NBER WORKING PAPER SERIES

RULES VS. DISCRETION:
TREATMENT OF MENTAL ILLNESS IN U.S. ADOLESCENTS

Emily Cuddy
Janet Currie

Working Paper 27890
http://www.nber.org/papers/w27890

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2020, Revised January 2024

The authors thank Armin Falk, John List, James Heckman, two referees and seminar participants at Berkeley, Bocconi, Boston University, ECARES, University of Chicago, University College London, the University of Lyon, University of Pennsylvania, and Tufts for helpful comments. Financial support from the NOMIS Foundation and Princeton's Center for Health and Wellbeing is gratefully acknowledged. We are grateful to Pankti Pathak of BCBS and Allen Campbell at IQVIA for assistance accessing data. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Rules vs. Discretion: Treatment of Mental Illness in U.S. Adolescents
Emily Cuddy and Janet Currie
NBER Working Paper No. 27890
October 2020, Revised January 2024
JEL No. I1,I11

ABSTRACT

Many mental health disorders start in adolescence and appropriate initial treatment may improve trajectories. But what is appropriate treatment? We use a large national database of insurance claims to examine the impact of initial mental health treatment on the outcomes of adolescent children over the next two years, where treatment is either consistent with FDA guidelines, consistent with looser guidelines published by professional societies ("grey-area" prescribing), or inconsistent with any guidelines ("red-flag" prescribing). We find that red-flag prescribing increases self-harm, use of emergency rooms, and health care costs, suggesting that treatment guidelines effectively scale up good treatment in practice.

Emily Cuddy
Department of Economics
Duke University
419 Chapel Drive
Durham, NC 27708
emily.cuddy@duke.edu

Janet Currie
Department of Economics
Center for Health and Wellbeing
185A Julis Romo Rabinowitz Building
Princeton University
Princeton, NJ 08544
and NBER
jcurrie@princeton.edu

I. Introduction

Adolescent mental health disorders are often more debilitating and harmful for a child's future than common physical health problems. They increase health care costs and the likelihood of being disabled while decreasing educational attainment and employment prospects (Currie et al., 2010, Smith and Smith, 2010, Goodman et al., 2011).[1] Many mental health disorders first manifest in adolescence, and appropriate initial treatment offers an important opportunity to intervene.[2] However, how is a clinician to know what treatment is appropriate?

The U.S. Food and Drug Administration (FDA) has approved a small number of drugs for specific psychiatric disorders in children and adolescents in specific age ranges. These drugs have a strong evidence base in terms of clinical trials, so a clinician might be reasonably confident that the results will "scale" up to their real-world practice (Al-Ubaydli et al., 2017a; Al-Ubaydli et al., 2017b; List, 2022). However, one reason so few drugs are approved for children is that in order to be FDA-approved for a particular disorder in a target population, a company must choose to undergo a costly application process involving pediatric clinical trials. Once drugs are available in generic formulations, costly clinical trials on pediatric patients may no longer be financially viable for innovator firms, and so the drugs do not receive FDA approval for use in children. Hence, a clinician using only psychiatric drugs that have been

The authors thank Armin Falk, John List, James Heckman, John List, Bentley MacLeod, two referees and seminar participants at Berkeley, Bocconi, Boston University, ECARES, Harvard University, University of Chicago, University College London, the University of Lyon, University of Pennsylvania, University of Washington, Tufts, and Wellesley College for helpful comments. Financial support from the NOMIS Foundation and Princeton's Center for Health and Wellbeing is gratefully acknowledged. We are grateful to Pankti Pathak of BCBS and Allen Campbell at IQVIA for assistance accessing data. All errors are our own. This paper was edited by John List.

[1] Currie et al. (2010) find that children with ADHD and conduct disorders in childhood are more likely (30–100% more likely depending on the child's age) to be on welfare after age 18. Smith and Smith (2010) adults who suffered from mental health problems before the age of 16 have family incomes 20% less than their siblings, with a lifetime difference of $300,000. Goodman et al. (2011) find that children with psychological problems in childhood had 28 percent lower family incomes by age 50.

[2] However, one of the few papers to show positive effects of treatment on long-run outcomes is Biasi, Dahl, and Moser (2021).

FDA-approved for specific uses in children typically has a very limited menu of options.

A second source of treatment guidance comes from professional societies which publish guidelines drafted by expert panels who both evaluate the evidence from the clinical literature and codify expert opinion. These guidelines cover an array of topics beyond drug treatment including recommendations for trials of therapy (Birmaher et al., 2007; Connolly et al., 2007). These guidelines usually also include a discussion of the "off-label" use of drugs; that is, the use of drugs that were FDA-approved for one disorder (e.g., depression) for treatment of another disorder (e.g., anxiety), or the use of drugs that are FDA-approved for one population (e.g., adults or adolescents) but not another (e.g., younger children).

Off-label prescribing is controversial. The evidence base for such prescribing is weaker, and off-label prescribing has been shown to present an elevated risk of adverse drug reactions (Syed et al., 2021). Yet with rare exceptions, (such as "scheduled drugs" like heroin), practitioners "retain the autonomy to prescribe drugs for any reason they believe, in their clinical judgment, will benefit their patients" (Syed et al, 2021). This means that neither FDA approval nor professional guidelines are binding on clinicians in practice.

Both the FDA approval process and guidelines from professional societies represent important potential pathways for information about the efficacy of treatment to be "scaled up" from myriad often small-scale studies in the research literature to actual clinical practice. This study addresses two questions about these types of guidance. First, how often is actual treatment consistent with either FDA-approved care or with broader professional guidelines? Al-Ubaydli et al. (2017b) and List (2022) discuss several reasons why relevant information such as that contained in guidelines is not always acted on. These reasons include psychological switching costs, hyperbolic discounting (if the costs of switching are borne up front while the benefits

appear only later), and complexity combined with limited cognitive capacity. Complexity may be especially salient in this context because once a practitioner goes "off label" there are many possible drugs to choose from.

Second, how do children's outcomes compare across the three different types of prescribing? This question is akin to asking whether either the FDA approval process or the publication of professional guidelines is actually helpful in terms of helping to scale up the findings from the research literature, or more broadly, do the findings from the research literature scale and which form of scaling (conservative guidelines based on the strictest evidence vs. broader guidelines that also incorporate expert opinion) produces better patient outcomes?

We focus on the care that children receive when they are first treated for anxiety, depression, or for adjustment disorders with features of anxiety or depression.[3] We chose to focus on anxiety and depression because they are very common—8.4% of U.S. children 3-17 have been diagnosed with anxiety or depression (CDC, 2021)—and because there are both FDA-approved drugs for the initial treatment of these disorders in children and guidelines from professional associations.

We focus on initial episodes for two reasons. First, this is a setting where treatment guidelines may be especially useful since clinicians will by definition have little patient history to inform their decision making. Second, there is persistence in mental health treatment choices over time, suggesting that the initial choice is especially important in determining treatment trajectories (Currie and MacLeod, 2020).

The sample includes 45,223 children who were covered by private health insurance for at

[3] We limit adjustment disorder diagnoses to those with depressed mood (ICD10 F43.21), with anxiety (ICD10 F43.22), or with mixed anxiety and depressed mood (ICD10 F43.23). While there are no specific guidelines for the treatment of adjustment disorders, when adjustment disorders have "features" such as anxiety and depression those features are to be treated according to the guidelines for those disorders.

least one year between 2012 and 2018, had their first mental illness claim between the ages of 10 and 17 and can be followed for at least two years after their initial mental health claim. The main outcomes we examine include health care costs, facility use (hospitalization and emergency room visits), and self-harming behaviors (self-harm, suicidal ideation, and suicide attempts). The latter are particularly important given concerns about rising youth suicide rates (Martínez-Alés et al, 2022).

Clearly, the child's initial treatment is not randomly assigned and is likely to be correlated with both observed and unobserved patient characteristics such as age, gender, and parental preferences. Measures of local provider supply and practice style offer a potentially exogenous source of variation in treatment (Currie and MacLeod, 2017; Cutler et al., 2019; Finkelstein, Gentzkow, and Williams, 2015).

We use measures of provider supply and practice style in the local health care market along with their interactions with patient characteristics as instruments. The rationale for including interactions is that practice style could have a greater impact on some patients than others. For example, there are fewer treatments that are FDA-approved for younger children, so whether a doctor is willing to prescribe off-label or not will have a greater impact on these children. However, using interactions generates a large set of potential instrumental variables raising potential concerns about weak instruments. Hence, we turn to the post-Lasso procedure proposed by Belloni et al. (2012) to guide our choice of instruments and assist in causal inference.

The results suggest that guidelines are useful in scaling research findings into clinical practice. Most doctors who prescribe drugs follow either FDA or professional guidelines, though a sizeable minority do not. Children whose doctors don't follow either guideline have a

higher probability of self-harm, are more likely to use the Emergency Room, and have higher medical costs. Children whose doctors stick to FDA-approved drugs have initially lower probabilities of self-harm but higher mental health care costs. However, over the next two years, children whose doctors follow professional guidelines, and hence can experiment with a broader range of treatments, have the lowest costs, the lowest facility use, and do not have a higher probability of self-harm.

The rest of this paper proceeds as follows. Section II provides background about treatment guidelines for children and about variations in provider supply and practice style. Section III describes our data. Section IV provides an overview of methods, and Section V presents our results. We conclude with a discussion and conclusions in Section VI.

II. Background

A. Guidelines for the Initial Treatment of Children with Anxiety and Depression

Therapy is recommended as a first-line treatment for children. If warranted, medication is specific to the mental illness being treated. Only one drug (Fluoxetine) has been approved for use in childhood depression, and only two are approved for depressed teens (Fluoxetine and Escitalopram). Only one drug (Duloxetine) is approved for treatment of Generalized Anxiety Disorder in children. One reason so few drugs are approved for children is that it is expensive and difficult to run the necessary clinical trials with children and children make up a relatively small share of the market. Moreover, since drugs may be prescribed off-label in any case, the incentive for drug-makers to seek FDA-approval may be further reduced. Having so few approved drugs can be problematic when an individual does not respond to them.

Professional guidelines could play an important role in scaling research findings into clinical

practice by summarizing the evidence in favor of a broader set of drugs. Guidelines from the American Academy of Family Physicians, the American Academy of Pediatrics, and the American Psychological Association are summarized in Appendix Table A1. A summary of recommendations from the UK National Institute for Health and Care Excellence and the World Health Organization are also included for additional context. It is interesting to note that the WHO guidelines, in particular, are very conservative with respect to medications for children.[4]

For depression, most of these guidelines suggest therapy or therapy combined with a trial of Selective Serotonin Reuptake Inhibitors (SSRIs) though the American Psychiatric Association is more specific and recommends Fluoxetine (the FDA-approved drug) as the first line treatment for depression in children since it has been the most studied. SSRIs are also specified as a first-line treatment for anxiety.

SSRIs are recommended because they have fewer side effects than older tricyclic antidepressants (TCAs) which can cause heart problems and are more likely to be fatal in overdose. However, SSRIs are required to carry a black-box warning stating that they may increase the risk of suicidality in persons under the age of 26. Side effects of psychiatric medications appear to be common in children, though it is difficult to obtain reliable data. Hilt et al. (2014) conducted a large mail survey of parents of pediatric patients and found that 84% reported that their children experienced side effects.

Adults with anxiety are frequently prescribed benzodiazepines though they are potentially addictive and dangerous in overdose. The American Academy of Adolescent Psychiatrist's guidelines note that "benzodiazepines have not shown efficacy in controlled trials in childhood

[4] Prescription of "red-flag" drugs to children seems to be lower in other countries, but it is rising, and concerns have been expressed about these trends. For example, Kalverdijk et al. (2017) discuss rising anti-psychotic prescribing in several European countries and conclude that "The implications of this study are that guidelines and practice parameters for AP use drugs need closer scrutiny."

anxiety disorders…Clinicians should use benzodiazepines cautiously because of the possibility of developing dependence" (Connolly et al., 2007). However, two sedatives, lorazepam and midazolam are approved for use in children and very short courses of these drugs are sometimes used for acute anxiety.

Based on these guidelines, we examine the following measures of treatment: 1) Did the child receive follow-up treatment in the three months following the initial mental health claim? 2) If the child received treatment in that window, did they receive drugs? (If a child was treated but did not receive drugs, then they received therapy only.) 3) If the child received drugs in the three-month window, did they receive drugs that were FDA-approved, grey-area drugs that were not FDA-approved but were consistent with professional guidelines, or red-flag drugs that were neither FDA-approved nor consistent with professional guidelines?[5]

The use of medical guidelines is controversial, especially in psychiatry. Meehl (1954), Grove et al. (2000) and Kahneman and Klein (2009) argue that in general an algorithm (i.e., treatment that followed a strict guideline) could do as least as well as a psychiatrist in the treatment of mental illness. Many family physicians who treat mental health may have less specialized knowledge than psychiatrists. In this case, the argument in favor of adherence to guidelines may be even stronger.

On the other hand, Frank and Zeckhauser (2007) argue that guidelines could result in care that is not sufficiently individualized. Again, this argument may be especially relevant in psychiatry given that optimal treatment for mental illness is known to involve trial and error. In

[5] The most common examples of red-flag drugs in our data include antipsychotics (first- and second-generation), other classes of antidepressants (e.g., tricyclic antidepressants), and long-term benzodiazepines. Of the 1,994 children with red-flag drug prescriptions, 37% (2%) receive second (first)-generation antipsychotics, 36% receive non-SSRI/SNRI antidepressants, and 25% receive long-term anxiety medications. The remaining 11% receive something else (e.g., mood stabilizer).

our framework, requiring that a treatment be FDA-approved corresponds to a strict guideline while a treatment that follows professional guidelines but is not FDA-approved would correspond to a protocol that allowed the clinician more flexibility and experimentation. Red-flag prescribing represents complete flexibility. Hence, our results speak to this debate about how much discretion in prescribing is desirable.[6]

B. *Supply-Side Variation in Access to Mental Health Treatment*

The fact that similar patients receive different treatment when they live in different places, has been extensively documented (c.f. Fisher et al., 2003a,b). Recently, several studies have argued that much of this variation reflects differences in provider behavior rather than differences in patient demand for medical services (Currie and MacLeod, 2017; Cutler et al., 2019).

For example, Finkelstein, Gentzkow, and Williams (2015) follow movers and find that the same patient receives different care in different areas and that at least half of the variation in treatment received by elderly Medicare patients is explained by their location (measured using Health Service Areas). Because HSAs are large, 50% is a lower bound on the amount of variation that is explained by the geographic area—as Currie and MacLeod (2017) and Schnell and Currie (2018) show, there is great variation in the treatments offered within areas and even between doctors working in the same hospital or practice location.

[6] Currie and MacLeod (2020) show that the patients of doctors who violate professional guidelines for adult patients have poorer outcomes. However, guidelines for adults are looser than for children and typically focus on transitions from one drug to another. One reason is that adults often have significant medical histories that can be used to personalize treatment. Guidelines for children cover initial treatment as well since there will likely be little history available to guide the physician's initial decision.

Many observers blame shortages of child psychiatrists for variations in the treatment of mental illness in children (Thomas and Holzer, 2006; McBain et al., 2019; Findling and Stepanova, 2008). According to the American Academy of Child and Adolescent Psychiatry, there are 8,300 practicing child and adolescent psychiatrists and approximately 15 million children in the U.S. who need their services.[7] There are also shortages of therapists trained to treat children. The National Institute of Health's Health Resources and Services Administration (2016) estimates that there is a shortage of more than 10,000 full-time equivalent mental health counselors, including school psychologists and counselors. Moreover, the available workforce tends to be concentrated in larger urban areas.

In the absence of sufficient mental health professionals, primary care physicians (PCPs) often prescribe psychiatric drugs. However, many report that they do not feel comfortable in this role—the majority feel that ideally PCPs should refer children with mental health problems to specialists for treatment. And yet, long waiting periods to see mental health professionals are cited as a significant barrier to treatment (Heneghan, 2008; Fremont et al., 2008). These considerations suggest that some variation in treatment could be a function of the supply of professionals who are available to treat the child.

In addition to variation in the numbers of providers, a large literature documents differences in practice styles. Frank and Zeckhauser (2007) cite survey data showing that doctors have favorite drugs for each condition and the favorite usually accounts for 60% of their prescriptions for that condition. In contrast, patient demographics have little explanatory power when it comes to the type of drug prescribed. Berndt et al. (2015) use data on prescriptions of antipsychotics.

[7] See https://www.aacap.org/AACAP/Resources_for_Primary_Care/Workforce_Issues.aspx.

They show that most doctors have a favorite drug and that on average 66% of their prescriptions are for this drug. Again, different doctors have different favorites.[8]

These differences in practice style could be due in part to ability or training. They mean that treatments could vary between small areas just because the doctors working in those areas make systematically different decisions for similar patients. For example, if most of the doctors in an area prescribe benzodiazepines for anxiety in adolescents, then a given adolescent seeking care in that area will be more likely to receive a benzodiazepine. Taken altogether, the literature suggests that variations in provider supply and in practice style are possible instruments for individual treatment choices.

III. Data

Our main source of data is insurance claims data for 2012 through 2018 from the Blue Cross Blue Shield Alliance for Health Research (BCBS), a collaborative effort involving most of the regional BCBS plans.[9] These data have many strengths. The most obvious include the large sample size, broad national coverage, and the fact that we can follow patients over time and see all their claims for inpatient, outpatient, and pharmaceutical treatments. Previous large-scale analyses of mental health treatment for adolescents rely on parent/caregiver reports which may be subject to recall errors or bias introduced by survey non-response. Moreover, the questions national surveys ask about the treatment of mental illness are very general (e.g., whether the

[8] Doctors also differ in terms of their diagnostic skill such that different doctors might diagnose the same symptoms differently, leading to different treatments (Chan et al. 2019; Currie and MacLeod, 2017; Currie et al., 2016). However, here we focus on children who already have a specific diagnosis and look at treatment conditional on diagnosis.

[9] This limited data set is made available through a secure data portal and is drawn from Blue Cross Blue Shield (BCBS) Axis®, the largest source of commercial insurance claims data in the U.S. Accessing insurance claims data often requires extended negotiations with individual insurance carriers, or with government entities. Further information about the BCBS Health of American Initiative, including information about their Axis® data base and contact information is available at: https://www.bcbs.com/the-health-of-america/about.

child has ever been diagnosed or treated) and do not include information about the setting or type of treatment. In contrast, the use of claims data with both diagnostic information and treatment allows us to say whether the treatment received is broadly consistent with evidence-based treatment guidelines.

Our focus on children with health insurance is also an advantage because it allows us to rule out lack of insurance coverage as a reason for variations in care and to focus on other determinants of such variation, such as differences in practice style.[10] Another advantage is that BCBS plans typically offer some coverage of visits to out-of-network providers so that most mental health service encounters are likely to be captured. A disadvantage is that children with private health insurance through BCBS tend to live in areas that are relatively younger, less racially diverse, and of higher socioeconomic status than the average American, as shown in Appendix Table A2.

We select children who are observed before age 11 (typically they are observed from age 7 or 8) who had their first insurance claim for depression, anxiety, or adjustment disorder with anxiety or depression between the ages of 10 and 17. The mean age of these children at the time of the first mental health claim is 11.62, with most first claims occurring between ages 10 to 14. We focus on first episodes because there are relatively clear guidelines about how these children should be treated, as discussed above.

The BCBS data include 4,356,831 children who have a master member ID (which means that they can be followed over time); have claims dates consistent with their coverage period; and meet the age criteria for our study: They must be observed before the age of 11 and for at least

[10] Most U.S. children have health insurance either as dependents on their parent's employer-provided health insurance or through various publicly provided plans. However, one critique of public health insurance is that people often face difficulties accessing specialist care.

one year between the ages of 10 and 17. In addition, they must have BCBS drug coverage over the time period that we observe them. The number of children who meet these criteria is 2,223,930. Children must also have valid geographic information, and consistent demographic information (age and sex) over the period that we observe them, leaving 2,201,566 children.

In this initial sample of 2,201,566 BCBS covered children, there are 202,066 with at least one claim related to mental illness for an overall rate of 9.18%, which is consistent with the rate of 8.4% cited by the CDC (2021). In order to identify claims related to mental illness, we include claims with a diagnosis code corresponding to a mental health disorder in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5); a procedure code indicating a mental health service, such as therapy; or the prescription of a psychiatric drug.[11] Restricting the sample further to those with non-missing provider information, a first mental health claim between the ages of 10 and 17,[12] and a follow-up period of at least 24 months, yields a sample of 97,306.[13]

Finally, we limit our analysis to those children with initial diagnoses of depression, anxiety, or adjustment disorders with anxiety or depression according to the DSM-5. Figure 1 shows that this is a closely related set of diagnoses and that children are often diagnosed with more than one condition. Our final sample consists of 45,223 children.

[11] In arriving at this subsample of 97,306, we exclude the following conditions from consideration as mental illnesses: intellectual disabilities, communication disorders, motor disorders, breathing-related sleep disorders, disruptive, impulse-control, and conduct disorders, neurocognitive disorders, and medication-induced movement disorders and other adverse effects of medication from consideration. This means that these 97,306 children have a mental illness such as depression or anxiety.

[12] We define the first mental health claim as the first that is not a neurodevelopmental disorder (e.g., ADHD).

[13] We would have liked to follow children longer, ideally for many years, but requiring longer time periods would drastically reduce the sample size. The maximum time any child can be followed in our sample is 77 months. This length is a function of both the date of diagnosis and of the length of the current data extract, i.e., children who have a first claim later on in our data mechanically have a shorter potential follow-up period. Of the 45,223 children in our final sample, we retain 25,759 (57.0%) with a 36-month follow-up, 13,738 (30.4%) with a 48-month follow-up, 6,232 (13.8%) with a 60-month follow-up, and 1,487 (3.3%) with a 72-month follow-up.

A. Children's Characteristics, Treatment, and Outcomes

Table 1 provides an overview of child characteristics, types of treatment and supply-side measures in the whole BCBS sample, the sample with any mental illness, and the target sample of children with anxiety and depression broken out by type of treatment that they received. Table 1 shows that 75% of the final sample of children were treated, with 57% receiving therapy alone, and a further 28% receiving drugs. Turning to the types of drugs prescribed, grey-area drugs represent the largest group at 10%. It is striking that only 4% of the sample receive FDA-approved drugs, while the same fraction receives red-flag drugs. Grey-area drugs that are not FDA-approved but reflect professional guidelines are prescribed to 10% of the sample children. These figures suggest that professional association guidelines are an important way that research findings are scaled up for use in clinical practice and that empirically they have more influence on provider behavior than FDA approval.

In terms of children's characteristics, the claims data record the child's age, gender, and address. Table 1 shows that girls are somewhat over-represented in the groups of children who are prescribed drugs, particularly FDA-approved drugs. We use the patient's address to assign children to zip-code tabulation areas (ZCTAs), and therefore to measure the supply-side environment that they face. However, we do not include characteristics of the areas such as percent of Black residents, or median income in our models because most of these characteristics are fairly fixed over the short period of time represented by our panel, and we include area fixed effects.

We can construct variables describing aspects of the child's medical history including the age of the 1st mental health episode, whether they were hospitalized or went to the ER for this 1st

episode, and whether the first mental health encounter was an evaluation. Table 1 shows that the mean age of the first episode is 11.62 in the sample of children with anxiety and depression, and that for about 6% of these children, that first episode involved a facility visit. More commonly though, the first episode involves an evaluation for a mental health condition (54.4% of the sample).

We also know whether a child had a neurodevelopmental condition such as Attention Deficit Hyperactivity Disorder (ADHD) or autism. These neurodevelopmental conditions generally manifest very early in life, much before the mental illnesses that are our focus here. The incidence of neurodevelopmental conditions is twice as high in our sample of children with anxiety and depression as it is in the whole BCBS sample of children, suggesting that children with these conditions are at increased risk of mood disorders. We treat neurodevelopmental conditions as an important potential control variable. Table 1 shows that children with these disorders are more likely to receive drug prescriptions for anxiety and depression and less likely to receive therapy alone.

The main outcomes we examine are: the total costs of medical care, whether the child ever visited the emergency room or was hospitalized, and whether the child ever had an incident involving self-harm. Means of these variables are shown in Table 1 for all months that each child was in the sample. We also present some supplementary analyses of costs for mental health care alone and number of days spent overnight in a hospital.

An advantage of using the claims data is that we see the actual costs of care: We use the contracted reimbursable amount, which is the total combined cost to the insurer and patient. We look at both total costs and mental health care costs because mental illness and some drug treatments can cause physical health problems like injuries or drug reactions.

Visits to the hospital or the emergency room are an important outcome for tracking mental health because people having a mental health crisis are universally advised to go to the nearest emergency room (ER), and once there, they may have to be hospitalized. This is one reason why the Centers for Disease Controls tracks children's mental health visits (Leeb et al., 2020). These visits may represent children who are not being successfully stabilized by mental health treatment or who may be having serious negative reactions to a medication. In addition, children who end up in in a facility because of a mental health crisis may suffer medical trauma (which can lead to post-traumatic stress disorder and depression) as a result of the treatment itself (Marsac et al., 2014; Paskarian et al., 2014). It can be confusing and very frightening for a child to spend many hours or days being held in a hospital emergency room or psychiatric unit.[14]

Self-harm is also an important outcome to look at in this context especially because most anti-depressants carry a "black-box" warning label indicating that they may increase the risk of self-harm in children and young people. FDA-approved drugs for children are generally those that are thought to be least likely to increase the risk of self-harm at least in the short run since this outcome is carefully monitored in clinical trials.

Table 1 shows that compared to children treated with therapy alone, children treated with drugs are more likely to have been hospitalized or taken to the ER, are more likely to self-harm, and have higher average health care costs. Within the group of children treated with drugs, those treated with grey-area drugs are less likely to use ERs or hospitals, less likely to self-harm, and have lower average cost than children prescribed either FDA-approved or red-flag drugs. Note that at this point, we are not making any causal statement, just describing patterns in the data.

[14] Some researchers have even suggested that in some cases hospitalization could cause rather than prevent suicide because of the associated "trauma, stigma and loss of social role" (Large et al. 2017).

These patterns could be due to selection if, for example, children who are more severely ill are more likely to receive a drug prescription—something we endeavor to account for in using our instrumental variables procedure.

We examine all these measures at a three-month, 12-month, and 24-month time horizon following the initial mental health episode. It is possible that a treatment that had positive short-term effects might have negative long-term consequences or might not be as efficacious as another drug over time.

Panel C of Table 1 shows the same breakdown by diagnosis. Table 1 shows that anxiety is the most common diagnosis in the analysis sample, followed by some form of adjustment disorder, and then depression. In terms of treatment, children with depression or anxiety are over-represented in the group that receive drugs, and under-represented in the group that receive therapy alone. Conversely, children with adjustment-disorder are over-represented in therapy alone. Given these clear differences in treatment by diagnosis, we control for diagnosis in what follows.

B. *Provider Supply and Practice Style*

Following Kessler and McClellan (2000) and Currie and MacLeod (2017), we define the market facing people who live in a particular zip code using information about where children in that zip code actually go to receive mental health services each year. For all children who live in a given ZCTA, we examine up to the 10 most common ZCTAs visited by BCBS children in order to receive mental health treatment over the entire sample period. For example, Figure 2 illustrates the definition of the market area for the ZCTA that includes Princeton, New Jersey. ZCTAs that have fewer than 10 BCBS children with a mental health problem are excluded because it is difficult to reliably compute our supply-side measures for very small cells.

Using the actual travel area has several advantages compared to defining a small area based on an arbitrary geographical definition such as a county or HSA. First, only providers who are actually available to treat BCBS children at some point over the sample period are included. Second, the measure scales naturally. For example, in a rural setting where people drive long distances to get to a grocery store, it may not be unusual to drive a long distance to see a psychiatrist. Third, providers can serve clients from more than one ZCTA and do not have to be arbitrarily assigned to one market or another. Fourth, the market definition is specific to psychiatric treatment.

In addition to the BCBS data, we also turn to an additional source of supply-side measures: National data on anti-depressants and anti-anxiety prescriptions filled at retail pharmacies is available from IQVIA's LRx data base.[15] An advantage of using the IQVIA data to supplement information from BCBS is that we can see prescriptions from all of the providers treating children in a particular area, even those who do not treat BCBS children. We focus on the first prescription by a given provider to a patient—such first prescriptions are identified in the IQVIA data.

In order to identify the types of mental health professionals who treat BCBS children, we first find mental health providers who provided treatment to at least one child in the BCBS claims data. Claims data can be merged with data from the National Plan and Provider Enumeration System using the provider's National Provider ID to recover the provider type.[16]

[15] IQVIA (formerly known as IMSQuintiles) is a public company specializing in pharmaceutical market intelligence The IQVIA data is available for purchase to qualified researchers. For further information, contact Allen.Campbell@iqvia.com.

[16] Psychiatrists include NPPES codes 2084P0800X and 2084P0804X. Primary care physicians providing mental health services are largely pediatricians, but may also include doctors in family medicine, general practice, adolescent medicine, or developmental/behavioral pediatrics (NPPES codes 208000000X, 2080A0000X, 2080P0006X, 207Q00000X, 208D00000X). Therapists include psychologists, social workers, and mental health counselors (NPPES codes 1041C0700X, 101YM0800X, 101YP2500X, 103TC0700X, 103T00000X, 106H00000X, 101Y00000X, 104100000X, 103TC2200X).

Leave-one-out supply-side measures are then calculated by removing the index child and then recalculating the supply measures.

By examining only psychiatrists and therapists who actually treat BCBS children, we ensure that we are focusing on the relevant group of mental health professionals for the children in our sample. For example, we rule out psychiatrists who only treat adults, or who are not actively practicing. It would not be possible to focus on this more relevant group of clinicians using other sources such as the National Plan and Provider Enumeration System (NPPES).

The ZCTA-level measures that we use as possible instrumental variables can be thought of as measures of provider supply and practice style and include the following eight measures:

(1) Of the BCBS providers providing mental health treatment to children, what share are psychiatrists?

(2) Of the BCBS providers providing mental health treatment to children, what share are primary care physicians (PCPs)?

(3) Among primary care physicians treating BCBS children what share of each physician's child caseload receives a psychiatric drug (averaged over all primary care physicians in the ZCTA)?[17]

(4) Of all IQVIA prescribers of first psychiatric drug prescriptions, what fraction were psychiatrists?

(5) What is the share of BCBS patients with anxiety or depression who are prescribed anti-psychotics?

(6) The number of psychiatrists per 1,000 BCBS children 10-17.

[17] When calculating this average, we weight shares by the physician's caseload, i.e., the number of patients that the physician sees. We also experimented with weighting by the number of prescriptions written by the physician, but results were very similar.

(7) The number of therapists per 1,000 BCBS children 10-17.

(8) The number of PCPs providing mental health treatment per 1,000 BCBS children 10-17.

While Table 1 provided a child-level overview of the data, Table 2 provides an area-level perspective. The first panel explores variability in treatment across areas. For example, the fraction of children who receive treatment within three months of their initial claim ranges from 50% in a ZCTA at the 10th percentile to 89% at the 90th percentile.[18] However, since our models include area fixed effects, what is more important is that there should be sufficient variation within areas. Table 2 shows that there is within-area variation, especially in red-flag prescribing. This variation could arise in a number of ways. For example, older practitioners could retire and be replaced by younger ones who are more versed in current guidelines. Whether there is sufficient within-area variation for these variables and interactions of these variables with children's characteristics to be valid instruments is an empirical question which we address below.

IV. Methods

The baseline linear probability/OLS models take the form:

[18] An interesting question raised by the first panel of Table 2 is whether the between area variations we observe are real—they could be partially a result of smaller cell sizes, for example. To investigate whether there is more variation in treatment than would be expected by chance, we bootstrapped the data, drawing children randomly from our sample to match only the number of children observed in each area (i.e., not the observed treatment pattern in each area). From the resulting 10,000 bootstrap samples, we then constructed the empirical distribution of treatment—the share untreated, treated with therapy only, treated with red-flag drug, grey-area drug, or FDA-approved drugs—under the null hypothesis that the distribution of treatments is the same in all areas. The actual data strongly reject this null hypothesis. As shown in Appendix Figure A1, the χ^2 test statistic exceeds that of the 99th percentile of the empirical bootstrap distribution generated by the random draws.

(1) $Outcome_{ymzi} = \beta_0 + \beta_1 * Treatment_{ymzi} + \beta_2 * Female_{ymzi} + \Omega * Age_{ymzi} + \beta_3 * (1st\ claim\ is\ hospitalization)_{ymzi} + \beta_4 * (1^{st}\ claim\ is\ an\ ER\ visit)_{ymzi} + \beta_5 * (1st\ claim\ is\ an\ evaluation)_{ymzi} + \beta_6 * (Any\ ER\ or\ hospital\ visit\ last\ 6\ months)_{ymzi} + \alpha * (Diagnosis\ codes)_{ymzi} + \beta_7 * (Neurodevelopmental)_{ymzi} + \mu_m + \gamma_y + \zeta_z + \varepsilon_{ymzi}$,

where *Outcome* is one of the outcome variables discussed above and $Treatment_{ymzi}$ is a zero/one indicator equal to one if child *i* in ZCTA *z*, year *y* and month *m* received a particular type of treatment. *Age* indicates a vector of single year of age fixed effects which are included to allow for the possibility of non-linear age effects. *Diagnoses* are captured using five indicators corresponding to whether a child has a diagnosis of depression, anxiety, adjustment disorder with anxiety, adjustment disorder with depression, or adjustment disorder with anxiety and depression. As discussed above, we are also controlling for neurodevelopmental conditions as well as for whether the first claim was for an evaluation, an ER visit, or a hospitalization, and whether the child had been hospitalized for any non-mental health reason in the last 6 months.

The indicators $\mu_m + \gamma_y + \zeta_z$ are month, year, and area fixed effects defined at the 3-digit zip code level.[19] The month and year fixed effects capture possible seasonal effects (such as those induced by the school calendar) and time trends in outcomes and treatment propensities. Standard errors are clustered at the level of the 3-digit zip code to allow for correlations between children facing similar supply-side measures. The area fixed effects control for relatively fixed differences across market areas such as urbanicity, racial composition, and median income. They will also capture the mean numbers of mental health professionals per capita over the

[19] Given that each child's "market" includes up to 10 5-digit ZCTAs, the three-digit zip code corresponds most closely to the market area than the 5-digit ZCTA on average.

sample period as well as area-level variables such as the average distance to a hospital or emergency room.

Establishing causality in such a model is difficult because even conditional on variables like diagnosis, age, gender, and the measures of initial severity that we can see in the claims data, there could be omitted unobserved variables that affect treatment choices. One possibility is that children who are sicker in an unobserved way will be more likely to be treated, and will be treated more aggressively (i.e., with drugs, and perhaps with red-flag drugs). Alternatively, it is possible that high SES parents demand prompter and more aggressive treatment for their children, even if their children are on average less sick. It is even possible that some parents demand certain drugs for their children, though that seems unlikely to be a widespread reason for treatment choice in these cases of initial mental health treatment where parents may not have a lot of experience with the condition or knowledge about the available treatments.

While we will estimate linear probability or ordinary least squares models as a baseline, causal inference requires identifying instruments that affect treatment decisions but are uncorrelated with these omitted variables. Area-level measures of practice style offer a possible solution to this problem. Indeed, many researchers have used physicians' propensities to prescribe as instruments for a patient's receipt of certain medications (see, for instance, Chorniy and Kitashima (2016) and Dalsgaard, Nielsen, and Simonsen (2014)). However, the potential endogeneity in the formation of the patient-physician match remains. In response, other researchers have turned to area-level measures of the propensity to prescribe (Currie and MacLeod, 2017, 2020). The logic behind such area-level instruments is that individuals living in areas with particular styles of prescribing are themselves more likely to be exposed to such prescribing styles.

In this vein, our ZCTA-level measures introduced in Section III.B. can be thought of as capturing area-level characteristics of mental health practice style which affect the type of care children in those areas are likely to receive. Measures (1) and (2) capture the composition of the providers serving BCBS children. Note that the total number of BCBS providers providing mental health treatment also includes therapists. One might expect that children living in areas with a greater share of physicians with specialized mental health training, i.e., psychiatrists, may be more likely to receive care which is consistent with existing treatment guidelines. Measure (4) is also likely to be affected by the supply of psychiatrists available to treat children but also captures prescribing behavior. It includes all prescribers, not only those seeing BCBS children. Measure (3) looks at whether PCPs who serve BCBS children with mental illness appear to specialize in mental health treatment. Even if there are few PCPs available to serve children with mental health needs, if each of them has a large share of such children in their practice, then they might develop more expertise than a PCP who treats few such patients. Measure (5) looks directly at a measure of red-flag prescribing in the area. Measures (6) to (8) measure the supply of professionals available to treat children in each market. Consistent with the literature on the widespread shortage of behavioral health specialists, measures (6) to (8) will capture directly how a lack (or excess) of local behavioral health specialists may affect a child's ability to access mental health care (see HRSA (2016) as an example).

To increase the power of our instruments, we can also interact them with measures of individual demographics, severity, and diagnosis. Substantively, this procedure is justified if we think the composition or practice style of the provider workforce could have a different impact on children with different baseline characteristics. For instance, if most prescribers of psychiatric drugs in a market prescribe antipsychotics to depressed children, then a depressed child living in

that area may be more likely to receive drugs than would a child with an anxiety disorder. In practice, we include interactions with patient age, sex, diagnosis, and disease severity, as these are key inputs into the existing practice parameters for adolescent mental health treatment. For example, certain drugs are considered FDA-approved for use at only certain ages and for only certain diagnoses (e.g., escitalopram for age 12+ with depression)—thus, if an 11-year old child with depression is living in an area with high adherence to FDA approval guidelines (i.e., low red-flag prescribing), then they are less likely to receive a prescription for escitalopram all else equal. The interactions with gender are included not on the grounds of treatment guidelines, but by the empirical observation that girls are more likely to be treated for depression and anxiety than boys, so that they may be more sensitive to area-level variations in treatment style at the margin.

Interacting the instruments and the controls leads to a very large set of potential instruments, raising concerns about bias from weak instruments in a two-stage least squares (TSLS) setting. Hence, we use the "post-Lasso" 2SLS estimator discussed in Belloni et al. (2012, 2014) and implemented in different contexts by Belloni et al. (2012) and Sands and Gilchrist (2016). This method involves using Lasso to select the instruments to be included in the first stage. The Lasso estimator selects instruments and estimates the first-stage regression coefficients via a shrinkage procedure. The post-Lasso estimator discards the Lasso coefficient estimates and uses the set of instruments selected by Lasso to refit the first-stage regression via ordinary least squares (OLS) to alleviate Lasso's shrinkage bias.

We require Lasso to select the following controls in the first stage: *Female*, the five *Age* bins, *1st claim is hospitalization, 1st claim is an ER visit, 1st claim is an evaluation, Any ER or hospital visit last 6 months*, the five *Diagnosis* codes, *Neurodevelopmental*, μ_m, γ_y, and ζ_z.

As potential instruments, we include a second-degree polynomial of the original vector of instruments as well as interactions of these instruments with child age, gender, a vector of diagnosis codes[20], and indicators for whether the first claim is a hospitalization, ER visit, or evaluation, whether the child had any hospitalization in the past 6 months, and whether they have a neurodevelopmental condition. We found that empirically neither the numbers of psychiatrists, therapists, or primary care physicians nor their interactions with child characteristics were ever selected as instrumental variables, so in what follows, we focus on the remaining five instruments and their interactions.[21] In the end, these remaining instruments and their interactions result in an instrument vector with 116 potential instruments.[22]

We use the efficient F-statistic discussed by Montiel Olea and Plueger (2013) to check that the instruments selected by Lasso pass a test for weak instruments. This test answers the question raised above of whether there is sufficient variation in the instruments to identify treatment effects in models that also include area fixed effects.

Once Lasso selects the first-stage instruments, we estimate the effect of treatment on a child's health outcomes via Two-Stage Least Squares (2SLS).[23] The 2SLS model is similar to the baseline model in equation (1) replacing the child's actual treatment status with predicted

[20] The diagnosis codes include three sets of indicators for diagnoses: (1) indicators for "only anxiety," "only depression," "only adjustment" (N=3), (2) indicators for whether they have a formal diagnosis of anxiety, depression, adjustment with depression, adjustment with anxiety, or adjustment with both anxiety and depression (N=5), and (3) and indicator for whether they have either anxiety or adjustment with anxiety (i.e. any anxiety) and an indicator for whether they have depression or adjustment with depression (i.e. any depression) (N=2). We also include two indicators for whether the person had more than one of these diagnoses ("any multiple") or a formal diagnosis of depression and anxiety specifically ("formal multiple") (N=2). Finally, we include an indicator for a neurodevelopmental disorder (N=1).

[21] This is consistent with Cuddy and Currie (2020) who also find that these measures of provider supply explain very little of the observed variation in treatment. However, we included them here because they are so commonly discussed in the literature.

[22] The full potential instrument vector includes a second-degree polynomial of the five instruments (N=20) and interactions with age and gender (N=6), four severity measures (N=20), and fourteen diagnosis indicators (N=70).

[23] Conventional standard errors are appropriate as long as the number of selected instruments is not close to the sample size (see Belloni, Chernozhukov, and Hansen, 2014).

treatment from the first-stage regression. Our main results are based on the set of instruments chosen using a refined data-driven penalty, as described in Belloni et al. (2012). In our application, this procedure selects between one and six instruments. We also present several robustness checks where we modify the instrument set and re-estimate the 2SLS models in order to investigate the stability of our estimates.

V. Estimation Results

A. First stage estimates

We begin with an analysis of the factors that influence the type of treatment received in the three months following an initial claim for mental illness. We use the full sample of children to estimate the probability of receiving any treatment in the three months after the initial claim. We then use the sample of children who were treated to examine the probability of drug treatment, and finally, we use the sample who received drug treatment to examine the probability of receiving different types of drug treatment conditional on any drug treatment.

An OLS first-stage regression for each treatment type is shown in Table 3. In each case, we include the instrumental variables chosen by Lasso. These models show that the probability of any treatment and of drug treatment conditional on treatment is higher for girls and also increases with the age of the child. Examining the type of drug conditional on drug treatment suggests some tendency for grey-area treatments to decline with age while FDA-approved ones tend to increase. However, there is no age-pattern in red-flag prescribing.

Children whose first claim resulted in a hospitalization are more likely to be treated within the next three months, are more likely to receive drug treatment, and are more likely to be prescribed red-flag drugs and less likely to be prescribed grey-area drugs. The same is true

among children who had any hospitalization (for a non-mental health condition) in the past six months. Children whose first claim was a mental health evaluation are more likely to be treated, but less likely to be prescribed drugs. Children with adjustment disorders are somewhat more likely to be treated than those with anxiety or depression, but those with anxiety or depression are more likely to receive drug treatment. Depressed children are more likely to receive FDA-approved drugs, while children with anxiety are less likely to receive them.

Table 3 shows that in places where a larger share of new prescriptions are written by psychiatrists, children with multiple diagnoses are more likely to receive any treatment within three months after a first claim, while children who have only an anxiety diagnosis are less likely to receive any treatment.[24] This finding suggests that more severely ill children will be more likely to receive prompt treatment in places where they are more likely to be treated by psychiatrists.

The share of mental health providers who are primary care physicians reduces the probability that a child whose first claim is an evaluation will be treated with drugs conditional on being treated, but increases the probability that a child with an anxiety diagnosis will be prescribed drugs. The share of mental health providers who are PCPs also reduces the probability that children who only have an adjustment disorder will be prescribed grey-area drugs and marginally increases the probability that these children will be prescribed red-flag drugs.

The share of mental health providers who are psychiatrists is only chosen when it is interacted with whether the child has any depression diagnosis. In this case, it increases the

[24] While we offer some interpretation of the chosen instruments here, in their survey of machine learning applications Mullainathan and Spiess (2017) specifically warn against "dangers of naively interpreting the estimated β parameters as indicating the discovered structure," because "Similar predictions can be produced using very different variables. Which variables are actually chosen depends on the specific finite sample."

probability of red-flag prescribing and reduces the probability of grey-area prescribing. This finding suggests that psychiatrists may actually feel more comfortable than non-specialists in prescribing drugs that are not FDA-approved and are not recommended by professional guidelines.

In contrast, in places where the share of primary care physician's patients who receive a mental health drug is high, depressed children are more likely to receive an FDA-approved drug. Children with multiple diagnoses are more likely to receive any treatment within three months and children whose first visit is for an evaluation are less likely to receive drugs. This is a provocative pattern of findings which suggests that shortages of psychiatrists are not the main driver of red-flag prescribing to children.

Finally, in places where a large share of depression, anxiety, and adjustment disorder patients receive anti-psychotics, children with adjustment disorders are especially likely to receive a red-flag drug (possibly an anti-psychotic) while those with a depression diagnosis are less likely to receive a grey-area drug.

The key issue, however, is whether in addition to being individually statistically significant, these variables jointly explain much of the treatment choices. This question is addressed in Appendix Table A3 which summarizes the R-squared and the Montiel Olea and Plueger effective F statistics (F_{EFF}) for the instruments for various potential models. Appendix Table A3 shows that neither our supply-side and practice-style instruments by themselves, or the full set of instruments interacted with child characteristics satisfy this test. However, the last column of the table shows that the optimally selected instrument set does pass a weak instruments test with the F_{EFF} statistics ranging from 14.93 to 30.28.

B. Estimates of the Effects of Treatments on Outcomes

Table 4 focuses on the log of overall total health care costs, whether a child visited an Emergency Room or was hospitalized, and whether there was self-harm. All of these variables are measured for periods of 3 months, 12 months, and 24 months after the initial claim in order to see both the short-and medium-term effects of treatment choices. In each panel, the first row of estimates shows OLS coefficients and the next row shows the post-Lasso 2SLS estimates.

Panel A shows the effects of receiving any treatment in the three months after an initial claim. Not surprisingly, treatment initially increases costs. It is more surprising that treatment also appears to increase facility use. However, since mental health patients are told to call 911 or go to an Emergency Room in the event of a mental health crisis, this increase may show that people follow this advice. Similarly, treatment is also estimated to increase the probability of self-harm, which could reflect an increased *detection* of self-harming behaviors for those who are in treatment, or a side effect of some mental health medications.

The post-Lasso 2SLS estimates in Panel A are all much larger in absolute value than the OLS estimates which suggests that the effect of treatment on the marginal person who is affected by the instruments is large. We examine one potential mechanism for heterogeneous treatment effects below. We also see that the estimated coefficients all tend to grow over time, suggesting higher cumulative values after 24 months than after three months. We might have expected that successful treatment, while initially costly, would reduce future costs, facility use, and self-harm in time so that the estimated coefficients might fall over time, but this does not appear to be the case over a time horizon of 24 months.

Panel B focuses on the subset of children who received treatment and investigates the effect of receiving drug treatment. OLS estimates suggest that costs, facility use, and self-harm

are all increased by drug treatment, while in the post-Lasso estimates the sign is reversed. These patterns suggests that the more severely ill children are the most likely to get drug treatment and that it is necessary to control for this selection into drug treatment. When selection is accounted for, Table 4 indicates that drug treatment reduces costs in the set of children who received any treatment without impacting facility use or self-harm, i.e., that drug treatment is on average cheaper than other forms of treatment.

The three parts of Panel C focus on the subset of children who received drug treatment and asks whether children who received one of the three types of drug treatments have better outcomes. The first row focuses on red-flag drug treatments. Relative to children who receive other drug treatments, children who are moved into red-flag treatment by our supply-side instruments have higher costs, higher facility use, and a higher incidence of self-harm.

The last row of Panel C looks at children who receive FDA-approved drugs compared to children who receive any other sort of drug. These children have higher costs and a lower probability of self-harm at three months and one year, though there is no significant effect on facility use and the effect on self-harm becomes insignificant over two years. These results may be consistent with the FDA prioritizing the avoidance of harmful side effects such as short-term increases in self-harming behavior when considering drug approvals.

The middle row of Panel C looks at children who receive grey-area drugs in comparison with children who received either red-flag or FDA-approved drugs. This may be a less natural comparison than the other two: One might expect FDA-approved drugs to be associated with better outcomes, and red-flag drugs to be associated with worse ones.[25] The results suggest that

[25] The weighted sum of the effects of the three types of drugs should add up to the overall effect of receiving drugs. We have not imposed this restriction. Nevertheless, estimating a model for grey-area drugs in the same format as the other two models allows us to show that this adding-up constraint roughly holds in an unconstrained model.

grey area drugs are associated with lower costs and lower facility use, but that they have no effect on the probability of self-harm.

Figures 3 to 5 trace out the implications of the point estimates in Table 4 for self-harm, facility use, and total costs, respectively. All the figures are based on the estimated coefficients for the relevant outcome after 24 months. The tree diagram starts with the first choice, to treat or not, then conditional on treatment shows the implication of choosing drug treatment, and then conditional on drug treatment, shows the effects of choosing different drugs.

Figure 3 shows the estimated impacts of different types of treatment on the probability of self-harm. Any treatment is estimated to increase the probability of self-harm and within the subset of treated children, which may be due to an increased probability of detection. Among children who are treated, those with drug treatment have somewhat lower probabilities of self-harm than those who receive only therapy (5.1% vs. 5.6%). Within the set of children with drug treatment, red-flag treatments have the largest probability of self-harm at 5.7%, which can be compared to a probability of 4.9% in the group treated with grey-area drugs and 3.8% in the group treated with FDA approved drugs.

Figure 4 traces out the same decision tree but considering facility use as the outcome. The figure illustrates that treatment is associated with a higher probability of future facility use (28.1% vs. 17.2%). Conditional on treatment, drug treatment is associated with modestly lower facility use (25.6% vs. 28.9%). But the effects of different types of drug treatment vary widely. Children prescribed red-flag drugs have the highest facility use of any group (33.6%) while those with grey-area treatments having the lowest probability of facility use among children who receive any treatment (18.6%). Children with FDA-approved drugs fall somewhere in the middle with a 26.8% probability of using an emergency room or being hospitalized.

Figure 5 shows that future total costs after two years are $5591.21 for treated children compared to $1889.28 for those who are not treated. Much of this higher cost is apparently due to children who receive only therapy: Conditional on being treated, children who receive drug treatment have costs of $3895.92 compared to $6161.31 for children who are treated without drugs. However, this favorable cost profile for drug treatment is entirely due to the grey-area drugs, i.e., those that are not FDA-approved but consistent with professional guidelines. These children have estimated future costs over the next two years of $1745.17. This estimate is lower than the future cost of non-treatment ($1889.28). Children treated with red-flag drugs or FDA-approved drugs both have costs that are much higher at $9557.12 and $9656.43 respectively.

Table 5 examines two supplementary outcome variables. The first three columns show estimates for the effects of different types of treatment on the log of mental health care costs. The effect of treatment on costs for mental health care are higher than those for total health care costs, suggesting that the spending incurred for treatment does reduce other types of health care costs (such as costs of self-harm). The estimated effect of drug treatment conditional on any treatment is strongly negative, confirming that drug treatment is cheaper than other forms of mental health treatment. This may be in part because most of the drugs used to treat these children are available as generics. However, conditional on drug treatment, the estimates in Panel C of Table 5 confirm the pattern seen in Table 4: Red flag drugs are associated with higher costs for mental health care while grey-area drugs are associated with the lowest costs.

The last three columns of Table 5 show estimated effects on the length of a facility stay. Someone who went to emergency and then was discharged without being hospitalized would have a length of stay of zero. Effects on length of stay are very imprecisely estimated,

suggesting that much of the precision in our estimates of facility use comes from changes in visits to the emergency room.

C. *Robustness*

Although we have focused on children with anxiety, depression, or adjustment disorders with features of anxiety and depression, a small number of children in our sample have other psychiatric diagnoses. There are 1,200 children (2.7%) who received an additional mental health diagnosis in the first three months of care which could have caused them to be prescribed what we are calling red-flag drugs legitimately. These additional conditions include schizophrenia spectrum and other psychotic disorders, bipolar and related disorders, autism spectrum disorder, and Tourette's disorder.

Given this possibility, we have re-estimated our models excluding these children. These estimates are shown in Table 6. For each model, we repeat the Table 4 post-Lasso coefficient for reference, and then we show the estimate from the smaller sample. The estimates in Panels B and C are quite similar. The main difference between Table 6 and Table 4 is in the estimated effects of treatment. When we drop the children with multiple diagnoses within the first three months, we no longer find that treatment has a positive effect on costs or on self-harm. We still find, however, that conditional on any treatment, drug treatment reduces costs. Conditional on drug treatment, grey-area treatment that is sanctioned by professional societies but not FDA-approved drives the reduction in costs and facility use. Red-flag treatment increases costs, facility use, and increases self-harm in the short and medium term, while FDA-approved drugs increase costs but reduce self-harm in the medium term.

In the discussion above, we have emphasized conclusions based on a post-Lasso 2SLS procedure with a set of instruments chosen using the refined data-driven discussed in Belloni

(2012). However, it is useful to think about how robust these conclusions are to changes in the instrument set. One experiment we tried was to limit the number of instruments to the "top three" instruments chosen by the Lasso procedure. These estimates are shown in Appendix Table A4 and are very similar to those shown in Table 4.[26] A second experiment was to reintroduce the three physician supply measures—PCPs, psychiatrists, and therapists per 1,000 BCBS children 10-17—and their interactions with patient characteristics to the potential instrument vector. These estimates are shown in Appendix Table A5 and again are very similar to those shown in Table 4.

Panels A, B, and C of Figure 6 summarize our estimates across all specifications.[27] It shows the point estimates from the main specification shown in Table 4 with their 95% confidence intervals, the estimates when we restrict to three instruments, the estimates when we include the "supply" instruments, and the estimates when we exclude children with other mental health diagnoses. Focusing on the nine models where we find a significant effect in our preferred specification, ("Main" estimates), only 2 of 27 estimated coefficients are outside of the 95% confidence interval of the main effect. Taken together, this suggests that our main results are robust to our assumptions along multiple dimensions.

VI. Discussion and Conclusions

We use health insurance claims data to explore the effects of variation in mental health treatment received in the first three months after a child's initial claim for mental illness on

[26] Note that Lasso uses a slightly different penalty function when we ask it to select only three instruments compared to when it is unconstrained. Therefore, even in cases where it chose three instruments unconstrained, forcing it to choose only three may result in a different set of instruments being chosen, in which case the TSLS estimates will be slightly different.

[27] See Appendix Table A6 for the full list of instruments chosen by Lasso in each model discussed in the paper.

future health outcomes. In particular, we focus on whether children who receive treatments that are FDA-approved or grey-area (not FDA-approved but consistent with professional guidelines) have better outcomes than those who receive unapproved and unrecommended red-flag drug therapies. These comparisons, in turn, shed light on the value of scaling up research findings via official government channels (i.e., FDA approval) or via the efforts of voluntary professional associations such as the American Academy of Child and Adolescent Psychiatrists.

Since treatment is likely to be affected by many unobserved factors that we cannot observe in claims data, we pursue an instrumental variables strategy. Our instruments are based on the idea that other things being equal, children living in different places face doctors with different practice styles, and that this is an exogenous source of variation in the treatment that they receive. It is conceivable that parents who have strong views about particular medications for their children could all live in the same place. But, we think our setting, in which children are being treated for the first time and we are looking at treatment over the first three months, makes it unlikely that this is the major source of variation in the treatments that are chosen.

Given a large number of potential instruments and concerns about bias due to weak instruments, we use the post-Lasso method suggested by Belloni et al. (2012) in which Lasso is used to select instruments from the large pool of potential variables. Our results add to a small but growing literature suggesting that such methods are a useful way to extract "signal" from large medical claims data sets.

We find that receiving follow-up mental health treatment in the first three months after an initial claim increases total health care costs over the next 24 months, which suggests that the costs of mental health treatment are not offset by reductions in other forms of care. However, these increased costs are driven by a small number of children with multiple diagnoses and

complex cases, and when these children are excluded, we find no increase in costs. Hence, for the majority of children increased costs of mental health care among treated children are offset by reductions in other health care costs.

Conditional on treatment, children treated with drugs have lower costs than those treated with therapy alone which may explain the popularity of drug treatments. However, this overall comparison between drug treatment and therapy alone masks considerable heterogeneity in the effects of different types of drugs. Conditional on drug treatment, costs are actually highest for children taking FDA-approved drugs or red-flag drugs and lowest for children taking grey-area drugs that are consistent with professional guidelines but not FDA-approved for use in these children.

Facility use is higher for treated children than for those who receive no treatment. This may well be the result of the standard advice to patients in treatment, which is that they proceed to the nearest emergency room in the event of a mental health crisis. Children prescribed drugs have a lower probability of facility use, conditional on receiving treatment. Turning to the type of drug, we find that red-flag drugs increase facility use most, followed by FDA-approved drugs. Children prescribed grey-area drugs have the lowest probability of facility use conditional on drug treatment. Hence, the patterns for facility use are very consistent with those for total costs.

Many may consider self-harm to be the most important outcome that we study, but since it is fortunately rare, our estimates are less precise. However, we find that children given red-flag prescriptions are more likely to self-harm. We find that FDA-approved drugs significantly reduce the probability of self-harm over a 3-month or 12-month window. At 24 months, the point estimate is still large, though not statistically significant. Given that the FDA approval

process places a lot of weight on making sure that approved drugs do not cause harm, it would not be unreasonable to expect that FDA-approved drugs would be better in this dimension.

Our results show that existing guidelines do usefully scale research findings in a way that improves clinical practice. Children whose clinicians do not follow guidelines—those who are prescribed red-flag drugs that are neither FDA-approved or professional sanctioned—consistently have the worst outcomes in terms of self-harm, facility use, and total costs. Such findings are consistent with the medical literature on the safety issues associated with the use of red-flag molecules. For instance, antipsychotics are associated with cardiovascular and metabolic changes, e.g., weight gain, diabetes, and stroke (see Daviss et al. (2016) for a summary); tricyclic antidepressants have been shown to cause seizures, coma, and cardiac arrhythmias in overdose (see Olgun (2009)); and benzodiazepines carry a high potential risk of both overdose and dependency upon long-term use (Riddle et al. 1999).

The results are less clear about which type of guideline is preferred and suggest a trade-off. Prescribing consistent with FDA-approvals produced the best patient outcomes in terms of reducing self-harm in the short- and medium-term though it is less clear that it is effective at the two-year horizon. FDA-approved drugs, however, have a relatively high cost in terms of greater facility use and higher total costs. Prescribing based on professional assessments of "best practice" had no significant effect on self-harm, but greatly reduced facility use and total health care costs, which are also important outcomes. This discrepancy may be a result of professional associations weighing more common outcomes like facility use and costs more heavily in formulating their guidelines.

Congress has contemplated several approaches to dealing with off-label prescribing ranging from practices that would likely discourage it (such as requiring physicians to explain to

patients why they are prescribing off label) to actions that would encourage off-label prescribing such as removing current prohibitions that prevent drug manufacturers from marketing their products for off-label purposes (Bodie, 2021). However, in at least one instance, government regulations endorse an approach that supports FDA-approved and grey-area prescribing and prohibits insurance reimbursement for red-flag prescribing. Under the Medicare program (public health insurance for all Americans 65 and over), the Social Security Act defines a "medically accepted indication" as one which is either FDA-approved or for which there is "supportive clinical evidence in peer reviewed medical literature."[28] Our results suggest that such an approach could be beneficial for child mental health patients as well.

One perhaps reassuring finding is that most of the physicians in our sample who prescribe to children prescribe either FDA-approved drugs (22%), or drugs that are consistent with professional guidelines (55%). This begs the question though of why the 22% who prescribe red-flag drugs fail to adopt *either* practice guideline. U.S. physicians do not usually directly profit from prescribing one drug over another, though a large literature suggests that they can be influenced by pharmaceutical marketing, and that restricting such marketing can improve compliance with guidelines (Larkin et al., 2014). Training and cohort effects may also be important (Schnell and Currie, 2018). Given the negative effects of red-flag prescribing, understanding the underlying reasons for it is an important topic for future research.

These results break new ground and contribute to several strands of literature. First, they focus on an important determinant of children's non-cognitive skills: Their mental health. Second, the results provide support for the guidelines themselves, contributing not only to the on-going controversy about "rules vs. discretion" in medicine but also demonstrating that, when adopted,

[28] Social Security Act §1861(t)(2)(B) [42 U.S.C. 1395x].

these guidelines scale from small samples to real world. At the same time, they raise questions about another scaling challenge: non adoption. The guidelines are easy to find and to understand and yet clinicians do not always use them. Third, this research suggests that analyses of large-scale claims data can provide a useful complement to clinical research studies in identifying best practices. Such analyses are now more feasible than ever before given advances in machine learning and econometrics.

Data Availability:

Our main source of data is insurance claims data for 2012 through 2018 from the Blue Cross Blue Shield Alliance for Health Research (BCBS), a collaborative effort involving most of the regional BCBS plans. This limited data set is made available through a secure data portal and is drawn from Blue Cross Blue Shield (BCBS) Axis®, the largest source of commercial insurance claims data in the U.S. Accessing insurance claims data often requires extended negotiations with individual insurance carriers, or with government entities. Further information about the BCBS Health of American Initiative, including information about their Axis® data base and contact information is available at: https://www.bcbs.com/the-health-of-america/research-alliance.

National data on anti-depressants and anti-anxiety prescriptions filled at retail pharmacies is available from IQVIA's LRx data base. IQVIA (formerly known as IMSQuintiles) is a public company specializing in pharmaceutical market intelligence The IQVIA data is available for purchase to qualified researchers. For further information, contact Allen.Campbell@iqvia.com.

We are not permitted to make the underlying data available, but we have made all our code available in the HarvardDataverse (https://doi.org/10.7910/DVN/KXAVHB; Cuddy and Currie 2023). Researchers should be able to use this code to create similar data using any standard U.S. medical claims data set.

References

Al-Ubaydli, Omar, John A. List, and Dana L. Suskind. 2017a. "What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results." *American Economic Review* 107 (5): 282-286.

———. 2017b. "Scaling for Economists: Lessons from the Non-Adherence Problem in the Medical Literature." *Journal of Economic Perspectives* 31 (4): 125-143.

Biasi, Barbara, Michael S. Dahl, and Petra Moser. 2021. "Career Effects of Mental Health." *Working Paper no.* 29031 (July), NBER, Cambridge, MA.

Bansak, Cynthia and Stephen Raphael. 2008. "The State Children's Health Insurance Program and Job Mobility: Identifying Job Lock among Working Parents in Near-Poor Households." *Industrial and Labor Relations Review* 61 (4): 564-579.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen. 2012. "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain." *Econometrica* 80 (6): 2369-2429.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28 (2): 29-50.

Berndt, Ernest, Robert Gibbons, Anton Kolotilin, and Anna Taub. 2015. "The Heterogeneity of Concentrated Prescribing Behavior: Theory and Evidence from Antipsychotics." *Journal of Health Economics* 40 (March): 26–39.

Birmaher, Boris, David A. Brent, and AACAP Work Group on Quality Issues. 2007. "Practice Parameter for the Assessment and Treatment of Children and Adolescents with Depressive Disorders." *Journal of the American Academy of Child and Adolescent Psychiatry* 46 (11): 1503-1526.

Bodie, Agate. 2021. "Off-Label Prescribing of Prescription Drugs." *Congressional Research Service Report* R45792.

Borghans, Lex, Angela Duckworth, James J. Heckman, and Bas Ter Weel. 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43 (4): 972-1059

Centers for Disease Control and Prevention. "Data and Statistics on Children's Mental Health." https://www.cdc.gov/childrensmentalhealth/data.html, accessed Nov. 27, 2021.

Cheung, Amy H., Rachel A. Zuckerbrot, Peter S. Jensen, Danielle Laraque, Ruth E. K. Stein, and GLAD-PC Steering Group. 2007. "Guidelines for Adolescent Depression in Primary Care (GLAD-PC): Part II. Treatment and Ongoing Management." *Pediatrics* 141 (3): 1-16.

Chorniy, Anna and Leah Kitashima. 2016. "Sex, Drugs, and ADHD: The Effects of ADHD Pharmacological Treatment on Teens' Risky Behaviors." *Labour Economics* 43 (December): 87-105.

Connolly, Sucheta, Gail Bernstein, and AACAP Work Group on Quality Issues. 2007. "Practice Parameter for the Assessment and Treatment of Children and Adolescents with Anxiety Disorders." *Journal of the American Academy of Child and Adolescent Psychiatry* 46 (2): 267-283.

Cuddy, Emily and Janet Currie. 2020. "Treatment of Mental Illness in the U.S. Adolescents Varies Widely Within and Across Areas." *Proceedings of the National Academy of Sciences* 117 (39): 24039-24046.

———. 2023. Replication data for: "Rules vs. Discretion: Treatment of Mental Illness in U.S. Adolescents." https://doi.org/10.7910/DVN/KXAVHB.
Currie, Janet, William B. MacLeod, and Jessica Van Parys. 2016. "Physician Practice Style and Patient Health Outcomes: The Case of Heart Attacks." *Journal of Health Economics* 47 (May): 64-80.
Currie, Janet and William B. MacLeod. 2017. "Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians." *Journal of Labor Economics* 35 (1): 1-42.
———. 2020. "Understanding Doctor Decision Making: The Case of Depression Treatment." *Econometrica* 88 (3): 847-878.
Currie, Janet, Mark Stabile, Phongsack Manivong, and Les Roos. 2010. "Child Health and Young Adult Outcomes." *Journal of Human Resources* 45 (3): 517-548.
Cutler, David, Jonathan Skinner, Ariel D. 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.
Dalsgaard, Søren, Helena Skyt Nielsen, and Marianne Simonsen. 2014. "Consequences of Medication Use for Children's Outcomes." *Journal of Health Economics* 37 (September): 137-151.
Daviss, William B., Erin Barnett, Katrin Neubacher, and Robert E. Drake. "Use of Antipsychotic Medications for Nonpsychotic Children: Risks and Implications for Mental Health Services." *Psychiatric Services* 67 (3): 339-41.
Findling, Robert L. and Ekaterina Stepanova. 2018. "The Workforce Shortage of Child and Adolescent Psychiatrists: Is It Time for a Different Approach?" *Journal of American Academy of Child and Adolescent Psychiatry* 57 (5): 300-301.
Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2015. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics* 131 (4): 1681-1726.
Fisher, Elliot S., David Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile Pinder. 2003a. "The Implications of Regional Variations in Medicare Spending." *Annals of Internal Medicine* 138 (4): 273-298.
———. 2003b. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Annals of Internal Medicine* 138 (4): 288-298.
Frank, Richard G. and Richard J. Zeckhauser. 2007. "Custom-Made Versus Ready-to-Wear Treatments: Behavioral Propensities in Physicians' Choices." *Journal of Health Economics* 26 (6): 1101-1127.
Fremont, Wanda P., Robert Nastasi, Nancy Newman, and Nancy Roizin. 2008. "Comfort Level of Pediatricians and Family Medicine Physicians Diagnosing and Treating Child and Adolescent Psychiatric Disorders." *International Journal of Psychiatry in Medicine* 38 (2): 153-68.
Ghandour, Reem M., Laura J. Sherman, Catherine J. Vladutiu, Mir M. Ali, Sean E. Lynch, Rebecca H. Bitsko, and Stephen J. Blumberg. 2019. "Prevalence and Treatment of Depression, Anxiety, and Conduct Problems in U.S. Children." *The Journal of Pediatrics* 206 (March): 256-267.
Goodman, Alissa, Robert Joyce, and James P. Smith. (2011). "The Long Shadow Cast by Childhood Physical and Mental Problems on Adult Life." *Proceedings of the National Academy of Sciences* 108 (15): 6032-6037.

Hazell, Philip. 2012. “Depression in Children and Adolescents.” *American Family Physician* 86 (12): 1138-1139.

Health Resources and Services Administration/National Center for Health Workforce Analysis; Substance Abuse and Mental Health Services Administration/Office of Policy, Planning, and Innovation. 2016. “National Projections of Supply and Demand for Behavioral Health Practitioners: 2013-2025.” Rockville, Maryland.

Heckman, James J. 2000. “Policies to Foster Human Capital.” *Research in Economics* 54 (1): 3-56.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. “*The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior*.” *Journal of Labor Economics* 24 (3): 411-482.

Heneghan, Amy, Andrew S. Garner, Amy Storfer-Isser, Karl Kortepeter, Ruth E.K, Stein, and Sarah M. Horwitz. 2008. “Pediatricians' Role in Providing Mental Health Care for Children and Adolescents: Do Pediatricians and Child and Adolescent Psychiatrists Agree?” *Journal of Developmental and Behavioral Pediatrics* 29 (4): 262-269.

Hilt, Robert J., Monica Chaudhari, Janice F. Bell, Christine Wolf, Kent Koprowicz, and Bryan H. King. 2014. “Side Effects from Use of One or More Psychiatric Medications in a Population-Based Sample of Children and Adolescents.” *Journal of Child and Adolescent Psychopharmacology* 24 (2): 83-89.

Kahneman, D. and Gary Klein. 2009. “Conditions for Intuitive Expertise: A Failure to Disagree.” *American Psychologist* 64 (6): 515-526.

Kessler, Daniel P. and Mark B. McClellan. 2000. “Is Hospital Competition Socially Wasteful?” *Quarterly Journal of Economics* 115 (2): 577-615.

Kalverdijk, Luuk J., Christian J. Bachmann, Lise Aagaard, Mehmet Burcu, Gerd Glaeske, Falk Hoffmann, Irene Petersen, Catharina Schuiling-Veninga, Linda P. Wijlaars, and Julie M. Zito. 2017. “A Multi-National Comparison of Antipsychotic Drug Use in Children and Adolescents, 2005-2012.” *Child and Adolescent Psychiatry and Mental Health* 11 (55): 1-9.

Large, Matthew Michael, Daniel Thomas Chung, Michael Davidson, Mark Weiser, and Christopher James Ryan. 2017. “In-Patient Suicide: Selection of People at Risk, Failure of Protection and the Possibility of Causation.” *BJPsych Open* 3 (3): 102-105.

Larkin, Ian, Desmond Ang, Jerry Avorn, and Aaron S. Kesselheim. 2014. “Restrictions on Pharmaceutical Detailing Reduced Off-Label Prescribing of Antidepressants and Antipsychotics in Children.” *Health Affairs* 33 (6): 1014-1023

Leeb, Rebecca T., Rebecca H. Bitsko, Lakshmi Radhakrishnan, Pedro Martinez, Rashid Njai, and Kristin M. Holland. 2020. “Mental Health-Related Emergency Department Visits Among Children Aged <18 Years During the COVID-19 Pandemic — United States, January 1-October 17, 2020.” *Morbidity and Mortality Weekly Report* 69 (45): 1675-1680.

List, John A. 2022. *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale*. New York: Penguin Random House.

Locke, Amy B., Nell Kirst, and Cameron G. Schultz. 2015. “Diagnosis and Management of Generalized Anxiety Disorder and Panic Disorder in Adults.” *American Family Physician* 91 (9): 617-624.

Marsac, Meghan L., Nancy Kassam-Adams, Douglas L. Delahanty, Keith F. Widaman, and Lamia P. Barakat. 2014. “Posttraumatic Stress Following Acute Medical Trauma in

Children: A Proposed Model of Bio-Psycho-Social Processes During the Peri-Trauma Period." *Clinical Child and Family Psychology Review* 17 (4): 399-411.
Martínez-Alés, Gonzalo, Tammy Jiang, Katherine M. Keyes, and Jaimie L. Gradus. 2022. "The Recent Rise of Suicide Mortality in the United States." *Annual Review of Public Health* 43 (April): 99-116.
McBain, Ryan K. Aaron Kofner, Bradley D. Stein, Jonathan H. Cantor, William B. Vogt, and Hao Yu. 2019. "Growth and Distribution of Child Psychiatrists in the United States: 2007-2016." *Pediatrics* 144 (6): 1-10.
McQuaid, John R. and Members of the Guideline Development Panel for the Treatment of Depressive Disorders. 2019. "Clinical Practice Guideline for the Treatment of Depression Across Three Age Cohorts." *American Psychological Association Guideline Development Panel for the Treatment of Depressive Disorde*rs.
Montiel Olea, Jose L. and Carolin E. Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business and Economic Statistics* 31 (3): 358-369.
Mullainathan, Sendhil and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87-106.
National Institute for Health and Care Excellence. 2005. "Depression in Children and Young People." Clinical guideline CG28. London: National Institute for Health and Care Excellence.
———. 2011. "Generalised Anxiety Disorder and Panic Disorder in Adults: Management." Clinical guideline CG113. London: National Institute for Health and Care Excellence.
———. 2014. "Anxiety Disorders." Quality standard QS53. London: National Institute for Health and Care Excellence.
———. 2019. "Depression in Children and Young People: Identification and Management." NICE guideline 134. London: National Institute for Health and Care Excellence. Olgun, Hasim, Zuhal Keskin Yldrm, Mehmet Karacan, and Naci Ceviz. 2009. "Clinical, Electrocardiographic, and Laboratory Findings in Children with Amitriptyline Intoxication." *Pediatric Emergency Care* 25 (3): 170-3.
Paksarian, Diana, Ramin Mojtabai, Roman Kotov, Bernadette Cullen, Katie L. Nugent, and Evelyn J. Bromet.. 2014. "Perceived Trauma During Hospitalization and Treatment Participation among Individuals with Psychotic Disorders." *Psychiatric Services* 65 (2): 266-269.
Riddle, Mark A., Gail A. Bernstein, Edwin H. Cook, Henrietta L. Leonard, John S. March, and James M. Swanson. 1999. "Anxiolytics, Adrenergic Agents, and Naltrexone." *Journal of American Academy of Child and Adolescent Psychiatry* 38 (5): 546-56.
Sands, Elizabeth G. and Duncan S. Gilchrist. 2016. "Something to Talk About: Social Spillovers in Movie Consumption." *Journal of Political Economy* 124 (5): 1339-1382.
Schnell, Molly and Janet Currie. 2018. "Addressing the Opioid Epidemic: Is There a Role for Physician Education?" *American Journal of Health Economics* 4 (3): 383-410.
Southammakosane, Cathy and Kristine Schmitz. 2015. "Pediatric Psychopharmacology for Treatment of ADHD, Depression, and Anxiety." *Pediatrics* 136 (2): 351-359.
Syed, Shariful, Birgham Dixson, Eduardo Constantino, and Judith Regan. 2021. "The Law and Practice of Off-Label Prescribing and Physician Promotion." *Journal of the American Academy of Psychiatry and the Law* 49:1.
Smith, James P. and Gillian Smith. 2010. "Long-Term Economic Costs of Psychological Problems During Childhood." *Social Science and Medicine* 71 (1): 110-115.

Thomas, Christopher R. and Charles E. Holzer. 2006 “The Continuing Shortage of Child and Adolescent Psychiatrists.” *Journal of the American Academy of Child & Adolescent Psychiatry* 45 (9): 1023-1031.

Walter, Heather J., Oscar G. Bukstein, A. Reese Abright, Helene Keable, Ujjwal Ramtekkar, Jane Ripperger-Suhler, and Carol Rockhill. 2007. “Clinical Practice Guideline for the Assessment and Treatment of Children and Adolescents With Anxiety Disorders.” *Journal of the American Academy of Child and Adolescent Psychiatry* 59 (10): 1107-1124.

World Health Organization. 2016. “mhGAP Intervention Guide for Mental, Neurological and Substance Use Disorders in Non-Specialized Health Settings: Mental Health Gap Action Programme (mhGAP): Version 2.0.” Geneva: World Health Organization.

Tables

Table 1
Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	All BCBS	**All MH BCBS**	**Sample**	**Treated**	**Therapy Only**	**Therapy and Drugs**	**Drugs**	**FDA-Approved Drugs**	**Grey-Area Drugs**	**Red-Flag Drugs**
A: Child Characteristics										
Number of Children	2,201,566	202,066	45,223	33,812	25,624	4,402	8,188	1,744	4,450	1,994
% of Sample	-	-	-	*75%*	*57%*	*10%*	*18%*	*4%*	*10%*	*4%*
Female	0.490	0.528	0.567	0.573	0.564	0.613	0.600	0.638	0.586	0.598
Age 1st Appearance in Sample	7.581	8.395	8.729	8.752	8.681	8.982	8.974	9.091	8.902	9.034
Neurodevelopmental Condition	0.173	0.378	0.367	0.375	0.345	0.497	0.466	0.417	0.462	0.517
Neuro Condition is ADHD	0.125	0.274	0.264	0.269	0.246	0.359	0.342	0.303	0.342	0.377
Age, 1st Mental Illness Episode	-	12.023	11.618	11.619	11.512	11.977	11.955	12.311	11.770	12.058
Hospitalized, 1st Mental Illness Episode	-	0.012	0.017	0.016	0.006	0.063	0.050	0.063	0.029	0.085
ER, 1st Mental Illness Episode	-	0.033	0.038	0.035	0.031	0.058	0.046	0.060	0.035	0.058
1st Mental Illness Episode Is Evaluation	-	0.421	0.544	0.616	0.720	0.449	0.289	0.300	0.287	0.283
B: Child Outcomes										
Hospitalized, Any Reason	0.001	0.003	0.003	0.004	0.002	0.009	0.008	0.008	0.005	0.013
Self-Harm	0.017	0.065	0.103	0.112	0.083	0.259	0.203	0.263	0.143	0.283
ER, Any Reason	0.009	0.014	0.018	0.019	0.017	0.030	0.025	0.026	0.021	0.032
Average Monthly Costs ($2018)	$157	$302	$322	$331	$278	$506	$494	$450	$428	$681
C: Prevalence of Lifetime Depression/Anxiety Mental Health Diagnoses in the Sample										
Depression	-	0.243	0.404	0.425	0.368	0.683	0.604	0.770	0.508	0.671
Anxiety	-	0.421	0.680	0.669	0.609	0.862	0.859	0.792	0.913	0.795
Adjustment with Anxiety	-	0.080	0.161	0.176	0.204	0.112	0.085	0.070	0.088	0.094
Adjustment with Depression	-	0.060	0.110	0.122	0.135	0.106	0.081	0.103	0.062	0.104
Adjustment with Anxiety/Depression	-	0.119	0.236	0.264	0.301	0.206	0.151	0.192	0.129	0.163

NOTE.—Data are from the BCBS Axis data base of insurance claims for 2012 to 2018. It covers children who have a valid master member ID, pharmacy coverage, valid geographic information, and who were observed both before age 11 and for at least one year between the ages of 10 and 18. Children in column 2 had at least one claim related to mental illness. Column 3 includes all children who can be followed for at least 24 months and have a diagnosis of depression, anxiety, or adjustment disorders with features of anxiety and depression; column 4 includes all children who can be followed for at least 24 months: this is our population of interest in the analysis. Column 4 includes those who received any follow up treatment in the 3 months following an initial claim. Children in column 5 received only therapy (no drugs) in the three months following the initial claim. Column 6 includes children who received both therapy and drugs in the three months following the initial claim. Column 7 includes children who received only drug treatment (no therapy) in the three months following the initial claim. Column 8 includes children who received only FDA-approved drugs in the 3 months following the first claim. Column 9 includes children who received grey-area drugs in the 3 months following the first claim. Column 8 includes children who received red-flag drugs in the 3 months following the first claim. Columns 8-10 are mutually exclusive categories.

Table 2
Small-Area Variation in Treatment, Provider Supply, and Practice Style

Percentiles of the Area-Level Distribution:	(1) 10th	(2) 25th	(3) 50th	(4) 75th	(5) 90th	(6) Within-ZCTA Variation	(7) Between-ZCTA Variation
A. Treatment (Source is BCBS)							
Child treated within 3 months	0.50	0.65	0.75	0.83	0.89	79%	21%
Therapy only (if any treatment)	0.50	0.65	0.76	0.85	0.93	63%	37%
Drugs & therapy (if any treatment)	0.00	0.06	0.12	0.17	0.25	73%	27%
Drugs only (if any treatment)	0.00	0.03	0.10	0.19	0.31	67%	33%
Drugs (if any treatment)	0.05	0.15	0.24	0.35	0.50	65%	35%
Red-flag drug (if any drug treatment)	0.00	0.00	0.20	0.33	0.50	79%	21%
Grey-area drug (if any drug treatment)	0.00	0.33	0.50	0.68	1.00	76%	24%
FDA-approved drug (if any drug treatment)	0.00	0.00	0.14	0.29	0.50	77%	23%
B. Provider Supply (Source is BCBS)							
Psychiatrists per 1,000 BCBS children 10-17	4.66	6.19	8.85	12.17	16.93	16%	84%
Therapists per 1,000 BCBS children 10-17	16.26	21.87	29.33	41.45	54.67	13%	87%
PCPs providing MH treatment per 1,000 BCBS children 10-17	24.42	34.06	47.38	63.07	87.07	28%	72%
C. Practice Style (Source is BCBS or IQVIA)							
BCBS: Share PCP's patients who receive a MH drug	0.04	0.05	0.07	0.09	0.12	21%	79%
BCBS: Share MH providers who are PCPs	0.13	0.15	0.18	0.21	0.23	12%	88%
BCBS: Share MH providers who are psychiatrists	0.08	0.10	0.12	0.14	0.17	14%	86%
IQVIA: Share new prescriptions by psychiatrists	0.03	0.05	0.07	0.09	0.13	33%	67%
BCBS: Share DAX patients with antipsychotic	0.06	0.07	0.08	0.09	0.11	51%	49%

NOTE.—This table is calculated by computing small area-level rates and then calculating percentiles of the distributions of those rates. Small areas are defined using information about where children in a particular zip code actually go to receive mental health care. Each row represents a separate distribution of ZCTAs. For example, places at the 90th percentile in terms of psychiatrists per capita could be at the 10th percentile in terms of the fraction of PCPs treating mental health. BCBS indicates that the variable is calculated using our main sample. IQVIA indicates that the data was calculated using that data base. The variables "within" and "between" ZCTA variation represent the decomposition of total variation in each treatment, supply, and style measure.

Table 3
First-stage with LASSO Selected Instruments

	(1) Treated	(2) Drugs	(3) Red-Flag	(4) Grey-Area	(5) FDA-Approved
Child Female	0.018***	0.007**	0.003	0.004	-0.007
	(0.004)	(0.004)	(0.011)	(0.011)	(0.009)
Age 11	0.010+	0.020***	-0.013	0.024	-0.012
	(0.005)	(0.006)	(0.014)	(0.016)	(0.012)
Age 12	0.003	0.047***	-0.007	-0.044**	0.051***
	(0.006)	(0.007)	(0.013)	(0.018)	(0.016)
Age 13	0.017**	0.077***	0.014	-0.062***	0.045***
	(0.007)	(0.007)	(0.015)	(0.018)	(0.014)
Age 14	0.021***	0.110***	0.022	-0.069***	0.044**
	(0.008)	(0.009)	(0.019)	(0.021)	(0.019)
Age 15	0.041***	0.162***	0.041	-0.076**	0.034
	(0.013)	(0.013)	(0.027)	(0.032)	(0.026)
1st Claim is Hospitalization	0.034**	0.203***	0.138***	-0.095***	-0.039
	(0.017)	(0.021)	(0.029)	(0.027)	(0.026)
1st Claim is ER Visit	-0.066***	-0.015	0.023	-0.037	0.015
	(0.023)	(0.015)	(0.026)	(0.030)	(0.026)
1st Claim is MH Evaluation	0.197***	-0.117***	-0.017	0.022+	-0.004
	(0.006)	(0.021)	(0.011)	(0.011)	(0.010)
Hospitalized Last 6 Months	-0.019	0.093***	0.225***	-0.208***	-0.019
	(0.029)	(0.030)	(0.054)	(0.056)	(0.043)
Depression	0.109***	0.200***	-0.001	-0.083**	0.127***
	(0.007)	(0.008)	(0.020)	(0.033)	(0.017)
Anxiety	0.108***	0.080***	-0.041**	0.115***	-0.045***
	(0.006)	(0.020)	(0.016)	(0.021)	(0.014)
Adjustment, with Anxiety	0.109***	0.007	-0.045**	0.055+	-0.005
	(0.009)	(0.008)	(0.023)	(0.032)	(0.018)
Adjustment, with Depression	0.107***	0.010	0.018	-0.001	-0.015
	(0.009)	(0.010)	(0.025)	(0.028)	(0.024)
Adjustment, with Depression/Anxiety	0.134***	-0.001	-0.082***	0.017	0.069***
	(0.008)	(0.008)	(0.019)	(0.023)	(0.022)
Neurodevelopmental	0.014***	0.054***	0.061***	-0.014	-0.048***
	(0.005)	(0.005)	(0.010)	(0.012)	(0.009)
Share new prescriptions by psychiatrists ×					
Multiple diagnoses	0.328***				

	(1) Treated	(2) Drugs	(3) Red-Flag	(4) Grey-Area	(5) FDA-Approved
continued from above...					
Share MH providers who are PCPs ×					
1st claim is evaluation		-0.583***			
		(0.108)			
Any depression diagnosis				-0.178	
				(0.177)	
Only adjustment diagnosis			0.590+	-0.681***	
			(0.300)	(0.188)	
Anxiety diagnosis		0.721***			
		(0.110)			
Only anxiety diagnosis				0.069	
				(0.172)	
Share MH providers who are psychiatrists ×					
Any depression diagnosis			0.541***	-0.471***	
			(0.170)	(0.181)	
Share PCP's patients who receive a MH drug ×					
1st claim is evaluation		-0.424***			
		(0.121)			
Multiple diagnoses	0.276***			-0.255	
	(0.083)			(0.207)	
Any depression diagnosis					0.899***
					(0.196)
Share DAX patients with antipsychotic prescription ×					
Multiple diagnoses	0.111				
	(0.085)				
Any depression diagnosis				-0.586**	
				(0.248)	
Only adjustment diagnosis			1.334+		
			(0.695)		
Constant	0.330***	0.295***	0.196***	0.655***	0.110***
	(0.016)	(0.017)	(0.036)	(0.049)	(0.027)
Observations	45223	33812	8188	8188	8188

	(0.092)	R-squared	0.126	0.288	0.077	0.153	0.127
Only anxiety diagnosis	-0.343***	F_{EFF}	22.448	30.276	24.928	14.929	20.935
	(0.119)	Mean Dependent Variable	0.748	0.242	0.244	0.543	0.213

NOTE.—This table presents first-stage results, where a child's treatment status is instrumented using patient characteristics and the set of instrumental variables selected by the LASSO procedure described by Belloni et al. (2012) using their refined data-driven penalty. The character × indicates an interaction between one of the six primary instruments and the child's diagnosis or severity. Each column is from a single regression model, where the sample is conditional on treatment in column 2 and conditional on drug treatment in columns 3-5. FEFF represent the effective F statistics for the instruments from Montiel Olea and Plueger (2013). All models include area, month, and year fixed effects, corresponding to the child's initial mental illness claim. Standard errors, clustered at the ZIP-3 level, are in parentheses. + p<0.10, * p<0.05, ** p<0.01.

Table 4
Effects of Treatment on Costs, Facility Use, and Self-Harm

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Log Total Costs			DV: Visited Facility?			DV: Any Harm Behavior?		
	3-Months	1-Year	2-Years	3-Months	1-Year	2-Years	3-Months	1-Year	2-Years
Panel A: Effect of Treatment									
Observations	45223	45223	45223	45223	45223	45223	45223	45223	45223
Mean Dependent Variable	6.926	7.846	8.448	0.112	0.182	0.254	0.025	0.036	0.052
OLS	0.563**	0.489**	0.381**	0.015**	0.024**	0.027**	0.013**	0.018**	0.020**
	(0.016)	(0.016)	(0.016)	(0.003)	(0.004)	(0.005)	(0.001)	(0.002)	(0.002)
R-squared	0.232	0.191	0.159	0.419	0.267	0.206	0.208	0.178	0.149
Post-Lasso 2SLS	0.884**	0.952**	1.085**	0.354**	0.468**	0.631**	0.213**	0.253**	0.273**
	(0.303)	(0.277)	(0.245)	(0.078)	(0.091)	(0.121)	(0.055)	(0.065)	(0.073)
Panel B: Effect of Drug Treatment									
Observations	33812	33812	33812	33812	33812	33812	33812	33812	33812
Mean Dependent Variable	7.079	7.976	8.550	0.113	0.184	0.258	0.029	0.042	0.058
OLS	0.383**	0.399**	0.385**	0.086**	0.104**	0.110**	0.034**	0.047**	0.052**
	(0.020)	(0.018)	(0.017)	(0.005)	(0.007)	(0.007)	(0.004)	(0.004)	(0.004)
R-squared	0.220	0.190	0.168	0.401	0.260	0.205	0.236	0.205	0.174
Post-Lasso 2SLS	-0.528*	-0.501*	-0.487*	0.021	-0.041	-0.115	-0.042	-0.005	-0.096+
	(0.218)	(0.217)	(0.233)	(0.058)	(0.074)	(0.093)	(0.040)	(0.048)	(0.057)
Panel C1: Effect of Red-Flag Drug Treatment									
Observations	8188	8188	8188	8188	8188	8188	8188	8188	8188
Mean Dependent Variable	7.478	8.388	8.945	0.226	0.315	0.394	0.085	0.112	0.136
OLS	0.562**	0.472**	0.405**	0.122**	0.111**	0.102**	0.051**	0.055**	0.055**
	(0.035)	(0.033)	(0.033)	(0.010)	(0.012)	(0.012)	(0.007)	(0.008)	(0.009)
R-squared	0.324	0.269	0.232	0.351	0.258	0.214	0.269	0.252	0.236
Post-Lasso 2SLS	1.133**	1.388**	1.458**	0.217**	0.266**	0.437**	0.186**	0.180*	0.149+
	(0.310)	(0.288)	(0.286)	(0.084)	(0.103)	(0.116)	(0.067)	(0.074)	(0.080)
Panel C2: Effect of Grey-Area Drug Treatment									
OLS	-0.316**	-0.241**	-0.218**	-0.074**	-0.067**	-0.066**	-0.024**	-0.030**	-0.033**
	(0.027)	(0.026)	(0.027)	(0.009)	(0.009)	(0.010)	(0.006)	(0.007)	(0.008)
R-squared	0.307	0.253	0.220	0.343	0.252	0.211	0.264	0.248	0.234
Post-Lasso 2SLS	-0.962**	-1.249**	-1.271**	-0.170*	-0.240*	-0.450**	-0.089	-0.097	-0.061
	(0.301)	(0.295)	(0.281)	(0.081)	(0.101)	(0.113)	(0.067)	(0.077)	(0.083)
Panel C3: Effect of FDA-Approved Drug Treatment									
OLS	-0.201**	-0.205**	-0.159**	-0.035**	-0.034**	-0.024	-0.025**	-0.021*	-0.017
	(0.033)	(0.031)	(0.031)	(0.011)	(0.012)	(0.015)	(0.008)	(0.009)	(0.011)
R-squared	0.299	0.249	0.216	0.338	0.249	0.207	0.264	0.247	0.232
Post-Lasso 2SLS	0.898	1.097+	1.625*	-0.234	-0.206	0.063	-0.299+	-0.378*	-0.301
	(0.576)	(0.606)	(0.646)	(0.178)	(0.195)	(0.208)	(0.169)	(0.190)	(0.210)

NOTE.—This table presents results from OLS and post-Lasso 2SLS models of mental health treatment on log total health costs (columns 1-3), facility use (columns 4-6), and any self-harm behavior (columns 7-9). The instruments in the post-Lasso 2SLS models are chosen using a refined data-driven penalty and are displayed in Appendix Table A6. Panel A includes 4 instruments; Panel B includes 3 instruments; and Panels C1, C2, and C3 include 3, 6, and 1 instrument respectively. Post-estimation results from the first-stage of each model are included in Table 3. All models include patient controls (age, female, severity, and diagnosis) as well as area, month, and year fixed effects, corresponding to the child's initial mental illness claim. Standard errors, clustered at the ZIP-3 level, are in parentheses. + p<0.10, * p<0.05, ** p<0.01.

Table 5
Effects of Treatment on Mental Health Costs and Facility Use

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Total MH Costs			Length of Facility Stay		
	3-Months	1-Year	2-Years	3-Months	1-Year	2-Years
Panel A: Effect of Treatment						
Observations	45223	45223	45223	45223	45223	45223
Mean Dependent Variable	6.082	6.566	6.890	0.437	0.775	1.240
OLS	0.897**	1.103**	1.100**	0.327*	0.537**	0.778**
	(0.022)	(0.022)	(0.023)	(0.160)	(0.182)	(0.209)
R-squared	0.308	0.298	0.266	0.026	0.033	0.028
Post-Lasso 2SLS	1.384**	1.585**	1.638**	0.959	2.791	7.253
	(0.361)	(0.379)	(0.370)	(2.475)	(3.024)	(4.799)
Panel B: Effect of Drug Treatment						
Observations	33812	33812	33812	33812	33812	33812
Mean Dependent Variable	6.352	6.881	7.201	0.501	0.888	1.412
OLS	-0.034	0.174**	0.294**	0.870**	1.272**	1.714**
	(0.027)	(0.025)	(0.027)	(0.250)	(0.319)	(0.405)
R-squared	0.239	0.214	0.198	0.026	0.034	0.030
Post-Lasso 2SLS	-1.589**	-1.472**	-1.390**	-0.174	-1.572	-6.217+
	(0.309)	(0.349)	(0.366)	(1.120)	(1.606)	(3.411)
Panel C1: Effect of Red-Flag Drug Treatment						
Observations	8188	8188	8188	8188	8188	8188
Mean Dependent Variable	6.388	7.098	7.522	1.655	2.581	3.656
OLS	0.464**	0.469**	0.444**	2.361**	3.478**	4.462**
	(0.040)	(0.041)	(0.041)	(0.792)	(0.725)	(0.938)
R-squared	0.425	0.376	0.338	0.062	0.081	0.084
Post-Lasso 2SLS	1.514**	1.553**	1.345**	1.610	3.649	13.130+
	(0.320)	(0.349)	(0.365)	(3.968)	(4.810)	(7.152)
Panel C2: Effect of Grey-Area Drug Treatment						
OLS	-0.164**	-0.124**	-0.074*	-1.218**	-1.681**	-2.109**
	(0.036)	(0.037)	(0.037)	(0.363)	(0.285)	(0.492)
R-squared	0.415	0.366	0.328	0.061	0.078	0.081
Post-Lasso 2SLS	-1.984**	-2.034**	-1.760**	-1.115	-4.516	-12.106*
	(0.348)	(0.374)	(0.380)	(3.304)	(3.774)	(5.874)
Panel C3: Effect of FDA-Approved Drug Treatment						
OLS	-0.306**	-0.369**	-0.413**	-1.006*	-1.644**	-2.179**
	(0.039)	(0.041)	(0.042)	(0.459)	(0.583)	(0.784)
R-squared	0.418	0.371	0.335	0.061	0.078	0.081
Post-Lasso 2SLS	1.267+	1.244+	1.278	-13.038	-15.917	-2.715
	(0.716)	(0.750)	(0.824)	(11.842)	(13.868)	(17.102)

NOTE.—This table presents results from OLS and post-Lasso 2SLS models of mental health treatment on log total mental health costs (columns 1-3) and length of facility stay (columns 4-6). The instruments in the post-Lasso 2SLS models are chosen using a refined data-driven penalty and are displayed in Appendix Table A6. Panel A includes 4 instruments; Panel B includes 3 instruments; and Panels C1, C2, and C3 include 3, 6, and 1 instrument respectively. Post-estimation results from the first-stage of each model are included in Table 3. All models include patient controls (age, female, severity, and diagnosis) as well as area, month, and year fixed effects, corresponding to the child's initial mental illness claim. Standard errors, clustered at the ZIP-3 level, are in parentheses. + p<0.10, * p<0.05, ** p<0.01.

Table 6
Effects of Treatment on Costs, Facility Use, and Self-Harm Among Children Without Psychosis, Bipolar, Autism, or Tourette's

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Log Total Costs			DV: Visited Facility?			DV: Any Harm Behavior?		
	3-Months	1-Year	2-Years	3-Months	1-Year	2-Years	3-Months	1-Year	2-Years
Panel A: Effect of Treatment									
Post-Lasso 2SLS, Trimmed Sample	-0.506	-0.087	0.239	0.400**	0.539**	0.687**	-0.060	-0.053	-0.024
	(0.329)	(0.289)	(0.247)	(0.067)	(0.076)	(0.100)	(0.060)	(0.064)	(0.074)
Table 4 Post-Lasso 2SLS Estimate	0.884**	0.952**	1.085**	0.354**	0.468**	0.631**	0.213**	0.253**	0.273**
	(0.303)	(0.277)	(0.245)	(0.078)	(0.091)	(0.121)	(0.055)	(0.065)	(0.073)
Panel B: Effect of Drug Treatment									
Post-Lasso 2SLS, Trimmed Sample	-0.486*	-0.503*	-0.482*	0.024	-0.039	-0.101	-0.046	-0.006	-0.083
	(0.220)	(0.219)	(0.232)	(0.053)	(0.069)	(0.089)	(0.037)	(0.044)	(0.052)
Table 4 Post-Lasso 2SLS Estimate	-0.528*	-0.501*	-0.487*	0.021	-0.041	-0.115	-0.042	-0.005	-0.096+
	(0.218)	(0.217)	(0.233)	(0.058)	(0.074)	(0.093)	(0.040)	(0.048)	(0.057)
Panel C1: Effect of Red-Flag Drug Treatment									
Post-Lasso 2SLS, Trimmed Sample	1.215**	1.374**	1.378**	0.236**	0.280**	0.442**	0.147*	0.132+	0.102
	(0.315)	(0.281)	(0.273)	(0.083)	(0.103)	(0.119)	(0.062)	(0.070)	(0.079)
Table 4 Post-Lasso 2SLS Estimate	1.133**	1.388**	1.458**	0.217**	0.266**	0.437**	0.186**	0.180*	0.149+
	(0.310)	(0.288)	(0.286)	(0.084)	(0.103)	(0.116)	(0.067)	(0.074)	(0.080)
Panel C2: Effect of Grey-Area Drug Treatment									
Post-Lasso 2SLS, Trimmed Sample	-0.951**	-1.183**	-1.188**	-0.215**	-0.260*	-0.486**	-0.097	-0.090	-0.055
	(0.308)	(0.285)	(0.276)	(0.083)	(0.102)	(0.119)	(0.066)	(0.075)	(0.084)
Table 4 Post-Lasso 2SLS Estimate	-0.962**	-1.249**	-1.271**	-0.170*	-0.240*	-0.450**	-0.089	-0.097	-0.061
	(0.301)	(0.295)	(0.281)	(0.081)	(0.101)	(0.113)	(0.067)	(0.077)	(0.083)
Panel C3: Effect of FDA-Approved Drug Treatment									
Post-Lasso 2SLS, Trimmed Sample	1.081+	1.255+	1.733*	-0.177	-0.258	0.015	-0.225	-0.344+	-0.310
	(0.628)	(0.679)	(0.758)	(0.175)	(0.192)	(0.210)	(0.157)	(0.187)	(0.205)
Table 4 Post-Lasso 2SLS Estimate	0.898	1.097+	1.625*	-0.234	-0.206	0.063	-0.299+	-0.378*	-0.301
	(0.576)	(0.606)	(0.646)	(0.178)	(0.195)	(0.208)	(0.169)	(0.190)	(0.210)

NOTE.—This table presents results from post-Lasso 2SLS models of mental health treatment on log total health costs (columns 1-3), facility use (columns 4-6), and any harm behavior (columns 7-9). The "Trimmed Sample" estimates exclude children with schizophrenia spectrum and other psychotic disorders, bipolar and related disorders, autism spectrum disorder, and Tourette's disorder from the sample whereas the "Table 4" estimates include the full sample. The instruments in the post-Lasso 2SLS models are chosen using a refined data-driven penalty described in Belloni et al. (2012) and are displayed in Appendix Table A6. For the "Trimmed Sample" models, Panel A includes 5 instruments; Panel B includes 3 instruments; and Panels C1, C2, and C3 include 5, 5, and 1 instrument respectively. All models include patient controls (age, female, severity, and diagnosis) as well as area, month, and year fixed effects, corresponding to the child's initial mental illness claim. Standard errors, clustered at the ZIP-3 level, are in parentheses. + p<0.10, * p<0.05, ** p<0.01.

Appendix Table A1
Guidelines for Treatment of Child and Adolescent Depression and Anxiety

Guideline	American Academy of Child and Adolescent Psychiatrists (2007)	American Academy of Family Physicians (2012)	American Academy of Pediatrics (2007)	American Psychological Association (2019)	U.K. National Institute for Health and Care Excellence (2005, 2019)	The World Health Organization (2016)
Depression Disorders						
Mild Depression	Education, support, and case management is sufficient.	Cognitive behavior therapy (CBT)	Consider support and monitoring before starting other evidence-based treatment.	Does not offer separate guidance for mild depression.	Suggest a period of 4 weeks of "watchful waiting" before moving to therapy.	Suggest psychosocial interventions with child, family, and school.
Moderate to Severe Depression	Drug treatment may be administered alone until the child is amenable to psychotherapy, or they can be combined. SSRIs have a relatively good response rate. Fluoxetine is the only FDA-approved medication for this indication in children. Tricyclic antidepressants should not be used as a first-line medication. They are associated with more side effects than the SSRIs and can be fatal after an overdose. Recommend lower initial doses than for adults. 2020 update: SNRIs can also be used. Need more research on Benzodiazepines.	CBT plus medication. Three SSRIs recommended as 1st line treatment. Tricyclic antidepressants should not be used.	Clinicians "should recommend scientifically tested and proven treatments (i.e. psychotherapies such as CBT or IPT (Interpersonal therapy) and/or antidepressant treatment such as SSRIs." Guidelines include a table of 6 SSRIs with a dosing schedule.	For children (<12) there is insufficient evidence for any recommendation. For adolescents, clinicians should offer either CBT or IPT. An SSRI (fluoxetine) is recommended as a first line medication. Insufficient evidence to choose between therapy and medication.	Providers should offer a specific psychotherapy (e.g. CPT, IPT, but others also mentioned), with the addition of fluoxetine if therapy is insufficient. In 2015 this guidance was updated to say that the combination of therapy and fluoxetine could be consider as an initial therapy. Other drugs should be prescribed only if the combination of fluoxetine and therapy is ineffective. Sertraline and Citalopram are recommended as alternatives.	Do not prescribe medication to children less than 12. If psychosocial intervention has been tried for at least six months, children 12 and over can be referred to a specialist and receive fluoxetine. They should not receive TCAs or other SSRIs. Patients prescribed fluoxetine should be monitored weekly for suicidal ideations.

Guideline	AACAP (2007, 2020)	AAFP (For adults,* 2015)	AAP (2015)	NICE (For adults,* 2011)	WHO (2016)
Anxiety Disorders	CBT or psychodynamic therapy should be combined with SSRIs as "the medication of choice." The safety and efficacy of other medications for childhood anxiety has not been established. Controlled trials of TCAs have not established efficacy for this use. Benzodiazepines have not shown efficacy in controlled trials in childhood anxiety. Clinicians should use benzodiazepines cautiously to avoid developing dependency.	Physical activity can be a cost-effective therapy. Psychotherapy can be as effective as medication (suggest CBT). SSRIs are the first line medication. Benzodiazepines do not improve longer-term outcomes. Because benzodiazepines are associated with tolerance, they should be used only short-term during crises.	"SSRIs are important treatment tools for moderate- severe depression and anxiety disorders." Includes a table with 5 SSRIs and a dosing schedule. Notes that one of the 5 SSRIs is not approved for use in children.	Begin with low intensity psychological counselling. Initiate higher intensity therapy such as CBT if condition persists. Consider adding an SSRI (recommend sertraline) if the condition persists. If SSRI is ineffective, try another SSRI or an SNRI. If these are ineffective consider pregabalin (an anti-seizure drug). Do not use benzodiazepines except for short-term stabilization. Do not offer an anti-psychotic.	No specific guidelines for anxiety. Guidelines refer to depression and other "emotional disorders." If after 6 months of psychosocial intervention the child has not improved, they should be referred to a specialist. They should not be prescribed medication.

NOTE.—A * symbol indicates that we were unable to find specific guidelines for treatment of anxiety in adolescents from these sources. But, in general, one would expect more conservative treatment for children and adolescents than for adults so we have included the adult guidelines here. Abbreviations in the table are defined as follows: CBT (cognitive behavioral therapy), IPT (interpersonal therapy), SNRI (serotonin-norepinephrine reuptake inhibitor), SSRI (selective serotonin reuptake inhibitor), and TCA (tricyclic antidepressant). Guideline sources are as follows: Walter et al. (2020), Connolly et al. (2007), and Birmaher et al. (2007) [for AACAP]; Hazell (2012) [for AAFP]; Locke et al. (2015) and Cheung et al. (2007) [for AAP]; Southammakosane and Schmitz (2015) and McQuaid and Members of the Guideline Development Panel for the Treatment of Depressive Disorders (2019) [for APA]; National Institute for Health and Care Excellence (2005, 2011, 2014, and 2019) [for NICE], and World Health Organization (2016) [for WHO].

Appendix Table A2
Area Characteristics of BCBS Children

	(1) National Average	(2) All BCBS	(3) All MH BCBS	(4) Sample
Total Population (000s)	31.5	27.1	26.8	26.5
Total Population: Female	51%	51%	51%	51%
Total population: 0-17	23%	23%	23%	23%
Total population: 10-17	10%	11%	11%	11%
White Alone	73%	80%	82%	83%
Black or African American Alone	13%	9%	8%	7%
Other Race	15%	11%	10%	10%
Hispanic or Latino	19%	12%	10%	9%
Family Households	67%	69%	68%	68%
Married (Not Including Separated)	48%	53%	53%	53%
Less than High School	13%	10%	9%	8%
High School Diploma	56%	56%	55%	54%
Bachelor's Degree or Better	31%	34%	36%	37%
Labor Force Participation Rate	63%	64%	65%	65%
Employment Rate	94%	95%	95%	95%
Average Household Income (In $2018)	$84,303	$90,742	$94,584	$96,813
Gini Index	0.44	0.43	0.43	0.43
Owner Occupied Housing Units	64%	71%	71%	72%
Families Below Poverty Level	11%	8%	8%	7%
Adult Poverty Rate	14%	11%	10%	10%
Average Commute to Work (In Min.)	26.7	25.6	25.5	25.6
Uninsured	9%	8%	7%	7%
Public Health Coverage	35%	32%	31%	31%
Private Health Insurance	67%	73%	75%	76%
Children Living with Single Parents	31%	26%	26%	25%
Number of ZCTAs	32799	27906	16487	10595

NOTE.—This table presents estimates from the American Community Survey 2018 5-Year files for ZCTA geographies. Each row represents a weighted average of each ZCTA measure, where the weights correspond to the relevant population in each ZCTA. Column 1 includes all ZCTAs in the country, and the weights correspond to the total reported population in the ZCTA. Column 2 includes all ZCTAs with at least one child aged 10-17 with BCBS coverage, and the weights correspond to the number of BCBS-insured children 10-17. Column 3 includes all ZCTAs with at least one child aged 10-17 with BCBS coverage and a mental health condition, and the weights correspond to the number of BCBS-insured children 10-17 with mental illness. Column 4 includes all ZCTAs with at least one child aged 10-17 with BCBS coverage, at least 2 years of continuous coverage after their initial mental health event, and a formal diagnosis of depression, anxiety, or an adjustment disorder with features of depression or anxiety--this column corresponds to our main analysis sample. The three health insurance variables ("Uninsured," "Public Health Coverage," and "Private Health Coverage") do not add to 100 given overlapping coverage.

Appendix Table A3
Test Statistics for Alternative First-Stage Regressions

	(1) OLS - No Instruments	(2) Main Instruments + Quadratic	(3) Full Interactions	(4) Post-Lasso: 3 Instruments	(5) Post-Lasso: All Instruments
Any Follow-Up Treatment in 3 Months After 1st Claim					
N	45223	45223	45223	45223	45223
R-squared	0.123	0.124	0.131	0.126	0.126
F_{EFF}	-	2.424	3.095	**26.312**	**22.448**
Drug Treatment in 1st 3 Months After 1st Claim					
Conditional on Treatment					
N	33812	33812	33812	33812	33812
R-squared	0.284	0.285	0.294	0.287	0.288
F_{EFF}	-	4.865	3.082	**23.984**	**30.276**
Red-flag Drug Treatment in 1st 3 Months After 1st Claim					
Conditional on Drug Treatment					
N	8188	8188	8188	8188	8188
R-squared	0.066	0.068	0.098	0.076	0.077
F_{EFF}	-	1.853	2.193	**19.040**	**24.928**
Grey-Area Drug Treatment in 1st 3 Months After 1st Claim					
Conditional on Drug Treatment					
N	8188	8188	8188	8188	8188
R-squared	0.143	0.144	0.170	0.149	0.152
F_{EFF}	-	0.392	1.948	**17.924**	**14.929**
FDA-Approved Drug Treatment in 1st 3 Months After 1st Claim					
Conditional on Drug Treatment					
N	8188	8188	8188	8188	8188
R-squared	0.122	0.123	0.146	0.126	0.125
F_{EFF}	-	0.993	1.452	**8.549**	**20.935**

NOTE.—This table presents post-estimation results from the first-stage model, where the instrument choice set varies across specifications. FEFF refers to the effective F-statistic of Montiel Olea and Plueger (2013) and is in bold when it exceeds the 5% critical value for testing the null hypothesis that the 2SLS bias exceeds 10% of the OLS bias. The main instruments include the shares of psychiatrists and PCPs providing mental health treatment to children, the average share of patients in PCPs caseloads receiving psychiatric drugs, the share of first psychiatric drug prescriptions prescribed by psychiatrists, and the share of patients with anxiety or depression who are prescribed anti-psychotics. In columns 1-3, the instruments are pre-selected: column 1 uses no instruments (N=0); column 2 uses the main instruments as well as their squared terms (N=12); and column 3 includes all potential instruments, i.e., a second-degree polynomial in the main instruments, as well as interactions with age, gender, severity (indicators for first event in ER, first event in hospital, first event is evaluation, and hospitalization in previous 6 months), and diagnosis (N=116). In columns 4 and 5, Lasso selects the instruments. Column 4 is a post-Lasso model with a three-instrument constraint (N=3). Column 5 is a post-Lasso model where a refined data-driven penalty is used for the Lasso ($1 \leq N \leq 6$). See Appendix Table A6 for a full list of the instruments selected in each model. All models include patient controls (age, female, severity, and diagnosis) as well as area, month, and year fixed effects, corresponding to the child's initial mental illness claim. Standard errors, clustered at the ZIP-3 level, are in parentheses.

Appendix Table A4
Effects of Treatment on Costs, Facility Use, and Self-Harm, Constraining Lasso to Select Three Instruments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Log Total Costs			DV: Visited Facility?			DV: Any Harm Behavior?		
	3-Months	1-Year	2-Years	3-Months	1-Year	2-Years	3-Months	1-Year	2-Years
Panel A: Effect of Treatment									
Post-Lasso 2SLS, 3 Instruments	0.862**	0.937**	1.095**	0.356**	0.473**	0.639**	0.210**	0.248**	0.270**
	(0.305)	(0.279)	(0.247)	(0.077)	(0.091)	(0.123)	(0.054)	(0.064)	(0.072)
Table 4 Post-Lasso 2SLS Estimate	0.884**	0.952**	1.085**	0.354**	0.468**	0.631**	0.213**	0.253**	0.273**
	(0.303)	(0.277)	(0.245)	(0.078)	(0.091)	(0.121)	(0.055)	(0.065)	(0.073)
Panel B: Effect of Drug Treatment									
Post-Lasso 2SLS, 3 Instruments	-0.860**	-0.422	-0.349	0.010	0.080	0.025	0.063	0.130*	0.117+
	(0.295)	(0.291)	(0.306)	(0.067)	(0.095)	(0.118)	(0.042)	(0.053)	(0.062)
Table 4 Post-Lasso 2SLS Estimate	-0.528*	-0.501*	-0.487*	0.021	-0.041	-0.115	-0.042	-0.005	-0.096+
	(0.218)	(0.217)	(0.233)	(0.058)	(0.074)	(0.093)	(0.040)	(0.048)	(0.057)
Panel C1: Effect of Red-Flag Drug Treatment									
Post-Lasso 2SLS, 3 Instruments	1.178**	1.379**	1.520**	0.264**	0.359**	0.493**	0.172**	0.175*	0.159*
	(0.329)	(0.300)	(0.309)	(0.093)	(0.115)	(0.129)	(0.059)	(0.069)	(0.076)
Table 4 Post-Lasso 2SLS Estimate	1.133**	1.388**	1.458**	0.217**	0.266**	0.437**	0.186**	0.180*	0.149+
	(0.310)	(0.288)	(0.286)	(0.084)	(0.103)	(0.116)	(0.067)	(0.074)	(0.080)
Panel C2: Effect of Grey-Area Drug Treatment									
Post-Lasso 2SLS, 3 Instruments	-1.254**	-1.491**	-1.630**	-0.270*	-0.399**	-0.584**	-0.207**	-0.214**	-0.206*
	(0.402)	(0.385)	(0.387)	(0.110)	(0.137)	(0.157)	(0.072)	(0.081)	(0.090)
Table 4 Post-Lasso 2SLS Estimate	-0.962**	-1.249**	-1.271**	-0.170*	-0.240*	-0.450**	-0.089	-0.097	-0.061
	(0.301)	(0.295)	(0.281)	(0.081)	(0.101)	(0.113)	(0.067)	(0.077)	(0.083)
Panel C3: Effect of FDA-Approved Drug Treatment									
Post-Lasso 2SLS, 3 Instruments	1.263*	1.395*	1.829**	-0.074	-0.029	0.130	-0.243	-0.233	-0.175
	(0.566)	(0.572)	(0.604)	(0.160)	(0.173)	(0.190)	(0.152)	(0.169)	(0.187)
Table 4 Post-Lasso 2SLS Estimate	0.898	1.097+	1.625*	-0.234	-0.206	0.063	-0.299+	-0.378*	-0.301
	(0.576)	(0.606)	(0.646)	(0.178)	(0.195)	(0.208)	(0.169)	(0.190)	(0.210)

NOTE.—This table presents results from post-Lasso 2SLS models of mental health treatment on log total health costs (columns 1-3), facility use (columns 4-6), and any self-harm behavior (columns 7-9). The "3 Instruments" estimates include instruments from a Lasso procedure, where Lasso is constrained to select only three instruments whereas the "Table 4" estimates include instruments from a Lasso procedure using the refined data-driven penalty described in Belloni et al. (2012). All instruments selected are displayed in Appendix Table A6. All models include patient controls (age, female, severity, and diagnosis) as well as area, month, and year fixed effects, corresponding to the child's initial mental illness claim. Standard errors, clustered at the ZIP-3 level, are in parentheses. + p<0.10, * p<0.05, ** p<0.01.

Appendix Table A5
Effects of Treatment on Costs, Facility Use, and Self-Harm, including Supply Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Log Total Costs			DV: Visited Facility?			DV: Any Harm Behavior?		
	3-Months	1-Year	2-Years	3-Months	1-Year	2-Years	3-Months	1-Year	2-Years
Panel A: Effect of Treatment									
Post-Lasso 2SLS, Incl. Supply Measures	1.132**	1.137**	1.302**	0.382**	0.439**	0.614**	0.264**	0.310**	0.338**
	(0.334)	(0.334)	(0.290)	(0.095)	(0.103)	(0.134)	(0.062)	(0.073)	(0.078)
Table 4 Post-Lasso 2SLS Estimate	0.884**	0.952**	1.085**	0.354**	0.468**	0.631**	0.213**	0.253**	0.273**
	(0.303)	(0.277)	(0.245)	(0.078)	(0.091)	(0.121)	(0.055)	(0.065)	(0.073)
Panel B: Effect of Drug Treatment									
Post-Lasso 2SLS, Incl. Supply Measures	-0.528*	-0.501*	-0.487*	0.021	-0.041	-0.115	-0.042	-0.005	-0.096+
	(0.218)	(0.217)	(0.233)	(0.058)	(0.074)	(0.093)	(0.040)	(0.048)	(0.057)
Table 4 Post-Lasso 2SLS Estimate	-0.528*	-0.501*	-0.487*	0.021	-0.041	-0.115	-0.042	-0.005	-0.096+
	(0.218)	(0.217)	(0.233)	(0.058)	(0.074)	(0.093)	(0.040)	(0.048)	(0.057)
Panel C1: Effect of Red-Flag Drug Treatment									
Post-Lasso 2SLS, Incl. Supply Measures	1.145**	1.359**	1.522**	0.240**	0.326**	0.482**	0.177**	0.186**	0.164*
	(0.337)	(0.308)	(0.321)	(0.092)	(0.113)	(0.130)	(0.060)	(0.070)	(0.077)
Table 4 Post-Lasso 2SLS Estimate	1.133**	1.388**	1.458**	0.217**	0.266**	0.437**	0.186**	0.180*	0.149+
	(0.310)	(0.288)	(0.286)	(0.084)	(0.103)	(0.116)	(0.067)	(0.074)	(0.080)
Panel C2: Effect of Grey-Area Drug Treatment									
Post-Lasso 2SLS, Incl. Supply Measures	-1.040**	-1.337**	-1.329**	-0.172*	-0.242*	-0.445**	-0.107	-0.120	-0.076
	(0.303)	(0.301)	(0.286)	(0.081)	(0.103)	(0.114)	(0.067)	(0.077)	(0.083)
Table 4 Post-Lasso 2SLS Estimate	-0.962**	-1.249**	-1.271**	-0.170*	-0.240*	-0.450**	-0.089	-0.097	-0.061
	(0.301)	(0.295)	(0.281)	(0.081)	(0.101)	(0.113)	(0.067)	(0.077)	(0.083)
Panel C3: Effect of FDA-Approved Drug Treatment									
Post-Lasso 2SLS, Incl. Supply Measures	0.898	1.097+	1.625*	-0.234	-0.206	0.063	-0.299+	-0.378*	-0.301
	(0.576)	(0.606)	(0.646)	(0.178)	(0.195)	(0.208)	(0.169)	(0.190)	(0.210)
Table 4 Post-Lasso 2SLS Estimate	0.898	1.097+	1.625*	-0.234	-0.206	0.063	-0.299+	-0.378*	-0.301
	(0.576)	(0.606)	(0.646)	(0.178)	(0.195)	(0.208)	(0.169)	(0.190)	(0.210)

NOTE.—This table presents results from post-Lasso 2SLS models of mental health treatment on log total health costs (columns 1-3), facility use (columns 4-6), and any self-harm behavior (columns 7-9). Both the "Incl. Supply Measures" and "Table 4" estimates are from post-Lasso 2SLS models using the refined data-driven penalty described in Belloni et al. (2012); however, the "Incl. Supply Measures" models have a larger potential instrument set, i.e., provider supply-side variables (and their interactions with patient characteristics) are added to the potential instrument set. These instruments are displayed in Appendix Table A6. For the "Incl. Supply Measures" models, Panel A includes 3 instruments; Panel B includes 3 instruments; and Panels C1, C2, and C3 include 2, 5, and 1 instrument respectively. All models include patient controls (age, female, severity, and diagnosis) as well as area, month, and year fixed effects, corresponding to the child's initial mental illness claim. Standard errors, clustered at the ZIP-3 level, are in parentheses. + p<0.10, * p<0.05, ** p<0.01.

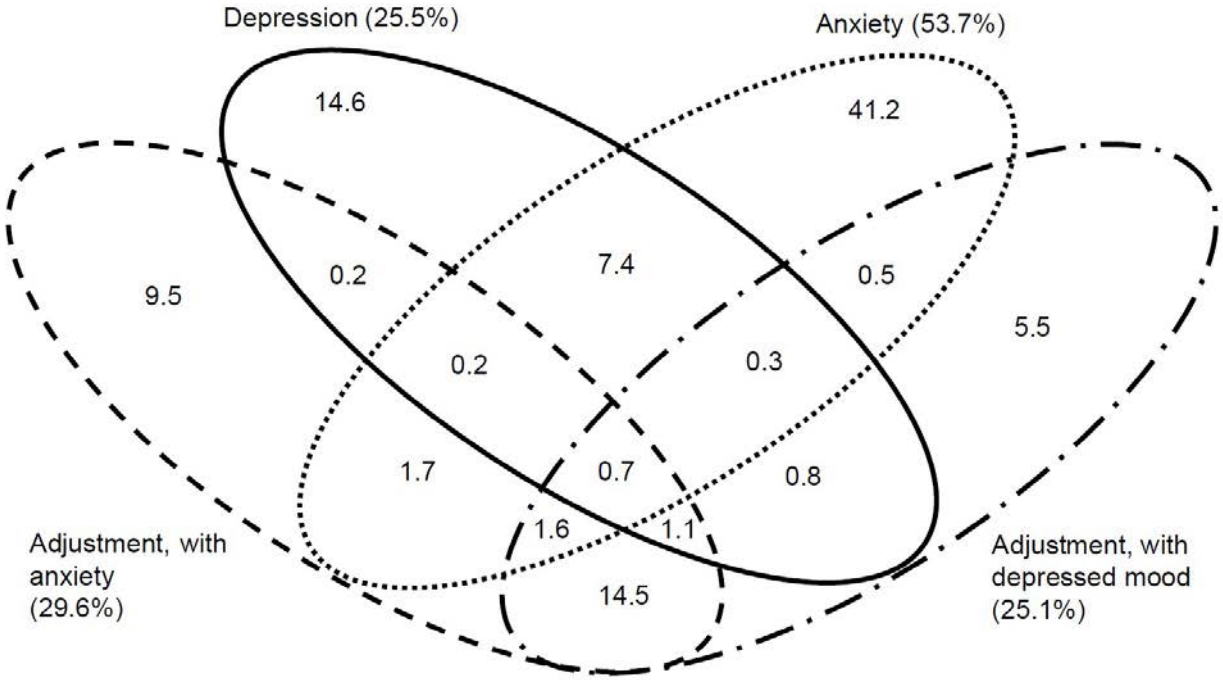
Appendix Table A6
Instruments Selected by Lasso

Specification	Probability of Any Follow-Up Treatment in 3 Months After 1st Claim	Probability of Receiving Drug Treatment in 1st 3 Months After 1st Claim, Conditional on Treatment	Probability of Receiving Red-Flag Drug Treatment in 1st 3 Months After 1st Claim, Conditional on Treatment	Probability of Receiving Grey-Area Drug Treatment in 1st 3 Months After 1st Claim, Conditional on Treatment	Probability of Receiving FDA-Approved Drug Treatment in 1st 3 Months After 1st Claim, Conditional on Treatment
1. Lasso, Data-driven Penalty	any_multiple#share_mh any_multiple#s1_by_psych any_multiple#share_ap only_anxiety#s1_by_psych	first_eval#share_mh first_eval#share_gp anxiety#share_gp	any_depression#share_psych only_adjustment#share_gp only_adjustment#share_ap	any_depression#share_psych any_depression#share_gp any_depression#share_ap any_multiple#share_mh only_anxiety#share_gp only_adjustment#share_gp	any_depression#share_mh
2. Lasso, Excluding Children with Additional MH Diagnosis	first_er#share_gp any_multiple#share_mh any_multiple#s1_by_psych any_multiple#share_ap only_anxiety#s1_by_psych	first_eval#share_mh first_eval#share_gp anxiety#share_gp	any_depression#share_psych only_anxiety#share_ap only_adjustment#share_psych only_adjustment#share_gp only_adjustment#share_ap	any_depression#share_psych any_depression#share_gp any_depression#share_ap only_adjustment#share_mh only_adjustment#share_gp	any_depression#share_mh
3. Lasso, Top 3 Instruments	any_multiple#s1_by_psych only_anxiety#s1_by_psych any_multiple#share_mh	first_eval#share_mh share_gp any_anxiety#share_mh	only_adjustment#share_ap only_adjustment#share_gp only_adjustment#s1_by_psych	only_adjustment#share_gp only_adjustment#share_psych only_adjustment#s1_by_psych	any_depression#share_mh share_psych#share_gp any_multiple#share_mh
4. Lasso, Including Physician Supply Measures	any_multiple#share_mh any_multiple#s1_by_psych any_multiple#share_ap	first_eval#share_mh first_eval#share_gp anxiety#share_gp	only_adjustment#share_gp only_adjustment#share_ap	any_depression#share_psych any_depression#share_ap any_multiple#share_mh only_anxiety#share_gp only_adjustment#share_gp	any_depression#share_mh

NOTE.—This table presents the instruments selected by Lasso in first-stage models, where the instrument choice set varies across specifications. In rows 1-3, Lasso is presented with the full 116-element potential instrument vector discussed in the text. In row 4, the Lasso is presented with an alternative potential instrument vector, which includes measures of physician supply. Row 1 lists the instruments selected in the main analysis, corresponding to Tables 4 and 5. Row 2 lists the instruments selected when the sample excludes children with schizophrenia spectrum and other psychotic disorders, bipolar and related disorders, autism spectrum disorder, and Tourette's disorder, corresponding to Table 6. Row 3 lists the instruments selected when Lasso is constrained to select three instruments, corresponding to Appendix Table 4. Row 4 expands the potential instrument vector to include three additional physician supply measures--psychiatrists, therapists, and PCPs per 1,000 BCBS children, corresponding to Appendix Table 5. The symbol # indicates an interaction between a main instrument and indicators of patient diagnosis or severity. The main instruments are defined as follows: "Share_psych" and "share_gp" refer to the shares of psychiatrists and PCPs providing mental health treatment to children; "share_mh" refers to the average share of patients in PCPs caseloads receiving psychiatric drugs; "share_ap" refers to the share of patients with anxiety or depression who are prescribed anti-psychotics the share of first psychiatric drug prescriptions prescribed by psychiatrists, and "s1_by_psych" refers to the share of psychiatrists among initial psychiatric prescribers. Diagnosis indicators are defined as follows: having multiple diagnoses ("any_multiple"); a sole diagnosis of anxiety ("only_anxiety") or adjustment ("only_adjustment"); a diagnosis of any form of depression, with or without adjustment, ("any depression") or anxiety, with or without adjustment, ("any_anxiety"); or a formal diagnosis of anxiety ("anxiety"). Severity indicators are defined as follows: a first mental health episode occurring in the ER ("first_er") or during a formal mental health evaluation ("first_eval").

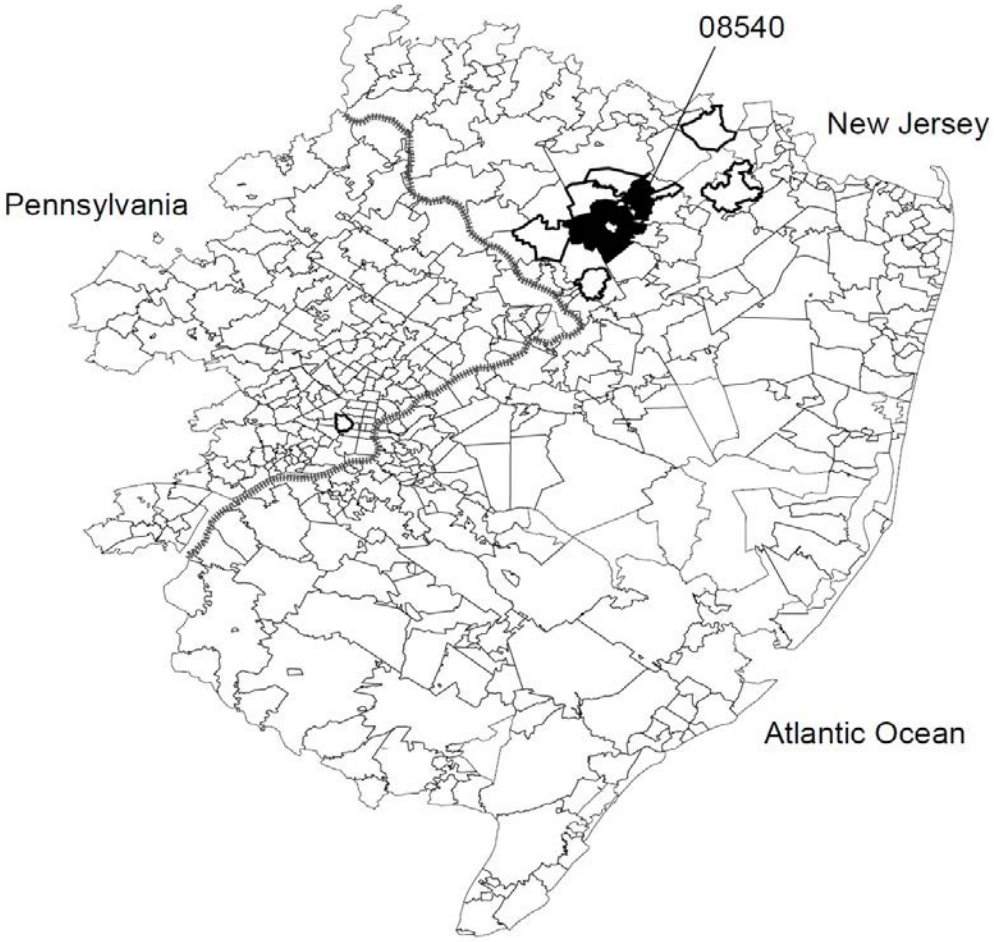
Figures

Figure 1
Prevalence of Depression, Anxiety, and Adjustment Disorders Among Sample



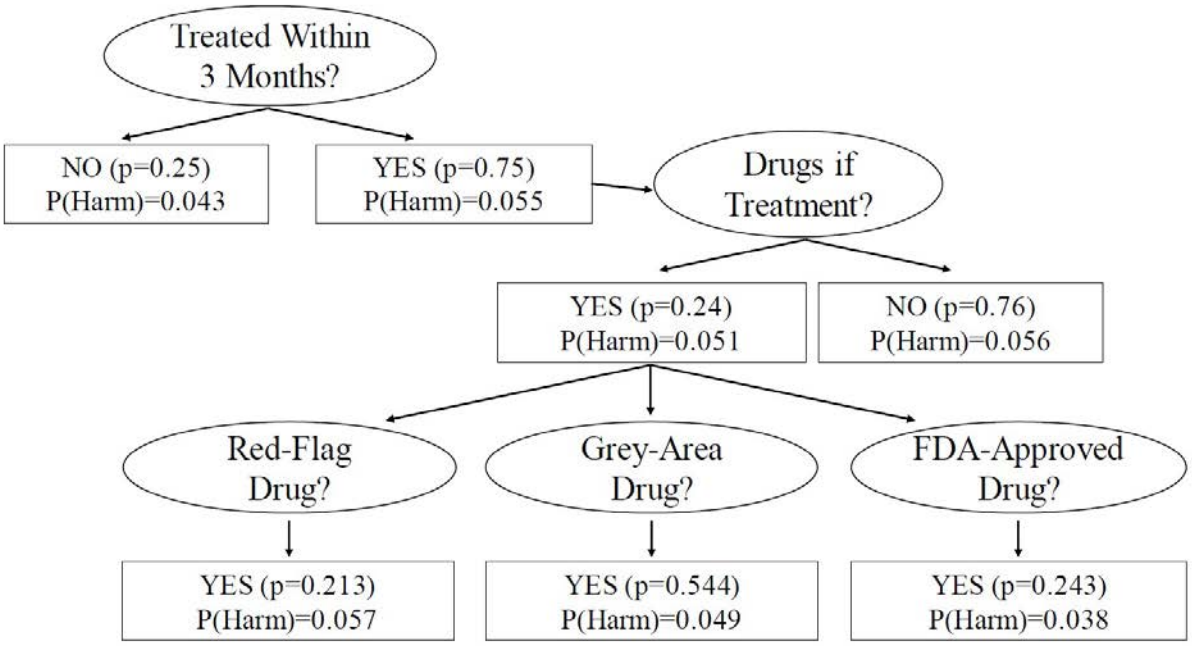
NOTE.—This figure shows the prevalence of depression (solid-line ellipse), anxiety (dotted-line ellipse), and adjustment disorders with features of depression (dash-dot ellipse) or anxiety (long-dashed ellipse) among the 45, 223 children in the sample. Intersecting ellipses correspond to sample children who have those corresponding diagnoses, where the number listed in each area represents the percentage of the total sample with that combination of diagnoses. Each diagnosis is defined according to the Diagnosis Manual of Mental Disorders (DSM-5).

Figure 2
Definition of the Market Area for a Specific Zip Code: Princeton, NJ 08540



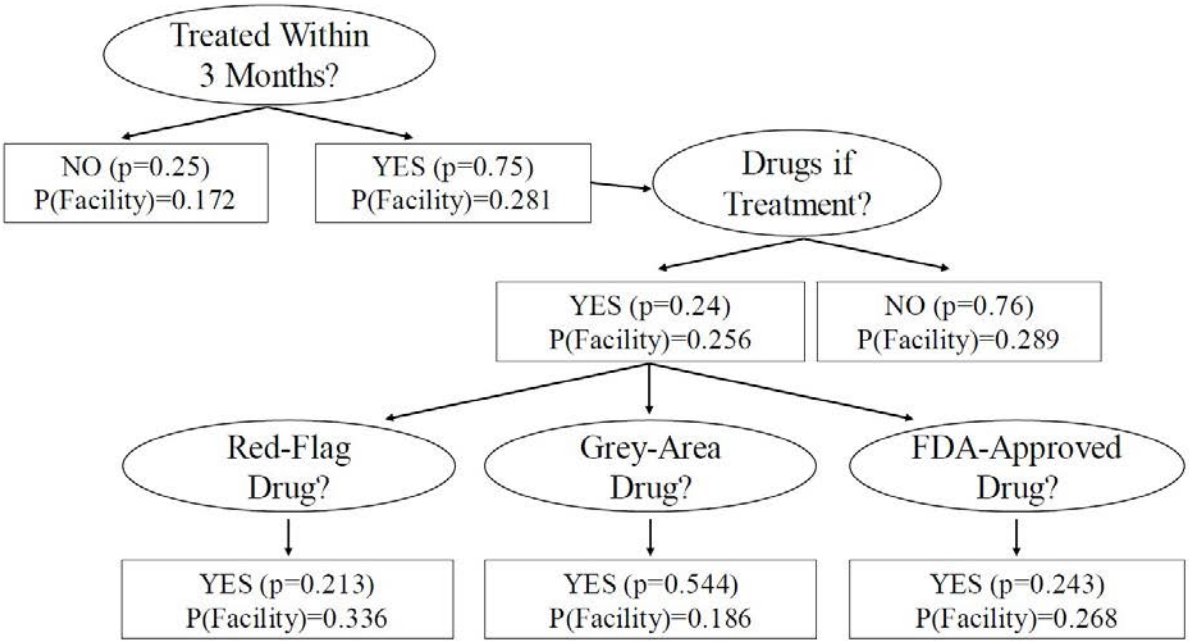
NOTE.—BCBS children residing in Princeton, NJ saw providers for mental health services in their own zip code (08540), in solid black, and also in the zip codes that are outlined in a thick black line. Hence, these zip codes together are considered the relevant market area for children in Princeton, NJ. Markets for all other zip codes are defined analogously so that providers in each zip code may form part of the market for children residing in multiple zip codes. We include up to 10 ZCTAs in each market, ranked in terms of the number of children who travel there (as in this example); in markets where children traveled to fewer than 10 ZCTAs, we only use those ZCTAs.

Figure 3
Estimated Effects of Treatment on the Probability of Self-Harm at 24 Months



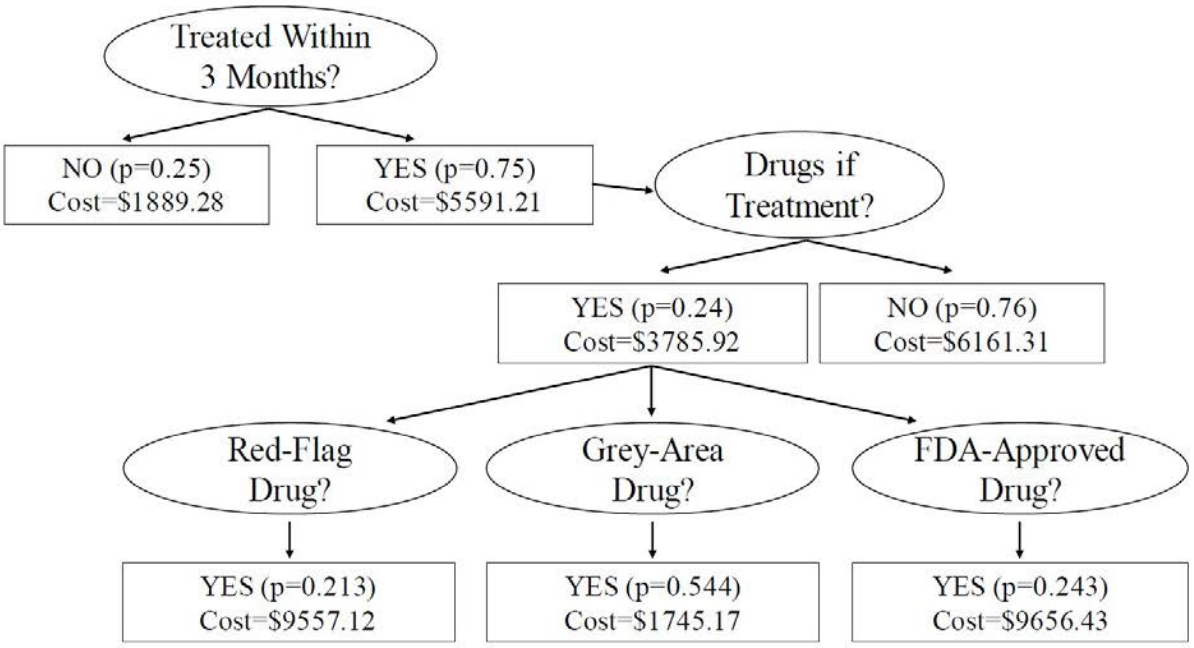
NOTE.—This figure traces out the implications of each initial treatment choice on the probability of self-harm at 24 months, where the share of children in each branch (p) appears in parentheses. Self-harm estimates are based on the post-Lasso 2SLS estimates shown in Table 4.

Figure 4
Estimated Effects of Treatment on the Probability of Facility Use at 24 Months



NOTE.—This figure traces out the implications of each initial treatment choice on the probability of facility use at 24 months, where the share of children in each branch (p) appears in parentheses. Facility use estimates are based on the post-Lasso 2SLS estimates shown in Table 4.

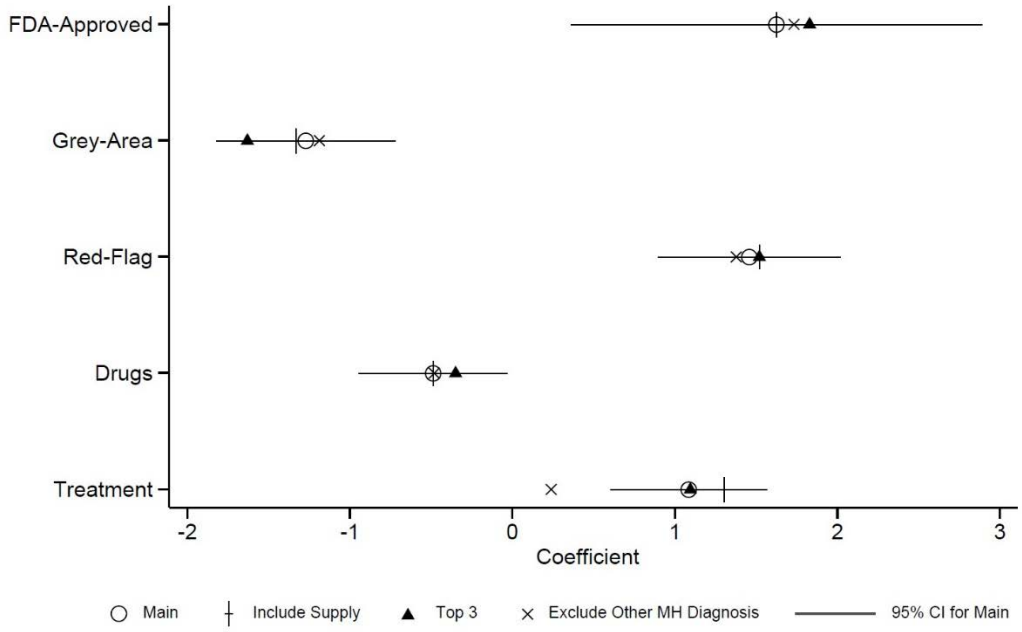
Figure 5
Estimated Effects of Treatment on Total Cost at 24 Months



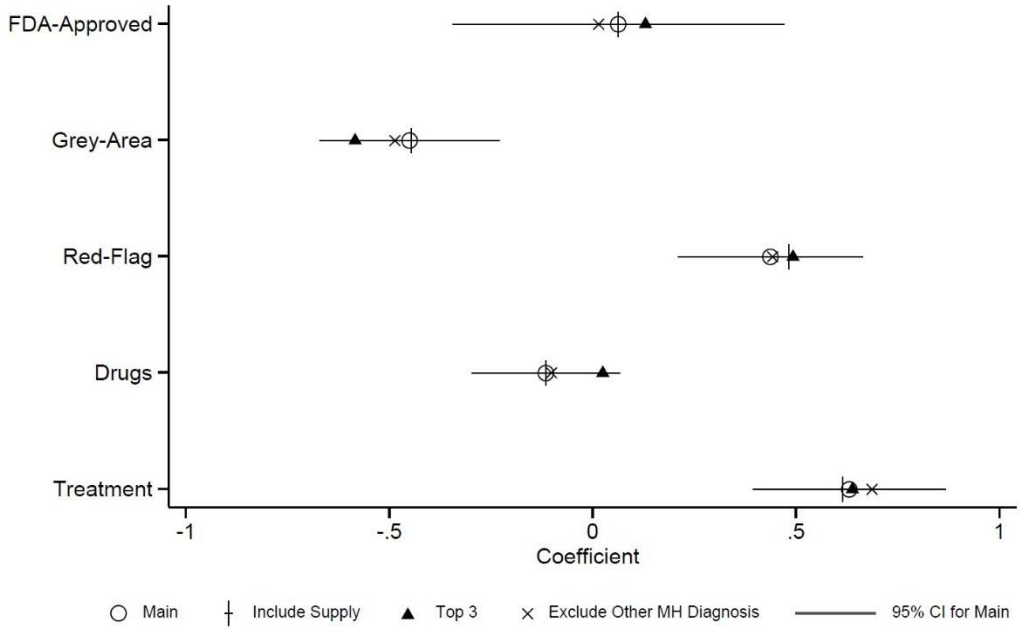
NOTE.—This figure traces out the implications of each initial treatment choice on log total costs at 24 months, where the share of children in each branch (p) appears in parentheses. Log total cost estimates are based on the post-Lasso 2SLS estimates shown in Table 4.

Figure 6
Summary of Estimated Effects of Treatment on Self-Harm, Facility Use, and Total Costs

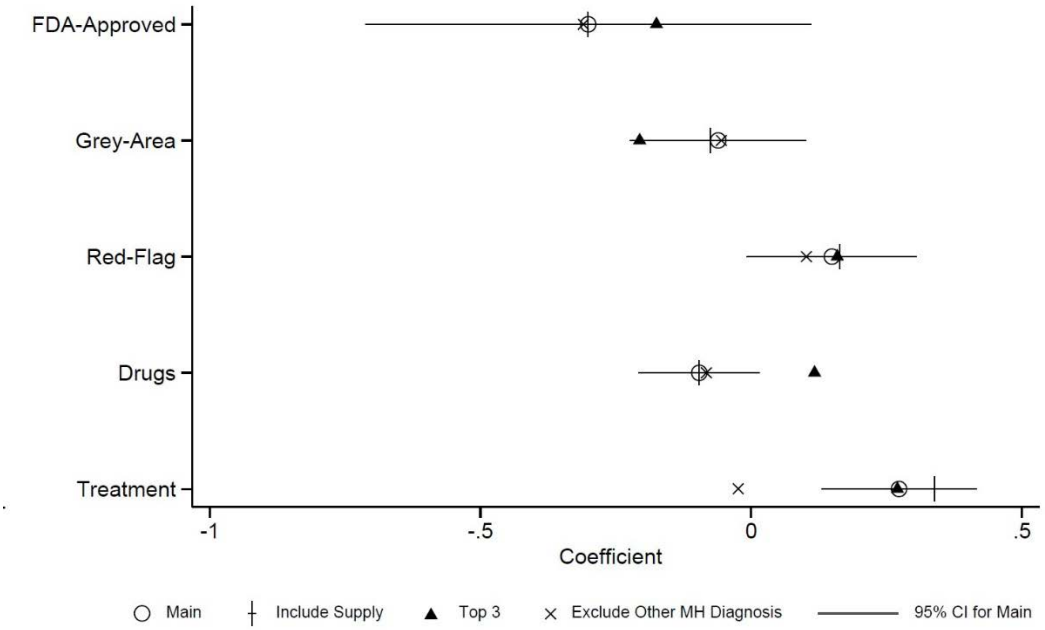
Panel A: Self Harm



Panel B: Facility Use

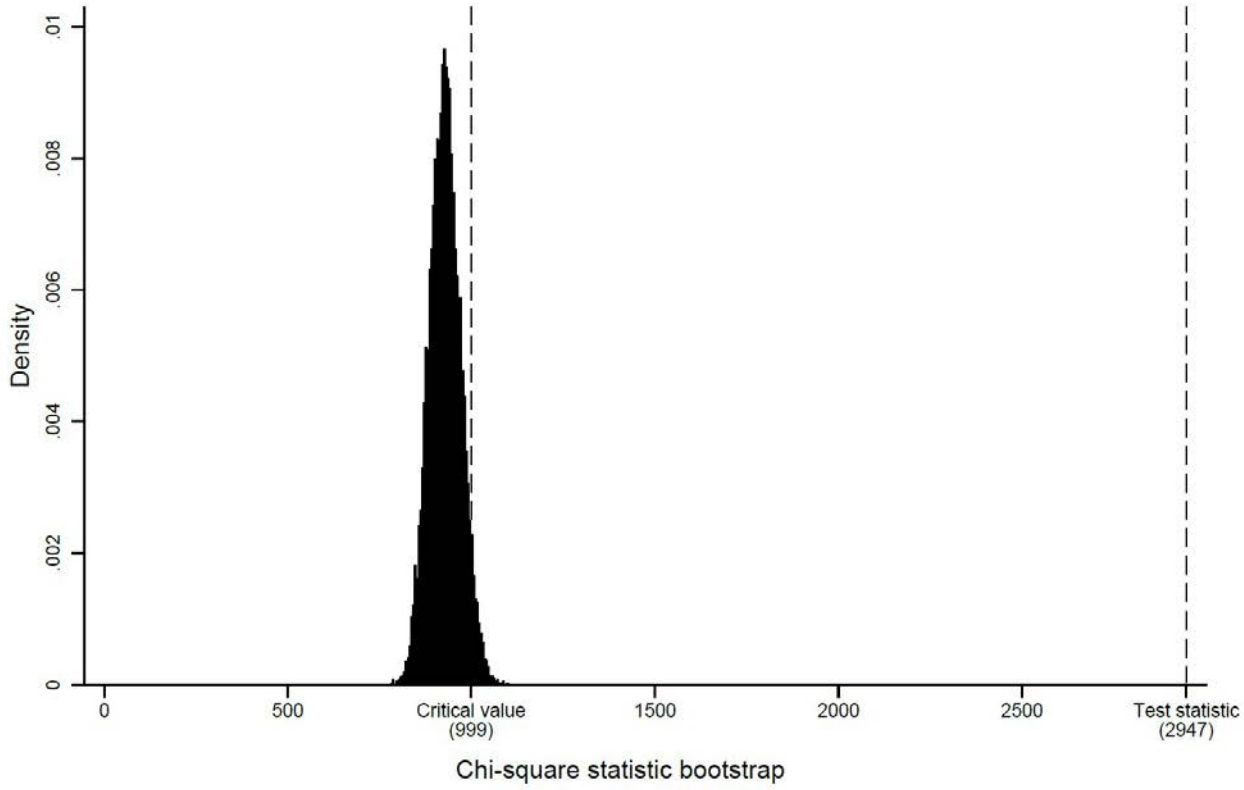


Panel C: Total Cost



NOTE.—This table presents results from the post-Lasso 2SLS models of mental health treatment on any self-harm behavior (panel A), facility use (panel B), and log total health care costs (panel C). "Main" (empty circle) represents the estimates from our preferred model (as shown in Table 4), with the full sample, 116-variable potential instrument vector, and Lasso procedure using the refined data-driven penalty described in Belloni et al. (2012). The "95% CI for Main" (thin black line) represents the 95% confidence interval of the "Main" effect coefficient. "Top 3" (triangle) represents the estimates from an alternative model (as shown in Appendix Table A4), where Lasso is constrained to select only three instruments. "Include Supply" (elongated cross sign) represents the estimates from an alternative model (as shown in Appendix Table A5), where a larger potential instrument set, i.e., provider supply-side variables (and their interactions with patient characteristics) is used in the Lasso procedure. "Exclude Other MH Diagnosis" (ex mark) represents the estimates from an alternative model (as shown in Table 6), where children with schizophrenia spectrum and other psychotic disorders, bipolar and related disorders, autism spectrum disorder, and Tourette's disorder are excluded from the sample.

Appendix Figure A1
Empirical Distribution of Bootstrap Chi-Square Values



NOTE.—The figure shows the empirical distribution of chi-square values from a bootstrap exercise, where we bootstrapped the treatment data, drawing children randomly from our sample to match only the number of children observed in each area (i.e., not the observed treatment pattern in each area). From the resulting 10,000 bootstrap samples, we constructed the empirical distribution of treatment—the share untreated, treated with therapy only, treated with red-flag drug, grey-area drug, or FDA-approved drugs—under the null hypothesis that the distribution of treatments is the same in all areas. The $\chi 2$ test statistic (2947) exceeds that of the 99th percentile of the empirical bootstrap distribution (999) generated by the random draws.