

# Are We Overdiagnosing Mental Illnesses? Evidence from Randomly Assigned Doctors

Marieke Bos (Swedish House of Finance, SSE) and  
Andrew Hertzberg\* (Philadelphia Fed)  
with Andrés Liberman (Stern, NYU)

July 2020

\*Disclaimer: The views expressed in this paper do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System

## Motivation: Diagnosis of Mental Illness is Prevalent

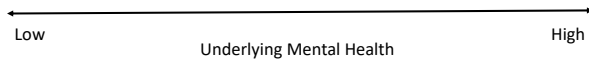
- ▶ EU: 17.3% of adults in 2018 were diagnosed with a mental health problem
  - ▶ Source: OECD/EU Health a Glance Report
  
- ▶ US: During 2011-2014, 12.7% of persons age 12 and over took antidepressant medication in the last month
  - ▶ Source: National Health and Nutrition Examination Survey

# The Question

Is mental illness over or under diagnosed?

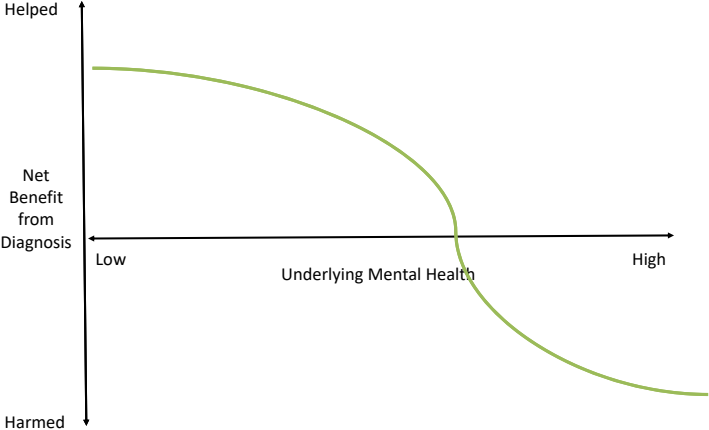
# Conceptual Framework

**Patients have heterogeneous  
underlying mental health**



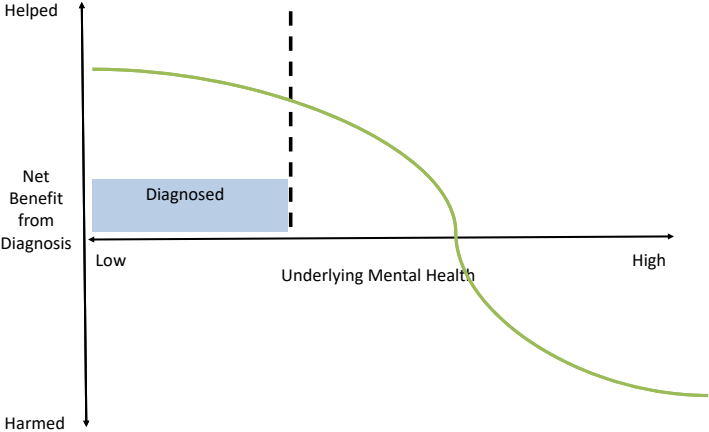
# Conceptual Framework

**Diagnosis is beneficial for patients with low mental health**



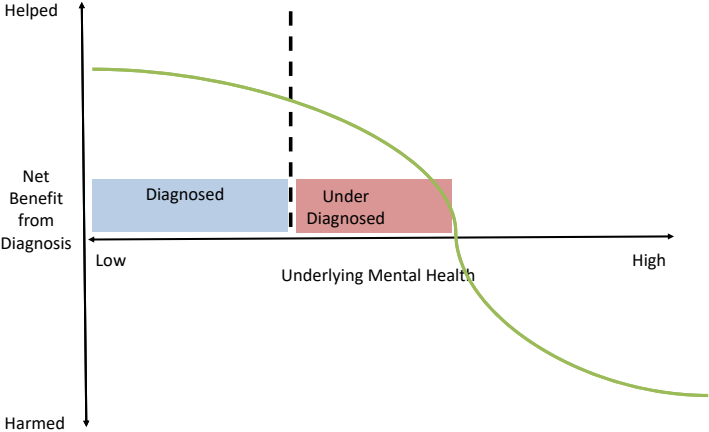
# Conceptual Framework

**The role of a doctor is to assess the underlying mental health of a patient and diagnose all who fall below a threshold**



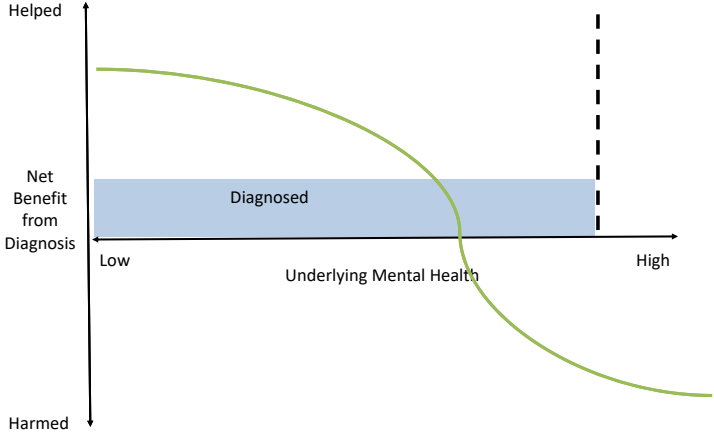
# Conceptual Framework

**Underdiagnosis: marginal patient has strictly positive benefit from diagnosis**



# Conceptual Framework

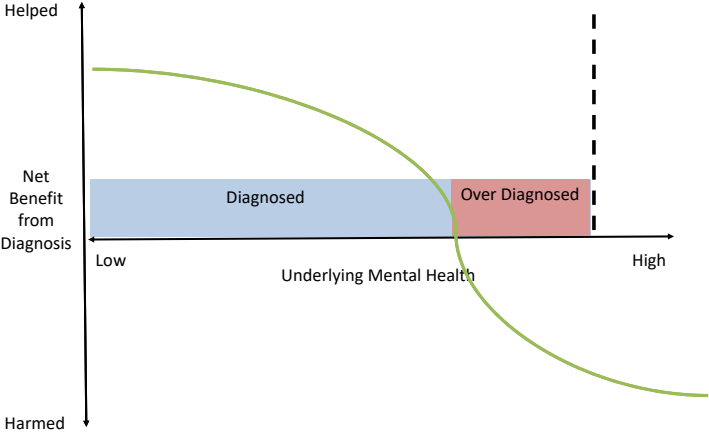
**Overdiagnosis: marginal patient is harmed by diagnosis**





# Conceptual Framework

**Overdiagnosis: marginal patient is harmed by diagnosis**



# Our approach to assessing under or over diagnosis

- ▶ Measure the causal effect of a mental illness diagnosis on measure associated with welfare of a “*marginal*” patient

# Our approach to assessing under or over diagnosis

- ▶ Measure the causal effect of a mental illness diagnosis on measure associated with welfare of a “*marginal*” patient
- ▶ Two challenges:
  1. How to measure the *causal effect on the marginal* (not average) patient
    - ▶ Exploit random assignment of doctors in Swedish military conscription

# Our approach to assessing under or over diagnosis

- ▶ Measure the causal effect of a mental illness diagnosis on measure associated with welfare of a “*marginal*” patient
- ▶ Two challenges:
  1. How to measure the *causal effect on the marginal* (not average) patient
    - ▶ Exploit random assignment of doctors in Swedish military conscription
  2. How to adequately measure the *welfare* of the marginal patient
    - ▶ Wide range of health, economic and family outcomes over a twenty-year window after diagnosis

## Preview of results

- ▶ Diagnosis has a detrimental effect on an 18-year-old man with marginal mental health

# Preview of results

- ▶ Diagnosis has a detrimental effect on an 18-year-old man with marginal mental health
- ▶ Worse life outcomes over 20-year span after diagnosis
- ▶ Health outcomes:
  - ▶ Increased morbidity
  - ▶ More sick days
  - ▶ Higher probability of admission to hospital
- ▶ Labor market and family outcomes:
  - ▶ More likely to be unemployed
  - ▶ Less likely to be married

# How can overdiagnosis harm a patient?

- ▶ Several channels are possible. For example:
  - ▶ Treatment (pharmaceutical or otherwise) may have unintended side effects

# How can overdiagnosis harm a patient?

- ▶ Several channels are possible. For example:
  - ▶ Treatment (pharmaceutical or otherwise) may have unintended side effects
  - ▶ Labeling
    - ▶ internally: changes self-view, change likelihood of seeking treatment in the future
    - ▶ externally: changes the way other doctors assess and treat patients health, and also family and friends.

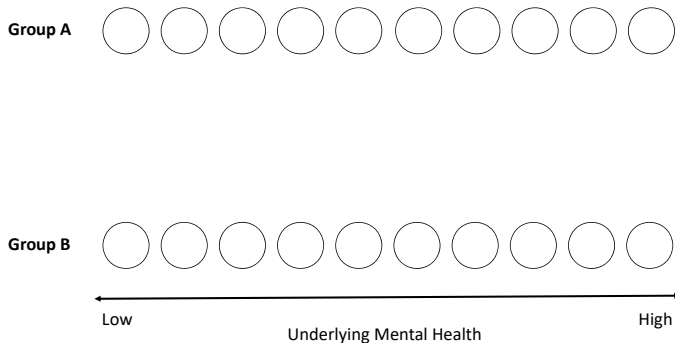


# How can overdiagnosis harm a patient?

- ▶ Several channels are possible. For example:
  - ▶ Treatment (pharmaceutical or otherwise) may have unintended side effects
  - ▶ Labeling
    - ▶ internally: changes self-view, change likelihood of seeking treatment in the future
    - ▶ externally: changes the way other doctors assess and treat patients health, and also family and friends.
  - ▶ Military service
    - ▶ We will rule this out as primary channel in our setting

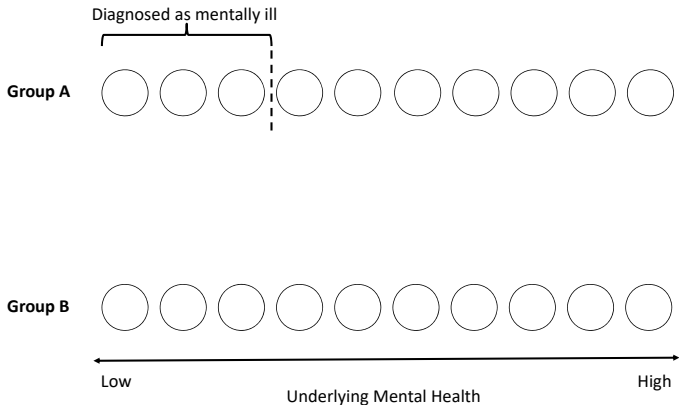
# Idealized experiment

Two identical groups containing people  
with heterogeneous underlying mental health



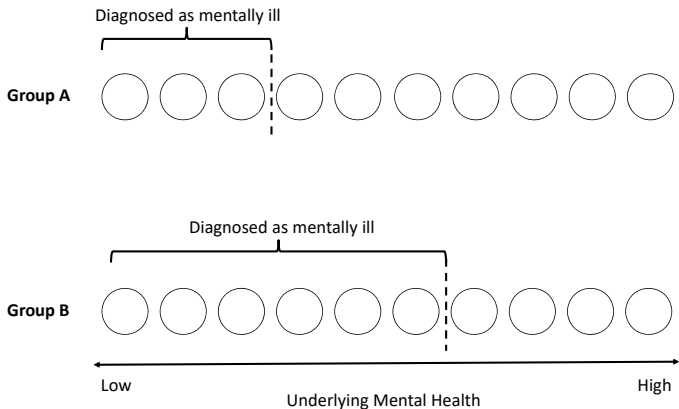
# Idealized experiment

**Group A:** Examined for mental health by Doctor A, who applies a strict standard for diagnosing mental illness



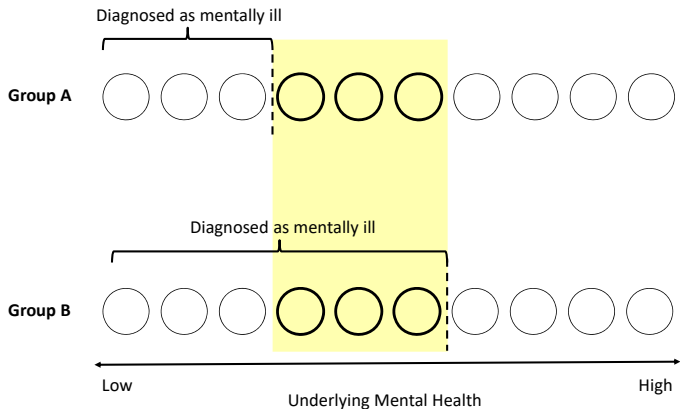
# Idealized experiment

**Group B:** Examined for mental health by Doctor B, who applies a lenient standard for diagnosing mental illness



# Idealized experiment

**Experiment:** Compare outcomes for groups A and B  
Any difference will be caused by  
differential diagnosis of marginal patients



# Requirements for ideal experiment

- ▶ **Counterfactual:** Groups A and B must be ex-ante **identical**
  - ▶ Random assignment of doctor will ensure this
  - ▶ Check empirically by comparing observable characteristics of each group

# Requirements for ideal experiment

- ▶ **Counterfactual:** Groups A and B must be ex-ante **identical**
  - ▶ Random assignment of doctor will ensure this
  - ▶ Check empirically by comparing observable characteristics of each group
  
- ▶ **Variation:** Doctors must vary in the mental illness threshold they apply
  - ▶ Check this empirically
  - ▶ Verify that leave-out propensity of diagnosing other patients predicts likelihood that a patient is diagnosed

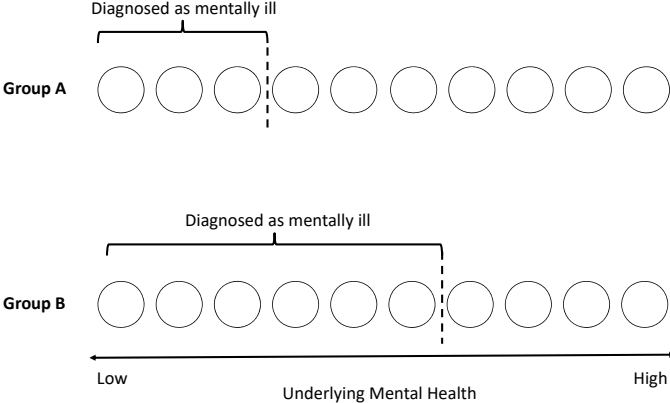
# Requirements for ideal experiment

- ▶ **Counterfactual:** Groups A and B must be ex-ante **identical**
  - ▶ Random assignment of doctor will ensure this
  - ▶ Check empirically by comparing observable characteristics of each group
- ▶ **Variation:** Doctors must vary in the mental illness threshold they apply
  - ▶ Check this empirically
  - ▶ Verify that leave-out propensity of diagnosing other patients predicts likelihood that a patient is diagnosed
- ▶ **Monotonicity:** Doctors agree on underlying ranking of mental health
  - ▶ Check this empirically
  - ▶ Why this matters....



# Monotonicity

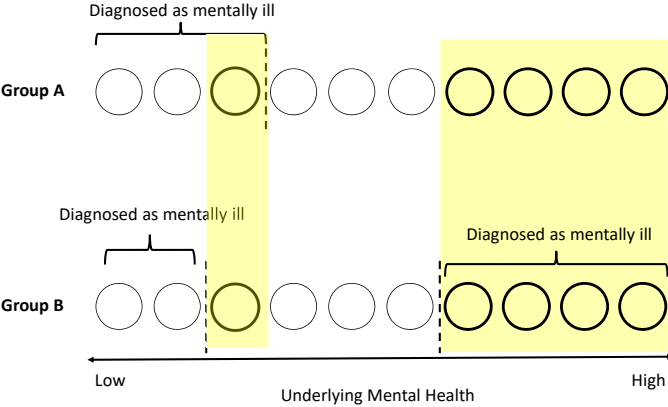
**Monotonicity:** Doctors agree on underlying ranking of mental illness, but apply a different threshold for diagnosis



# Monotonicity

## Experiment without monotonicity

Will potentially confound effect of failure to diagnose with diagnosis of non-marginal patients



## Setting: Swedish military service

- ▶ Sweden had mandatory military conscription from 1901 to 2010

## Setting: Swedish military service

- ▶ Sweden had mandatory military conscription from 1901 to 2010
- ▶ All male citizens report to their regional test office shortly around turning 18
  - ▶ Two days of cognitive and physical tests to determine if conscript was fit to serve

## Setting: Swedish military service

- ▶ Sweden had mandatory military conscription from 1901 to 2010
- ▶ All male citizens report to their regional test office shortly around turning 18
  - ▶ Two days of cognitive and physical tests to determine if conscript was fit to serve
  - ▶ As part of this process every conscript is examined by a doctor (GP) who assesses his physical and mental health

# The Diagnosis Process

- ▶ Assignment of conscript to doctor is random:
  - ▶ Several doctors work in each regional office
  - ▶ After completing several other tests, conscripts place records in a box and are called in order by the next available doctor

# The Diagnosis Process

- ▶ Assignment of conscript to doctor is random:
  - ▶ Several doctors work in each regional office
  - ▶ After completing several other tests, conscripts place records in a box and are called in order by the next available doctor
- ▶ Doctors go through a standardized protocol (questions, vital signs) and use this to diagnose mental illness
  - ▶ The diagnosis includes a code which records the **type** of mental illness
  - ▶ Doctors also assign a **severity score** to their diagnosis on a scale of 1 to 9

# The Diagnosis Process

- ▶ Assignment of conscript to doctor is random:
  - ▶ Several doctors work in each regional office
  - ▶ After completing several other tests, conscripts place records in a box and are called in order by the next available doctor
- ▶ Doctors go through a standardized protocol (questions, vital signs) and use this to diagnose mental illness
  - ▶ The diagnosis includes a code which records the **type** of mental illness
  - ▶ Doctors also assign a **severity score** to their diagnosis on a scale of 1 to 9
- ▶ Doctors do not provide treatment
  - ▶ If a conscript is diagnosed, he is informed of the diagnosis and is referred to a specialist outside of the military



# The Diagnosis Process

- ▶ Assignment of conscript to doctor is random:
  - ▶ Several doctors work in each regional office
  - ▶ After completing several other tests, conscripts place records in a box and are called in order by the next available doctor
- ▶ Doctors go through a standardized protocol (questions, vital signs) and use this to diagnose mental illness
  - ▶ The diagnosis includes a code which records the **type** of mental illness
  - ▶ Doctors also assign a **severity score** to their diagnosis on a scale of 1 to 9
- ▶ Doctors do not provide treatment
  - ▶ If a conscript is diagnosed, he is informed of the diagnosis and is referred to a specialist outside of the military
- ▶ Diagnosis lowers the probability of serving from 74% to 37%

## Types of mental illnesses diagnosed

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%

- ▶ Diagnosis rate in our sample: **2.84%**
- ▶ Examples:
  - ▶ **Psychosomatic disorders:** Stress induced physical ailment such as ulcers and high blood pressure
  - ▶ **Psychological development disorders:** Autism, attention deficit/hyperactivity disorder
  - ▶ **Personality disorders:** Narcissistic personality disorder, paranoia

# Sample and data

- ▶ Sample of Swedish males called to enlist between 1989 and 2001
  - ▶ Chosen to ensure consistency of diagnosis protocol throughout sample

# Sample and data

- ▶ Sample of Swedish males called to enlist between 1989 and 2001
  - ▶ Chosen to ensure consistency of diagnosis protocol throughout sample
- ▶ Only include a doctor in year  $t$  if she saw at least 500 conscripts in that year
  - ▶ Only include conscripts who saw this set of doctors

# Sample and data

- ▶ Sample of Swedish males called to enlist between 1989 and 2001
  - ▶ Chosen to ensure consistency of diagnosis protocol throughout sample
- ▶ Only include a doctor in year  $t$  if she saw at least 500 conscripts in that year
  - ▶ Only include conscripts who saw this set of doctors
- ▶ The resulting analysis sample contains 410,146 conscripts assessed by 102 doctors

# Sample and data

- ▶ Sample of Swedish males called to enlist between 1989 and 2001
  - ▶ Chosen to ensure consistency of diagnosis protocol throughout sample
- ▶ Only include a doctor in year  $t$  if she saw at least 500 conscripts in that year
  - ▶ Only include conscripts who saw this set of doctors
- ▶ The resulting analysis sample contains 410,146 conscripts assessed by 102 doctors
- ▶ We link each individual in the draft data to
  - ▶ National medical board records (diagnosis, prescriptions, death, hospital admission)
  - ▶ Statistics Sweden data on wealth, family and labor market outcomes

# Empirical strategy

# Doctor leniency

We construct our instrument using a residualized, annual leave-out mean doctor leniency measure similar to that used to exploit variation in judge propensities for:

- ▶ Sentence length (Kling 2006)
- ▶ Juvenile incarceration (Aizer and Doyle 2015)
- ▶ Pretrial detention (Dobbie et al. (2018))

We account for two sources of non-random variation in the construction of our instrument:

- ▶ variation in diagnosis rates across recruitment centers
- ▶ variation in diagnosis rates over time



## Doctor leniency

- ▶ Let the diagnosis of mental illness after removing the effect of enlistment-center-by-year fixed effects  $X_{ct}$  be denoted by

$$Draft\_Diagnosis_{ict} = \gamma X_{ct} + \varepsilon_i$$

# Doctor leniency

- ▶ Let the diagnosis of mental illness after removing the effect of enlistment-center-by-year fixed effects  $X_{ct}$  be denoted by

$$\text{Draft\_Diagnosis}_{ict} = \gamma X_{ct} + \varepsilon_i$$

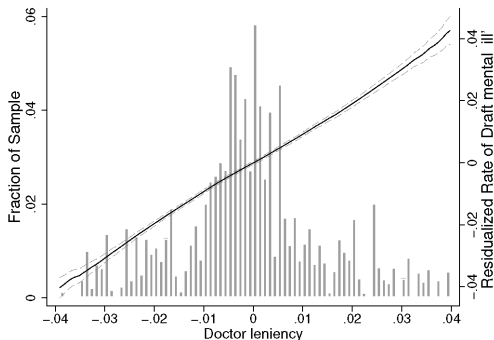
- ▶ Then, we define  $Z_{ict}$  as doctor  $j$ 's tendency to diagnose a mental illness for each individual  $i$  in center  $c$  in year  $t$  as

$$Z_{ict} = \frac{\sum_{k \in N_{j,t}} \varepsilon_k - \varepsilon_i}{N_{c,j,t} - 1}$$

- ▶ where  $N_{cjt}$  is the total number of draftees  $k$  attended by doctor  $j$  in center  $c$  in year  $t$

# Variation in Doctor leniency

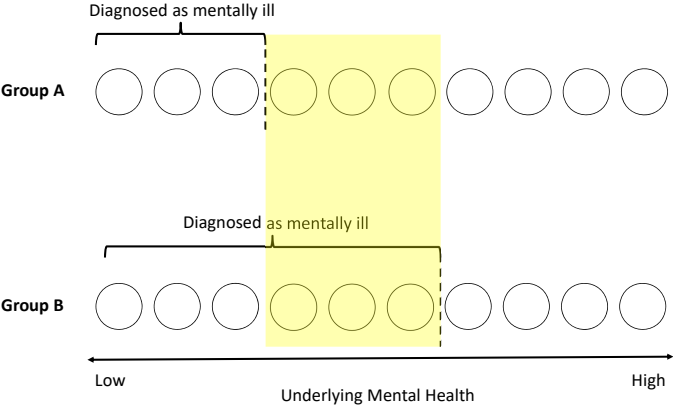
Figure 1: First Stage for All Draftees



- ▶ This figure reports the first stage relationships between draftee mental illness diagnosis during conscript and the numerical value of Doctor leniency.
- ▶ The solid line represents a local linear regression of mental illness on Doctor leniency

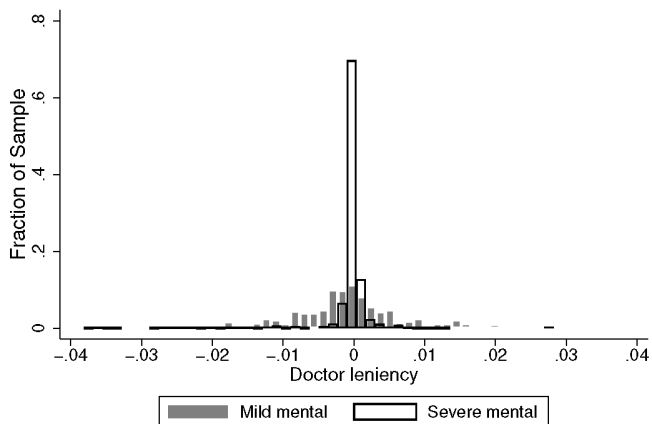
# Monotonicity

**Monotonicity:** Variation in leniency should come from different tendency to diagnose marginal cases



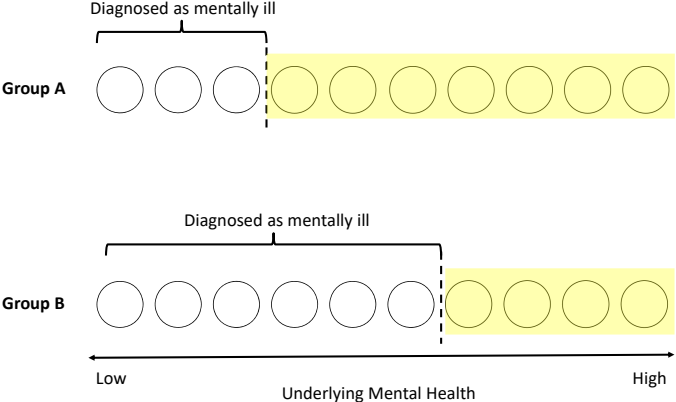
## Monotonicity: Leniency calculated separately on severe and mild diagnoses

Distribution of leniency for severe and mild diagnoses



# Monotonicity

**Monotonicity:** Among all people who are not diagnosed, those who saw the lenient doctor should have higher average health



## Monotonicity: Average mental health of subsample of undiagnosed conscripts

Subsample not diagnosed as mentally ill at the draft

VARIABLES	Mental Illness Diagnosis over next 10 years
Above median leniency	-0.00288*** (0.000652)
Observations	393,285
Dep. var mean	0.04369

# Main Results



## The effect of diagnosis on health at age 30 (2SLS)

VARIABLES	Death up to age 30	Complete suicide up to age 30
Age	(1)	(2)
Draft diagnosis	0.00764*	0.00118
Mental illness	(0.00460)	(0.00250)
Observations	407,162	405,273
% change	123	73
Dep. Var mean	0.0062	0.0016
Time x center FE	Yes	Yes
Nr of clusters	102	102

## The effect of diagnosis on health at age 30 (2SLS)

VARIABLES	Outpatient at age 30	Inpatient at age 30	Sick days at age 30
Age	(3)	(4)	(5)
Draft diagnosis Mental Illness	0.204*** (0.0587)	0.0388* (0.0226)	10.74*** (2.846)
Observations	404,909	404,909	397,566
% change	88	127	244
Dep. Var mean	0.2315	0.0305	4.4078
Time x center FE	Yes	Yes	Yes
Nr of clusters	102	102	102

# The effect of diagnosis on labor market at age 30 (2SLS)

VARIABLES	Unemployed at age 30	Income from work at age 30	Years of schooling at age 30
Age	(1)	(2)	(3)
Draft diagnosis	0.153***	-42,020	-0.551
Mental Illness	(0.0503)	(30,920)	(0.644)
Observations	397,566	402,839	397,440
% change	136	-16	-4.3
Dep. Var mean	0.1125	266799.7621	12.7694
Time x center FE	Yes	Yes	Yes
Nr of clusters	102	102	102

# The effect of diagnosis on wealth at age 30 (2SLS)

VARIABLES	Wealth at age 28 (4)	1(Home owner >0) mean over age 31-46 (5)
Draft diagnosis	-382.4	-0.0482
Mental Illness	(36,590)	(0.149)
Observations	296,258	10,395
% change	-.53	-7.2
Dep. Var mean	71766.4522	0.6723
Time x center FE	Yes	Yes
Nr of clusters	74	102

► Wealth = Financial Assets at Market Value

# The effect of diagnosis on family structure at age 30 (2SLS)

VARIABLES	Married	Divorced
Age	at age 30	at age 30
	(7)	(8)
Draft diagnosis	-0.101**	0.00964
Mental Illness	(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

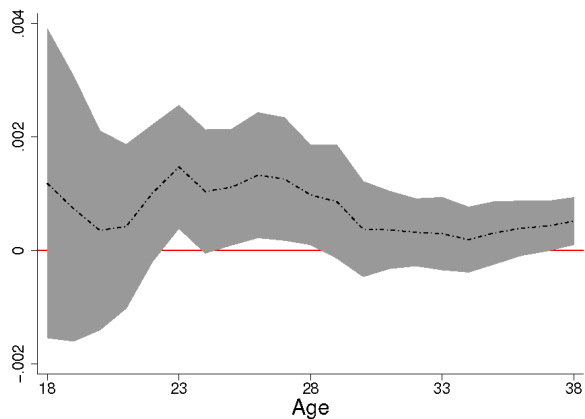
# Results So Far

- ▶ Diagnosis at age 18 makes marginal patient worse off at age 30

# Results So Far

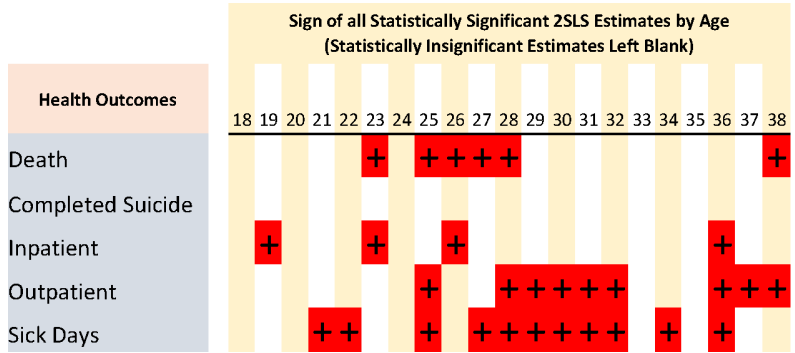
- ▶ Diagnosis at age 18 makes marginal patient worse off at age 30
- ▶ Next: outcomes at other ages

## Effect of diagnosis ages 18 to 38: death (2SLS)

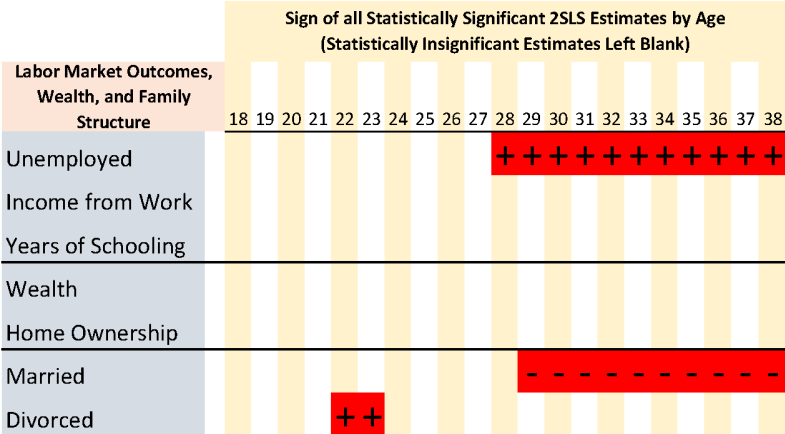




# Effect of diagnosis ages 18 to 38: health outcomes (2SLS)



# Effect of diagnosis ages 18 to 38: other outcomes (2SLS)



# Summary of Results

- ▶ Diagnosis at age 18 makes marginal patient worse off at all ages
  - ▶ Health outcomes are worse
    - ▶ Mortality
    - ▶ Admission to hospital as an inpatient or outpatient
    - ▶ Sick days
  - ▶ Higher unemployment
  - ▶ Lower probability of being married

How does diagnosis affect life outcomes?

## Is the mechanism: diagnosis alters the probability of serving in the military?

- ▶ Conscripts diagnosed as mentally ill are 38.5 percentage points less likely to serve in the military
  - ▶ Is this why diagnosis affects life outcomes?

# Is the mechanism: diagnosis alters the probability of serving in the military?

- ▶ Conscripts diagnosed as mentally ill are 38.5 percentage points less likely to serve in the military
  - ▶ Is this why diagnosis affects life outcomes?
- ▶ We separately measure the causal effect of serving in the military on the same set of outcomes
  - ▶ Borrow the identification strategy first used by:
  - ▶ *“Randi Hjalmarsson, Matthew J Lindquist, The Causal Effect of Military Conscription on Crime, The Economic Journal, Volume 129, Issue 622, August 2019, Pages 2522–2562”*

# Is the mechanism: diagnosis alters the probability of serving in the military?

- ▶ Conscripts diagnosed as mentally ill are 38.5 percentage points less likely to serve in the military
  - ▶ Is this why diagnosis affects life outcomes?
- ▶ We separately measure the causal effect of serving in the military on the same set of outcomes
  - ▶ Borrow the identification strategy first used by:
  - ▶ *“Randi Hjalmarsson, Matthew J Lindquist, The Causal Effect of Military Conscription on Crime, The Economic Journal, Volume 129, Issue 622, August 2019, Pages 2522–2562”*
- ▶ Exploits random assignment of conscripts to officiator and variation in the influence of each officiator
  - ▶ Caveat: effect of service may be different for conscript with marginal mental health

## The effect of Military Service, 2SLS

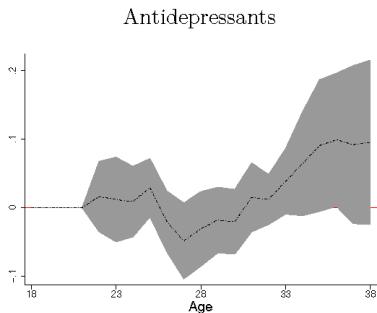
	Sick days at age 30	Unemployed at age 30
1(Military service>0)	-1.323 (1.840)	-0.0618*** (0.0139)
Observations	256,770	256,770
% change	-30	-58
Dep. Var mean	4.4513	0.1068
Nr of clusters	70	70

- ▶ Multiplying these estimates by 38.5% and subtracting from original 2SLS estimates only partially offsets our estimated effects
  - ▶ Example: Effect of diagnosis on number of sick days (unmediated by military service) becomes 10.2 days



# Does the harmful effect of diagnosis come through increased exposure to antidepressants?

Figure 15: Event time evolution of 2SLS estimates for **effect of mental illness diagnosis on future medical treatment**



- ▶ Prozac was first prescribed in Sweden in 1991

## Conclusion: Evidence points to overdiagnosis of mental illness

- ▶ First evidence of long-term effects of a mental illness diagnosis for the marginal patient
- ▶ Being diagnosed mentally ill at 18 has harmful effects on mortality, health, employment
  - ▶ Same conclusion at any horizon in the 20 years after diagnosis
  - ▶ This effect remains after removing the effect of diagnosis mediated by military service

## Conclusion: Evidence points to overdiagnosis of mental illness

- ▶ First evidence of long-term effects of a mental illness diagnosis for the marginal patient
- ▶ Being diagnosed mentally ill at 18 has harmful effects on mortality, health, employment
  - ▶ Same conclusion at any horizon in the 20 years after diagnosis
  - ▶ This effect remains after removing the effect of diagnosis mediated by military service
- ▶ Applying our results to the diagnosis of mental illness outside of the military
  - ▶ Diagnosis rate in our sample: 2.84%
  - ▶ In the Swedish general population in 2014: 7.7% of 18-year old men were diagnosed with a mental illness

## Conclusion: Evidence points to overdiagnosis of mental illness

- ▶ First evidence of long-term effects of a mental illness diagnosis for the marginal patient
- ▶ Being diagnosed mentally ill at 18 has harmful effects on mortality, health, employment
  - ▶ Same conclusion at any horizon in the 20 years after diagnosis
  - ▶ This effect remains after removing the effect of diagnosis mediated by military service
- ▶ Applying our results to the diagnosis of mental illness outside of the military
  - ▶ Diagnosis rate in our sample: 2.84%
  - ▶ In the Swedish general population in 2014: 7.7% of 18-year old men were diagnosed with a mental illness
- ▶ Open questions:
  - ▶ Other ages? What about for women?

# Are We Overdiagnosing Mental Illnesses? Evidence from Randomly Assigned Doctors

Marieke Bos (Swedish House of Finance, SSE) and  
Andrew Hertzberg\* (Philadelphia Fed)  
with Andrés Liberman (Stern, NYU)

July 2020

\*Disclaimer: The views expressed in this paper do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System