TI 2021-094/IV
Tinbergen Institute Discussion Paper

Are We Overdiagnosing Mental Illnesses? Evidence from Randomly Assigned Doctors

Marieke Bos[1]
Andrew Hertzberg[2]
Andres Lieverman[3]

[1] Stockholm School of Economics, Vrije Universiteit Amsterdam, Tinbergen Institute
[2] Federal Reserve Bank of Philadelphia
[1] Betterfly

Are We Overdiagnosing Mental Illnesses? Evidence from Randomly Assigned Doctors

Marieke Bos Andrew Hertzberg Andres Liberman*

This draft: August 2021
First draft: January 2020

Abstract

Almost two in 10 adults in the U.S. and Europe are, at any moment in time, diagnosed with a mental illness. This paper asks whether mental illness is over- (or under-) diagnosed, by looking at its causal effect on individuals at the margin of diagnosis. We follow all Swedish men born between 1971 and 1983 matched to administrative panel data on health, labor market, wealth and family outcomes to estimate the impact of a mental illness diagnosis on subsequent outcomes. Exploiting the random assignment of 18-year-old men to doctors during military conscription, we find that a mental illness diagnosis for people at the margin increases the future likelihood of death, hospital admittance, being sick from work, and unemployment, while lowering the probability of being married. Using a separate identification strategy, we measure the effect of military service on the same set of outcomes to rule out that the effect of diagnosis in our setting is primarily mediated by altering the probability of serving. Our findings are consistent with the potential over-diagnosis of mental illness.

Keywords: Mental Illness, Long-run Effect of a Diagnosis, Economics
JEL Classification Codes: D03, D12, I18, L51, L66

*Bos is at The Swedish House of Finance at the Stockholm School of Economics and a Visiting Scholar at the Federal Reserve Bank of Philadelphia, e-mail: marieke.bos@hhs.se, Hertzberg is at the research department of the Federal Reserve Bank of Philadelphia, e-mail: andrew.hertzberg@phil.frb.org, and Liberman is at Betterfly. We thank Anders Anderson, Laurent Bach, Johannes Haushofer, Christoph Kronenberg, Daniel Paravisini, Matthew Ridley, Frank Schilbach, James Vickery and the seminar participants of the 2020 NBER summer institute, Karolinska Institute Medical University, Virtual monthly Mental health Economics Seminar Series, HEC Paris, Philadelphia FED, ESSEC Paris for helpful comments. We also thank Johannes Haushofer for pointing out the availability of doctor IDs. Simon Almerström Przybyl, Jesper Böjeryd and Johan Orrenius provided excellent research assistance. Funding from VINNOVA is gratefully acknowledged.

1 Introduction

"When the doctor thinks he is diagnosing madness as a phenomenon of nature, it is the existence of this threshold that enables him to make such a judgement. Each culture has its own threshold, which evolves with the configuration of that culture."

— Foucault 1954, p.78

Motivation and Question

Mental illness diagnoses are prevalent around the developed world. For example, in 2018, 17.3% of adults in the European Union were diagnosed with a a mental health problem (OECD/EU (2018)). As an indicator within the U.S., 13.2% of persons age 18 and above surveyed during the period 2015 to 2018 reported taking a prescribed antidepressant medication in the previous 30 days (Brody and Gu (2020)). Given the prevalence of diagnosis, a natural question is whether mental illness is being over- or under-diagnosed. That is the question we address in this paper.

Conceptual Framework

Our answer to this question starts with the fact that a person's true underlying mental health exists on a continuous scale. As a result, under- or over-diagnosis is not a question of accuracy since, by necessity, diagnosis involves applying an arbitrary threshold to an assessment on this continuum of underlying health. Instead, the approach we take is to ask whether the threshold for diagnosis is being set too high or low. To make this assessment, we assume that the net benefit of diagnosis varies based on a person's underlying mental health. A person with poor underlying mental health is helped by a diagnosis because this gives them access to treatment. Conversely, a person with high underlying mental health is negatively affected by a diagnosis (owing, for example, to side effects of medication or the stigma of being diagnosed). Applying a standard principle of optimization, the question of under- or over-diagnosis hinges

upon whether a person who is diagnosed right at the threshold is made better or worse off as a causal result of the diagnosis. If she is made strictly better (worse) off, then the threshold of diagnosis is sub-optimally low (high) and mental illness is being under- (over-) diagnosed.

Challenges to Answering the Question

Providing an empirical answer to this question faces two key challenges that we address with our empirical strategy. First, it requires measuring the *causal* effect of diagnosis on the *marginal* (as opposed to *average*) patient. Second, it requires forming a sufficiently comprehensive welfare measure to assess whether the marginal patient is clearly better or worse off.

Challenge 1: Causal Effect of Marginal Diagnosis

To overcome the identification challenge, we implement an examiner-type empirical strategy that exploits the fact that between 1986 and 2006, Sweden had a mandatory military service conscription for all 18-year-old men. As a part of this program, all individuals were required to undergo an assessment of their mental health as part of determining their fitness to serve in the military. We exploit the quasi-random assignment of conscripts to doctors who vary in their tendency to diagnose a mental illness as an instrumental variable for the diagnosis. Following Dobbie, Goldin and Yang (2018), we measure the tendency to diagnose mental illnesses with a leave-out measure based on the diagnoses of all conscripts reviewed by each doctor at the same regional conscription office within the same year. This measure is highly predictive of the diagnosis of a mental illness at age 18 but is uncorrelated with other baseline characteristics, such as IQ test results and parents' education. The tendency of doctors to diagnose mental illnesses is also uncorrelated with other contemporaneous diagnoses, which validates its use as an instrumental variable for the diagnosis of

a mental illness.[1] This approach yields estimates of local average treatment effects (LATE) of diagnosis among compliers of the instrument; that is, individuals at the margin of diagnosis who are diagnosed only because they are assigned to a doctor who is more prone to diagnose a mental illness.

Challenge 2: Measure Welfare

We measure the welfare effects of diagnosis by linking the Swedish enlistment data to six administrative datasets that track health, labor market, and family outcomes for the universe of the Swedish population from 1960 to 2016. This enables us to provide a comprehensive account of the effect of diagnosis by using a wide range of outcomes measured over age horizons from 18 to 38 years of age.[2]

Results Summary

Overall, our results indicate that the diagnosis of a mental illness at age 18 lowers the welfare of the marginal patient. We start by measuring the effect on outcomes at age 30.[3] We find a marginally significant 0.7 percentage-point increase in death between age 18 and age 30. Diagnosis increases the likelihood of illness at age 30: a 20.4 percentage-point increase in the likelihood of being admitted to hospital as an outpatient and an additional 10.74 sick days from work (both results are significant at the 1% level). We also find a 15.3% increase in the likelihood of having a spell of unemployment at age 30. Diagnosis lowers the probability of being married by 10.1% at the same horizon. Other point estimates (suicide, divorce, wealth, homeownership,

[1]This research strategy is similar to that used by Kling (2006) to estimate the impact of incarceration in the United States; Chang and Schoar (2008), Dobbie and Song (2015), Dobbie, Goldsmith-Pinkham and Yang (2017) to estimate the propensity of judges to grant bankruptcy protection; Doyle (2007, 2008) to estimate the effects of foster care; and Liberman, Paravisini and Pathania (2016) to estimate the effect of high-cost debt on credit scores, among other applications.

[2]By contrast, the data available to previous researchers who study the economics of mental health is limited to a narrow range of outcome measures (e.g., suicide (Berndt et al. (2015) and Ludwig, Marcotte and Norberg (2009)) and emergency room visits Currie and MacLeod (2020)).

[3]We choose this age as a starting point because other research has shown a high correlation between outcomes at age 30 and subsequent lifetime labor-market outcomes (see Böhlmark and Lindquist (2006) for Sweden and Haider and Solon (2006) the US).

income) also indicate a welfare reduction but are not statistically significant. We then measure the effect of diagnosis on our set of outcomes at each age from 18 to 38. All statistically significant estimates indicate a reduction in welfare (increased morbidity, increased illness, higher unemployment, reduced likelihood of marriage, and an increased likelihood of divorce). While none of these results, in isolation, provides a definitive measure of welfare, in combination, they indicate that the marginal patient is made worse off by diagnosis.

Robustness of Empirical Strategy and Identifying Assumptions

In order to interpret our coefficients as local average treatment effects we make the identifying assumption of monotonicity. In our setting this translates into assuming that every conscript diagnosed as mentally ill by a strict doctor would also be diagnosed by a more lenient doctor *and* every conscript judged to be mentally well by a lenient doctor would also be judged well by a stricter doctor. We provide four pieces of evidence to support the assumption of monotonicity. First we take advantage of the fact that when a doctor diagnoses a mental illness, she also records an assessment of the severity of the illness. We show that most of the variation in doctors' tendency to diagnose comes from differences in the rate at which they diagnose less severe cases. In contrast there is little variation across doctors in the rate at which they diagnose severe mental illness. Second, and related, we show that doctors who have a higher tendency to diagnose have less severe diagnoses on average. Third, we show that the tendency to diagnose a mental illness is positively correlated across ex-ante observably different subsamples that vary on their propensity to be mentally ill within doctors. This suggests that differences across doctors in the propensity to diagnose does not come from doctors placing different weight on the range of information they use to diagnose each patient. Finally, we show that the contemporaneous diagnosis of a physical illnesses is uncorrelated with the tendency to diagnose a mental illness,

which is inconsistent with the hypothesis that doctors vary in their ability to diagnose *any* illness.

Ruling out Alternative Interpretations of Results

By design, the diagnosis of a mental illness reduces the chance of undertaking military service by approximately 30%. In turn, military service can potentially alter economic outcomes, for example by affecting individuals' education or human capital, or by creating personal networks.[4] We rule out the possibility that our effects are primarily mediated by the (avoidance of) military service by separately estimating the causal effect of military service on the full set of outcomes in our primary analysis. We do this by exploiting the fact that capacity constraints limit the yearly total military service enrollment, but the distribution of enrollment depends on a negotiation performed by military officers. As a result, the probability of enrollment of military service conscripts is correlated with the negotiation ability of their quasi-randomly assigned division officers. This strategy, which was first used by Hjalmarsson and Lindquist (2019), produces local average treatment effects of the military service on a population that differs from our baseline estimates of the effects of a diagnoses. In general, we can rule out effects of military service that represent more than an 5% of the baseline estimates of the causal effect of the diagnosis of mental illness with a 95% level of confidence. This finding is robust to specifications that estimate heterogeneous effects among individuals who go on to do the military service even after a diagnosis of a mental illness.

The fact that the military service interview is highly structured and is conducted only in order to determine eligibility for military service is a distinct advantage in isolating the causal effect of diagnosis. This highly structured setting limits the pos-

[4]See, e.g., Angrist (1990); Angrist and Chen (2011); Imbens and Klaauw (1995); Grenet, Hart and Roberts (2011); Bauer et al. (2012); Card and Cardoso (2012).

sibility that differences in doctor skill explain the variation in leniency. Furthermore, doctors in our setting do not undertake any treatment or prescribe any medication; instead, all diagnoses are referred to physicians outside the military. This allows us to focus purely on the effect of diagnosis independent from the effect of treatment skill, which is unobservable and is potentially correlated with the tendency of a doctor to diagnosis mental illness. As Currie and MacLeod (2020) stress, both of these factors will affect patient outcomes and are ambiguously correlated.

Mechanism

Three mechanisms could explain our findings. First, the diagnosis of a mental illness leads to prescription and treatment of the illness, which can itself have a direct causal effect. For example, a medical literature has demonstrated potential negative effects of the use of Prozac in certain populations, particularly adolescents (e.g., Hodes et al. (2009)). However, contrary to this explanation, we fail to find a statistically significant increased use of antidepressants before age 36, and at some horizons the point estimate is negative (although not statistically significant). A second possible mechanism is that diagnosis of mental illness may cause that individual to become aware of the fact that he is ill. This may in turn modify that individual's beliefs with changes in future behavior and outcomes. And third, the diagnosis may cause a stigma, either explicitly through medical notes becoming observable by future doctors or other agents, or implicitly through the patient's relatives and community. As yet, we are unable to provide clear evidence to support or reject these possible mechanisms.

Related Literature

Our paper fits in a literature that investigates the causal effects of diagnoses at the margin. We contribute to this literature by providing well-identified estimates of the effect of a diagnosis of depression at the margin, which is by itself a first-order health policy question. We innovate methodologically in this literature by linking

marginal estimates obtained from leave-out instrumental variables.

Paper Structure

The remainder of the paper is structured as follows. Section 2 provides a brief overview of the empirical setting, describes the data, and presents summary statistics. Section 3 describes our empirical strategy, and Section 4 presents the results. Section 5 describes the tests to disentangle the effects of military service and other heterogeneous tests, and Section 6 concludes. Additional tables and figures are available in an Online Appendix.

2 Setting, Data and Summary Statistics

In this section, we describe the Swedish military service conscription process, discuss the definition and diagnosis of mental illnesses, discuss our data sources, and present summary statistics.

2.1 The Swedish Military Service Enlistment Process

Our empirical strategy is based on the mandatory military enlistment process for Swedish males. Enlistment was instituted in Sweden in 1901, operated throughout the years of our sample period, and remains in place today. Under this program, all non-disabled Swedish males were required to report to an assigned regional test office shortly after turning 18 to determine if they would be required to serve in the Swedish military and, if so, in what service category. The specific reporting date and office was based only on the conscript's month and year of birth, municipality of residence at age 17, and, in some cases, the expected date of high school graduation.[5]

[5]A fraction of men deferred enlistment by a year. Between 1997 and 2001, the fraction who deferred draft by one year was 20%. While deferment is endogenous, it does not threaten our identification strategy because it occurs prior to being assigned to a doctor and hence affects the

The testing procedure typically took two days. Day one began with a meeting to inform conscripts about both their rights and obligations as well as the testing procedure. Each conscript then took part in a set of written tests measuring verbal, spatial, logical and technical ability and an a one-on-one telegraph test with one of several testers.[6] Each conscript carried a folder containing personal records through each of these steps. When a conscript completed these tests, he placed his folder in a box to wait for a medical examination with a doctor. Within each regional test office, multiple doctors were on duty to conduct these examinations. The assignment of a conscript to a doctor within the office was orthogonal to the attributes of the conscript and was simply determined by the timing of when conscripts had completed their other tests and placed their file in the box, in combination with which doctor happened to be free to pick up the next file. We confirm the orthogonality of conscript attributes to doctor assignment formally in Section 3.2.2.

2.2 The Diagnosis of a Mental Illness

The medical examination assessed the physical and mental health of each conscript in order to determine if he was fit to serve. Prior to the exam, measures of physical attributes, such as vision, strength, height, weight, and blood pressure of each conscript were recorded by a nurse and provided to the doctor. Based on our conversations with the conscription office, the average time spent in an examination was fifteen minutes with the doctor spending longer on conscripts who showed signs of illness based on initial diagnostics. The doctor did not have access to the conscript's prior medical history, although the conscript was permitted to bring a note or records from a personal physician. Each examination followed a detailed medical protocol that dictated

pool of conscripts seen by strict and lenient doctors equally

[6]Conscripts were asked to listen to a series of Morse code (dashes and dots) and then to repeat these back to the tester using a telegraph key.

the the methods and questions to be used in the exam, although the doctor had some discretion to ask follow-up questions based on the conscript's initial answers.[7]

Based on the conscript's answers, the doctor would record his diagnosis of any type of physical and/or mental illness(es) with a code from the International Classification of Diseases, Ninth Revision (ICD-9), combined with an assessment of the severity of each illness. All doctors used the same standardized system of diagnosis codes and severity scores. While each diagnosis code and associated severity score was not disclosed to the conscript, doctors are obliged to inform the conscript of any mental or physical illness diagnosis. In the case of a mental illness, the doctor would refer the conscript to see a psychiatrist outside the military. The assessing doctor did not prescribe any medication and did not schedule any follow-up appointments with the conscript. If a conscript was diagnosed with an illness but was ultimately selected to serve in the military, the doctor responsible for the conscript's assigned unit would be informed of the diagnosis.

The doctors who conducted these medical examinations were all fully accredited general practitioners. None of the doctors were psychiatrists. In Sweden, general practitioners are trained to diagnose both mental and physical illness, and their diagnosis is the typical initial step in a person seeing a mental health specialist for further diagnosis and treatment. Doctor pay is not tied to diagnosis or conscript health outcomes. Our conversations with representatives from the Swedish conscription process indicate that doctors are largely motivated by intrinsic incentives to perform their job well and the possibility that an obvious and ongoing history of inaccurate diagnoses would lead to losing the job and a loss of professional reputation within the close-knit Swedish medical community. The doctors do not track the health of the conscripts they examine over time and are not in any form of ongoing contact with them.

[7]The medical protocol is not publicly available.

On the second day of testing, each conscript met with a psychologist for an interview to assess his personality type. Finally, each conscript met individually with a test officiator who decided whether the conscript was fit to serve based on the previous evaluations, and if so, his assigned service category (i.e., medic, infantry soldier, tank driver, etc.). Several test officiators worked in each regional office, and conscripts were assigned based on the order their folder happened to be placed in a bin combined with which officiator happened to be free to pick up the next folder in the queue. In other words, assignment of conscripts to officiators was orthogonal to conscript attributes (a fact we will formally verify in Section 3.2.2). The officiator's decision as to whether a conscript was fit to serve was based mostly on the medical doctor's diagnosis of the type and severity of any mental and/or physical illness. Each man judged fit to serve left the center with an assigned service category and a starting date, expecting to serve. However in a typical year, many of the above-mentioned service categories were over-subscribed. To decide which recruits to release from service from within each service category, each officiator spoke for or against their own recruits in quarterly meetings across and within draft offices. Officiators vary in who is more (less) decisive and influential in these meetings than others, thereby placing a relatively high (low) share of their own recruits into actual service. Following the empirical strategy first proposed by Hjalmarsson and Lindquist (2019), we exploit the as-good-as-random assignment of conscripts to officiators as an exogenous source of variation in the likelihood that a conscript will serve in order to separately measure the effect of serving in the military on long-term conscript outcomes we study in the rest of the paper. In Section 5.1, we use this second source of exogenous variation to show that very little of the effect of diagnosis that we measure is induced by altering the chance that a conscript serves in the military.

2.3 Data Sources

Our main analysis data are an individual-level merge of military service enlistment outcomes to health, educational, marriage, and economic outcomes. Data for the Swedish enlistment process are maintained by the Swedish Military Archive for enlistments occurring before 1997 and by the Swedish Defense Recruitment Agency for 1997 and later. The data correspond to multiple cross-sections of 18-year-old individuals and include a conscription center identifier and identifiers for the doctor and psychologist assigned to conduct the assessment.

The data on the Swedish enlistment process is matched based on the conscripts' social security numbers to six distinct datasets.[8] First, the military service enlistment data are merged to a dataset from Statistics Sweden that contains information on the conscript's parents' education, the conscript's education, labor market outcomes including employment and wages, and wealth. Second, the data are merged to a dataset from the National Board of Health and Welfare, which includes individual-level diagnoses and prescriptions. Third, the data are merged to the National Death Registry, which includes dates and causes of death. Fourth, the data are merged to the Credit Bureau, which includes monthly snapshots of credit reports, arrears and delinquencies, and credit broken down by type (e.g., mortgages and consumer credit). Finally, the data are merged to a dataset from the National Enforcement Agency, which includes data on outstanding private and government claims.

[8]The merge process is conducted internally by the Swedish government agencies to preserve the confidentiality of the data.

2.4 Sample Selection

We make five sample restrictions to the data. First, we drop conscripts who lack information on their military service doctor.[9] Second, we we select individuals who were drafted between 1987, the year ICD-9 was implemented in the Swedish military service, and 2001, since the ratio of individuals diagnosed as mentally ill decreases sharply after that year owing to the implementation of a new decision process where a draftee is excused from further testing if he scored below the threshold of the previous test. We also drop observations from the 1997 draft, because many of the mental-illness diagnoses are missing from the data in that year. Third, we drop the relatively very small group of female draftees. Fourth, we drop individuals who have any of these four variables missing: education of parents, IQ test, height, and muscle strength. Missing observations on these variables are due to errors of transcription and uncorrelated with the baseline tendency to diagnose.

Fifth, we restrict our sample to include individuals assigned to doctors who see a minimum of 500 conscripts per year.[10] We also drop doctors who, after the 500 restriction is applied, are the only remaining doctor at a draft center in that year. After these restrictions, the analysis sample contains 457,037 unique individuals who are assessed by 130 doctors over 14 years.

2.5 Summary Statistics

Table 1 presents summary statistics of the main variables used in our analysis for the full analysis sample (Column 1), and broken down for individuals who were not diagnosed with a mental illness at military service (Column 2) and those who were

[9] The proportion of draftees who are not associated to a doctor ID is stable over time throughout our sample, at around 2%.

[10] We show in the Appendix that our results are robust to restricting our sample to doctors who see a minimum of 100 conscripts per year.

diagnosed (Column 3).

The proportion of individuals who are diagnosed with a mental illness at conscription is 3.15%.

While approximately 77.6% of all enlistees go on to do military service, this number drops to 37.6% for individuals who are diagnosed with a mental illness. This raises the possibility that the causal effect of diagnosis operates by altering the probability of serving. This motivates the additional analysis in Section 5.1, which rules this out as a primary explanation for our results.

We measure most outcomes at age 30, 12 years after conscription for most individuals, except for variables that are available on select years, which we measure on years that are as close as possible to age 30. This procedure allows us to measure outcomes for all cohorts in our sample. The general pattern that emerges from the table is that conscripts who are diagnosed with a mental illness go on to have worse outcomes late in life across a broad range of measures. They are more than twice as likely to be diagnosed with a mental illness at age 30, are twice as likely to have died and are almost twice as likely to have committed suicide by age 30. A similar pattern is obvious in the labor market, where those diagnosed obtain one fewer year of education, earn lower income from work, are more likely to receive government welfare, take more than twice the amount of sick leave, and are 59% more likely to be unemployed. They are also less likely to be married at age 30 while being 79% more likely to be divorced by the same age. These worse outcomes combine the causal effect of diagnosis along with the effect of the underlying mental health of the conscripts at the time of diagnosis. Our empirical strategy is designed to disentangle these two effects and focus only on the effect of diagnosis.

Table 2 presents the distribution of diagnoses of mental illness in our data, broken down for the entire sample and for severe and less severe diagnoses.

3 Research Design

In this section, we present the empirical strategy to measure the causal effect of a mental illness diagnosis for marginal patients using the Swedish military service evaluation process.

3.1 Preliminaries

Consider an empirical model that links future outcomes such as future diagnosis, morbidity, and earnings to a variable that indicates whether individual i was diagnosed with a mental illness at age 18,

$$Y_i = \alpha + \beta Diagnosis18_i + \omega X_i + \epsilon_i, \tag{1}$$

where $Diagnosis18_i$ is an indicator for diagnosis of a mental illness at age 18, X_i is a vector of predetermined characteristics, and ϵ_i is an error term. The coefficient of interest is β, which measures the average difference in outcome Y_i for individuals who are diagnosed and those who are not.

The fundamental identification problem of equation (1) is that individuals diagnosed with a mental illness at age 18 differ from those who are not diagnosed in ways that the econometrician cannot control for. For example, if a mental illness impairs individuals' cognitive faculties, the diagnosis of a mental illness would be associated with lower lifetime earnings irrespective of whether the diagnosis has a direct causal effect by itself. As a result, naïve OLS estimates of β will be biased.

To address this issue, we exploit variation in the diagnosis tendencies of quasi-randomly assigned doctors during military service enlistment as an instrumental variable for diagnosis. We use this instrumental variable to construct two-stage least-

squares (2SLS) estimates of the causal effect of diagnosis. In the 2SLS specifications, we interpret any difference in the outcomes of individuals assigned to doctors with a higher tendency to diagnose as the effect of the change in the probability of a diagnosis of mental illness scaled by the difference in the probability of diagnosis; that is, a standard Wald estimator.

As with any instrumental variable, we assume a first stage, an exclusion restriction, and monotonicity. Given these assumptions, the empirical design identifies a local average treatment effect (LATE), i.e., the causal effect of a mental illness diagnosis for individuals on the margin of being diagnosed. In the next subsections, we explain the construction of the instrumental variable and discuss the three identification assumptions.

3.2 Instrumental Variable Calculation

We construct our instrument using a residualized, leave-out measure of measure of a doctor's diagnosis tendency (e.g., Kling (2006), Aizer and Doyle (2015), Dobbie, Goldin and Yang (2018)). We account for two sources of non-random assignment in the construction of our instrument. First, individuals are selected to enlist in their local conscription centers. If geographic variation is correlated with the propensity to be diagnosed with a mental illness and with life outcomes, a comparison of enlistees in different conscription centers will conflate this selection effect with the causal effect of the diagnosis. Second, time trends in diagnosis tendencies and economic outcomes are likely to induce differential selection effects across cohorts.

In order to account for these two sources of potential selection bias, we first use enlistment center c by enlistment year t fixed effects to construct residuals of the diagnosis of a mental illness at age 18 at the individual level. Let the diagnosis of

mental illness for individual i, after removing the effect of enlistment center-by-year fixed effects X_{ct}, be denoted by:

$$Diagnosis^{\star}_{ict} = Diagnosis_{ict} - \gamma X_{ct}. \tag{2}$$

We define Z_{ict} as doctor j's tendency to diagnose a mental illness for individual i as:

$$Z_{ict} = \frac{\sum_{k \in N_{cjt}} Diagnosis^{\star}_{ict} - 1}{N_{cjt} - 1}, \tag{3}$$

where N_{cjt} is the total number of military service enlistees attended by doctor j in center c in year t.

Our sample includes 130 doctors. In any given year, there are a median of 25 doctors working in the six enlistment centers across Sweden. The median number of conscripts per doctor is 1,870.5 during the sample period. Figure 8 shows the distribution of the number of doctors in a given center by year. On average, each center by year has 7.34 doctors, while the median center by year has 5 doctors. We restrict our sample to include those doctors who see at least 500 individuals.[11]

Figure 1 presents a histogram of distribution of our residualized doctor leniency measure for mental illness diagnosis at the doctor-by-center-by-year level, winsorized between -4% and 4%. The histogram demonstrates a substantial degree of variation in doctors' tendency to diagnose a mental illness, in particular given that the average diagnosis rate is 3.2%.

For Z_{ict} to be a valid instrument of the diagnosis of mental illness, we assume, 1) that Z_{ict} predicts the individual-level diagnosis (a relevant first stage), 2) that Z_{ict} affects individual i's outcomes only through the diagnosis (an exclusion restriction),

[11] We have verified that our results are robust to alternative definitions of this cutoff, including 100 individuals and 1,000 individuals.

and 3) that a higher (lower) value of Z_{ict} never makes a diagnosis less (more) likely (monotonicity). Below, we examine each assumption.

3.2.1 First Stage

To examine the first-stage relationship between a doctor's tendency to diagnose mental illnesses and an individual's propensity to be diagnosed with a mental illness, we estimate the following equation for individual i, assigned to doctor d working at center c in year t using a linear probability model:

$$Diagnosis_{ict} = \pi Z_{ict} + \gamma X_{ct} + \epsilon_{ict}. \tag{4}$$

The coefficient of interest is π, which measures the change in the probability of a mental illness diagnosis from being assigned to a doctor with zero tendency to diagnose to a doctor with tendency equal to one. Figure 1 presents the first-stage relationship between the residualized measure of doctor's tendency to diagnose a mental illness and the residualized probability of diagnosis, accounting for center-by-year fixed effects. The figure shows a local linear regression of an indicator for diagnosis of mental illness at conscription on Z_{ict}, together with a 95% confidence interval. The figure suggests that assignment to a doctor with a higher tendency to diagnose a mental illness leads to a higher probability of diagnosis, consistent with a strong first stage.

Table 3, column 1 formalizes the intuition conveyed by Figure 1, presenting the results of regression model (4). The results suggest a strong first stage: assignment to a doctor with a an average Z_{ct} index that is 10 percentage points higher leads to a 0.9 percentage point higher probability of being diagnosed, a 28% increase over the mean diagnosis rate, a result that is statistically significant at the 1% level.

3.2.2 Exclusion Restriction

The second assumption for identification is that doctor assignment only affects outcomes through the diagnosis of a mental illness. There are two potential violations of this assumption. First, strict exogeneity would be violated if doctors' tendency to diagnose is correlated with unobservable determinants of future outcomes. To investigate this possibility, column 2 of Table 3 reports the results of a regression of Z_{ict} on variables that are observable for an individual at the time of military service conscription, controlling for center-by-year fixed effects. These variables are parents' education, IQ test results, height, and muscle strength. We find that doctors who vary on their tendency to diagnose a mental illness are assigned individuals who are observably equivalent, as noted by the p-value of 0.178 of the F-test of joint significance of these variables. Muscle strength loads negatively and is significant at the 5% level, but the coefficient is economically small.

A second violation of the exclusion restriction would arise if doctors with differing tendencies to diagnose a mental illness can affect outcomes through channels other than the actual diagnosis. We argue that although this assumption is fundamentally untestable, it is quite reasonable in our setting. Military service doctors meet with patients only once for a brief interview and with the explicitly stated objective of finding out whether they are fit for military service.

Given that doctors conduct a physical and a mental evaluation of the conscript, one possible alternative channel through which doctors can affect outcomes is through the diagnosis of a physical illness. That is, the exclusion restriction would be violated if doctors who are more likely to diagnose a mental illness were also more likely to diagnose a physical illness. We investigate this possibility by regressing the tendency to diagnose a mental illness on contemporaneous diagnosis of physical illnesses. The

results are presented in Column 3 of Table 3. Although the diagnoses of asthma and joint problems load significantly at the 10% level, the p-value of 0.254 implies that we cannot reject the hypothesis that the coefficients on all physical diagnoses are not significantly different from zero.

3.2.3 Monotonicity

The monotonicty assumption in our setting requires that every individual who is assigned to a doctor with a higher (lower) tendency to diagnose is always more (less) likely to be diagnosed. In general in the examiner design literature, this assumption is contentious because of the possibility that tendency to diagnose is correlated with skill. In that case, individuals who are mentally ill would be more likely to be diagnosed by a more skilled doctor, but those who are not ill would be less likely to be diagnosed, a clear violation of monotonicity. Hence, the 2SLS estimates would not be interpreted as local average treatment effects.

We provide several pieces of evidence to support the assumption of monotonicity. First, Figure 2 shows the histograms of measures of doctors' tendency to diagnose severe and less severe mental illnesses. The tendency to diagnose severe illnesses exhibits less dispersion and concentrates more than 70% of its mass at 0. This suggests that doctors are significantly more likely to coincide in the diagnosis of severe illnesses but may disagree in less severe illnesses. Moreover, it suggests that the marginal diagnosis is less severe, consistent with monotonicity.

Second, Figure **??** plots the relationship between a doctor's tendency to diagnose a mental illness in a given year and the mean severity of the diagnosis, conditional on diagnoses. The figure shows that doctors with a higher tendency to diagnose have slightly less severe diagnoses (note: the convention used in the Swedish military is that a higher Severity score corresponds to less severe diagnosis), although this

relationship is not statistically significant at conventional levels. Again, this suggests that doctors with different tendencies coincide on their assessment of more severe conscripts, and that the marginal individual is diagnosed with a less severe illness.

Third, Figure 3 plots the relationship between doctors' tendencies to diagnose a mental illness across subsamples split by the median IQ (top right), parents' education (top left), muscle strength (bottom right), and height (bottom-left). Although the actual probability of a mental illness is likely to vary across these subsamples, the mean tendency of a doctor to diagnose a mental illness in any given year is never negatively correlated across subsamples, which is consistent with the assumption of monotonicity.

4 Main Results

In this section, we present the main results of our identification strategy using the diagnosis tendencies of randomly assigned doctors and present robustness tests.

4.1 The Effect of Diagnosis of a Mental Illness at Age 18

We present the estimates from the 2SLS regression model (1) where the first stage is given by equation (4). The coefficients can be interpreted as the causal effect of a diagnosis of a mental illness at military service conscription on each outcome for individuals at the margin of a diagnosis. We start by focusing on outcomes for conscripts at age 30 (unless noted otherwise). We choose this age for two reasons. First, other research has shown a high correlation between outcomes at age 30 and subsequent lifetime labor-market outcomes (see Böhlmark and Lindquist (2006) for Sweden and Haider and Solon (2006) the US). Second, as a practical matter, this is the horizon at which we can measure all outcomes for every cohort of conscripts in

our study. After presenting the results for conscripts at age 30, we will estimate the effect on all outcomes at a wide range of ages.

4.1.1 Health Outcomes at Age 30

Table 4 shows that diagnosis during military service has a statistically and economically significant deleterious effect on health outcomes at age 30. Column 1 suggests that diagnosis increases morbidity between age 18 and 30 by 0.7 percentage point, although this is only marginally significant. As we show below, this effect is statistically significant when measured at other horizons. Column 2 indicates that we do not find a statistically significant effect on the probability of a completed suicide. The point estimate on the effect on suicide is approximately one-seventh the effect on morbidity, suggesting that the increase in the probability of dying does not come through suicide.

Diagnosis has a more pronounced effect on other health outcomes at age 30. As a result of diagnosis, conscripts are 20.4 percentage points more likely to be admitted to hospital as an outpatients (Column 3) and 3.9 percentage points more likely to be admitted as inpatients (although the effect on inpatient admission is only marginally significant). Both increases are economically large relative the unconditional incidence of each type of hospital admission. Column 5 shows that diagnosis is to increase the number of sick days taken at age 30 by 10.74 days.

The top panel in Figure 9 shows these estimates and associated confidence intervals expressed as a fraction of the standard deviation of each outcome among non-diagnosed draftees. Diagnosis increases the likelihood of outpatient admission by 48.4 percent and increases the number of sick days by 37.5 percent of the respective standard deviation for each outcome. Taken together, the results in Table 4 show that the marginal diagnosis harms the expected health of the conscript at age 30.

4.1.2 Labor Market and Household Outcomes at Age 30

Table 5 shows that the harmful effect of diagnosis during military service extends beyond health measures to a much broader set of life outcomes at age 30. Column 1 shows that diagnosis increases the likelihood of unemployment by 15.3 percentage points. This increase represents 48.8 percent of a standard deviation of unemployment among non-diagnosed draftees (see Figure 9). The point estimate of the effect on income from work and years of schooling (Columns 2 and 3) are both negative but statistically insignificant.

Columns 4 to 6 look at the effect of diagnosis on measures of household wealth. While the point estimates are are negative in each case, none of these estimates are statistically significant. This may reflect the operation of the robust social insurance system in Sweden, which limits the effect of poorer labor market outcomes on household wealth. It could also reflect the possibility that those diagnosed offset lower income with lower consumption, although we cannot assess this empirically. Certainly, there is no evidence that diagnosis *increases* the wealth of conscripts with marginal mental health.

Finally, Columns 7 and 8 of Table 5 measure the effect of diagnosis on family structure at age 30. Diagnosis lowers the likelihood of being married at age 30 and does not have a statistically significant effect on the likelihood of being divorced. The effect on the likelihood of being married represents a 25.7 percent reduction relative to the standard deviation of this outcome among non-diagnosed draftees (see Figure 9).

Taking the results of Table 4, Table 5 and Figure 9 together, all outcomes point toward diagnosis having a negative effect on the life of the marginal patient at age 30. It is always difficult to ensure that any set of outcomes fully captures a person's total

welfare. It is possible, for example, that while diagnosis causes people to have higher unemployment that they consume more leisure and are happier as a result. It is also theoretically possible that lower rates of marriage are welfare improving. However, it is hard to square these possibilities with other outcomes, such as the increased likelihood of death, hospital admission and sick days from work. We therefore conclude that all the outcomes taken in combination indicate that for the marginal patient, diagnosis at age 18 makes the marginal conscript worse off at age 30.

4.1.3 Outcomes at Other Ages

Having looked at a broad range of outcomes for conscripts at age 30, it is possible that diagnosis has beneficial effects at shorter or longer horizons. To the extent that our data allow, we address this possibility in Figures 11, Figure 12, and Figure 13, where we estimate the 2SLS estimates for the same full set of outcome measures in both Table 4 and Table 5 at each age for the first 20 years after diagnosis (ages 18 to 38). Figure 14 summarizes these results by reporting the sign (positive indicated by "+", negative indicated by "-") of all point estimates for which zero lies outside of the associated 95-percent confidence interval. Statistically insignificant estimates are left blank. Across all health outcomes, the only statistically significant point estimates at any horizon for each variable indicate a deleterious effect of diagnosis. The elevated incidence of morbidity as a result of diagnosis is statistically significant at age horizons of 23, 25 to 28, and 38. The increased incidence of sick days is statistically significant at 11 of the 21 age horizons for which we have data, and admission as an outpatient is significant at 9 different age horizons. In total, across all five health measures, a harmful statistically significant effect is present in 30 age-outcome regressions out of a possible 105. In contrast, there is no health outcome for which there is a statistically significant health improvement as a result of diagnosis.

The results are similar when looking at labor market, wealth and family structure outcomes. Diagnosis has a statistically significant increase on unemployment from age 28 to 38, lowers the likelihood of being married from age 29 to 38 and increases the incidence of divorce at age 22 and 23. The effect on stock-market participation, an indirect measure of wealth, is mixed: positive and statistically significant at ages 25 and 26 and negative and statistically significant from ages 35 to 37. If we interpret increased stock market participation at ages 25 and 26, at the same ages diagnosis increases morbidity, increases the likelihood of hospital admission and increases the number of sick days. It is implausible to argue that diagnosis increases welfare at this horizon. Across all the non-health outcomes there are 26 statistically significant estimates out of a possible 168 indicating a negative effect and only two estimates indicating a benefit.

In combination, looking across all outcomes and all horizons in the twenty years after diagnosis, the consistent theme is that for the marginal patient, diagnosis is welfare reducing. We therefore conclude that within the military, mental illness is over diagnosed for 18 year old Swedish military service conscripts.

5 Mechanisms

5.1 The Effect of Serving in the Military

Since the diagnosis of mental illness lowers the probability that a draftee serves in the military (see Figure 7), the effect of a diagnosis could simply come from the effect of military service on outcomes later in life. To rule this out as the primary explanation of our results, we use an empirical strategy based on the approach first employed by Hjalmarsson and Lindquist (2019) to separately measure the effect of

military service on the same outcomes measured in our main results. Their method instruments military service using the random assignment to a test officiator who influences if a draftee serves in the military.

5.1.1 Empirical Implementation

We measure officiator leniency in the same way we measured doctor leniency in our primary results. Center-by-year fixed effects are removed in equation 5

$$Service^{\star}_{ict} = Service_{ict} - \gamma X_{ct}. \tag{5}$$

We define Λ_{ict} as test officiator $j's$ tendency to assign individual i to a service as:

$$\Lambda_{ict} = \frac{\sum_{k \in N_{cjt}} Service^{\star}_{ict} - 1}{N_{cjt} - 1}, \tag{6}$$

where N_{cjt} is the total number of military service enlistees attended by officiator j in center c in year t.

The first stage then becomes

$$Service_{ict} = \pi \Lambda_{ict} + \gamma X_{ct} + \epsilon_{ict}. \tag{7}$$

5.1.2 Data

The situation exploited in Hjalmarsson and Lindquist (2019) only stretches from 1990-1996, and therefore, that time span is an restriction for the analysis of the effect of military service. To follow Hjalmarsson and Lindquist (2019), we drop observations that have no county during the draft. We also drop individuals who see an officiator who sees fewer than 100 draftees that year or do not have an officiator in the registry. Further individuals who do not have an end year of their service or finish their military

service after 23 are dropped. Finally, we only analyze those individually initially assigned a service category.

5.1.3 The Effect on Further Medical Treatment

Column 1 of Table 6 shows that a diagnosis of mental illness at age 18 causes a 1.9 percentage point increase in the probability of a subsequent diagnosis by age 30. This number implies a total rate of diagnosis that more than doubles the unconditional probability of diagnosis for the entire sample. Columns 2 and 3 show that this pattern extends to diagnosis of depression at age 30, as a subset of all mental illness diagnoses, and the prescription of antidepressants by age 35.

6 Conclusion

This paper asks whether mental illness is over- (or under-) diagnosed, by looking at its causal effect on individuals at the margin of diagnosis. We follow all Swedish men born between 1971 and 1983 matched to administrative panel data on health, labor market, wealth and family outcomes to estimate the impact of a mental illness diagnosis on subsequent outcomes. Exploiting the random assignment of 18-year-old men to doctors during military conscription, we find that a mental illness diagnosis for people at the margin increases the future likelihood of death, hospital admittance, being sick from work, and unemployment, while lowering the probability of being married. We find a similar pattern of negative effects at age horizons from age 18 to 38 and therefore conclude that for the marginal patient, a diagnosis of mental illness is harmful and is therefore over-diagnosed. Using a separate identification strategy, we measure the effect of military service on the same set of outcomes to rule out that the effect of diagnosis in our setting is primarily mediated by altering

the probability of serving. The fraction of men diagnosed with a mental illness in the military conscription process is lower than the fraction of the population with a diagnosis in the general population in Sweden, Europe and the U.S. Therefore, our findings are consistent with potentially widespread over-diagnosis of mental illness.

References

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

Angrist, Joshua D. 1990. "Lifetime earnings and the Vietnam era draft lottery: evidence from Social Security administrative records." *American Economic Review*, 313–336.

Angrist, Joshua D, and Stacey H Chen. 2011. "Schooling and the Vietnam-era GI Bill: Evidence from the draft lottery." *American Economic Journal: Applied Economics*, 3(2): 96–118.

Bauer, Thomas K, Stefan Bender, Alfredo R Paloyo, and Christoph M Schmidt. 2012. "Evaluating the labor-market effects of compulsory military service." *European Economic Review*, 56(4): 814–829.

Berndt, Ernst R, Robert S Gibbons, Anton Kolotilin, and Anna Levine Taub. 2015. "The heterogeneity of concentrated prescribing behavior: Theory and evidence from antipsychotics." *Journal of Health Economics*, 40: 26–39.

Böhlmark, Anders, and Matthew J Lindquist. 2006. "Life-cycle variations in the association between current and lifetime income: Replication and extension for Sweden." *Journal of Labor Economics*, 24(4): 879–896.

Brody, Debra J., and Qiuping Gu. 2020. "Antidepressant use among adults: United States, 2015-2018." *NCHS Data Brief*, 377.

Card, David, and Ana Rute Cardoso. 2012. "Can compulsory military service raise civilian wages? Evidence from the peacetime draft in Portugal." *American Economic Journal: Applied Economics*, 4(4): 57–93.

Chang, Tom, and Antoinette Schoar. 2008. "Judge Specific Differences in Chapter 11 and Firm Outcomes." *Unpublished Working Paper*.

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

Dobbie, Will, and Jae Song. 2015. "Debt Relief and Debtor Outcomes: Measuring the Effects of Consumer Bankruptcy Protection." *American Economic Review*, 105(3): 1272–1311.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review*, 108(2): 201–40.

Dobbie, Will, Paul Goldsmith-Pinkham, and Crystal S. Yang. 2017. "Consumer Bankruptcy and Financial Health." *Review of Economics and Statistics*, 99(5): 853–869.

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

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

Grenet, Julien, Robert A Hart, and J Elizabeth Roberts. 2011. "Above and beyond the call. Long-term real earnings effects of British male military conscription in the post-war years." *Labour Economics*, 18(2): 194–204.

Haider, Steven, and Gary Solon. 2006. "Life-cycle variation in the association between current and lifetime earnings." *American Economic Review*, 96(4): 1308–1320.

Hjalmarsson, Randi, and Matthew J Lindquist. 2019. "The Causal Effect of Military Conscription on Crime." *Economic Journal*, 129(622): 2522–2562.

Hodes, Georgia E, Lillian Yang, Jay Van Kooy, Jessica Santollo, and Tracey J Shors. 2009. "Prozac during puberty: distinctive effects on neurogenesis as a function of age and sex." *Neuroscience*, 163(2): 609–617.

Imbens, Guido, and Wilbert Van Der Klaauw. 1995. "Evaluating the Cost of Conscription in the Netherlands." *Journal of Business & Economic Statistics*, 13(2): 207–215.

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

Liberman, Andres, Daniel Paravisini, and Vikram Pathania. 2016. "High-Cost Debt and Perceived Creditworthiness: Evidence from the U.K." *Unpublished Working Paper.*

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

OECD/EU. 2018. *Health at a Glance: Europe 2018.*

7 Figures and Tables

Figure 1: First Stage for All Draftees

This figure reports the first-stage relationship between draftee outcomes and the measure of doctors' tendency to diagnose a mental illness. Doctor tendency to diagnose a mental illness is estimated using data from other draftees assigned to a doctor following the procedure described in Section 3. The solid line represents a local linear regression of an indicator for a mental illness diagnosis on doctor's tendency to diagnose. The indicator for a mental illness diagnosis and doctor's tendencies to diagnose are residualized using center-by-year fixed effects. Standard errors are clustered at the doctor level. The back plot shows a histogram of the distribution of the tendency to diagnose a mental illness, with equal weight on each doctor.

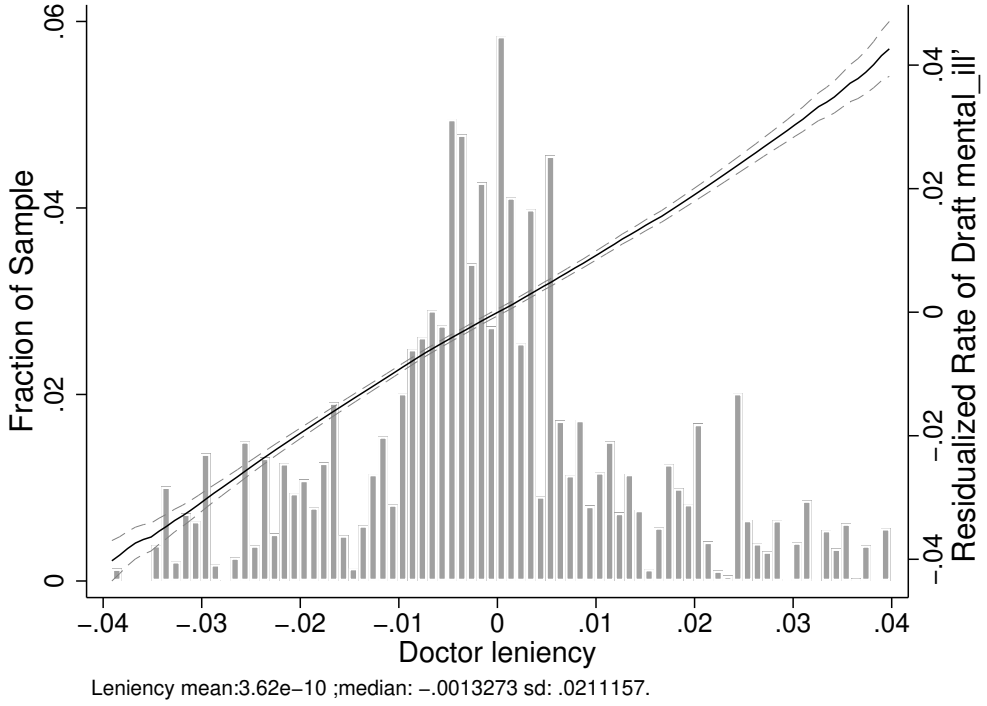
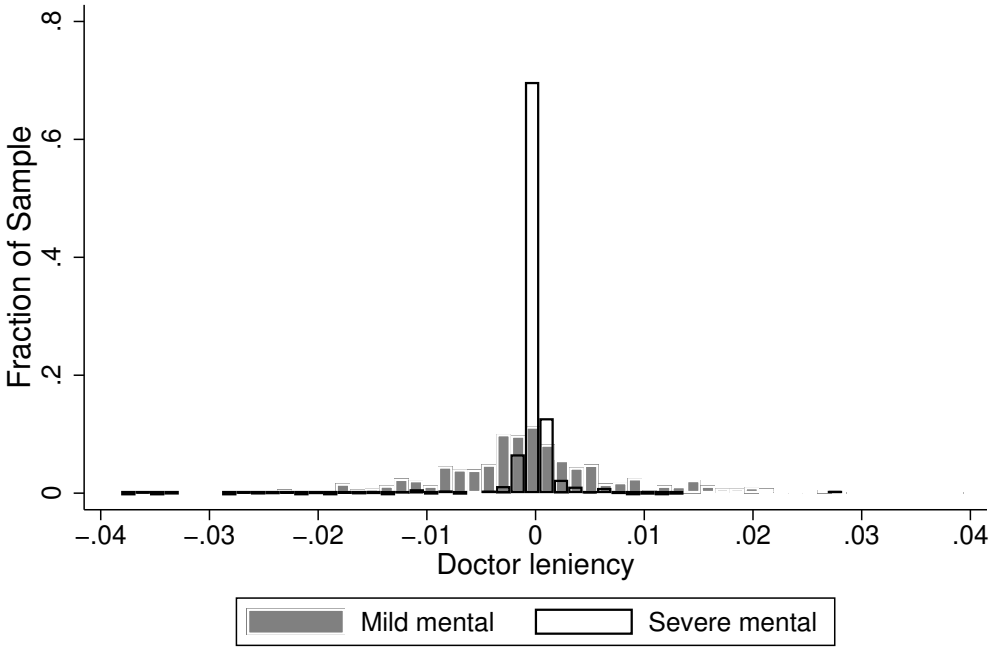


Figure 2: Monotonicity 1, Tendency To Diagnose A Mental Illness by Severity of the Diagnosis

This figure illustrates how the variation of the measure of tendency to diagnose varies by severity of the diagnosis. The figure superimposes the histograms of tendency to diagnose calculated with severe (severity 10) and mild diagnoses (severity 1-3). Severity is defined according to the military severity system explained in the text above.

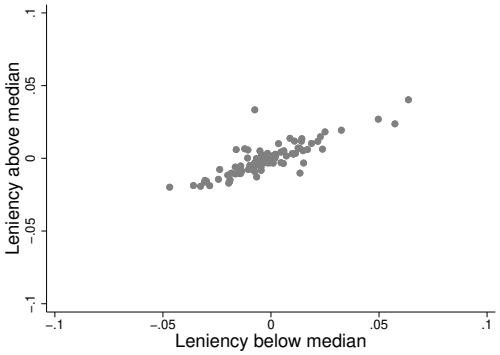


Leniency mild mean:−3.83e−11 ;median: −.0006612 sd: .0087113.
Leniency severe mean:3.03e−11 ;median: −7.37e−17 sd: .0081711.

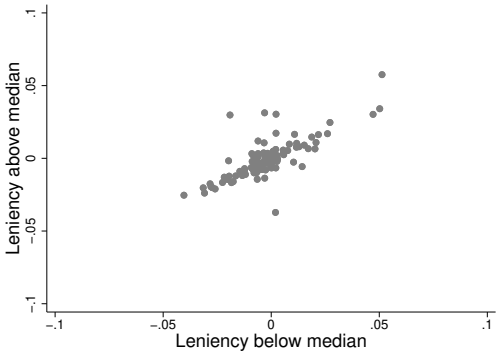
Figure 3: Monotonicity 3, Correlation of Tendency to Diagnose a Mental Illness Across Sub-Samples within Doctor by Year

These figures plot mean doctors' tendency to diagnose across subsamples, within doctor by year splits. The sample is split above and below the median of each characteristic, from top to bottom and left to right: IQ, education of parents, muscle strength, and height. The tendency to diagnose is computed for each subsample. The mean tendency of doctors in each subsample is plotted against the opposite subsample for each characteristic.

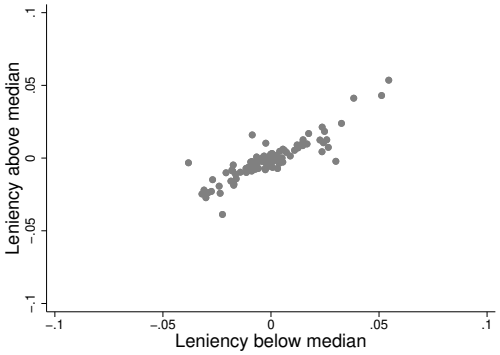
(a) IQ



(b) Education



(c) Muscle strength



(d) Height

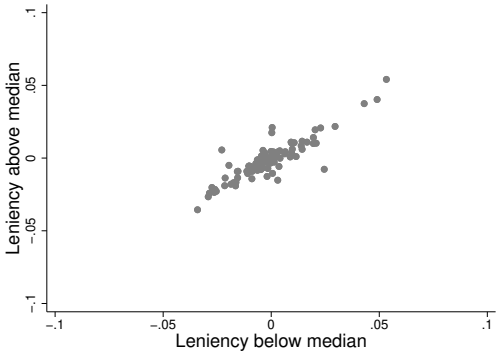


Figure 4: Monotonicity 4, First-Stage Estimate for Different Levels of Severity

This figure shows the estimates from the first-stage equation 4, calculated for different levels of severity of mental illness. The severity measure is the military's evaluation of how the condition will reduce the draftees' ability to serve in the military.

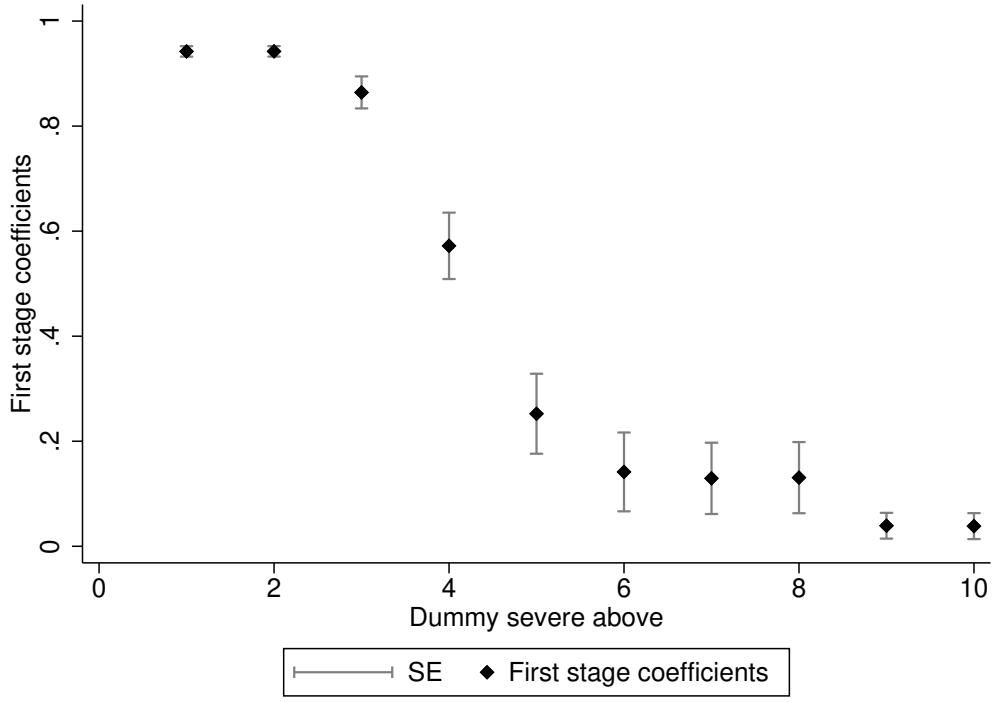


Figure 5: Monotonicity 5, Histograms for Quantities and Median Split of Leniency

The graphs show the distribution of gross income and income from work split by the median and quantiles of the distribution of doctor's tendency to diagnose a mental illness. We winsorize at the 1st and 99th percentiles.

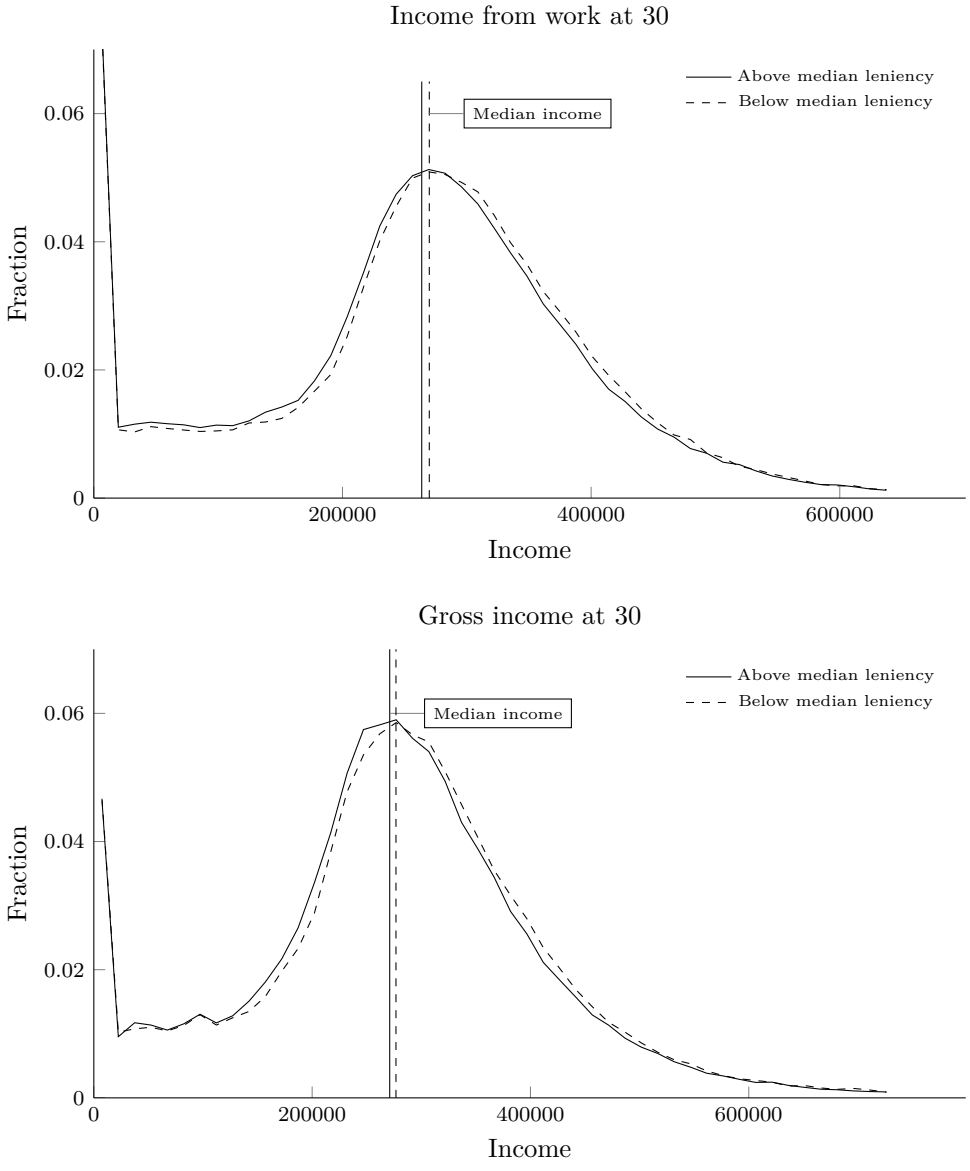


Figure 6: Evolution of the Testing Procedure

The graph plots the share of 18-year-old males who were participating in the military conscription over time. We plot the share of 18-year-old who; were tested, assigned to serve and actually served in the military. In 1995 the agency responsible for the military conscription in Sweden was changed "Värnpliktverket" to "Pliktverket", which had implications for the assignment of the reserve service categories. Before 1997 the data is delivered by the Swedish National Archive: after 1997, they are delivered by the Swedish Defence recruitment agency "Rekryteringsmyndigheten." The reason that the data for assigned and served coincide up until 1987 is data restrictions on the part of who did service, so it is set to be those assigned service.

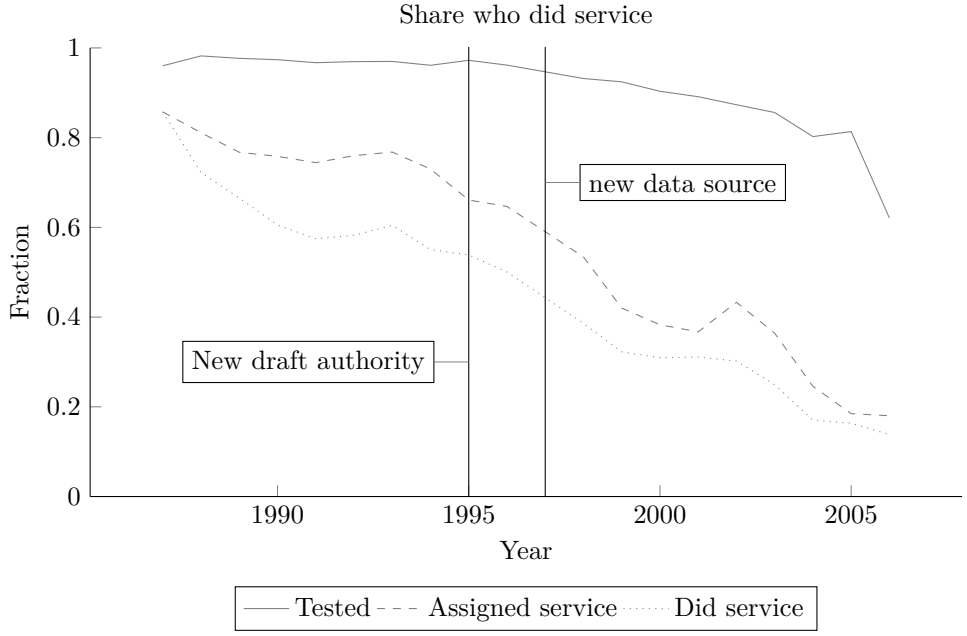


Figure 7: Correlation Between Severity and the Likelihood to Serve in the Military

Severity is a measure of your ability to serve in the military, this is only an empirical check of that statement.

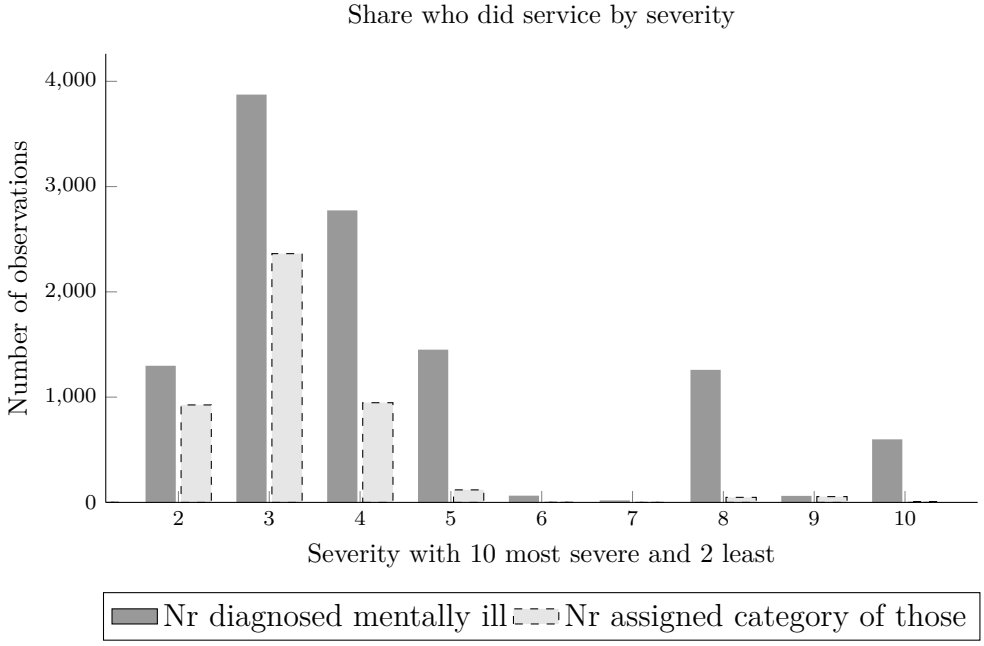


Figure 8: Distribution of Draftees Over Doctors Over Time

The number of doctors who see at least 500 patients per year is varying over time, but large enough to include all the years 1997-2001 in the sample.

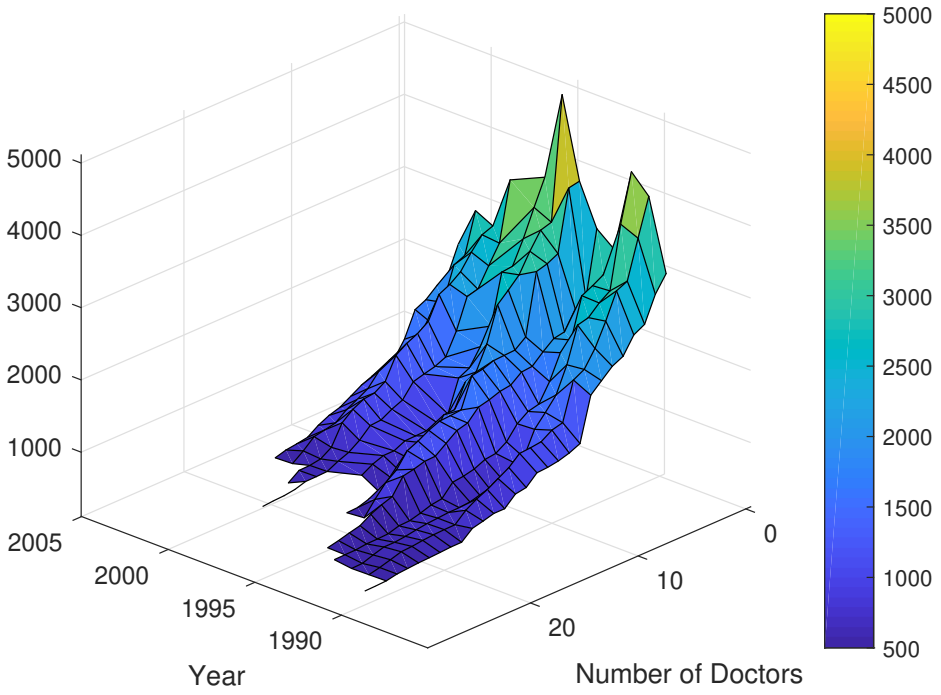


Figure 9: Effect of Diagnosis on Outcomes at Age 30 Relative to Standard Deviation of Outcome Among Non-Diagnosed

This figure reports 2SLS estimates of the effect of a mental illness diagnosis during the mandatory draft in Sweden on subsequent outcomes. We instrument for mental illness using the leave-out leniency measure constructed using the procedure described in Section 3. All specifications control for center-by-year of enrollment fixed effects. Standard errors are clustered at the doctor level. The outcomes are measured at a fixed age, specified below for each variable. The scale used is: *100 * coefficient / standard deviation for non-diagnosed*. The horizontal lines around the point estimates indicate 95% confidence intervals.

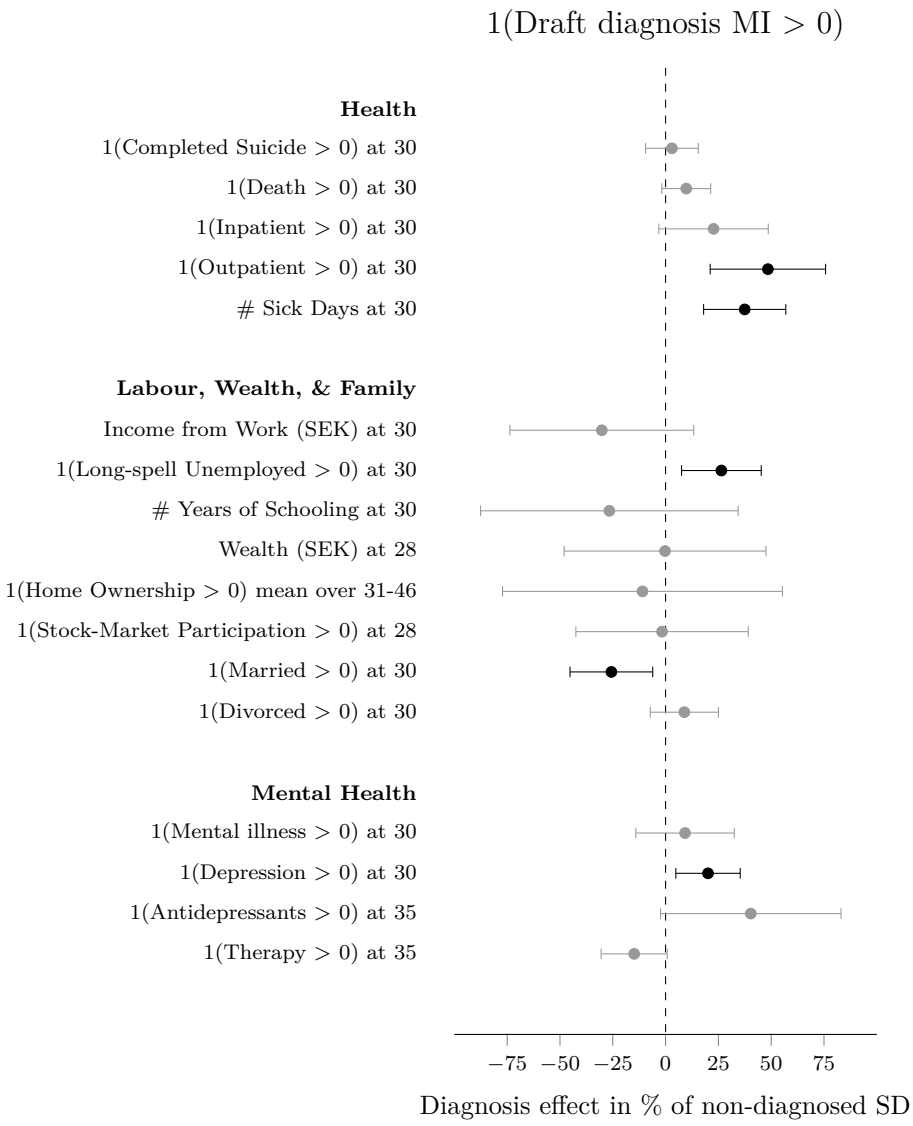


Figure 10: Effect of Serving in the Military

This figure reports 2SLS estimates of the effect of participating in military service on subsequent outcomes. All specifications control for center-by-year of enrollment fixed effects. Standard errors are clustered at the doctor level. The outcomes are measured at a fixed age, specified below for each variable. The scale used is: *100 * coefficient / standard deviation for non-diagnosed*. The horizontal lines around the point estimates indicate 95% confidence intervals. Black and grey indicate the draft diagnosis effect from Figure 9 while red is obtained by regressing each outcome on military service participation. These military service results (coefficients and confidence intervals) have then been further scaled by the probability of serving in the military conditional on receiving a mental illness diagnosis during the draft (38.5%).

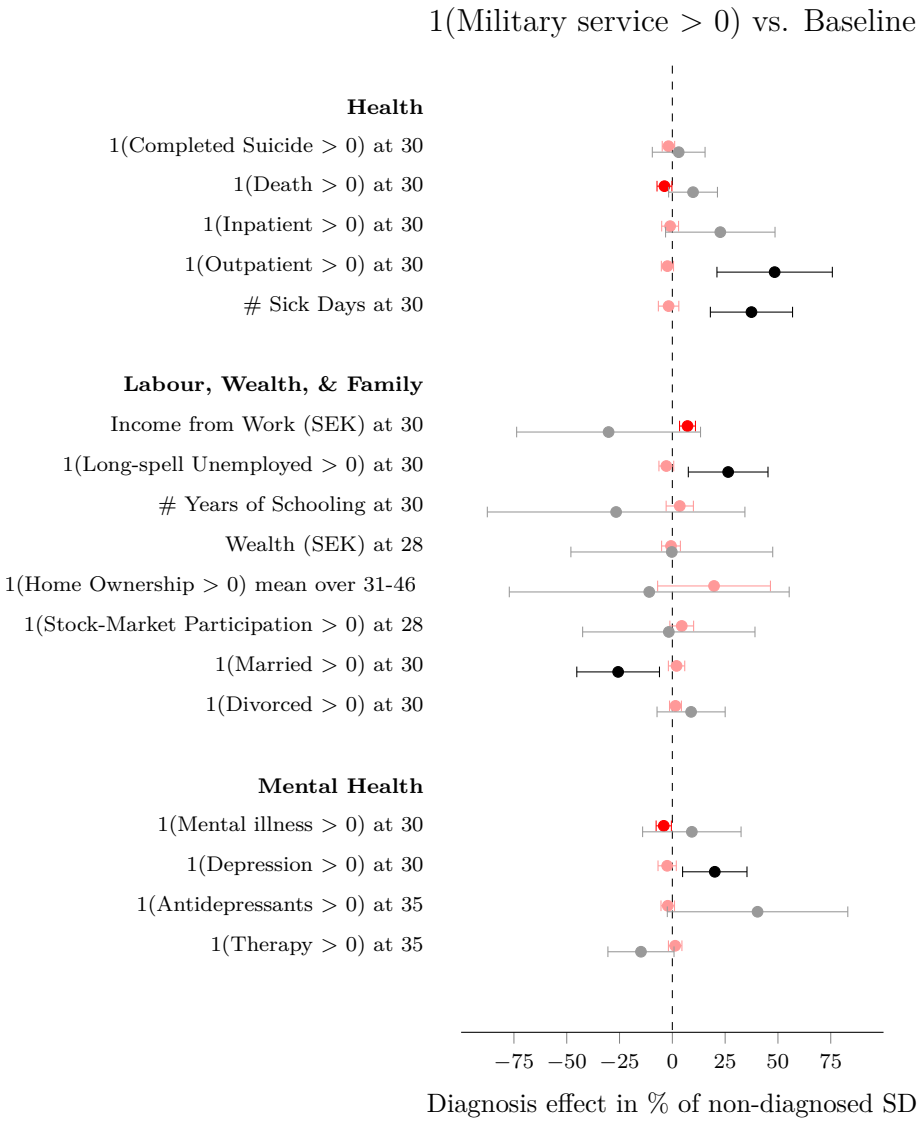
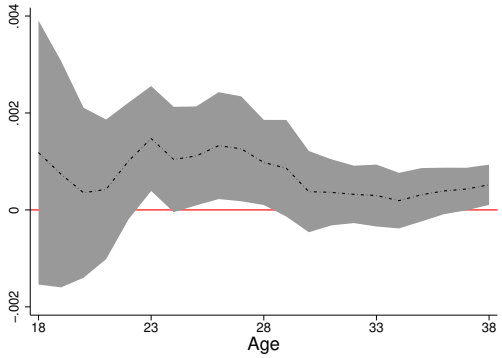


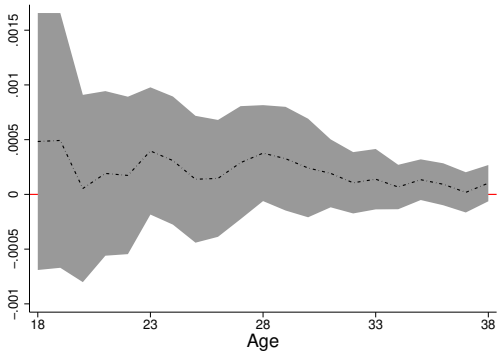
Figure 11: Event-time Evolution of 2SLS Estimates for the Effect of Mental Illness Diagnosis on Future Health

Note: Each of these panels is the graphical representation of 20 different 2SLS estimations that correspond to the regressions presented in Table 4, but here, the age of the conscript at the time the outcome is measured varies from 18 to 38. The solid line shows the point estimate at each horizon, and the associated 95 percent confidence interval is shown surrounding it.

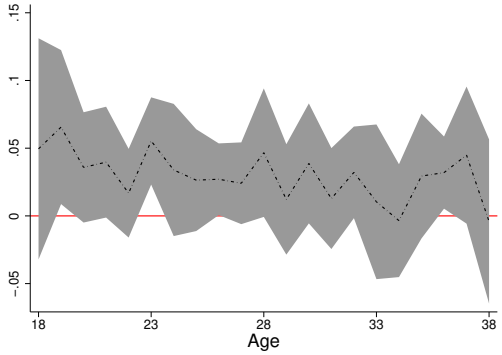
(a) Death (at each age)



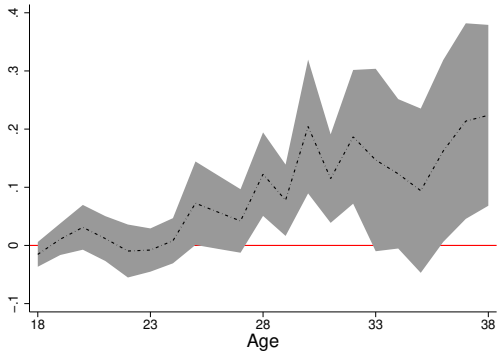
(b) Completed suicide (at each age)



(c) Inpatient



(d) Outpatient



(e) Sick days

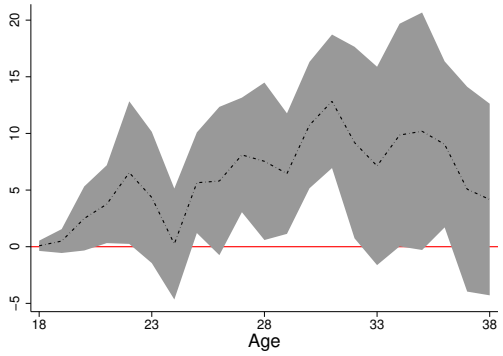
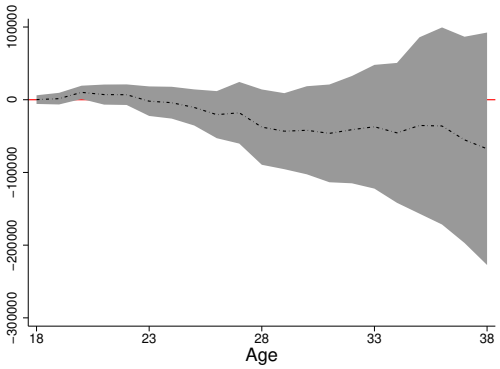


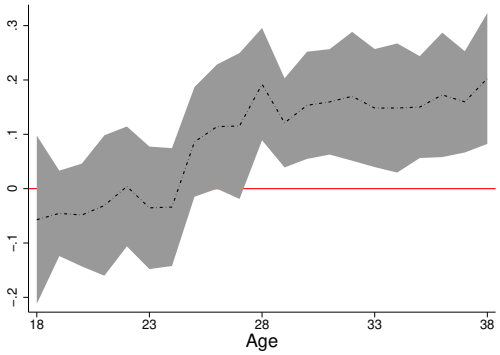
Figure 12: Event-time Evolution of 2SLS Estimates for the Effect of Mental Illness Diagnosis on Labor Market Outcomes

Note: Each of these panels is the graphical representation of 20 different 2SLS estimations that correspond to the first three regressions presented in Table 5, but here, the age of the conscript at the time outcome is measured varies from 18 to 38. The solid line shows the point estimate at each horizon, and the associated 95 percent confidence interval is shown surrounding it.

(a) Income from work (SEK)



(b) 1(Unemployed>0)



(c) Years of schooling

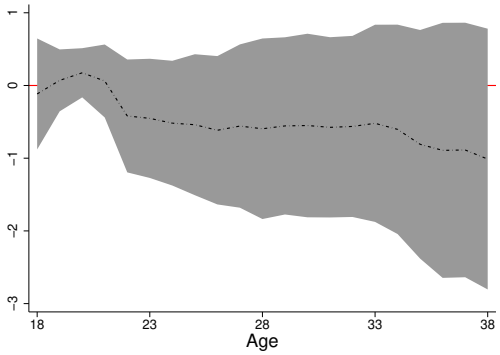
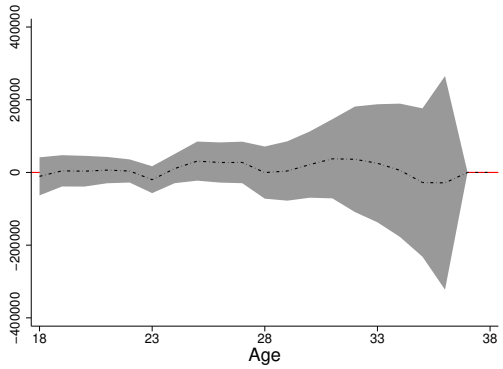


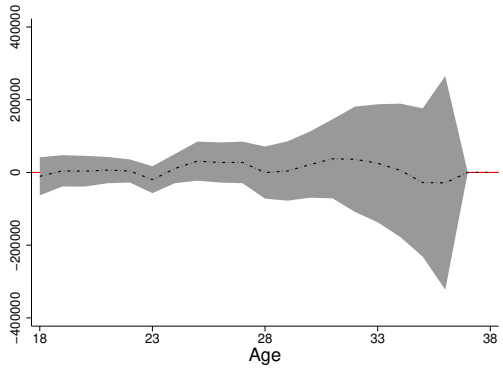
Figure 13: Event-time Evolution of 2SLS Estimates for the Effect of Mental Illness Diagnosis on Wealth and Family Formation

Note: Each of these panels is the graphical representation of 20 different 2SLS estimations that correspond to the last five regressions presented in Table 5, but here, the age of the conscript at the time outcome is measured varies from 18 to 38. The solid line shows the point estimate at each horizon, and the associated 95 percent confidence interval is shown surrounding it.

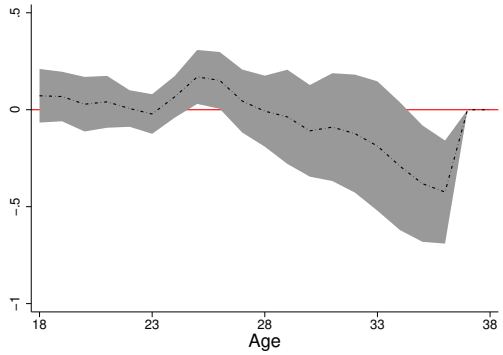
(a) Wealth



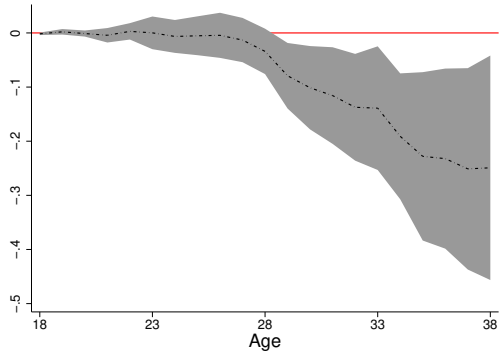
(b) 1(Homeownership>0)



(c) Stock-market participation



(d) 1(Married>0)



(e) 1(Divorced>0)

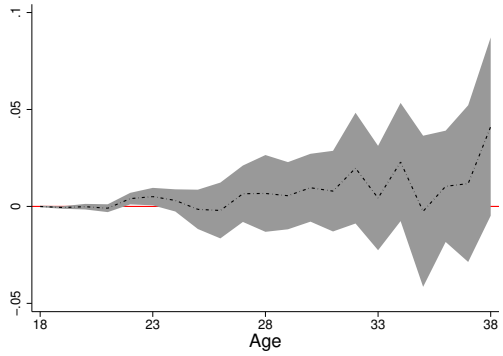


Figure 14: Summary of Event-time Evolution of 2SLS Estimates for All Outcomes

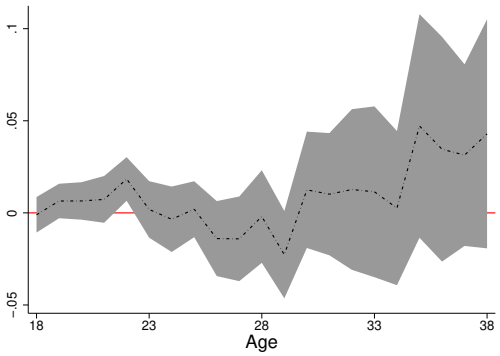
This summarizes the event time 2SLS estimates for all outcome variables at all age horizons from Figures 11, 12, and 13. If an estimate is statistically significant (i.e., zero lies outside the 95 percent confidence interval) then the sign of the point estimate is recorded. Otherwise, when the estimate is statistically insignificant, the cell is left blank.

Outcome	Sign of all Statistuically Significant 2SLS Estimates by Age (Statistcally Insignificant Estimates Left Blank)																				
Health (Table 4 / Figure 10)	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Completed Suicide																					
Death						+		+	+	+	+										+
Inpatient		+				+			+										+		
Outpatient								+			+	+	+	+	+				+	+	+
Sick Days				+	+			+		+	+	+	+	+	+		+		+		
Labor Market Outcomes, Wealth, and Family Structure (Table 5/ Figures 11 and 12)	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Income from Work																					
Unemployed											+	+	+	+	+	+	+	+	+	+	+
Years of Schooling																					
Wealth																					
Home Ownership																					
Stock-Market Particiaption								+	+									-	-	-	
Married												-	-	-	-	-	-	-	-	-	-
Divorced					+	+															

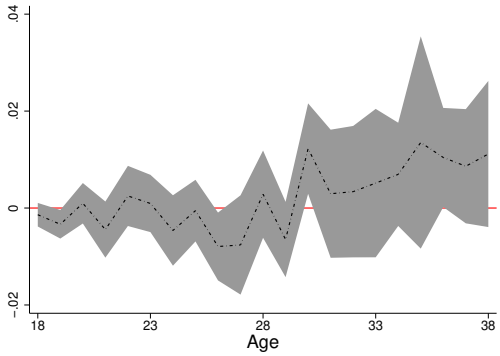
Figure 15: Event-time Evolution of 2SLS Estimates for the Effect of Mental Illness Diagnosis on Future Mental Health Treatment

Note: Each of these graphs is the graphical representation of 20 different 2SLS estimations. The age of the conscript at the time outcome is measured varies from 18 to 38. The solid line shows the point estimate at each horizon, and the associated 95 percent confidence interval is shown surrounding it.

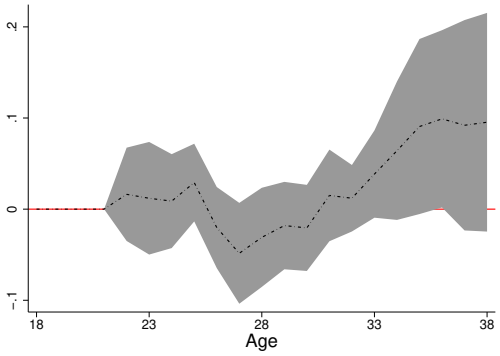
(a) Mental illness



(b) Depression



(c) Anti depressants



(d) Therapy

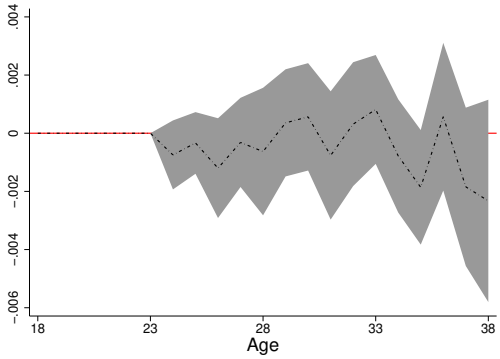


Table 1: Descriptive statistics

	All	Not diagnosed	Diagnosed
A: Military Conscription			
1(Diagnosed with a Mental Illness>0)	.0284	0	1
1(Served in the Military>0) (draft records)	.7615	.7724	.3874
1(Served in the Military>0) (tax records)	.6019	.6113	.2789
IQ Test (normalized)	5.153	5.178	4.293
Education of Parents (years)	12.31	12.32	11.88
Muscle Strength (normalized)	5.512	5.53	4.867
Height (normalized)	6.348	6.352	6.199
B: Future Health			
1(Was an Inpatient>0) at age 30	.03048	.03006	.04471
1(Was an Outpatient>0) at age 30	.2315	.2305	.2647
1(Died>0) up to age 30	.006232	.006156	.008871
1(Completed Suicide>0) up to age 30	.001596	.001569	.002538
Sick days at Age 30	4.408	4.298	8.165
C: Labor Market			
Education (in years) at Age 30	12.77	12.79	11.98
Income from work at Age 30 (KR)	266,807	267,985	226,538
1(Unemployed>0) at Age 30	.1125	.1106	.1757
D: Household			
1(Married>0) at Age 30	.1908	.1917	.1602
1(Divorced>0) at Age 30	.0121	.01193	.01794
Wealth at Age 28 (KR)	71,766	72,469	48,900
Stock-market Participation at 28	.2768	.2797	.1823
1(Homeowner >0) Average at Age 31-46	.6722	.6756	.5574
E: Mechanism			
1(Diagnosed with a Mental Illness>0) at age 30	.01918	.0185	.04222
1(Diagnosed with a Depression>0) at age 30	.003862	.003768	.007051
1(Taking Antidepressants>0) at age 35	.05414	.05264	.1034
1(Prescribed Therapy>0) at age 35	.000173	.0001655	.0004191
Observations	410,146	398,499	11,647

Note: This table reports the means of selected variables for individuals included in the full analysis sample (column 1), and for the subsample of individuals who were not diagnosed with a mental illness as part of the military service conscription (column 2), and individuals who were diagnosed with a mental illness as part of the military service conscription (column 3).

Table 2: Distribution of mental illness diagnoses

Diagnosis	All	Severe	Intermediate	Less severe
Depression	35%	23%	47%	30%
Psychosomatic disorders	29%	4%	52%	44%
Psychological development disorders	15%	84%	14%	2%
Neurosis, anxiety disorders	12%	7%	30%	63%
Personality disorders	3%	83%	16%	1%
Addiction	3%	64%	31%	5%
Other	3%	65%	32%	2%

Note: This table reports the distribution of diagnoses of mental illness at military service conscription, for all diagnoses (column 1), severe diagnoses (column 2), and less severe diagnoses (column 3). We use the ICD9 categorization. Severity is defined as follows; Severe is 10, intermediate is 4-9 and less severe is 1-3.

Table 3: First-Stage and Balance Tests

	Draft diagnosis	Doctor Leniency	Doctor Leniency
	(1)	(2)	(3)
Doctor leniency	0.93969***		
	(0.00972)		
Education of parents	−0.00041***	0.00004*	
	(0.00016)	(0.00002)	
IQ test results	−0.00623***	−0.00004	
	(0.00066)	(0.00006)	
Height	0.00031	0.00003	
	(0.00020)	(0.00002)	
Muscle strength	−0.00470***	−0.00007**	
	(0.00059)	(0.00003)	
389 (Hearing loss)			0.00162*
			(0.00091)
477 (Hay fever)			0.00025
			(0.00041)
370 (Eye problems)			0.00040
			(0.00077)
493 (Asthma)			0.00110**
			(0.00050)
724 (Back pain)			0.00048
			(0.00075)
728 (Disorder of muscle ligament and fascia)			−0.00119
			(0.00243)
719 (Joint problems)			0.00323*
			(0.00184)
692 (Dermatitis (skin))			0.00030
			(0.00074)
845 (Sprained ankle or foot)			0.00043
			(0.00078)
346 (Migraine)			0.00063
			(0.00070)
Dep. Variable Mean	0.028	0.000	0.000
Observations	410073	410073	410073
p-value on Joint F-test	[0.000]	[0.171]	[0.157]
Clusters	102	102	102

Note: This table reports first-stage results and balance tests. The regressions are estimated on the sample described in the notes to Table 1. Doctor tendency to diagnose a mental illness is estimated using data from other draftees assigned to each doctor following the procedure described in Section 3. Column 1 reports estimates from an OLS regression of a diagnosis of mental illness on the variables listed. Columns 2 and 3 report estimates from an OLS regression of the measure of doctors' tendency to diagnose mental illness on the variables listed. The p-value reported at the bottom of the columns is for an F-test of the joint significance of the variable listed in the rows. All specifications control for center-by-year fixed effects. Standard errors clustered at the doctor level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 4: 2SLS Estimates of the Effect a Mental Illness Diagnosis on Future Health Outcomes

	(1)	(2)	(3)	(4)
VARIABLES	Death	Complete suicide	Death (Internal)	Death (Circulatory)
Draft diagnosis MI	0.00658	0.000832	0.0124***	0.00658***
	(0.00477)	(0.00245)	(0.00305)	(0.00232)
Observations	399,857	399,857	399,857	399,857
% change	106	52	884	2195
Dep. Var mean	0.0062	0.0016	0.0014	0.0003
Time x center FE	Yes	Yes	Yes	Yes
Nr of clusters	102	102	102	102
	(1)	(2)	(3)	
VARIABLES	Outpatient	Inpatient	Sick days	
Draft diagnosis MI	0.205***	0.0397*	10.61***	
	(0.0604)	(0.0232)	(2.890)	
Observations	399,851	399,851	387,659	
% change	89	130	242	
Dep. Var mean	0.2314	0.0305	4.3847	
Time x center FE	Yes	Yes	Yes	
Nr of clusters	102	102	102	

Note: This table reports 2SLS estimates of the effect of a mental illness diagnosis during the mandatory draft in Sweden on subsequent health outcomes. We instrument for mental illness using the leave-out leniency measure constructed using the procedure described in Section 3. All specifications control for center-by-year of enrollment fixed effects. Standard errors clustered at the doctor level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level. Socialstyrelsen classifies death as having underlying external causes of death or not, our variable Internal Death is defined as the negation: Internal Death equals one when there are no underlying external causes of death. The variable I-Indicator equals one when the main cause of death starts with an I (diseases of the circulatory system in ICD-10).

Table 5: 2SLS Estimates of the Effect a Mental Illness Diagnosis on Labor Market Outcomes, Wealth, and Family Formation

VARIABLES Age	Long-spell unemployed at age 30	Income from work at age 30	Years of schooling at age 30
	(1)	(2)	(3)
Draft diagnosis MI	0.0486***	-42,020	-0.551
	(0.0177)	(30,920)	(0.644)
Observations	387,659	402,839	397,440
% change	139	-16	-4.3
Dep. Var mean	0.0351	266799.7621	12.7694
Time x center FE	Yes	Yes	Yes
Nr of clusters	102	102	102

VARIABLES Age	Wealth at age 28	1(Homeowner >0) mean over age 31-46	Stock-market participation at age 28
	(4)	(5)	(6)
Draft diagnosis MI	-382.4	-0.0482	-0.00751
	(36,590)	(0.149)	(0.0935)
Observations	296,258	10,395	291,662
% change	-.53	-7.2	-2.7
Dep. Var mean	71766.4522	0.6723	0.2768
Time x center FE	Yes	Yes	Yes
Nr of clusters	74	102	74

VARIABLES Age	Married at age 30	Divorced at age 30
	(7)	(8)
Draft diagnosis MI	-0.101**	0.00964
	(0.0393)	(0.00893)
Observations	397,566	397,566
% change	-53	80
Dep. Var mean	0.1907	0.0121
Time x center FE	Yes	Yes
Nr of clusters	102	102

Note: This table reports 2SLS estimates of the effect of a mental illness diagnosis during the mandatory draft in Sweden on income outcomes measured at a specific age listed in each column. We instrument for mental illness using the leave-out leniency measure constructed using the procedure described in Section 3. All specifications control for center-by-year of enrollment fixed effects. Standard errors clustered at the doctor level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 6: 2SLS Estimates of the Effect a Mental Illness Diagnosis on Future Mental Illness Diagnosis and Treatment

VARIABLES Age	Mental illness at age 30	Depression at age 30	Anti-depressants at age 35	Therapy at age 35
	(1)	(2)	(3)	(4)
Draft diagnosis MI	0.0125	0.0123**	0.0907*	-0.00186*
	(0.0161)	(0.00476)	(0.0490)	(0.00100)
Observations	404,909	404,909	323,602	323,602
% change	65	314	168	-930
Dep. Var mean	0.0192	0.0039	0.0541	0.0002
Time x center FE	Yes	Yes	Yes	Yes
Nr of clusters	102	102	98	98

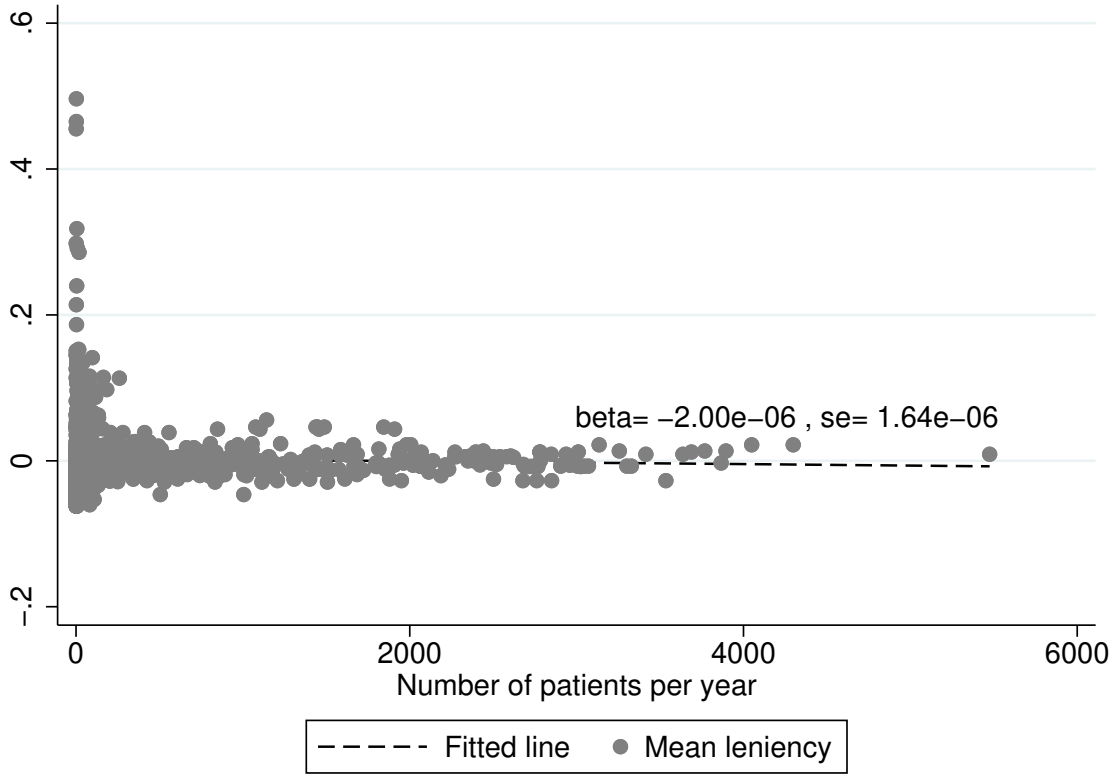
Note: This table reports 2SLS estimates of the effect of a mental illness diagnosis during the mandatory draft in Sweden on income outcomes measured at a specific age listed in each column. We instrument for mental illness using the leave-out leniency measure constructed using the procedure described in Section 3. All specifications control for center-by-year of enrollment fixed effects. Standard errors clustered at the doctor level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 7: **2SLS Estimates, Split by the Social Economic Background of the Parents of Draftees**

Category	Coefficient			Number of Observations		
Variable	Whole Sample	Low SES	High SES	Whole Sample	Low SES	High SES
Health						
1(Completed Suicide > 0) at 30	0.000832	0.00914**	-0.00208	399,857	120,489	266,188
	(0.00245)	(0.00457)	(0.00297)			
1(Death > 0) at 30	0.00658	0.00741	0.00470	399,857	120,489	266,188
	(0.00477)	(0.0110)	(0.00596)			
1(Inpatient > 0) at 30	0.0397*	0.0411	0.0334*	399,851	120,485	266,186
	(0.0232)	(0.0448)	(0.0175)			
1(Outpatient > 0) at 30	0.205***	0.183**	0.197***	399,851	120,485	266,186
	(0.0604)	(0.0922)	(0.0562)			
# Sick Days at 30	10.61***	6.361	12.28***	387,659	117,792	257,255
	(2.890)	(4.197)	(2.938)			
Labour, Wealth, & Family						
Income from Work (SEK) at 30	-37,121	-49,683	-31,278	392,774	118,699	261,242
	(31,578)	(39,611)	(29,176)			
1(Long-spell Unemployed > 0) at 30	0.0486***	0.00934	0.0698***	387,659	117,792	257,255
	(0.0177)	(0.0334)	(0.0161)			
# Years of Schooling at 30	-0.488	-1.036*	-0.304	387,535	117,737	257,194
	(0.660)	(0.618)	(0.589)			
Wealth (SEK) at 28	3,309	-674.0	7,522	291,094	92,583	187,619
	(37,479)	(23,869)	(48,252)			
1(Home Ownership > 0) mean over 31-46	-0.107	-0.0385	-0.245	10,123	3,096	6,675
	(0.152)	(0.275)	(0.246)			
1(Stock-Market Participation > 0) at 28	-0.00144	-0.119*	0.0865	286,567	91,382	184,535
	(0.0958)	(0.0681)	(0.113)			
1(Married > 0) at 30	-0.101**	-0.0933	-0.114**	387,659	117,792	257,255
	(0.0418)	(0.0578)	(0.0480)			
1(Divorced > 0) at 30	0.0113	0.0366**	0.00126	387,659	117,792	257,255
	(0.00913)	(0.0178)	(0.00985)			
Mental Health						
1(Mental Illness > 0) at 30	0.0133	-0.0104	0.0331*	399,851	120,485	266,186
	(0.0163)	(0.0200)	(0.0194)			
1(Depression > 0) at 30	0.0124**	0.00818	0.0136**	399,851	120,485	266,186
	(0.00493)	(0.00772)	(0.00639)			
1(Antidepressants > 0) at 35	0.0905*	0.0470	0.111**	322,297	101,179	209,540
	(0.0493)	(0.0584)	(0.0529)			

A Additional Results

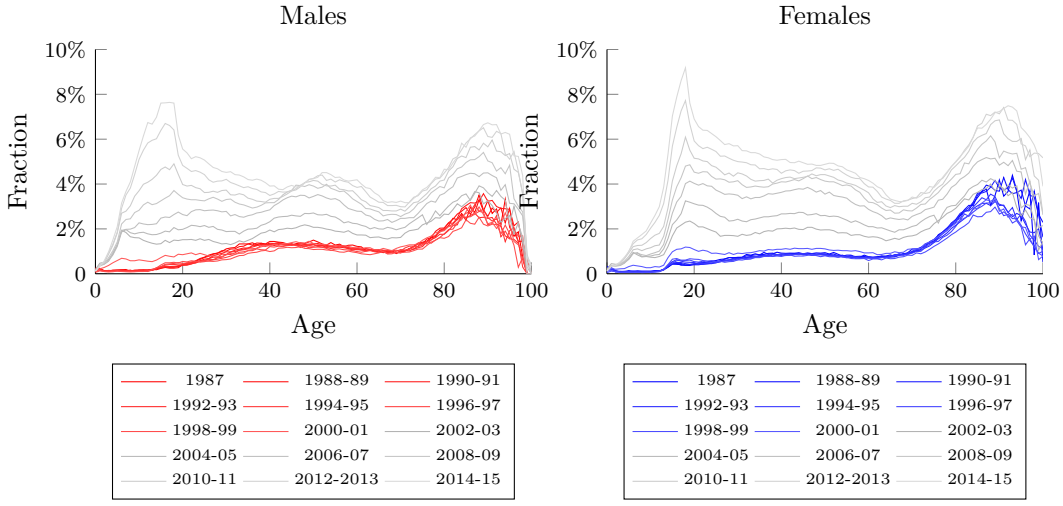
Appendix Figure A1: Monotonicity 6, Correlation Between the Tendency to Diagnose and the Number of Conscripts by Doctor and Year



Note: This figure shows the relation between mean doctor tendency to diagnose by year and the number of conscripts evaluated by the doctor each year.

Appendix Figure A2: Share of Individuals Diagnosed With a Mental Illness in the General Population by Birth Cohort

The figures plot the share of males (Panel A), and females (Panel B), that are diagnosed with a mental illness in the Swedish population, by age and cohort. In red (males) and blue (females), we indicate the years that are included in our analysis. In light grey are the shares for the later years not included in our analysis. From the figure, you can see an overall large increase in the percentage of individuals diagnosed with a mental illness, but the increase is particular large for individuals below 20 years of age.

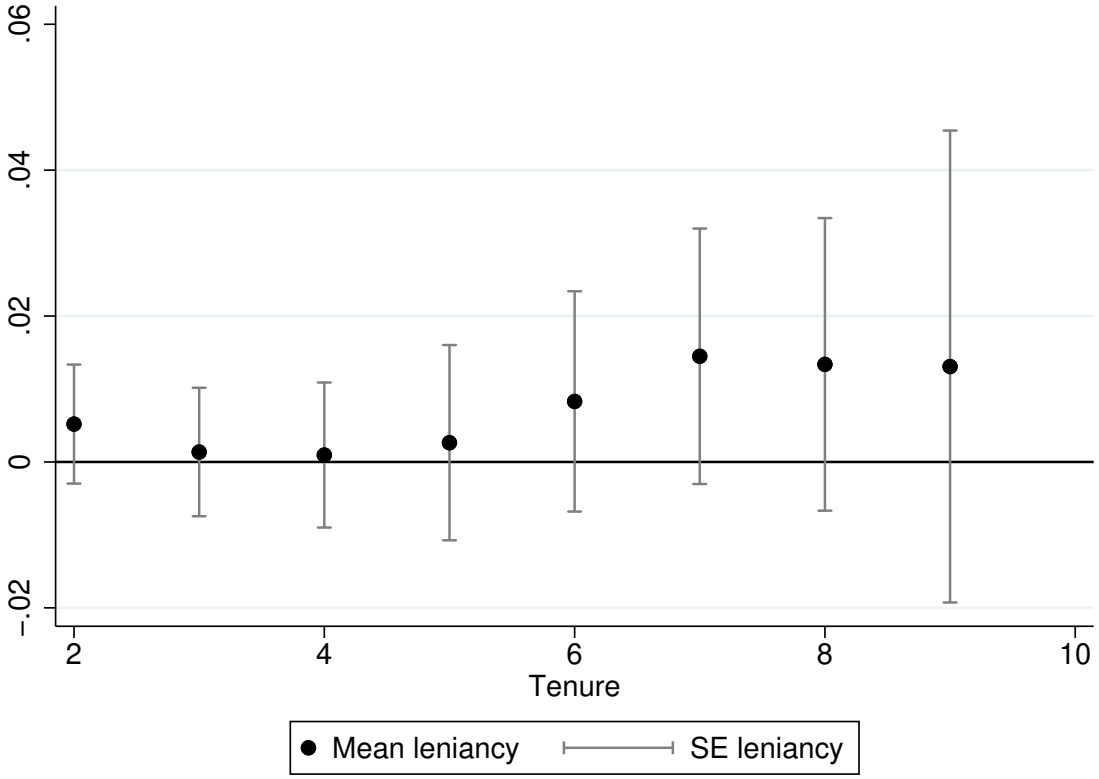


Appendix Figure A3: Monotonicity 7, Correlation Between Mean Doctor Leniency and Tenure

This figure shows the relation between yearly mean doctor tendency to diagnose by year and the tenure for that doctor. We estimated the following equation:

$$y_{i,t} = \beta_{t1} \times Tenure + \omega_t + \epsilon_{i,t}$$

, where ω_t is year fixed effects for doctor i in year t. Each coefficient is the estimated difference between that doctor and a doctor who only has one year of tenure.



Appendix Table A1: Monotonicity 8, Above, Below the Median Difference for Sub-samples

	Parents' education		IQ test	
	Above median	Below median	Above median	Below median
	(1)	(2)	(3)	(4)
Dr leniency	0.73318***	1.06460***	0.65746***	1.13928***
	(0.01504)	(0.01801)	(0.03298)	(0.02090)
Dep. Variable Mean	0.022	0.032	0.018	0.036
Observations	151876	258197	171560	238513
Clusters	102	102	102	102

	Height		Muscle strength	
	Above median	Below median	Above median	Below median
	(1)	(2)	(3)	(4)
Dr leniency	0.85333***	1.01237***	0.90115***	1.05489***
	(0.01338)	(0.01605)	(0.01171)	(0.02468)
Dep. Variable Mean	0.025	0.031	0.027	0.033
Observations	185377	224696	304914	105157
Clusters	102	102	102	102

Note: This table shows the first-stage coefficient for different sub-samples.

Appendix Table A2: Monotonicity check for subsample not diagnosed as mentally ill at the draft, part I.

VARIABLE	Mental illness		
Age	at age 30	at age 19-23	at age 19-28
	(1)	(2)	(3)
Above median leniency	-0.00114***	-0.000660	-0.00288***
	(0.000430)	(0.000417)	(0.000652)
Observations	393,285	393,285	393,285
Dep var mean	0.01850	0.01739	0.04369

VARIABLE	Depression		
Age	at age 30	at age 19-23	at age 19-28
	(1)	(2)	(3)
Above median leniency	0.000158	-0.000280	-0.00171***
	(0.000195)	(0.000189)	(0.000341)
Observations	393,285	393,285	393,285
Dep var mean	0.00377	0.00351	0.01156

VARIABLE	Outpatient		
Age	at age 30	at age 19-23	at age 19-28
	(1)	(2)	(3)
Above median leniency	-0.00206	-0.0260***	-0.0315***
	(0.00134)	(0.00120)	(0.00159)
Observations	393,285	393,285	393,285
Dep var mean	0.23050	0.17086	0.46352

Note: This table shows the regression of different outcomes on the dummy of seeing a doctor with above the median leniency.

Appendix Table A3: Monotonicity check for subsample not diagnosed as mentally ill at the draf, part II.

VARIABLE	Inpatient		
Age	at age 30	at age 19-23	at age 19-28
	(1)	(2)	(3)
Above median leniency	0.000203	0.00203*	0.00297**
	(0.000545)	(0.00118)	(0.00140)
Observations	393,285	393,285	393,285
Dep var mean	0.03006	0.16350	0.25891

VARIABLE	Anti-depressants		
Age	at age 35	at age 19-23	at age 19-28
	(1)	(2)	(3)
Above median leniency	-0.00211***	-0.000966	-0.00350***
	(0.000797)	(0.00132)	(0.00111)
Observations	314,060	47,755	177,415
Dep var mean	0.05263	0.02132	0.05796

VARIABLE	Complete Suicide		
Age	at age 35	at age 19-23	at age 19-28
	(1)	(2)	(3)
Above median leniency	-7.13e-06	1.64e-05	1.64e-05
	(0.000126)	(2.52e-05)	(2.52e-05)
Observations	393,853	398,393	398,393
Dep var mean	0.00157	0.00033	0.00033

Note: This table shows the regression of different outcomes on the dummy of seeing a doctor with above the median leniency.

ICD-9-CM Code	Definition
296.2	Major depressive disorder, single episode- unspecified
296.21	Major depressive disorder, single episode- mild
296.22	Major depressive disorder, single episode- moderate
296.23	Major depressive disorder, single episode- severe, without mention of psychotic behavior
296.24	Major depressive disorder, single episode- severe, specified as with psychotic behavior
296.25	Major depressive disorder, single episode- in partial or unspecified remission
296.26	Major depressive disorder, single episode- in full remission
296.3	Major depressive disorder, recurrent episode- unspecified
296.31	Major depressive disorder, recurrent episode- mild
296.32	Major depressive disorder, recurrent episode- moderate
296.33	Major depressive disorder, recurrent episode- severe, without mention of psychotic behavior
296.34	Major depressive disorder, recurrent episode- severe, specified as with psychotic behavior
296.35	Major depressive disorder, recurrent episode- in partial or unspecified remission
296.36	Major depressive disorder, recurrent episode- in full remission
296.51	Bipolar I disorder, most recent episode (or current) mild depression
296.52	Bipolar I disorder, most recent episode (or current) moderate depression
296.53	Bipolar I disorder, most recent episode (or current) severe without mention of psychotic behavior
296.54	Bipolar I disorder, most recent episode (or current) severe specified as with psychotic behavior
296.82	Atypical depressive disorder
296.82	Atypical depressive disorder
296.9	Unspecified episodic mood disorder
296.99	Other specified episodic mood disorder
300.4	Dysthymic disorder
309.1	Prolonged depressive reaction
309.28	Adjustment disorder with mixed anxiety and depressed mood
311	Depressive disorder, not elsewhere classified
296.6x	Bipolar I disorder, most recent episode (or current) mixed
298.0	Other non-organic depressive psychoses
309.0	Adjustment disorder with depressed mood

ICD-10-CA Code	Definition
F31.31	Bipolar affective disorder, current episode mild or moderate depression
F31.32	Bipolar affective disorder, current episode mild or moderate depression
F31.4	Bipolar affective disorder, current episode severe depression without psychotic symptoms
F31.5	Bipolar affective disorder, current episode severe depression with psychotic symptoms
F31.6x	Bipolar affective disorder, current episode mixed
F31.9	Bipolar affective disorder, current episode unspecified
F32.0	Mild depressive episode
F32.1	Moderate depressive episode
F32.2	Severe depressive episode without psychotic symptoms
F32.3	Severe depressive episode with psychotic symptoms
F32.4	Depressive disorder, single episode, in partial remission
F32.5	Depressive disorder, single episode, in full remission
F32.8	Other depressive episodes
F32.9	Depressive episode, unspecified
F32.9	Depressive episode, unspecified
F33.0	Recurrent depressive disorder, current episode mild
F33.1	Recurrent depressive disorder, current episode moderate
F33.2	Recurrent depressive disorder, current episode severe without psychotic symptoms
F33.3	Recurrent depressive disorder, current episode severe with psychotic symptoms
F33.41	Recurrent depressive disorder, currently in remission
F33.42	Recurrent depressive disorder, currently in remission
F33.8	Recurrent depressive disorder, other
F33.9	Recurrent depressive disorder, unspecified
F34.1	Dysthymia
F34.8	Other persistent mood disorders
F34.9	Persistent mood disorder, unspecified
F38.0	Other single mood disorders
F38.1	Other recurrent mood disorders
F38.8	Other specified mood disorders
F39	Unspecified mood disorder
F41.2	Mixed anxiety and depressive disorder
F43.2	Adjustment disorders
F43.2	Adjustment disorders
F45.0	Somatization disorder
F45.1	Undifferentiated somatoform disorder
F45.9	Somatoform disorder, unspecified
F48.8	Other specified neurotic disorders
F48.9	Neurotic disorder, unspecified
F60.6	Anxious [avoidant] personality disorder
F60.8	Other specific personality disorders
F61	Mixed and other personality disorders
F62.8	Other enduring personality changes
F68.8	Other specified disorders of adult personality and behaviour
F99	Mental disorder, not elsewhere specified

Appendix Table A4: 2SLS Estimates of the Effect Military Service on Future Health Outcomes

VARIABLES Age	Inpatient at age 30	Outpatient at age 30	Sick days at age 30
	(3)	(4)	(5)
1(Military service>0)	-0.00469	-0.0254	-1.323
Tax records	(0.00905)	(0.0161)	(1.840)
Observations	261,928	261,928	256,770
% change	-17	-11	-30
Dep. Var mean	0.0283	0.2234	4.4513
No. of clusters	70	70	70

VARIABLES Age	Complete suicide up to age 30	Death up to age 30
	(1)	(2)
1(Military service>0) pay	-0.00188	-0.00756**
Tax records	(0.00153)	(0.00364)
Observations	261,450	262,596
% change	-134	-130
Dep. Var mean	0.0014	0.0058
No. of clusters	70	70

Note: This table reports 2SLS estimates of the effect of a mental illness diagnosis during the mandatory draft in Sweden on subsequent diagnoses of mental illness. We instrument for mental illness using the leave-out leniency measure constructed using the procedure described in Section 3. All specifications control for center-by-year of enrollment fixed effects. Standard errors clustered at the doctor level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A5: 2SLS Estimates of the Effect Military Service on Labour Market Outcomes, Wealth, and Family Formation

VARIABLES Age	Years of schooling at age 30	Income from work at age 30	Unemployed at age 30
	(1)	(2)	(3)
1(Military service>0)	0.189	25,961***	-0.0618***
Tax records	(0.177)	(6,954)	(0.0139)
Observations	256,729	259,365	256,770
% change	1.5	10	-58
Dep. Var mean	12.7943	259948.5325	0.1068
Nr of clusters	70	70	70

VARIABLES Age	Married at age 30	Divorced at age
	(4)	(5)
1(Military service>0)	0.0204	0.00416
Tax records	(0.0202)	(0.00405)
Observations	256,770	256,770
% change	11	37
Dep. Var mean	0.1874	0.0114
Nr of clusters	70	70

VARIABLES Age	Wealth at age 28	1(Home owner >0) mean over age 31-46	Stock-market participation at age 28
	(6)	(7)	(8)
1(Military service>0)	-2,486	0.226	0.0519
Tax records	(8,847)	(0.156)	(0.0335)
Observations	264,475	6,810	260,443
% change	-3.4	31	18
Dep. Var mean	73575	0.7342	0.2948
Nr of clusters	70	70	70

Note: This table reports 2SLS estimates of the effect of military service in Sweden on labour market outcomes, wealth and family formation. We instrument for military service using the leave-out measure of negotiations due to capacity constraints, constructed using the procedure described in Section 5. All specifications control for center-by-year of enrollment fixed effects. Standard errors clustered at the doctor level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A6: 2SLS Estimates of the Effect Military Service on Future Mental Illness Diagnosis and Treatment

VARIABLES Age	Mental illness at age 30	Depressed at age 30	Anti-depressants at age 35	Therapy at age 35
	(1)	(2)	(3)	(4)
1(Military service>0)	-0.0143**	-0.00390	-0.0127	0.000436
Tax records	(0.00648)	(0.00351)	(0.00952)	(0.000530)
Observations	261,928	261,928	261,928	261,928
% change	-95	-115	-25	218
Dep. Var mean	0.0151	0.0034	0.0499	0.0002
Nr of clusters	70	70	70	70

Note: This table reports 2SLS estimates of the effect of military service in Sweden on future mental illness diagnosis and treatment. We instrument for military service using the leave-out measure of negotiations due to capacity constraints, constructed using the procedure described in Section 5. All specifications control for center-by-year of enrollment fixed effects. Standard errors clustered at the doctor level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.