (the predicted values from an OLS regression of the indicator for AD use on all the control variables) and of a non-linear index (the predicted values from a similar probit regression). The MTE curve is assumed to be a polynomial of second degree in unobserved resistance to treatment and is estimated over the range of common support trimmed by 1% at both tails.



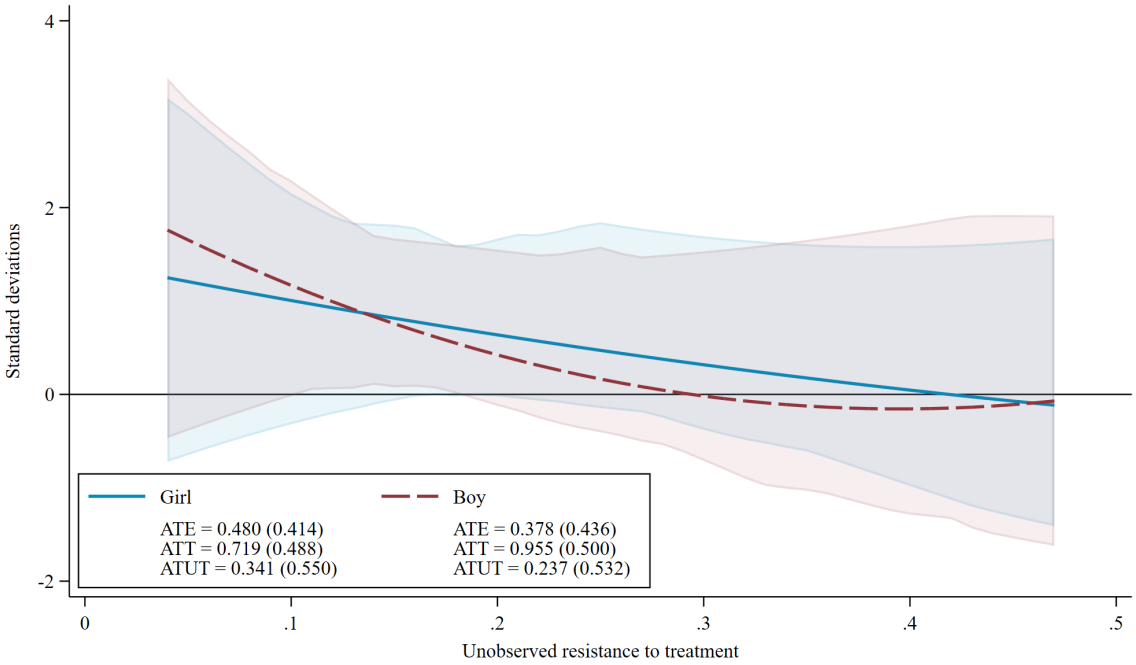
(a) Math



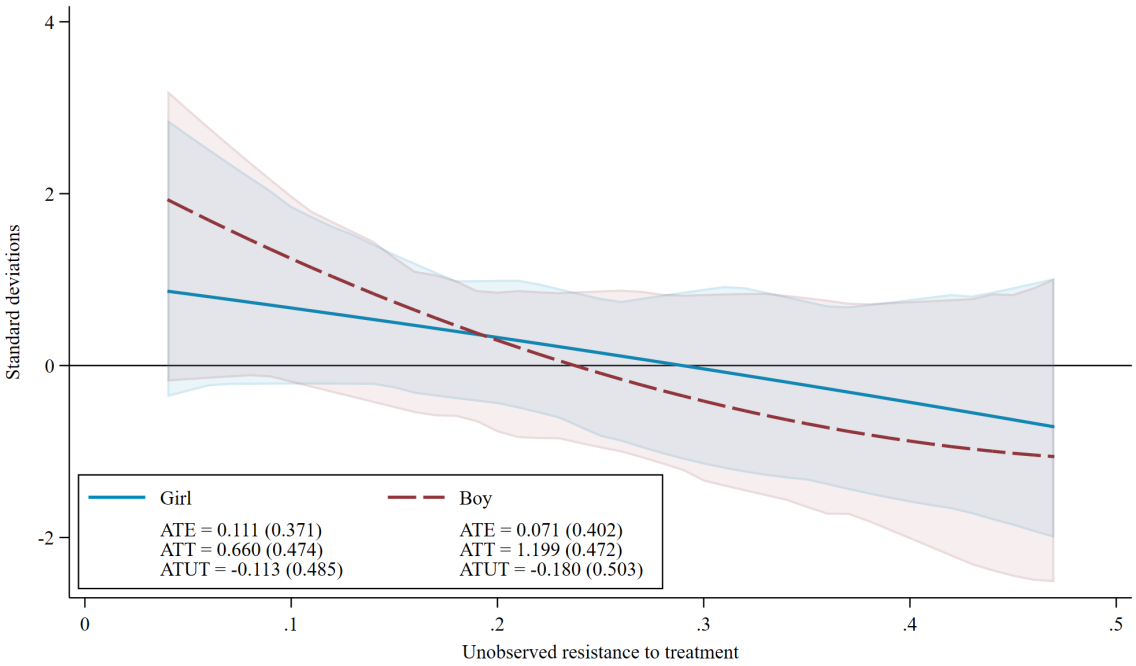
(b) Danish

Appendix Figure E4: Marginal treatment effects and the weights used to calculate the LATE

Notes: The solid line in each panel plots the marginal treatment effects (MTE) curve for the standardized test score indicated in the panel and the shaded areas represent the 95% confidence interval obtained from 100 replications of a Bayesian bootstrap clustered at the clinic-year level. The dots indicate the weights used to calculate the local average treatment effect (LATE) assigned to observations at each point in the distribution of unobserved resistance to treatment. The horizontal dashed line shows the average treatment effect calculated by averaging the MTEs over the relevant range. The MTE is estimated via local instrumental variables from a specification including all the control variables described in the notes to Table 2. The MTE curve is assumed to be a polynomial of second degree in unobserved resistance to treatment and is estimated over the range of common support trimmed by 1% at both tails.



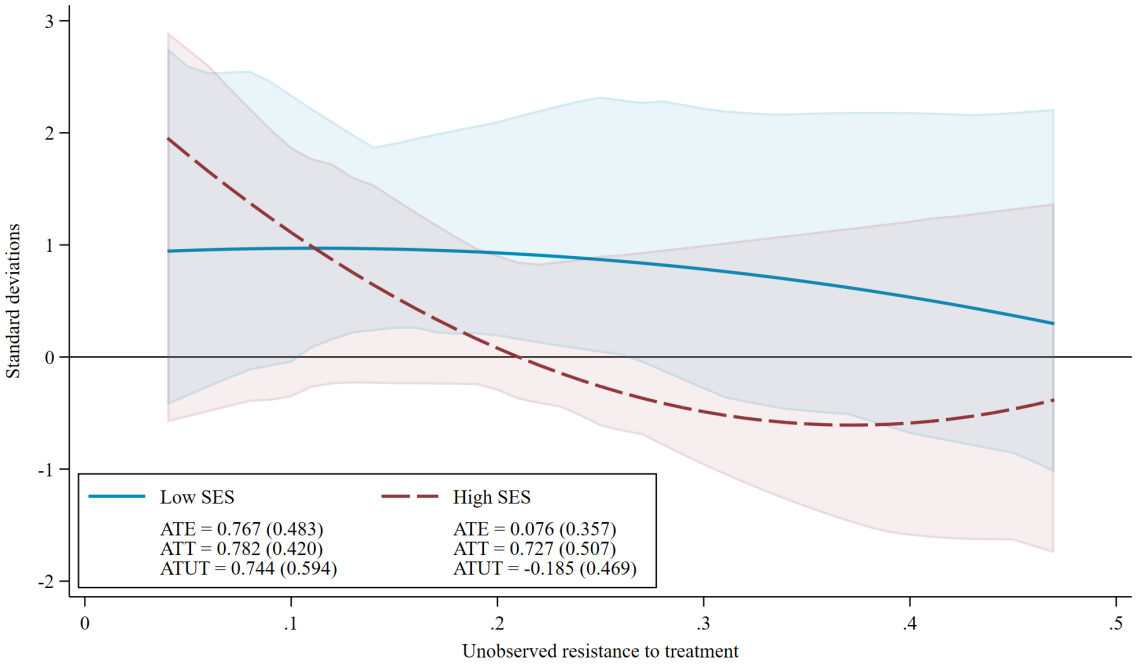
(a) Math



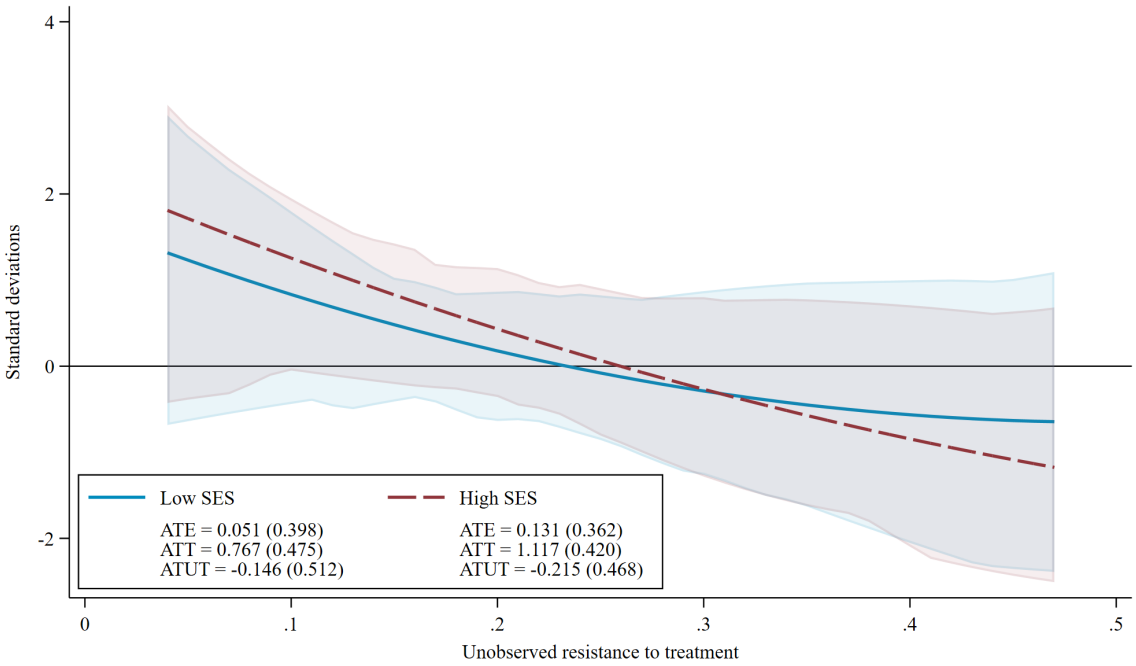
(b) Danish

Appendix Figure E5: Marginal treatment effects and average treatment effects by gender

Notes: The lines in each panel plot the marginal treatment effects (MTE) curve for girls (solid) and boys (dashed) for the standardized test score indicated in the panel and the shaded areas represent the 95% confidence intervals obtained from 100 replications of a Bayesian bootstrap clustered at the clinic-year level. All these measures are estimated via local instrumental variables from a specification including all the control variables described in the notes to Table 2. The MTE curves are estimated via local instrumental variables from a specification including all the control variables described in the notes to Table 2. The MTE curves are assumed to be polynomials of second degree in unobserved resistance to treatment and are estimated over the range of common support trimmed by 1% at both tails. Each panel also reports the average treatment effect (ATE), the average treatment effect on the treated (ATT), and the average treatment effect on the untreated (ATUT) within each group, along with the corresponding bootstrap standard errors.



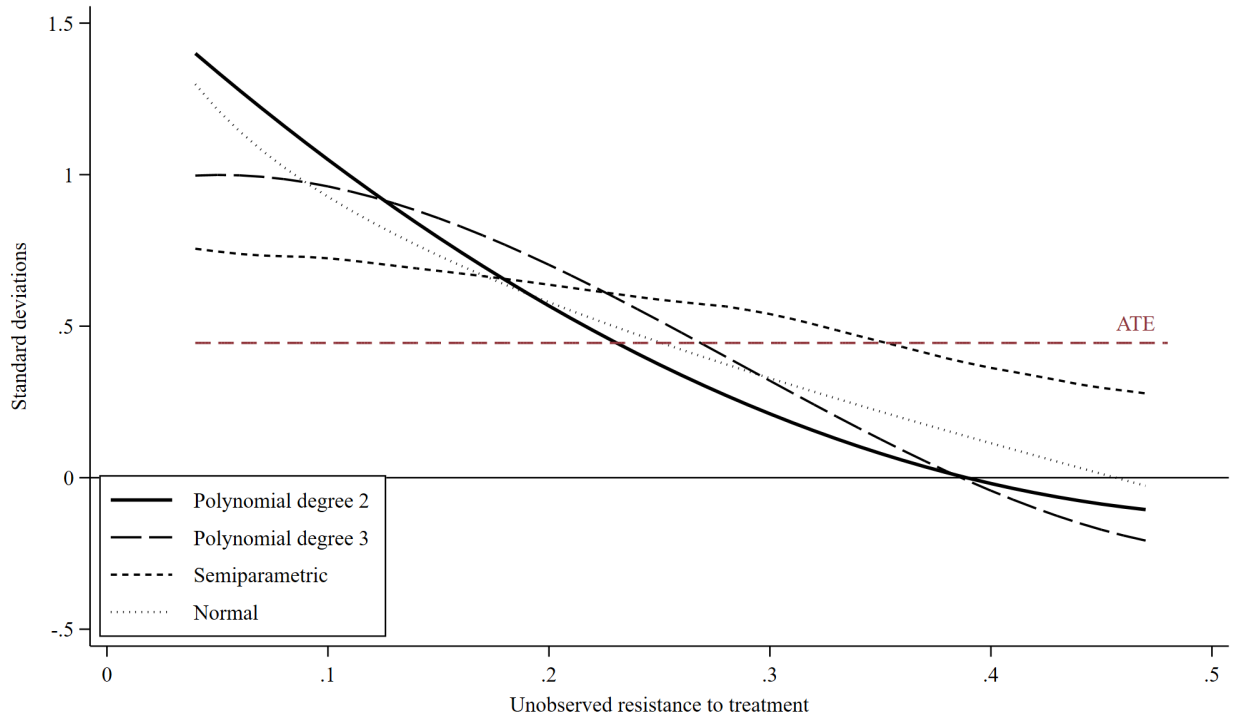
(a) Math



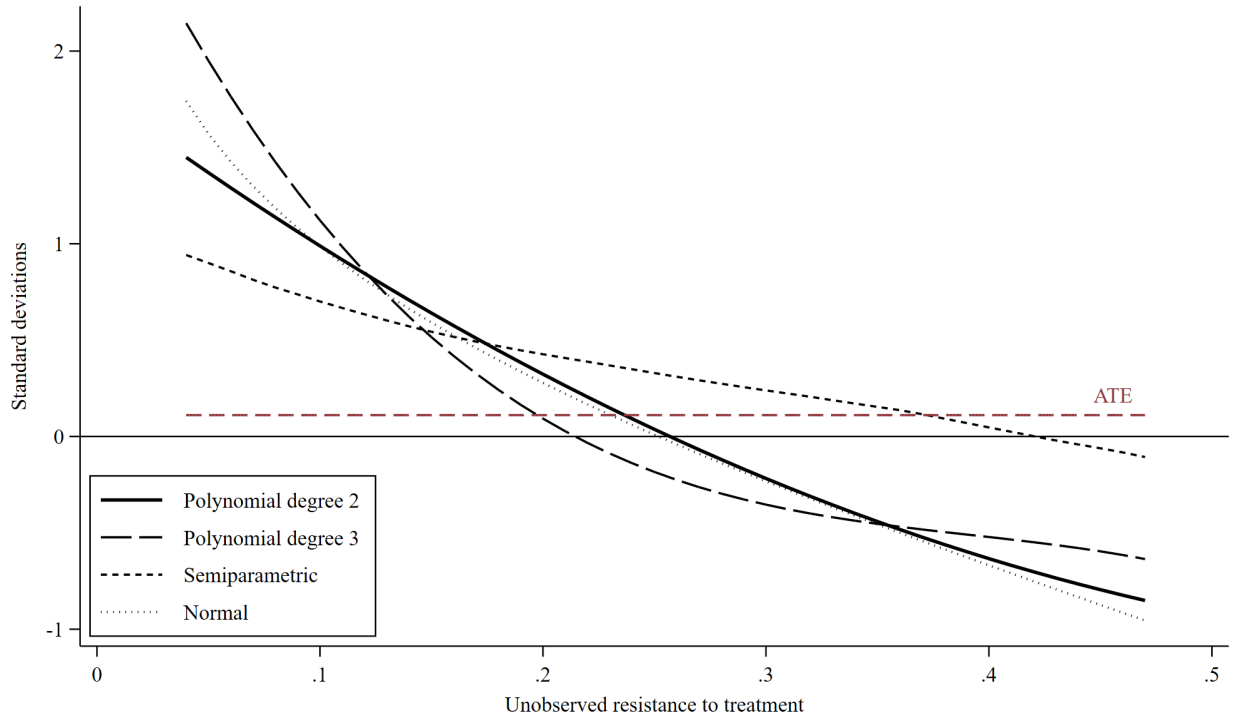
(b) Danish

Appendix Figure E6: Marginal treatment effects and average treatment effects by SES

Notes: The lines in each panel plot the marginal treatment effects (MTE) curve for low- (solid) and high-SES (dashed) for the standardized test score indicated in the panel and the shaded areas represent the 95% confidence intervals obtained from 100 replications of a Bayesian bootstrap clustered at the clinic-year level. All these measures are estimated via local instrumental variables from a specification including all the control variables described in the notes to Table 2. The MTE curves are estimated via local instrumental variables from a specification including the full set of controls listed in the notes to Table 2. The MTE curves are assumed to be polynomials of second degree in unobserved resistance to treatment and are estimated over the range of common support trimmed by 1% at both tails. Each panel also reports the average treatment effect (ATE), the average treatment effect on the treated (ATT), and the average treatment effect on the untreated (ATUT) within each group, along with the corresponding bootstrap standard errors.



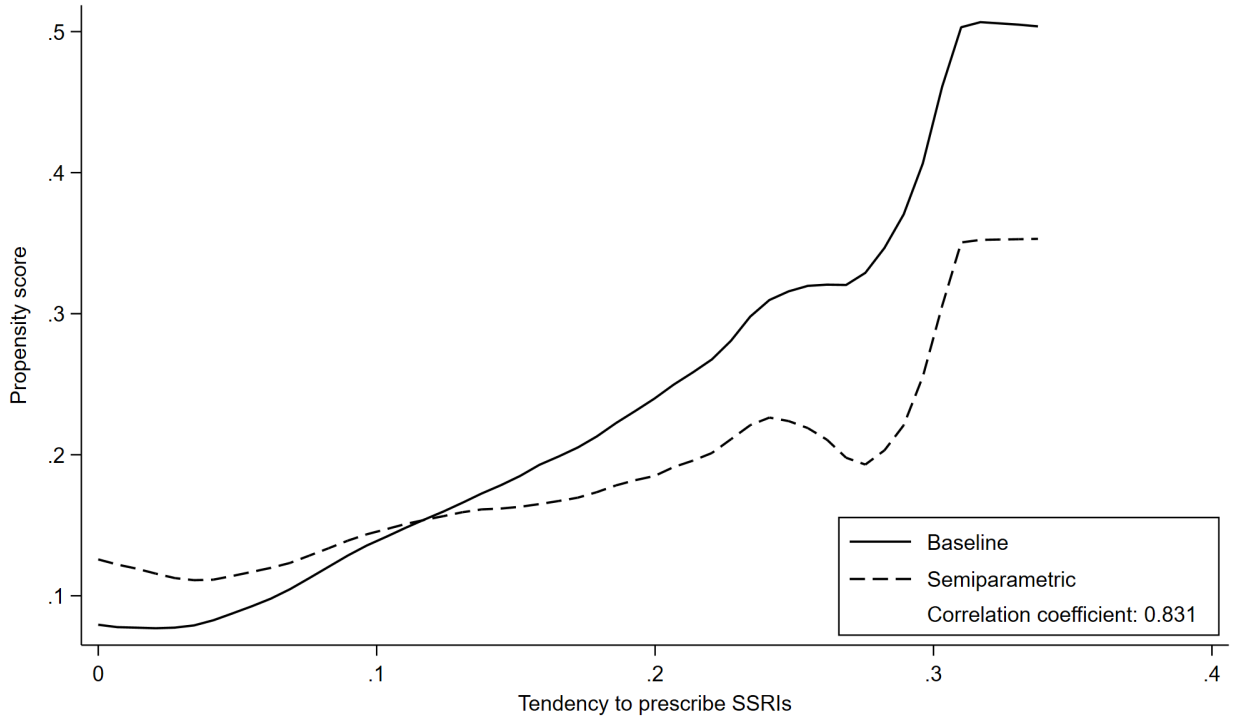
(a) Math



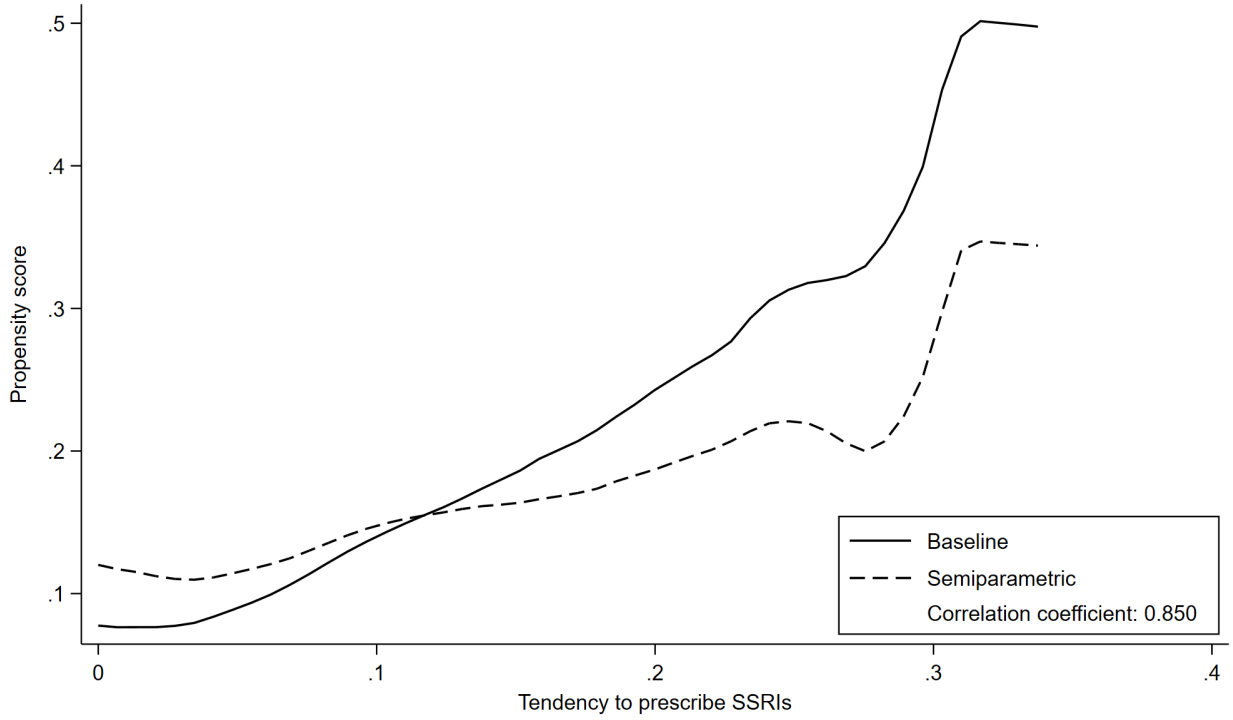
(b) Danish

Appendix Figure E7: Marginal treatment effect curves estimated under different hypotheses

Notes: Each panel plots the MTE curves for the standardized test score indicated in the panel, estimated under four different functional form assumptions: the baseline second-degree polynomial (solid line), a third-degree polynomial (long-dashed line), a semiparametric specification with a second-degree polynomial (short-dashed line), and a parametric normal distribution (dotted line). See Appendix E for details. All the MTE curves are estimated from specifications including the full set of controls listed in the notes to Table 2 over the range of common support trimmed by 1% at both tails.



(a) Math

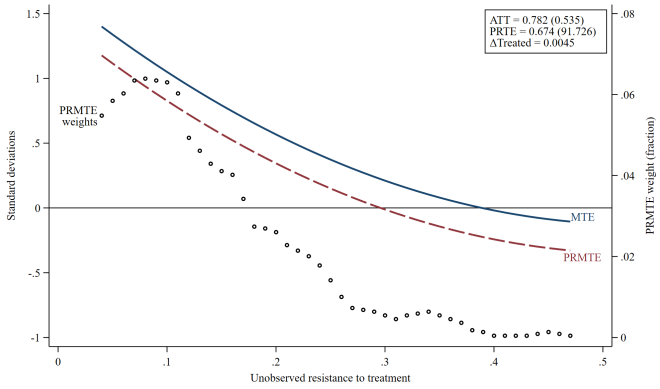


(b) Danish

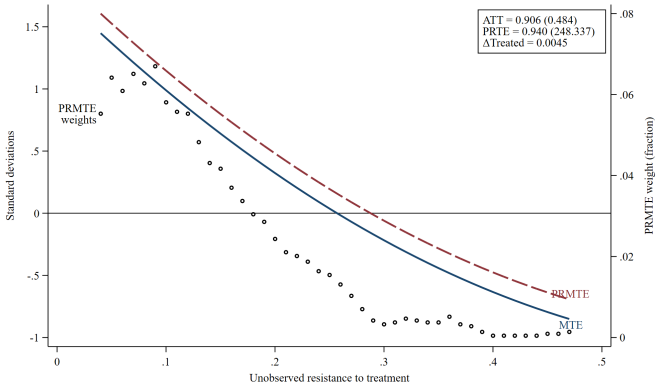
Appendix Figure E8: Robustness of the estimated propensity score to the first-stage specification

Notes: Each panel plots two estimated propensity scores for using AD in the sample of children taking the test for the specified subject. The solid line represents the propensity score estimated from our baseline specification. The dashed line is obtained from a nonparametric specification as described in Appendix E.3. Both curves are estimated from a specification including the full set of controls listed in the notes to Table 2 over the range of common support trimmed by 1% at both tails. Both panels also report the correlation coefficient between the two propensity scores.

I. Impose a floor of the prescribing rate of psychiatrists at the 25th percentile of the prescribing distribution

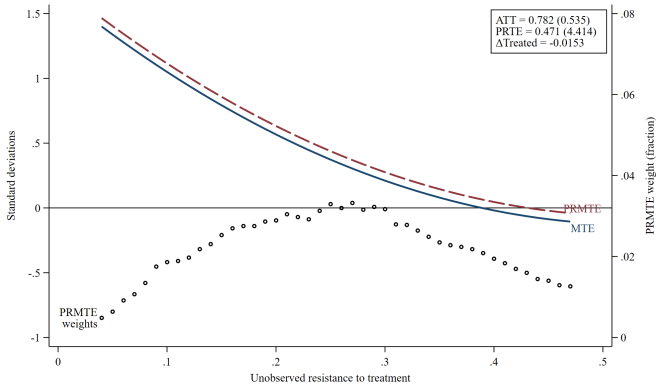


(a) Math

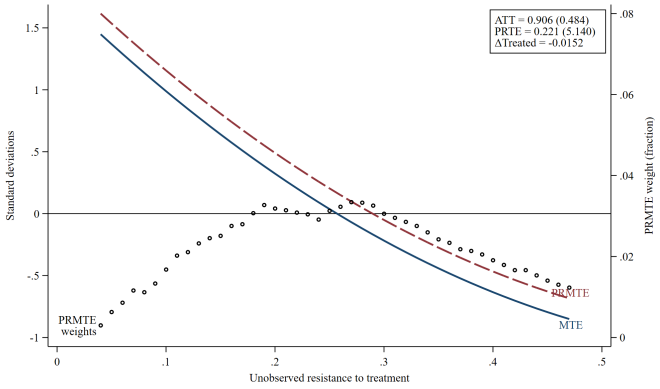


(b) Danish

II. Cap the prescribing rate of psychiatrists to the 75th percentile of the prescribing distribution



(c) Math



(d) Danish

Appendix Figure E9: Policy experiments

Notes: The two lines in each panel plot the marginal treatment effects (MTE) curves under the *status quo* (solid line) and for the children whose treatment status is changed by the new policy (dashed line), for the standardized test score indicated in the panel. The MTE curves are estimated as detailed in the notes to Figure 3. The dots show the fraction of the children affected by the policy at each level of unobserved resistance to treatment. Each panel reports the average treatment effect on the treated (ATT) under the *status quo*, the average treatment effect among the children pushed into or out of treatment with AD by the policy (PRTE), and the fraction of children affected by the policy (ΔTreated).

Appendix Table E1: 2SLS Estimates of the Effect of AD Use on Test Scores and the Frandsen et al. (2023) Joint Test of Exclusion and Monotonicity

	Full analysis sample	Children treated by psychiatrists with at least 20 newly diagnosed patients
	(1)	(2)
Standardized test score, math	0.520**	0.592**
	(0.266)	(0.315)
	[0.028, 1.063]	[0.014, 1.231]
Mean of outcome	−0.305	−0.304
Observations	5,495	4,346
Frandsen et al. (2023) test p-value	**0.166**	**0.298**
Standardized test score, Danish	0.397*	0.275
	(0.253)	(0.294)
	[−0.070, 0.913]	[−0.266, 0.873]
Mean of outcome	−0.241	−0.264
Observations	5,383	4,248
Frandsen et al. (2023) test p-value	**0.126**	**0.066**

Notes: Column 1 repeats our baseline estimates from the analysis sample as described in the notes to Table 1. Column 2 restricts the analysis sample to children whose first psychiatrist was also the first psychiatrist for at least 19 other children in the analysis sample in that same year. Each cell presents the results from a separate regression of the outcome listed in the row estimated in the sample indicated in the column. All specifications include the full set of controls listed in Table 2. The mean of the outcome is calculated among children not using SSRI. In addition to the coefficient estimate, each cell reports in square brackets Anderson-Rubin 95% confidence intervals that are robust to weak instruments (Andrews et al., 2019; Sun, 2018), as well as the p-value for the joint test of strict monotonicity and exclusion proposed by Frandsen et al. (2023). The test is conducted conservatively by assigning a weight of 1 to the fit p-value and a weight of 0 to the slope. A p-value below 0.1 indicates a violation of the exclusion and/or strict monotonicity assumption. Standard errors are clustered at the clinic-year of diagnosis level. ***, ** and * indicate significance based on Anderson-Rubin confidence intervals at the 1%, 5%, and 10% levels, respectively.