THE UNIVERSITY OF CHICAGO

WHAT GETS MEASURED (GETS DONE): ESSAYS ON HEALTHCARE, POLICY, AND PERILS OF THE SUBJECTIVE MADE OBJECTIVE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE IRVING B. HARRIS
GRADUATE SCHOOL OF PUBLIC POLICY STUDIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
YIHE NINA GAO

CHICAGO, ILLINOIS
JUNE 2018

I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.

–William Thomson, Baron Kelvin, 3 May 1883

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

For paper on the RBRVS, thanks to Christopher Berry, B. Pablo Montagnes, and Robert Lalonde for their guidance with this project while it was a nascent proposal originating in Political Economy 3. Further thanks for comments from Damon Jones, Dan Black, and David Meltzer.

The papers pertaining to mental health would not have existed without the clinical teaching of Michael Marcangelo. Dan Black, Damon Jones, Neale Mahoney, and Robert Gibbons have supplied incredible methodological guidance. My ignorance on the topic of psychiatric payment and services has been partially ameliorated thanks to the time and patience dedicated by Marcus Swan, Sam Scharff, Daniel Yohanna, and Peter Nierman. Owing only to the existence of these two independent groups of experts, I have thus far managed to pass as half-competent in front of the each, borrowing the knowledge from the other.

ABSTRACT

The topic of this defense is the extent policy studies can be used to address current problems in healthcare policy. As the subtitle states, the dissertation will present three essays on the topic of the subjective made objective. The first and second papers investigate the ramifications of two present policies, which attempted to reshape payments to fairly reflect two subjective quantities: patient need for inpatient psychiatric care and physician work, respectively. In both cases, unintended consequences can impact our interpretation of these policies and their repercussions, though these unintended consequences require some canniness to measure. The third paper is a departure from this tact. It will critique a commonplace measure for subjective mental health and give my suggested improvements.

In the first paper, I examine the assertion that there is substitutability in the populations of psychiatric hospitals, medical hospitals, and jails. I suggest that abrupt changes in psychiatric bed supply in local-area time series can supply a source of exogenous variation to examine the effects of psychiatric beds on hospitalizations and jail populations in the short-run. I present some evidence in favor of a previous claim that psychiatric bed reductions result in patient spillovers across settings, however, I also present some caveats. Some patients are much more likely to spill between settings than others and the effect of bed additions is not simply the inverse effect of bed reductions.

In the second paper, I investigate the extent to which medical specialty representation on a committee may have affected medical payments. The committee, the Resource-Based Relative Value Scale Update Committee, is charged with assigning a price for subjective physician work in Medicare. I examine whether rotating specialty representation correlates with specialty-specific payments, potentially exacerbating a generalist-specialist income gap. I find that rotating representation and specialist payments are indeed correlated, and in such a way that specialties prefer to raise the reimbursements most sharply on the procedures that are most unique to them.

In the third paper, I describe some limitations of the current measures of mental health.

I perform some tests to reveal properties of mental health metrics that could be useful for investigators. Most importantly, I recommend more adaptable metrics be used for policy studies with methods of analysis borrowed from the psychometrics or clinical psychiatry literature rather than the wholesale borrowing of instruments. In other words, I wade into the perils of subjective measurement myself.

CHAPTER 1

WHERE HAVE ALL THE INPATIENTS GONE? SPILLOVERS AND PROVIDER DISCRETION IN PSYCHIATRIC ADMITTING.

1.1 Introduction

Between 1955 and 2000, the number of state hospital psychiatric beds declined from 339 per 100,000 population to 22 per 100,000. Over the same time period, the fraction of mentally ill inmates within the criminal justice system rose dramatically: the National Inmate Survey in 2011-2012 found that 44 percent of jail inmates had been previously diagnosed with a mental illness and 63 percent in 2007-2009 met Diagnostic and Statistical Manual IV (DSM-IV) criteria for drug dependence or abuse [23, 24].[1] As a result, jails in many localities have expanded psychiatric care services, although the cost of such services has additionally garnered attention both for its moral implications and its inefficient use of public finances [35]. Publicly funded expansions to psychiatric inpatient capacity have been proposed as a potentially cost-effective redress toward the observed mental health caseloads in the criminal justice system, under the assumption that care for the mentally ill should be primary demesne of specialized psychiatric care providers and not the criminal justice system [96].

In addition to the potential criminal justice spillovers, general medical hospitals have seen an increase in psychiatric case load [2, 9]. Because the Emergency Medical Treatment and Labor Act of 1986 (EMTALA) prohibits medical providers from rejecting patients on the basis of ability to pay, some writers have proposed that emergency rooms and hospitals have become mental health providers of penultimate or ultimate resort [69, 73]. Care at general medical hospitals is more expensive compared to psychiatric specialty care, and thus, psychiatric care is provided to mentally ill patients at general hospitals comes at significant

1. By comparison, the Bolton study in 1976 of county jail inmates in California reported 6.7 percent of detainees were judged to have psychotic mental illness and 9.3 percent nonpsychotic mental illness [63].

public cost [72]. Again, an expansion to psychiatric bed supply has been proposed as one potential solution [9, 69, 73]. However, in order for expansions to psychiatric bed supply to meet any of these expectations, it must be true that greater psychiatric bed supply effectively prevents mentally ill patients from "spilling over" into non-psychiatric settings such as jails.

The contribution of this paper is two-fold. First, I provide a concise model for thinking about psychiatric bed policy. That is, while many current models abstract away from provider decisions, I propose that providers possess a level of discretion in the production of spillovers. I speculate that providers may potentially reject expensive or unfavorable patient populations. This model leads to a more nuanced picture of how spillovers should be appropriately measured and where. In particular, it makes the case that a large empirical study that averages the effects of bed supply changes is sensible.

Second, I conduct just such an empirical study. In contrast to previous studies that have treated psychiatric bed supply as exogenous across even long periods of time, I assume only that abrupt, short-run changes are exogenous. I find discrete changes to the time series of psychiatric bed supply within a local area and follow the short-run effects for psychiatric admissions at general hospitals and jail populations. My identification technique is reasonable because psychiatric hospitals face large fixed costs per ward and are therefore likely to abruptly decrease bed supply in the short-run. This identification is also favorable because it permits an investigation into the immediate local effect of psychiatric bed capacity changes, which contends both with the problem of underlying time trends and possible reverse causality, both problems in the existing literature. At the same time, by aggregating many repeated local experiments, I am able to study the general effects due to large psychiatric bed reductions, rather than concentrating on a single local example.

More generally, this paper presents a methodological contribution to the empirical field. Oftentimes systemic changes are gradual, making empirical identification difficult to achieve. I demonstrate that using motivated local-level changes represents a feasible redress to the empirical problem in the absence of some large top-down policy. As such, the methodology opens many new research questions up for empirical investigation.

The content of the remainder of this paper is as follows. Section 2 will discuss psychiatric context, particular with regard to the medical context of psychiatric practice. Section 3 will propose a simple model for understanding psychiatric spillovers. Section 4 will discuss the data and methodology used for the empirical analyses. Section 5 will present some preliminary findings, which will be discussed further in Section 6.

1.2 Background

This idea of that spillover effects could exist from psychiatric policy is not new. Penrose famously hypothesized that psychiatric commitment could act as a functional substitute for incarceration [82]. Subsequently, Penrose's hypothesis has yielded numerous examinations and reexaminations [56, 36]. In spite of the opinions espoused by advocates, however, the results from historical studies have been considerably more equivocal. Recently, [85] sought to detect potential spillovers between psychiatric hospitals and prisons by comparing of the demographics of psychiatric inpatients and prisoners in the decades following deinstitutionalization, a psychiatric policy movement during the 1960s that drastically reduced the aupply of state psychiatric beds. [85] found that within the initial wave of deinstitutionalization, between 1950 and 1980, spillover rates were not detectable. Subsequently from 1980-2000, a period of much slower psychiatric capacity decline, they found a much higher potential for spillover [85]. Many other studies have made use of nationally aggregated or state-level data and have exhibited similar variability in findings [56, 77].

Studies of potential psychiatric bed effects on general medical hospital admissions comprise a briefer, and more recent literature. Study effects from this literature have been similarly equivocal. [93] examine changes following bed closures of a single public psychiatric hospital closure in San Francisco and [73] examine the effects from the Medicaid Psychiatric Emergency Demonstration Project. Neither study found spillover effects from psychiatric capacity increases or reductions into the general medical setting, however, in both cases, the investigators focused on a relatively small number of local beds relative to the population size. [103], on the other hand, conducts a historical study similar in setup to studies in the criminal justice literature and finds a positive effect.

Given that many studies currently exist, the need for an additional study with improved identification comes from a default common to many of the previous study designs. Changes in psychiatric technology confound many of the above historical studies, which treat psychiatric bed supply as exogenous. Chlorpromazine, the first widely used antipsychotic medication, was released in 1955. Chlorpromazine had already transformed inpatient state psychiatric institutions before the passage of the Community Mental Health Act in 1963 and deinstitutionalization [92, 76]. Over the course of the ensuing decades, it, and other psychoactive medications, obviated the necessity for many patients to occupy psychiatric beds long-term by making community and outpatient treatments possible via medication [84]. These large changes in treatment practices mean that, over the long time periods used by many historical studies, psychiatric bed supply is almost certainly endogenous with period technology.

At present, psychiatric stays account for a large volume of hospital stays overall, and, in particular, psychiatric stays account for a large share of overall stays for patients with Medicaid. According to the 2012 National Inpatient Sample, a national sample of hospital discharges, mood disorders accounted for 847,000 inpatient hospital stays (12% of total discharges), schizophrenia for 383,000 inpatient stays (5%), and aggregated hospital inpatient stays for substance use accounted for between 300,000 and 400,000 stays (4-6%). By comparison, myocardial infarctions (heart attacks) accounted for 608,000 (9%) and pneumonia for 1,007,410 (14%). Among the subpopulation of patients with Medicaid, mood disorder was the single most common pathological diagnosis at discharge, schizophrenia was number five [48].[2]

2. I say pathological diagnosis here to distinguish diagnoses due to some medical or psychiatric pathology from admissions secondary to live births, which comprise the vast majority of Medicaid stays.

Current psychiatric inpatient care can be divided into two broad categories of care: crisis care and non-emergency care. Crisis care is targeted toward addressing psychiatric concerns chiefly pertaining to the assessment of a patient who may be acutely dangerous to either himself or to others. The goal of crisis care is stabilization of the patient. Crisis patients will often be triaged from an emergency department and transported either to an available in-facility psychiatric bed or transported to a separate psychiatric facility if no local beds are available. By contrast, non-emergency care often results from an outpatient referral to a specialized treatment center as a continuation or escalation of outpatient care for an ongoing psychiatric problem. For example, a non-emergency treatment center may specialize in alcoholism and accept referrals from physician's offices in order to provide short-term inpatient detoxifications or longer-term stays for patients who have not been successful at abstaining from alcohol in an outpatient setting. While such programs have declined in popularity since the early 1990s, many are still in existence.

Many historical commentaries on psychiatric inpatient service provision have focused their attention on the lengthy, involuntary commitments once common in the United States and other countries during the former half of the twentieth century [82, 36, 91]. Since this time, however, treatment trends have changed significantly. Currently, involuntary commitment statutes are variable by state, but frequently require proof that the patient is an imminent danger to himself or others due in part to the precedent established by O'Connor v. Donaldson (422 U.S. 563). Involuntary stays now comprise only the minority of psychiatric stays. In correspondence with the decline in involuntary commitments, the average psychiatric length of stay has fallen significantly since 1950, and has continued to shorten over the past 20 years. The average length of stay for patients from aggregated privately insurance claims from 1993-1995 was 13.1 days and an average of around 35 days for patients admitted at the Department of Veteran's Affairs or at a state hospital [66, 95]. By 2006, target lengths of stay had shortened considerably: 3-7 days for crisis management or 10-14 days for longer term stays [26, 98].[3]

While some of the observed changes in psychiatric hospitalization may have resulted from legal or judicial changes, in my interpretation, the origin of the observed changes in hospitalization patterns stem from more gradual transformations in attitudes, psychiatric guidelines, and practice norms. For example, from 1995 to 2005, practice guidelines for inpatient stays reflected a change in attitude toward the standard of care for substance users. In accordance with the American Psychiatric Association practice guidelines for substance users, "residential treatment of 3 months or more is associated with better long-term outcome in [appropriate patients]" [5]. By its practice guideline update in 2006, statements regarding the appropriateness of residential treatment had been significantly redacted and revised:

> Randomized, controlled trials have demonstrated that some individuals who would ordinarily referred to residential- or hospital-level care do just as well in [an outpatient program with daytime programming of 20 hours per week]... The duration of residential treatment should be dictated by the length of time necessary for the patient to meet specific criteria that would predict his or her transition to a less structured, less restrictive treatment setting [6].

Once in place, established practice guidelines could be enforced by payers through reimbursements. Because clinician assessments are subjective and not readily confirm-able via laboratory or radiographic testing, and psychiatric inpatient providers are more often reimbursed per diem than by diagnosis-related group, psychiatric hospitals had an incentive to keep patients for lengthy stays. During the 1990s, many managed care payers were successful at enforcing guidelines for a shorter standard-of-care. Subsequently, managed care organizations became dominant in the commercial psychiatric payer marketplace.

There is a styled fact in psychiatry that inpatient psychiatric care is inherently unprofitable, however, the truth is more subtle. While inpatient psychiatric care is reimbursed at lower rates in comparison to inpatient medical care, the former is generally also less costly

3. It may be worth noting that while the large reductions in stay from the 1950s onwards were observed not only as the isolated phenomenon in the United States, but occurred broadly throughout the Western World [59].

to provide. The profitability of inpatient psychiatric care varies significantly across practice environments. [50], which is often cited to demonstrate psychiatric service unprofitability, illustrates some of this heterogeneity in profitability across psychiatric services: while alcohol and substance abuse inpatient services were more likely to be offered by for-profit hospitals in comparison to non-profit and government hospitals, the opposite was true for psychiatric emergency services. As suggested by the previous discussion on managed care, inpatient psychiatric care prior to the 1990s can also be understood to have been quite profitable for providers in comparison to the current time.

Recently, psychiatric bed policy has returned the national consciousness through two recent trends. The first is the discussion of emergency room "boarding" by psychiatric patients [2, 9, 69, 40]. The 2011 Medicaid Emergency Psychiatric Demonstration is one an example of national policy motivated to reduce the observed psychiatric burden on emergency rooms. The second reason for the discussion is the rise in opioid deaths. This triggered Congressional introduction of the Medicare Coverage for Addiction Recovery Expansion Act bill, which also seeks to increase the psychiatric bed supply [33]. In particular, both these policies seek to address the Medicaid Institutions for Mental Disease (IMD) Exclusion, a longstanding historical rule that prohibited federal Medicaid funds from going to free-standing psychiatric hospitals.

Due to the discussion surrounding the Medicaid IMD exclusion, there are several studies that seek to determine a single number that would represent the psychiatric bed supply per population. Estimates from such papers very widely, from 22 to 40 per 100,000, and are often derived directly from psychiatric bed wait times of patients at general medical hospitals [28, 62]. However, existing psychiatric bed supply, estimated at 28 beds per 100,000, does not differ significantly from these recommendations [97].

1.3 Model

1.3.1 Model Parameters

Imagine there are two settings, a psychiatric setting and a non-psychiatric setting, each of which has its own admission criteria. In this example, I will consider a general medical hospital as the non-psychiatric setting, although, in principle, the same short-run model can be used to as a framework for thinking about spillovers into jails or other displacement settings. The criteria across the settings have some natural overlap, however, the settings are less than perfectly substitutable. Consider a population of N individuals, indexed by $i = 1, ..., N$. Each individual has some latent physical health, ϕ_i, mental health ψ_i, and willingness-to-pay for psychiatric services, Y_i. Now define the admission criteria. Call the psychiatric admission score $S_i = g(\psi_i, Y_i)$, where

$$\frac{\partial S_i}{\partial \psi_i} \geq 0 \quad \text{and} \quad \frac{\partial S_i}{\partial Y_i} \leq 0. \tag{1.1}$$

A hospital must make an admission decision for each individual $A_i^S \in \{0, 1\}$. Patients will be admitted to psychiatric hospital, $A_i^S > 0$, if their score falls below some threshold, $S_i < \bar{S}$. For each pair of patients i and j, with psychiatric admission scores $S_i < S_j$, the hospital prefers to admit i over j. In the short-run, the number of beds for each type is fixed at B.

In the short-run, the psychiatric hospital triages patients by their preference such that the patients with the lowest S_i are admitted first. I.e.

$$\min_{A_1^S, ..., A_N^S} \sum_{i=1}^{N} S_i A_i^S \tag{1.2}$$

subject to the constraints

$$A_i^S \in \{0,1\} \qquad \text{for all } i \tag{1.3}$$

$$\sum_{i=1}^{N} A_i^S \leq B. \tag{1.4}$$

Here A_i^S indicates that fraction of time for which i has been admitted to psychiatric hospital, B is the number of effective beds. Ceteris paribus, the relationship between S_i, ψ_i, Y_i this implies that patients with worse mental health are more likely to be admitted to psychiatric hospital and patients with higher willingness-to-pay for psychiatric services are more likely to be admitted to psychiatric hospital.

Notice, given a small number of psychiatric beds relative to the total population, $B << N$, thresholds $\bar{S}$ is implied directly from bed supply. This is because, given a fixed bed number and per-diem reimbursements, it is in the hospital's incentive to keep the ward full. The shadow price of admission is an individual-level constant equal to the psychiatric admission rank. Thus, at the optimum, the psychiatric hospital will rank all individuals to generate the ordered list $(S_{(1)}, S_{(2)}, ..., S_{(N)})$ such that $S_{(1)} < S_{(2)} < ... < S_{(N)}$.

The definition for the admission criteria for general hospitals is similarly structured. Call the general hospital admission score $H_i = h(\phi_i, \psi_i)$. Unlike psychiatric hospitals, medical hospitals in the U.S. are reimbursed by diagnosis-related group. In this model, I assume that more severe illness also corresponds to higher payments, thus, there is no parameter for medical willingness-to-pay. Patients are admitted to general hospital if their general hospital admission score falls below some threshold, $\bar{H}$, and patients with lower H_i are more likely to admitted to general hospital. That is, for each pair of patients i and j, with psychiatric admission scores $H_i < H_j$, the hospital prefers to admit i over j. Let

$$\frac{\partial H_i}{\partial \phi_i} \geq 0 \quad \text{and} \quad \frac{\partial H_i}{\partial \psi_i} \geq 0. \tag{1.5}$$

Here the problem is analogous to the psychiatric case, except the number of hospital beds is B^H. Given the relationship between H_i, ϕ_i, ψ_i, patients with worse mental health and worse physical health are more likely to admitted to general hospital. As in the psychiatric case, general hospitals choose whether to admit a patient, $A_i^H \in \{0,1\}$. General hospitals admit patients whenever $H_i < \bar{H}$.

1.3.2 Candidates for Admission in Both Settings

The length of stay is very short in both locations in comparison to the period length, then hospitalization in one setting does not obviate hospitalization at the other. Thus

$$\begin{aligned} A_{(k)}^{S*} &= 1 \quad \text{when } 1 \leq k \leq B \\ A_{(j)}^{H*} &= 1 \quad \text{when } 1 \leq j \leq B^H \end{aligned} \tag{1.6}$$

where k is the ordinal index for individual i by psychiatric admission rank and j is the ordinal index for individual i by the general hospital admission rank. All other cases receiving $A_i^{S*} = A_i^{H*} = 0$. Notice that, the admission thresholds for each setting are then, correspondingly

$$\begin{aligned} \bar{S} &= S_{(k)} \quad \text{where } k = \sup\{x \in \mathbb{N} | x < B\} \qquad \text{and} \\ \bar{H} &= H_{(j)} \quad \text{where } j = \sup\{x \in \mathbb{N} | x < B^H\}. \end{aligned} \tag{1.7}$$

However, if a patient cannot be admitted both to general and psychiatric hospital, then we have a difficulty, which is to contend with the group of patients who are eligible for admission in both settings, but who cannot be in two places at once. That is, patients with $S_{(k)}$ such that $1 \leq k \leq B$ and $H_{(j)}$ such that $1 \leq j \leq B^H$ or, in threshold terms, $S_i < \bar{S}$ and $H_i < \bar{H}$. It is this group of patients that are responsible for generating the entirety of the spillover effect, yet the criteria used by providers in these circumstances is unknown.

I suggest a simple illustration to understand substitutability for this group through examining two tail cases. First, imagine that all of the patients who fall into this category are admitted to the general hospital. One possible justification for this is to say that general hospitals have priority in admitting patients since they provide emergency departments. Under this condition, we find that the optimal values of A_i^H are unchanged, and the threshold for admission is only shifted by $d := F_{SH}(\bar{S}, \bar{H}) \cdot N$, where F_{SH} is the cumulative distribution function for S_i and H_i. In this case,

$$\begin{aligned} A_{(j)}^{H*} &= 1 \quad \text{when } 1 \leq j \leq B^H \\ A_{(k)}^{S*} &= 1 \quad \text{when } 1 \leq k \leq B + d \text{ and } A_i^{H*} = 0. \end{aligned} \tag{1.8}$$

We can see immediately that there is still no cross dependence of general hospital admissions with psychiatric bed numbers.

Now, imagine that all of the patients who fall into this category are admitted to the psychiatric hospital. Under this condition, we find that the optimal values of A_i^S are unchanged, and the threshold for admission is only shifted by $d := F_{SH}(\bar{S}, \bar{H}) \cdot N$, where F_{SH} is the cumulative distribution function for S_i and H_i. In this case,

$$\begin{aligned} A_{(k)}^{S*} &= 1 \quad \text{when } 1 \leq k \leq B \\ A_{(j)}^{H*} &= 1 \quad \text{when } 1 \leq j \leq B^H + d \text{ and } A_i^{S*} = 0. \end{aligned} \tag{1.9}$$

However, now the number d creates cross-dependence of A_i^{H*} on B.

When the number of psychiatric beds, B increases, this implies an increase to the threshold $\bar{S}$. This can affect hospital admissions in one particular way. As $\bar{S}$ increases, presumably a larger proportion of patients who would have formerly qualified for general hospital admission only now meet criteria for both settings. Since we presumed that patients who are eligible for admission in both settings are hospitalized at the psychiatric hospital, these patients may "spill" over from the general hospital to the psychiatric hospital.

The probability that patients switch settings can be written as:

$$\Pr\left[A_i^{S*}(B_1) = 1, A_i^{H*}(B_0) = 1\right] \tag{1.10}$$

where B_0 is the number of beds in the pre-period and B_1 is the number of beds in the post-period and $B_1 > B_0$. Alternatively this can be written as:

$$\Pr\left[(S_i < \bar{S}_1 \cup H_i < \bar{H}) \cap (S_i > \bar{S}_0 \cup H_i < \bar{H})\right] = F_{SH}(\bar{S}_1, \bar{H}) - F_{SH}(\bar{S}_0, \bar{H}). \tag{1.11}$$

Here $\bar{S}_1$ and $\bar{S}_0$ represent the thresholds implied by B_1 and B_0. From the representation of this effect in the form of the cumulative distribution function between S and H, we see that this value is non-negative if $\bar{S}_1 > \bar{S}_0$.

1.3.3 Model Implications and Extensions

First, in order to have cross-dependence in admissions, the model necessitates that a patient cannot be admitted to both a psychiatric hospital and medical hospital within the same period. While historical accounts illustrate that lifetime psychiatric institutionalization was previously common, psychiatric stays are now much shorter than in previous decades. Is it plausible that a 5-7 day psychiatric stay out of a 6 month period could prevent a general hospital admission? In fact, this mechanism is plausible due to the acute-on-chronic picture of many mental illnesses. For example, consider a patient with bipolar disorder who becomes a candidate for hospitalization each time he experiences a manic episode. His manic episodes recur with some average periodicity, during which time the eligibility for third party coverage of a hospital stay renews. If we define the total period length to be the duration of one manic episode, which may last perhaps 7 days, a relatively short 5-7 day long stay at a psychiatric hospital may now appear to be a significant proportion of the total period length. Many psychiatric disorders treated by psychiatric settings might exhibit this acute-on-chronic disease course including bipolar disorder, major depression, and schizophrenia, also known by inpatient psychiatric providers as “the big three” due to the proportion of psychiatric admissions attributable to these three disorders [68].

An interesting extension of this principle of length of stay is to consider a dynamic case, in which the admission decision must be repeated. Specifically, consider the case of the overlap in population between psychiatric and criminal justice settings (i.e. jails or prisons). Here, I will borrow the same short-run model as for general hospitals, only we can consider ϕ_i as some parameter of criminality rather than physical health. Unlike hospital stays, for which the length of stay can be enumerated in days, jail or prison admissions can result in stays months, years, or decades in duration. Additionally, while the total number of beds is large, the number of free beds is small and defined by some growth rate in the number of beds and the small number of prisoners who are released each period. The single-period model can be extended to consider a Markov process in which, during each period, B^H is defined by some number of prisoners that are released each period or some small number of beds that are added. N decreases to reflect the total number of eligible patients excluding the total number of prison beds. The expected spillover may, therefore, take on a shape distinct from the general hospital spillover case were, due to a small number of available beds, the spillovers from a psychiatric bed reduction may take multiple periods to be absorbed and to reach new equilibrium levels. Once these levels are reached, however, an abrupt increase in the number of psychiatric beds does not absorb these patients immediately, since they have been removed from the common pool. If B^H is driven more by bed expansion than outward transitions, this model also implies that the effect of psychiatric bed reductions and additions is inherently asymmetric.

With respect to a test that is relevant to our empirical investigation, this model makes several suggestions. Previous literature such as [73] and [93] found that opening psychiatric beds did not significantly impact spillovers to general hospitals. To what extent can these studies be used to perform inference regarding the value of psychiatric beds and the distributional implications? In order for a spillover to occur to general hospitals, there must be some population of patients who would have qualified both for admission at both the psychiatric and the general hospital. If no such population exists, for example, because medical and psychiatric catchments are defined as disjoint everywhere, then we should expect a measured spillover reduction of zero. On the other hand, no spillovers might occur if psychiatric hospitals disproportionately admit patients who are higher income rather than lower mental health, since these mental health and income are positively correlated in the population [78, 51, 83]. It is possible to distinguish between these two stories using data. In general, if psychiatric hospitals triage patients with respect to the severity of their mental health, then the average mental health of admitted patients should be lower when the number of beds is low. If psychiatric hospitals triage patients with respect to their ability-to-pay, the population correlate should drive the correlation in the opposite direction; that is, as the number of beds rises, the average mental health for an admitted psychiatric patient decreases.

1.4 Methods

1.4.1 Data

Psychiatric bed data at the provider level is available for providers of services to the Centers for Medicare and Medicaid Services (CMS) within the Provider of Services (POS) dataset 1991-2015. Additional provider-level data from the American Hospital Association (AHA) Annual Survey from 1972-2014 was used to supplement data on hospital ownership, the number of psychiatric beds, and the availability of psychiatric services. All provider-level data was aggregated to the HRR-level using files available from the Dartmouth Atlas.

Claims data for patients are reported from two sources. The first is the National Inpatient Sample from the Healthcare Cost and Utilization Project (NIS) managed by the Department of Health and Human Services. It is a database of hospital discharge data available for years 1988-2014. A second source of claims is MarketScan 2003-2014, a large-scale commercial claims database including complete longitudinal information from enrollees during the time period of data availability including inpatient use, emergency room use not leading to admission and pharmacy. Data for inpatients are reported at an admission or visit level and then aggregated to the hospital referral region (HRR)-level. HRRs are defined by the Dartmouth Atlas.

Jail data was available from the Annual Survey of Jails or Census of Jail Inmates for the years 1985-2014, subject to data availability in any given year. Jail data was intermittently conducted at the individual, jail or jurisdiction level. All levels of data were aggregated to the HRR-level for analysis.

1.4.2 Identifying Discontinuities

Several possible sources of psychiatric bed variation may be seen as sources for a quasi-experimental study on the repercussions of psychiatric bed reductions. In particular, there are reasons to use the psychiatric capacity reductions from a full sample rather than state psychiatric hospital beds alone, as has been previously suggested [97]. Consider examining the sequelae of state psychiatric hospital closures. By the 1990s and early 2000s public provision of psychiatric inpatient services comprised only half of total inpatient service provision. Thus, an investigation only on state hospitals would not permit us to examine the effects from changes in psychiatric bed supply in half the market. Additionally, state psychiatric hospitals were disproportionately built during an era when psychiatric hospital catchment was defined more broadly than current local markets. When the large, state-wide catchment area of state psychiatric hospitals is considered, a specification relying only on state hospital supply variation leaves few observations per year for the analysis of psychiatric bed reduction effects. As an alternative, it is possible to include psychiatric beds capacity reported by privately operated hospitals. Such psychiatric capacity reductions are not as widely reported in news sources compared to state hospital closures, however, they nevertheless allow us to examine the number of beds available at an HRR-level.

There are reasons to believe that abrupt changes in the number of psychiatric beds should exist within a time series of secular psychiatric capacity decline. Firstly, the operational costs for keeping an inpatient unit in a hospital are weighted heavily toward fixed costs rather than cost per marginal patient. On the other hand, revenue is received on the treatment of the marginal patient day. This cost structure is true too for psychiatric inpatient units [84]. As a result, while the unit is in operation, the hospital's incentive dictates that each bed should be occupied. When a unit is closed, the hospital's incentive dictates that the entire unit should be closed at once in order to avoid the fixed cost. The result is that within a time series for a given HRR, discrete bed additions and bed reductions can be found.

Discontinuities within HRR were identified as follows. Medicare POS and AHA data sets contain zipcode information for hospitals, but not matched hospital identifiers. Hospitals were matched on the basis of their location and the number of reported beds. For hospitals with missing psychiatric bed data, the last available psychiatric bed number and the first available psychiatric bed number following the gap were identified. To avoid the identification of missing data as a discontinuity, a line was fit from the last available to the first available date across gaps, following [85]. Subsequent discontinuities were tagged using the dataset of these predicted values substituting for missing data. At total of 4230 hospital-year pairs were filled in this way.

For each HRR present in the psychiatric beds data, kernel smoothing was performed on the number of psychiatric beds by HRR using an Epanechinikov kernel and bandwidth of 2. Such a low bandwidth was selected in order to minimize false positives, which would attenuate findings in the second stage. Kernel estimates were made for each year on [3,T-2] from both above (years previous) and below (years following). In other words, kernel estimates from below for psychiatric capacity in HRR i and year t, $\hat{f}^{-}(PsychCap_{it})$, were obtained for each $t \in [3, T-2]$ as:

$$\hat{f}^{-}(PsychCap_{it}) = \frac{1}{h}\sum_{s=1}^{t-1} k\left(\frac{PsychCap_{is} - PsychCap_{it}}{h}\right) \tag{1.12}$$

Correspondingly, kernel estimates from above, $\hat{f}^{+}(PsychCap_{it})$, were found by:

$$\hat{f}^{+}(PsychCap_{it}) = \frac{1}{h}\sum_{s=t}^{T} k\left(\frac{PsychCap_{is} - PsychCap_{it}}{h}\right) \tag{1.13}$$

where $k(u) = \frac{3}{4}(1-u^2)1[|u| \leq 1]$ and h is the bandwidth.

The difference between above and below estimates was then taken. A local T-statistic was constructed to test for significant discontinuities existing between limit from above and limit from below. More specifically, the locally defined T-statistic in HRR i in year t was defined as:

$$T_{it} = \frac{\hat{f}^{+}(PsychCap_{it}) - \hat{f}^{-}(PsychCap_{it})}{\frac{\sigma_{+}^{1/2}}{h+1} + \frac{\sigma_{-}^{1/2}}{h+2}} \tag{1.14}$$

where σ_{+} is the standard deviation of $PsychCap_{it}$ on $t \in [t, t+h]$ and σ_{-} is the standard deviation of $PsychCap_{it}$ on $t \in [t-h, t-1]$. If the p-value obtained on any given year was found to be less than 0.01, the year was marked as a discontinuity for that HRR.

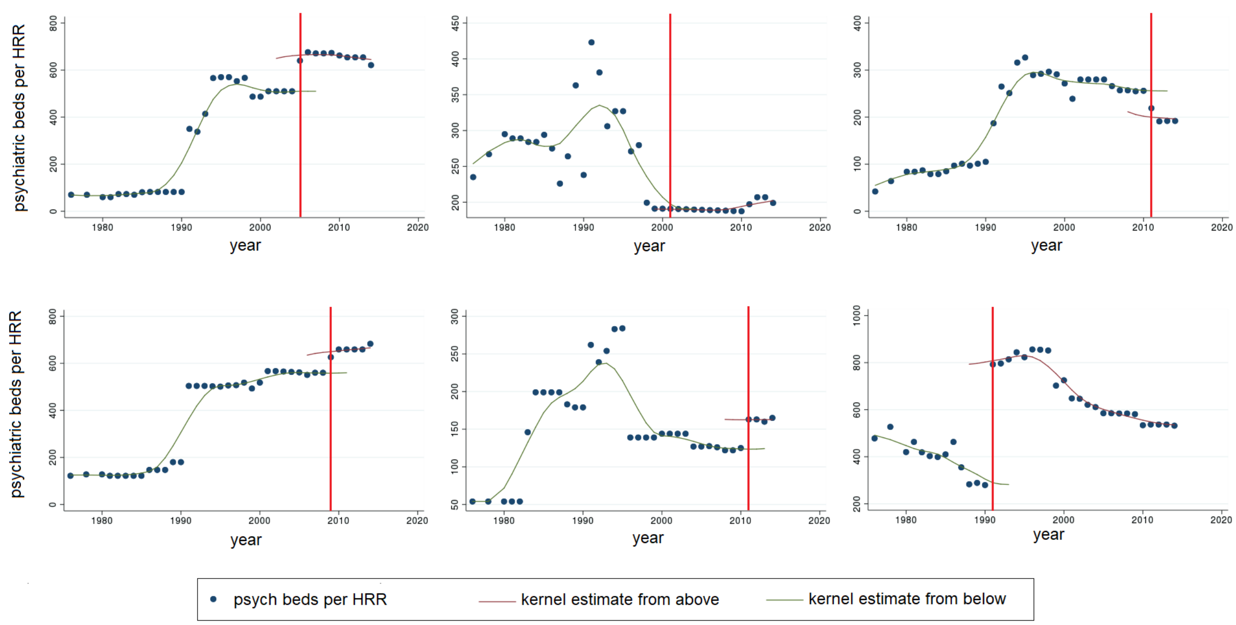
Because POS data begins in 1991, many discontinuities were identified in the year 1991 due to a large increase in the number of hospitals. To resolve the problem of erroneous tagging due to the data merge, any bed addition or reduction discontinuities tagged in 1991 were removed. In total, 84 abrupt increases in psychiatric bed capacity were marked and 109 abrupt decreases in psychiatric bed capacity were marked.

Figure 1.1 presents graphs for six discontinuities, selected at random, and shows the HRR-year psychiatric bed capacity overlaid with kernel estimates from above and below for the identified discontinuity.

1.4.3 Outcome Variables

Given the extensive literature on the relationship between psychiatric care provision and jail populations, I first examine the effect of psychiatric bed reductions and increases on jail populations. A survey for the number of inmates requiring substance use or psychiatric

Figure 1.1: Randomly sampled identified discontinuities at the HRR-level.



Source: Author's calculations of the number of psychiatric beds within an HRR using a merged psychiatric bed numbers provided in the Medicare Provider of Services Data 1991-2014 and the American Hospital Association Survey 1980-2014. Each panel displays a single HRR-year discontinuity from the set of identified discontinuities and may not reflect the total number of discontinuities in the series. The green line represents Epanechnikov kernel estimated from below the year of the identified discontinuity. The red line represents the Epanechnikov kernel estimated from above. Bandwidth specified at two years.

services was not available at an inmate level, however, jails and jurisdictions were surveyed as to whether alcohol or other substance use services are provided. While a number of factors may affect whether a given jurisdiction may provide substance use or mental health services, there are reasons to believe that the probability that a jail will supply substance use services is increasing in the proportion of inmates with mental illness or substance use disorders.[4] I interpret changes in the probability that a jail offers substance use or mental illness service as crudely indicative of changes in inmate composition.

Additionally, I examine several outcomes available in either aggregated commercial payer or hospital discharge data to evaluated the effects of spillovers into general hospital populations. I look at several broad specifications of diagnoses coded by the International Classification of Diseases version 9 (ICD-9). I aggregate psychiatric outcomes into a single outcome variable, however, results for individual outcomes are also presented.

To begin, I examine patients receiving a diagnosis of mental illness. Specifically, patients with an ICD-9 code commencing in "29" were classed as psychosis. Patient receiving a diagnosis of ICD-9 code commencing in "30" and "31" were considered together has having a non-psychotic mental illness. Further, I consider patients with a principal visit procedure of having received a psychiatric evaluation recorded in their inpatient stay record.

While medical practitioners receive training in diagnosing and recognizing psychiatric diagnoses, however, they may be reluctant to record definitive psychiatric diagnoses for a patient with great specificity. Alternatively, if no psychiatric beds are available and a patient is instead admitted for medical reasons, the patient may never receive a psychiatric evaluation or a defined psychiatric diagnosis. To answer the question of whether psychiatric patients may be admitted to general hospitals (albeit with alternative diagnoses), I examine several medical diagnostic categories that may be seen as suspicious for a secondary medical diagnosis. First, I consider patients with a diagnosis of alcoholic pancreatitis, cirrhosis of

4. The Supreme Court ruling in Estelle v. Gamble (429 U.S. 97) in 1976 established that deliberate indifference to a prisoner's medical needs would constitute "cruel and unusual punishment" under the U.S. Constitution. This ruling was subsequently extended to include treatment of mental illness.

the liver, alcoholic cardiomyopathy, and infective endocarditis. Infective endocarditis a rare complication of injection drug use. Each of these diagnoses is associated more with a stock of accumulated poor psychiatric health than with a short term effect of lack of treatment, however, we may wish to consider the effect of short term bed supply changes for those with well-establish psychiatric histories. Due to their association with discrete pathological findings, these diagnoses are easier to interpret compared with the other classes of diagnoses I will consider.

I also look for diagnoses coded under the header "Supplementary Classification of Factors Influencing Health Status and Contact with Health Services." In particular, I aggregate patients coded as having a received a diagnosis of "V60 Housing, household, and economic circumstances," "V62 Other psychosocial circumstances", and "V63 Unavailability of other medical facilities for care", and "V71 Observation and evaluation for suspected conditions not found". This aggregated outcome variable I term "V-Codes." Lastly, I examine patients who have received a diagnosis of "791.9 Other nonspecific findings on examination of urine". V-codes are a relevant outcome for consideration if we believe that patients with mental illness may be displaced if there is a low stock of inpatient psychiatric care. Further, I consider the outcomes of observation status and abnormal findings on urine exam because psychiatric patient populations may be particularly at risk for drug use. When a patient with altered mental status suspicious of substance use is examined in the emergency department, it is standard practice to administer a urine drug screen. If it is determined that a patient's altered mental status is due to the influence of a substance, the patient may then be admitted to a short stay in observation rather than a full medical admission. Although patients may experience displacement, be admitted for observation, or have nonspecific findings on urine exam for reasons other than psychiatric illness, I argue that there is little reason to believe that these outcomes should increase relative to medical diagnoses following changes in psychiatric bed availability other than for psychiatric reasons.

As a control variable for psychiatric outcomes, I construct an index of control diagnoses that should be common across HRRs, but not directly related to psychiatric services. In this case, I choose the cardiac related diagnoses flagged if ICD-9 code is commences with "41", chronic obstructive pulmonary disease with ICD-9 code starting in "49", and pneumonia with ICD-9 code starting in "49".

Variables were demeaned and standardized to have a standard deviation of one. Indexes for both psychiatric and control diagnoses were then computed as the average of these standardized individual outcome variables in accordance to Kling, Liebman and Katz [57]. For clarity:

$$\text{Index Score}_{it} = \frac{1}{J}\sum_{j=1}^{J} \frac{x_{it}^{j} - \bar{x}_{t}^{j}}{\sigma_{x_t^j}} \tag{1.15}$$

where x_{it}^j is the jth variable for HRR i in year t. $\bar{x}_t^j = \frac{1}{N}\sum_{i=1}^{N} x_{it}^j$ is the mean over i of x_i^j and $\sigma_{x_t^j}$ is the standard deviation of x_{it}^j.

1.4.4 Event Study

In the event study portion, I regress outcomes on a series of lagged dependent variables ranging from J years prior to the abrupt change in the number of psychiatric beds to J years after the change. Regressions contain HRR-level and year-level fixed effects.

$$\text{Outcome}_{it} = \alpha + \sum_{J}^{-J} \xi_j L^j Discont_{it} + \eta_t + \mu_i + \epsilon_{it} \tag{1.16}$$

where $L^j Discont_{it} = Discont_{i,t-j}$ for $t \in [1988, 2014]$ for hospital discharge outcomes and $t \in [1985, 2014]$ for jail outcomes. Outcome_{it} is the outcome value for to HRR i in year t, $Discont_{it}$ is an indicator for whether year t was tagged as a year of approach change in psychiatric bed number for HRR i, η_t is the year-level fixed effect for year t, μ_i is the HRR-level fixed effect for HRR i and ϵ_{it} is the error term. Event study figures plot estimates $\hat{\xi}_j$ on the y-axis versus j on the x-axis.

1.4.5 Two-Stage Least Squares

As a summary measure for the overall effect of psychiatric bed capacity on outcomes of interest, I first formulate the ordinary linear specification of:

$$\text{Outcome}_{it} = \alpha + PsychCap_{it}\delta + X_{it}\zeta + \eta_t + \mu_i + \epsilon_{it} \tag{1.17}$$

where Outcome_{it} is the outcome value for to HRR i in year t, $PsychCap_{it}$ is a measure of the inpatient psychiatric capacity, the aggregated number of inpatient psychiatric beds in HRR i in year t, η_t is the year-level fixed effect for year t, μ_i is the HRR-level fixed effect for HRR i and ϵ_{it} is the error term. X_{it} represents a vector of HRR-year level covariates. As discussed in the introduction, there may be reasons to believe that a strong underlying time trend exists across all HRRs, which may induce a correlation between error term ϵ_{it} and inpatient psychiatric capacity. To address the resulting bias, I use a two-stage least squares specification (2SLS) approach using the tagged discontinuities in psychiatric bed provision as an instrument for psychiatric bed number.

There are several reasons to believe that outcome fluctuations in response to abrupt changes in psychiatric capacity represent an improvement over the ordinary least squares model. While from the period 1990s through present has not been characterized by large technological changes or large precipitous declines in psychiatric beds, the general trend of capacity reduction in psychiatric beds nevertheless occurred in nearly all geographic localities [63]. This is partially due to the expansion of managed care in psychiatry and the resulting decline in profitability accompanied by decreases in psychiatric lengths of stay. A naive comparison of changes in psychiatric beds will therefore have a tendency to conflate time trends in hospitalization and incarceration with psychiatric bed declines.

More problematic for identification may be the assumption of pseudo-random variation in private hospital decisions to either expand or decrease psychiatric bed capacity. That is, slow-moving underlying trends do not move abruptly and concurrently with psychiatric bed capacity changes. In order to illustrate that this is not the case, I also present a summary of the regression coefficients from a specification using a randomly tagged placebo discontinuity in Appendix A.4.

All specifications used herein use clustered standard errors at the HRR-level.

1.5 Results

Table 1.1 shows a table of descriptive statistics. Due to differences in overlap for years with psychiatric beds data as well as variability of data reporting at various levels, the number of observations differs across years. Overall, the average number of psychiatric beds available within an HRR was 305 over the study period. Changes within HRR, across time accounted for the majority of observed standard deviation in psychiatric bed provision. A similar trend can be observed for psychiatric service provision. In general, the average volume observed per HRR-year in hospital discharge data exceeds the volume observed from commercial claims. Variation observed across diagnostic categories further differs in that hospital discharges demonstrate a larger between-HRR variation in psychiatric diagnoses versus control diagnoses compared to that, which is found from commercial claims.

Figure 1.2 shows the average level of psychiatric beds per HRR before and after each marked discontinuity. Bed reductions and bed additions are considered separately with bed reductions show in solid line on the left panel and bed additions shown in dashed line on the right panel. Standard errors are shown by the error bars. At the zeroth year, the marked discontinuity in number of beds occurs. In both the case of bed reductions and that of bed additions, there is no significant pre-trend or post-trend within a five-year window for the event. This is to say, the tagged discontinuities appear on average to occur abruptly and not in correspondence to some observed secular trend. The average size for a discontinuous psychiatric bed reduction is roughly 80 beds. The average size for a discontinuous psychiatric bed addition is also roughly 80 beds. Viewed in context of the long run supply presented in Appendix A.1, these discontinuities represent the equivalent of fairly large changes in psychiatric bed numbers, the equivalent of 3-4 simultaneous ward closures.

To check that the number of psychiatric bed changes did not occur in correspondence with global changes in medical supply or abrupt demographic shifts, I examine the pre- and post- trends for HRR population, medical bed supply, and outpatient psychiatric supply in Figures A.5, A.6 and A.7, included in the Appendix A.3. In general, we do not see strong trends in HRR population around the time of the discontinuity, however, psychiatric bed additions appear to be associated in time with medical bed additions, although the size of this effect is not statistically significant. The latter may indicate that some psychiatric bed additions may occur in the context of hospital openings and, therefore, may effect on medical care delivery as well as psychiatric care.

Figure 1.3 shows the aggregated effect by HRR of psychiatric bed reductions and additions at nearby hospitals. All hospitals shown by the solid line, hospitals without psychiatric beds shown in the dotted lines. The left panel shows bed reductions at year 0, with bed additions at year 0 shown on the right. Each line represents the average difference between the psychiatric index value and medical index value for each HRR. Here we see that in the year previous to the reported discontinuity, the average number of psychiatric admissions (relative to medical admissions) across all hospitals in the HRR begins to decline in HRRs with a tagged bed reduction. This is sensible, as for these large reductions in psychiatric bed capacity, hospitals may not report a reduction in bed capacity until all the patients on the psychiatric ward have been discharged. In the dotted lines in the left panel of Figure 1.3 shows opposite effect, indicating that the relative number of psychiatric admissions at hospitals without psychiatric beds rose as psychiatric admissions overall fell.

The right panel of Figure 1.3 shows the analogous effect by HRR, this time around the timing of psychiatric bed additions. Again, relative psychiatric admissions to all hospitals are shown by the solid line, relative psychiatric admissions to hospitals without psychiatric beds is shown in the dotted lines. From the solid line, we see the expected increase in the relative psychiatric admissions within the HRR following the index period, $t = 0$, as we would expect for an HRR that reported bed additions in $t = 0$. Unlike in the case of bed reductions, however, there appear to be no trend change coinciding with the change for hospitals without psychiatric beds.

The formatting for Figure 1.4 is similar, only this time the dotted line gives trends for public hospitals as a potential site for spillovers. Again, the left panel shows bed reductions at year 0, with bed additions at year 0 shown on the right. Each line represents the average difference between the psychiatric index value and medical index value for each HRR. All hospitals are again shown by the solid line, this time the dotted lines show the event study for public hospitals. Here the estimates are noisier, however, there also appears to be a small, statistically insignificant increase in the relative number of psychiatric admissions in public hospitals around the time that overall psychiatric admissions decline on the left. On the right, again we see that the increase in psychiatric bed supply is not associated with any marked decrease in the relative number of psychiatric admissions at public hospitals.

Figures 1.5 and 1.6 decompose the effect of psychiatric bed reductions into the effect for individual diagnoses. The outcome here is the indexed number of each psychiatric diagnosis differenced with the index for medical admissions. For all hospitals (left panels), the decline in admissions associated with psychiatric bed reductions is most marked for psychiatric diagnoses (both psychotic and non-psychotic diagnoses) and admissions for an unspecified findings on urine exam. The study is noisy and difficult to interpret for other outcomes. For the subset of hospitals with no psychiatric beds (right panels), relative admissions for psychosis increase in correspondence to decreases found in overall admissions, while admissions for non-psychotic mental illnesses decrease. Rates of psychiatric evaluation increase along with small increases in rates of admissions for abnormal urine screens. While admissions appear to increase for V-Codes, the particular effect of psychiatric bed capacity is difficult to interpret as there appears to be a surge in the timing of this diagnosis within the period prior to declines in overall admissions.

Figures 1.7 and 1.8 demonstrate the mechanism for spillovers in the bed reductions or additions. Figure 1.7 shows differential effects derived from hospital discharges versus from data derived from private insurance claims. In each case, the effect of bed reductions and admissions respectively are of the same sign for privately insured patients as they are for patients overall. That is, when the number of beds is reduced, the number of admissions for privately insured patients falls. Similarly, when the number of beds increases, admissions increase for privately insured patient as well as for patients overall. Relative to the baseline level of admissions within each insurance type, however, bed additions lead to larger average increases for privately insured patients than for patients overall.

Figure 1.8 tests for the ordering of admission with respect to patient psychiatric severity. When bed reductions take place, a higher proportion of the remaining hospitalizations should be comprised of patients with severe mental illness (i.e. psychosis) if patients are ordered on the basis of mental health. However, we do not observe any large increase in average severity at the point of reductions. When psychiatric bed additions take place, marginal patients should have higher mental health if providers prioritize the sickest patients. We do observe a small decline in the proportion of patients with psychotic illness at the point of bed additions, however, this effect could also be explained by a preference for less severe patients more generally. Overall, however, ordering on the basis of psychiatric severity does not appear to be as salient as in the case for insurance.

Figures 1.9 and 1.10 show the effect of psychiatric bed reductions and additions on the number of jail inmates and the probability that a jail offers alcohol of drug treatment. With respect to bed reductions, there may be a small increase in the number of inmates and in the probability of reporting an jail alcohol or drug program one period after psychiatric bed reductions take place, however, these changes are small relative to the degree of error. For psychiatric bed additions, there is little evidence that psychiatric bed numbers change either the number of inmates or the probability of reporting an jail alcohol or drug program.

To present the aggregated effects, I separate the first stage results in Table 1.2 from the second stage results in Tables 1.3, 1.4, and 1.5. In each of the aggregated results, the instrument is adjusted to reflect psychiatric beds per 100,000. Thus each of the second stage results can be interpreted as the annual effect on the outcome, in standard deviations, averaged across the first four years after the change in the bed number. Table 1.2 shows that, in general, the bed reductions tagged are associated with an average 4.31 decrease in psychiatric beds per 100,000 population. Bed additions are associated on average with an average increase of 3.42 beds per 100,000 population. The corresponding F-stats are given in Table 1.2, indicating that bed additions may yield second stage estimates with weak instrument bias, i.e. estimates are biased toward OLS. In general, we can define the outcome period of either the four years subsequent to the discontinuity including the year of the tagged discontinuity or not without substantial changes to the interpretation.

Summaries for IV results are shown in Tables 1.3, 1.4, and 1.5. For each of the regressions considering spillovers, I consider the period following a capacity change considered to be the year of the reported change up to three years following the change. From Table 1.3, we see that in general, capacity changes do not substantially increase or decrease the relative number of psychiatric admissions at hospitals overall, however, if we consider only psychiatric bed reductions in column (4), we find that a decrease of one psychiatric bed per 100,000 is associated with a -.0199 standard deviation change in relative psychiatric admissions at hospitals without psychiatric beds. Table 1.4 demonstrates that this spillover effect may be concentrated in public hospitals, where, in column (4), we find that a decrease of one psychiatric bed per 100,000 is associated with a -.0167 standard deviation change in relative psychiatric admissions. Finally, Table 1.4, columns (6), (7), and (8), demonstrate that, compared to the population of all patients where additional psychiatric beds do no substantially increase admissions, psychiatric admissions for patients with private insurance is positively correlated to the number of available psychiatric beds.

Figure 1.11 summarizes the absolute magnitude of the effects for each diagnosis. Here the number of admissions per HRR are regressed against population quintiles, year fixed effects, HRR-level fixed effects in a 2SLS specification where the instruments are discontinuous bed reductions only. The size of the bars represent the number of excess diagnoses per year associated with an additional bed per 100,000 population. In Fig. 1.11 we see that, when all hospitals are considered, increases in psychiatric bed numbers are associated with marginally significant increases in admissions for heart disease and COPD, two common medical diagnoses. This may be consistent with a picture that hospitals interested in eliminating or reducing the size of their psychiatric wards may also be increasing the size of other wards and vice versa. For many psychiatric diagnoses, increases in the supply of psychiatric beds is associated with admissions rates not distinguishable from zero overall. This effect is inclusive of both a reduction in psychiatric capacity decreasing psychiatric admissions as well as the effect of spillovers, which may increase psychiatric admissions. The second bar in each group draws this distinction by looking only the the effect of psychiatric admissions on hospitals with no reported psychiatric beds. Here we see that there are significant increases in the number of psychosis admissions at hospitals without psychiatric beds on the order of 200 excess admissions for a reduction of 1 bed per 100,000.

In Table 1.5, I show that the absolute magnitude of the effect for jail outcomes was not found to be statistically significant for either outcome variable. For the number of inmates, an additional psychiatric bed per 100,000 was associated with on average with -.002 standard deviation reduction in the number of jail inmates and an average .002 standard deviation increase in the probability of offering alcohol or drug treatment services.

1.6 Discussion

There appears to be some evidence that short-run spillovers occur from psychiatric to non-psychiatric settings when bed supply is reduced. In particular, the evidence is strongest for spillovers between psychiatric and general hospitals. Although the relative number of cases increases more substantially for public hospitals, with respect to overall magnitude, such spillovers can be detected in both private and public general hospitals. The evidence presented in this analysis lends credence to the hypothesis that spillover probability from bed reductions is decreasing with increasing mental health. Patients with psychotic illness might be conceived as the patients with the most severe mental illness; they are also the most likely patients to spill over between settings (Fig. 1.5). Because the baseline number of admissions for psychotic illness is also large, the measured spillover effect for psychosis is also large in comparison to other diagnosis groups: a reduction of one psychiatric bed per 100,000 is associated with approximately 200 additional admissions for psychosis at nearby hospitals (Fig. 1.11).

While bed reductions appeared to be associated with some level of psychiatric spillovers into general medical hospitals, bed additions did not seem to have the effect of reducing these spillovers. Although it is perhaps surprising that the findings with respect to general hospital spillovers are not symmetric, there are several possible explanations.

The first possible explanation for the asymmetric effect is, while divestment in the form of psychiatric bed reductions may occur immediately, investment may occur more gradually and require contracting with insurers, notifying nearby providers of service availability to enable patients transfer, and other transitional activities. From the form of the model presented, we know that the extent of spillovers is determined by the number of patients that meet eligibility for both settings. As such, one hypothesis may be that, in areas where psychiatric care is scarce, general hospitals may take all patients that meet their admissions criteria and be slow in changing this policy when a new psychiatric hospital opens.

There are several reasons that this reasoning may be uncompelling. From Figure 1.3, we see that the number of psychiatric admissions rises immediately after psychiatric beds are added and stay elevated at the new rate, suggesting that the new psychiatric beds reach capacity within a year of the bed addition. Given a context where psychiatric stays in both general and psychiatric hospitals are designed to be acute with short lengths of stay, it is sensible that the capacity of a new ward could be reached quickly, in a period of time much shorter than one year.

Another possible explanation may be given, in the context of Figures 1.7 and 1.8, is secondary to the psychiatric provider discretion, which prioritizes patients with higher ability-to-pay over patients with higher acuity mental illness. That is, although I have described the marginal revenue received per patient-day and the marginal cost of providing care per patient-day as a function of the number of patients alone, all patients may not be the same. That is, providers may receive higher marginal revenues per patient-day from private insurers, and providers may bear higher marginal costs per patient-day provided from disruptive patients or patients with very low baseline mental health. One prior is providers cannot reject patients who meet criteria for admission under EMTALA, however, Figure 1.7 shows that privately insured patients appear to be admitted preferentially while Figure 1.8 also suggests that patients are not strongly ordered over mental health alone.

Psychiatric provider discretion in admitting is not unheard of and can occur in a variety of ways. One possibility is beds may have been introduced at private providers who did not maintain emergency rooms, necessitating a transfer to take place between providers before a patient is admitted. This transfer could then be an opportunity for the provider to reject an undesirable patient.[5] Additionally, from [50], we know that different types of psychiatric units are differently profitable. Therefore, another opportunity to avoid unprofitable patients may be to disproportionately reduce psychiatric crisis-care beds while adding inpatient beds for substance use treatment.

With respect to the findings on jail populations, no statistically significant differences in the number of inmates can be found both following bed additions and bed reductions could be detected. In spite of this fact, the findings presented in Figs. 1.9 and 1.10 appear to present a pattern in mean values that are asymmetric as anticipated by the model extensions. Table 1.5 summarizes that the findings summarized in Figs. 1.9 and 1.10 are not statistically significant. Here, I was limited by the noise inherent in the data and the inability to observe the subset of mentally-ill inmates, which may have provided a less noisy measure. At a

5. In practice, such transfers are often facilitated via insurance networks, meaning that a psychiatric hospital is unlikely to reject any individual patient, but may contract preferentially with insurers in order to avoid uncompensated patients or patients with Medicaid coverage alone.

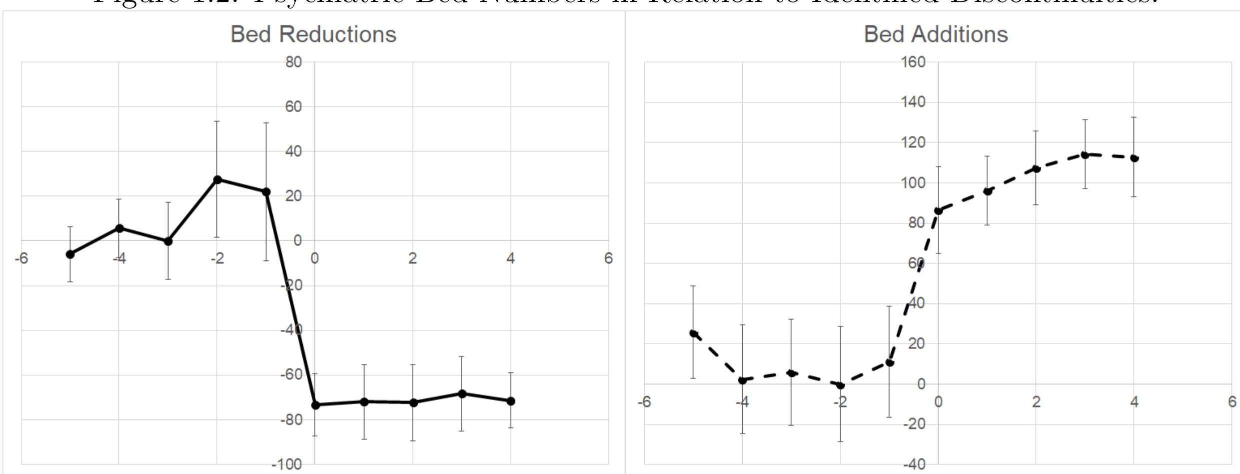
minimum, it is difficult to assert from these data that there is any evidence that psychiatric bed additions reduce the number of inmates in jails in the short-run.

This study has several implications for public policy. In a traditional formulation, patients with mental illness have been conceptualized as existing along a single demand curve: those with the most inelastic demand should be admitted first and at the highest price (i.e. a person with psychosis and complicating medical condition being admitted to the medical ward when psychiatric beds quantity equals zero), and, as the number of available beds increases, patients with decreasing willingness-to-pay are also admitted. Because demand is decoupled from willingness-to-pay by insurance status, however, providers have an incentive to discriminate between different consumers. The observed negative correlation between mental health and income in practice may induce some trade-off between these two types of patients [78, 51].

The model of hospital discretion is at odds with several perspectives put forth in the literature and in policy. For example, a simple computation of psychiatric bed need from those waiting for a public psychiatric bed may invariably produce estimates higher than current quantity; hospitals choose the number of beds in the long run and admit in the short-run so that the beds will be full [62]. These admission practices would imply that, in equilibrium, we should expect some number of patients waiting for psychiatric beds wherever those beds are provided. Similarly, in recent years, the substantial rise in opioid-related deaths has led to a discussion regarding possible repeal of the Medicaid IMD Exclusion as a means of expanding the availability of psychiatric services [33]. This suggestion, too, seems to a favor a view of psychiatric hospitals suggests that they will look favorably toward taking patients at low per-diem rates provided by Medicaid, which may not be the case.

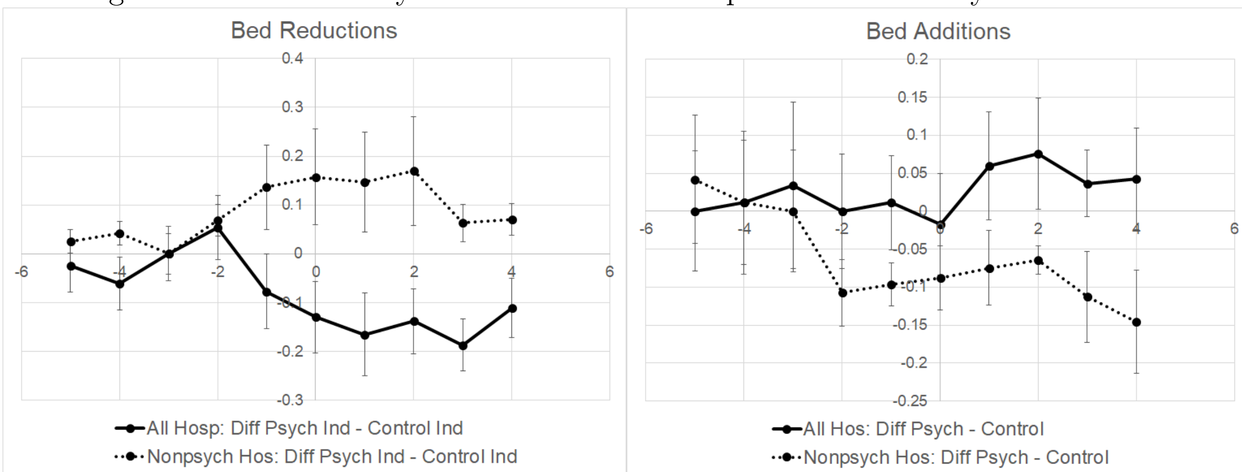
In part, the effectiveness of managed care in cost containment in psychiatry was premised on the idea that psychiatric providers favored excessively long stays for patients. However, in the context of provider discretion, reimbursement limitations may have had the side-effect of exacerbating spillovers between settings. At least partially, the failure for traditional managed care incentives to properly account for psychiatric spillovers lies in the substantial degree of heterogeneity within psychiatric populations coupled with the inherent subjectivity of provider assessment. In the shorter term, better characterization of heterogeneity within psychiatric patient populations may be a useful first step towards threading the needle of cost control and spillover prevention. Proportional payments could also give providers a greater incentive to provide care for more severely ill patients. Integration of medical and psychiatric records, public records of benefits receipt, and criminal justice system involvement may also lend objective reinforcement to subjective provider claims of patient illness severity in a context where reimbursement practice is similarly heterogeneous to patient populations. However, for the patients who are at greatest spillover risk, an alternative solution may lie in a departure from the acute event model of psychiatric care entirely. That is, for a subset of patients with severe chronic illness and poor social supports, a longer term, lower acuity setting may provide more benefit than even a large increase in the psychiatric bed supply.

Figure 1.2: Psychiatric Bed Numbers in Relation to Identified Discontinuities.



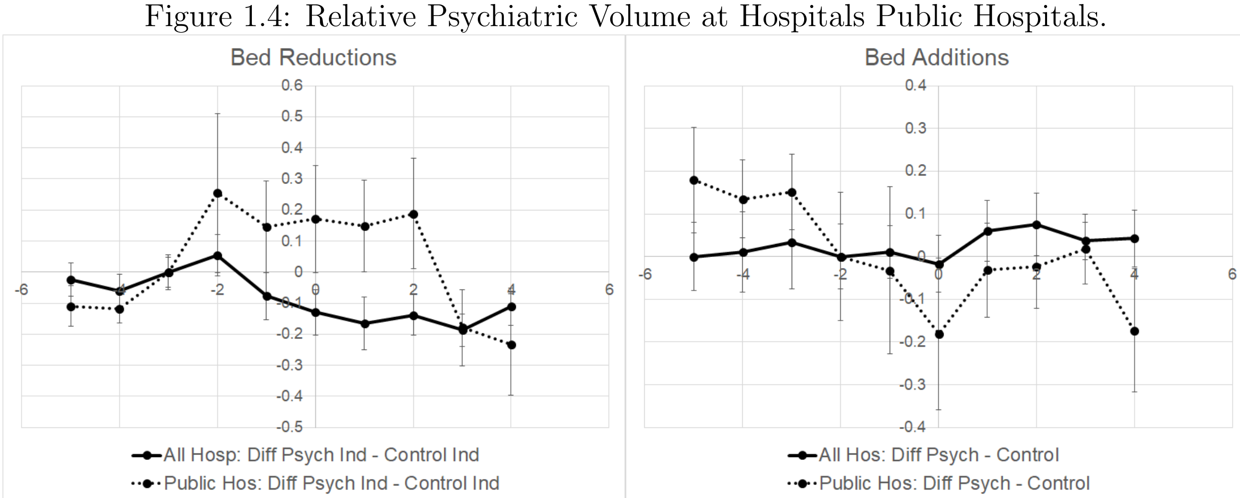
Author's calculations. Number of psychiatric beds by HRR versus years from identified discontinuities. Left: Discontinuous bed reductions. Right: Discontinuous bed additions. The zero point in both graphs is the year of the identified discontinuity. The plotted line is the average number of psychiatric beds within each HRR. The x-axis points -5, -4, -3, -2, -1 correspond to the averages 5, 4, 3, 2, and 1 years before the discontinuity respectively. X-axis points 1, 2, 3, 4 correspond to the years 1, 2, 3 and 4 years following the discontinuity respectively. Standard errors given by the error bars. Change at t=0 is significant at the 95% confidence level.

Figure 1.3: Relative Psychiatric Volume at Hospitals Without Psychiatric Beds.



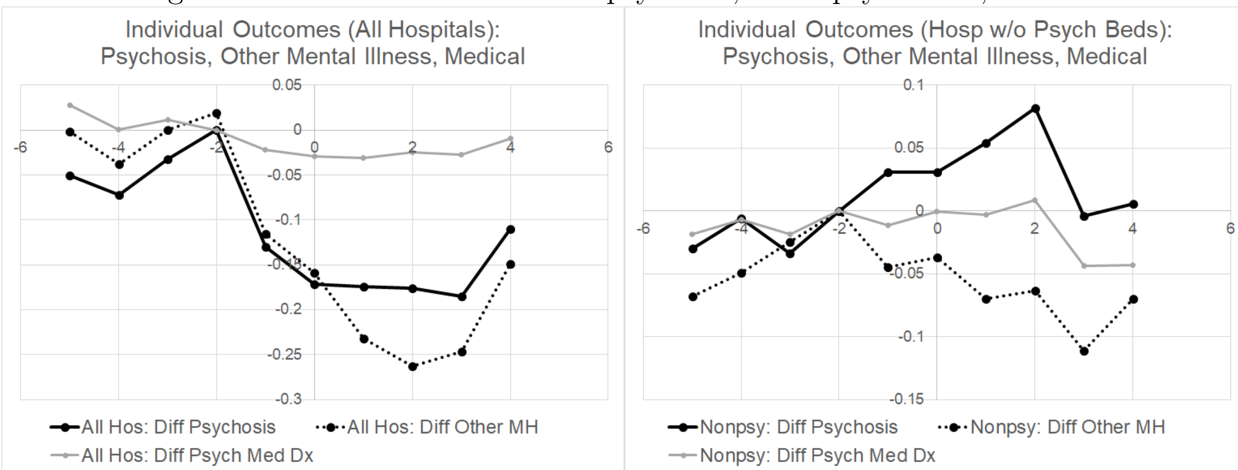
Author's calculations. The plotted lines represent average admissions volumes for psychiatric versus control diagnoses within an HRR relative to the year of the identified discontinuity (t=0). Solid line: All hospitals in HRR. Dotted line: public hospitals. Left: Bed reductions. Right: Bed additions.The x-axis points -5, -4, -3, -2, -1 correspond to the averages 5, 4, 3, 2, and 1 years before the discontinuity respectively. X-axis points 1, 2, 3, 4 correspond to the years 1, 2, 3 and 4 years following the discontinuity respectively. Standard errors given by the error bars. All 95% confidence intervals contain zero.

Figure 1.4: Relative Psychiatric Volume at Hospitals Public Hospitals.



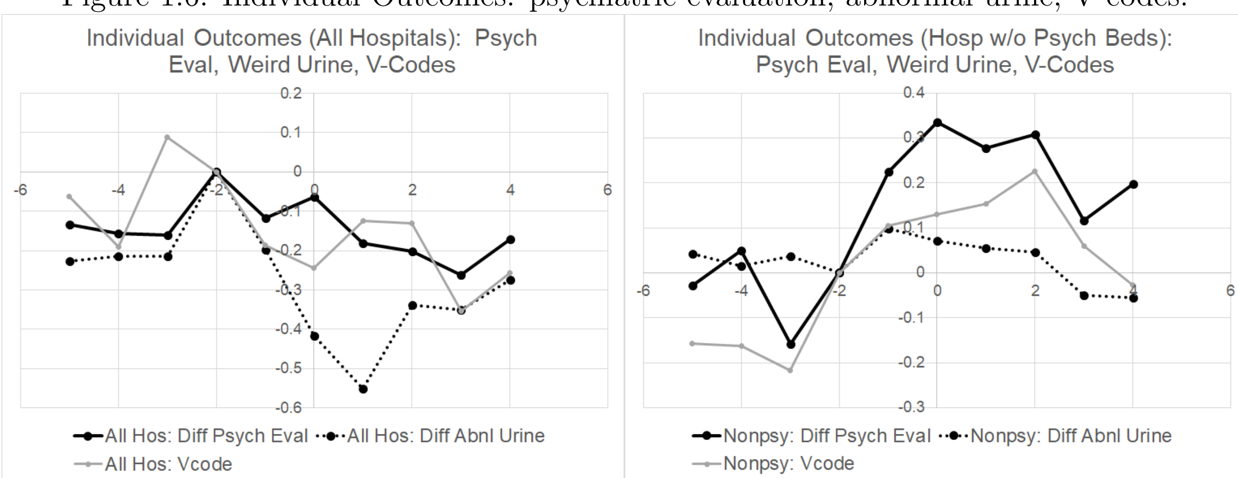
Author's calculations. The plotted lines represent average admissions volumes for psychiatric versus control diagnoses within an HRR relative to the year of the identified discontinuity (t=0). Solid line: All hospitals in HRR. Dotted line: public hospitals. Left: Bed reductions. Right: Bed additions. Standard errors given by bars. All 95% confidence intervals contain zero.

Figure 1.5: Individual Outcomes: psychosis, other psychiatric, medical.



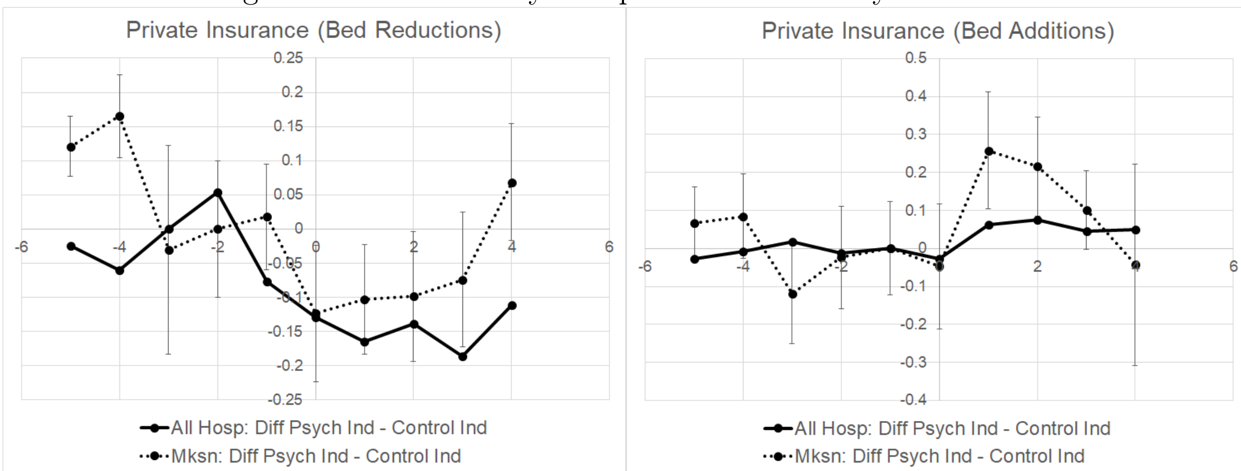
Author's calculations. The plotted lines represent average admissions volumes for psychiatric versus control diagnoses within an HRR relative to the year of the identified discontinuity (t=0) for all observed admissions within an HRR. Solid black line: Psychosis admissions. Dotted black line: Other mental illness admissions. Gray line: Medical sequelae of psychiatric illness. Left panel: All hospitals in HRR. Right panel: Hospitals without psychiatric beds.

Figure 1.6: Individual Outcomes: psychiatric evaluation, abnormal urine, V-codes.



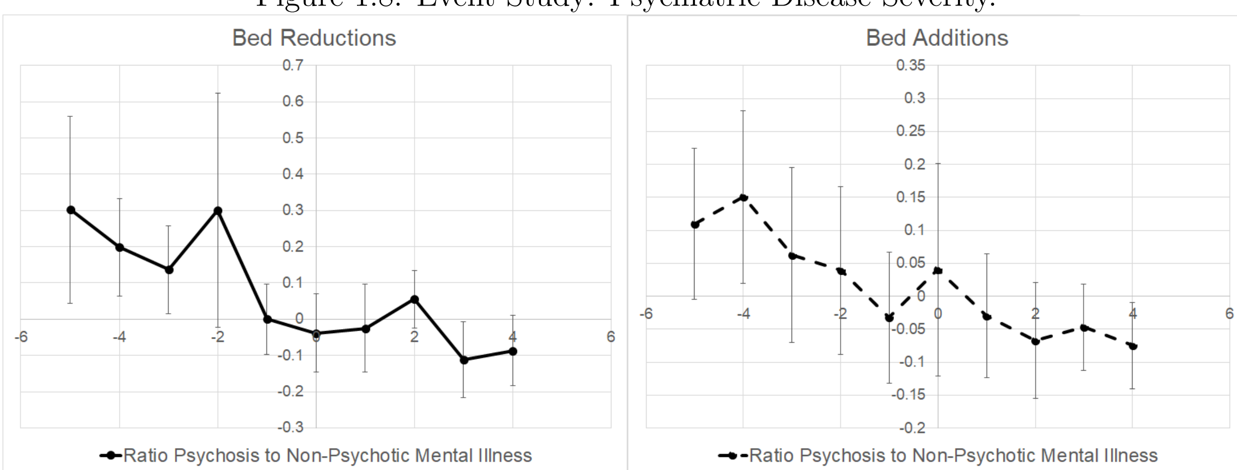
Author's calculations. The plotted lines represent average admissions volumes for psychiatric versus control diagnoses within an HRR relative to the year of the identified discontinuity (t=0) for all observed admissions within an HRR. Solid black line: Psychiatric evaluation. Dotted black line: Admissions for abnormal urine. Gray line: V-codes. Left panel: All hospitals in HRR. Right panel: Hospitals without psychiatric beds.

Figure 1.7: Event Study: Hospital Admissions by Insurance.



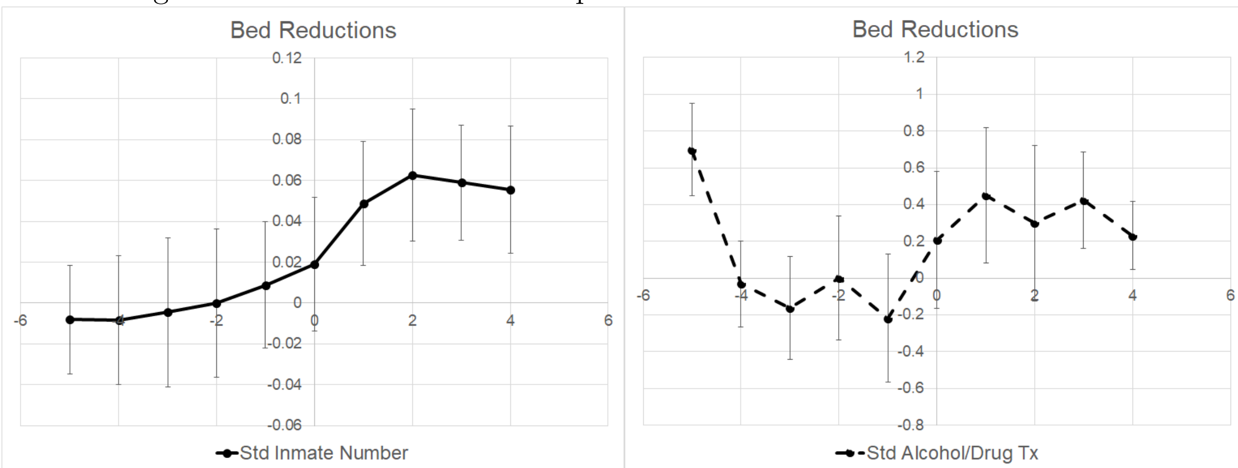
Author's calculations. The plotted lines represent average admissions volumes for psychiatric versus control diagnoses within an HRR relative to the year of the identified discontinuity (t=0). Solid line: all hospitals. Dotted line: hospitals without psych beds. Left: Bed reductions. Right: Bed additions. Standard errors given by bars. All 95% confidence intervals contain zero.

Figure 1.8: Event Study: Psychiatric Disease Severity.



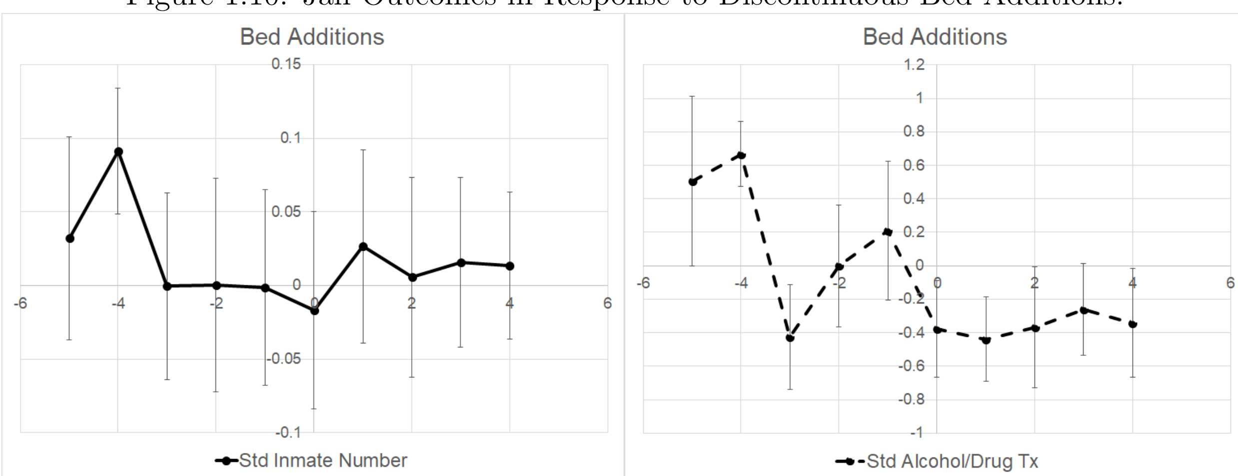
Author's calculations. The plotted lines represent the ratio of psychosis diagnoses to non-psychotic psychiatric diagnoses within an HRR relative to the year of the identified discontinuity (t=0). Solid line: all hospitals. Dotted line: hospitals without psych beds. Left: Bed reductions. Right: Bed additions. Standard errors given by bars. All 95% confidence intervals contain zero.

Figure 1.9: Jail Outcomes in Response to Discontinuous Bed Reductions.



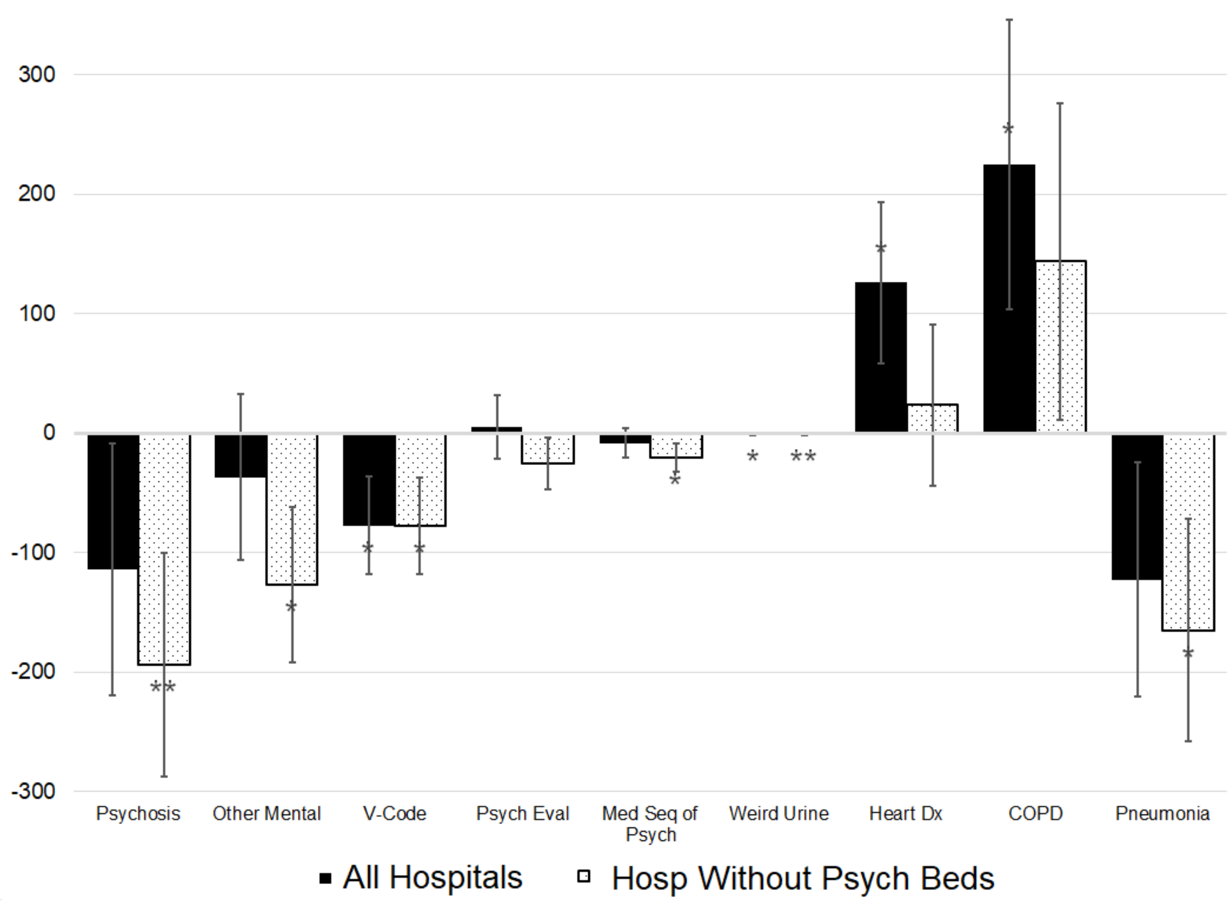
Author's calculations. Left: Outcome is HRR-level number of jail inmates, standardized to have a mean 0 and a standard deviation of 1. Right: Outcome is the proportion of jails offering alcohol or drug treatment. Timeline with respect to discontinuous bed additions in t=0. Standard errors given by bars. All 95% confidence intervals contain zero.

Figure 1.10: Jail Outcomes in Response to Discontinuous Bed Additions.



Author's calculations. Left: Outcome is HRR-level number of jail inmates, standardized to have a mean 0 and a standard deviation of 1. Right: Outcome is the proportion of jails offering alcohol or drug treatment. Timeline with respect to discontinuous bed additions in t=0. Standard errors given by bars. All 95% confidence intervals contain zero.

Figure 1.11: Absolute Magnitude of Effects on Individual Diagnoses.



Author's calculations. Here the number of admissions per HRR are regressed against population quintiles, year fixed effects, HRR-level fixed effects in a 2SLS specification. The instrumental variable are the discrete, identified local-area reductions in psychiatric bed capacity. The size of the bars represent the number of excess admissions per year associated with an additional 1 bed per 100,000 population. Standard errors are shown in error bars. * Indicates significance at the .10 level. ** Indicates significance at the .05 level.

Table 1.1: Summary Statistics at the HRR-year level.

	(1) mean	(2) overall SD	(3) within SD	(4) between SD	(5) N
Discrete Bed Additions	0.04	0.20	0.20	0.05	3328
Discrete Bed Reductions	0.03	0.18	0.17	0.04	3328
Psych Beds per HRR	304.73	417.86	109.05	379.71	3328
All Beds per HRR	2528.64	3362.18	923.00	3035.35	3328
Index of Psych Outcomes (HCUP)	0.00	0.92	0.38	0.74	3593
Index of Control Outcomes (HCUP)	0.00	0.99	0.34	0.82	3595
Admits for Psychosis per HRR	2624.18	10553.24	6543.36	7276.59	3593
Admits for Oth Psych Dx per HRR	3599.23	15579.16	9862.22	10582.64	3593
Admits with Psych Eval per HRR	617.10	1764.03	963.08	1351.24	3593
Admits for V-Code per HRR	157.21	1290.13	1119.64	562.68	3593
Admits for Weird Urine per HRR	6.31	37.53	26.87	22.92	3593
Admits Med Seq of Psych Dx per HRR	605.40	2603.85	1545.95	1837.18	3595
Admits for Cardiac Dx per HRR	4930.55	18577.04	9985.19	13751.49	3595
Admits for COPD per HRR	4438.20	19481.64	11603.17	13717.21	3595
Admits for Pneumonia per HRR	2383.26	10883.39	6389.71	7717.35	3595
Jail Inmates per HRR	1927.05	6041.67	5544.35	1815.80	7516
Jail Has Alcohol/Drug Treatment	0.08	0.28	0.15	0.23	4368
Population by HRR	959989	1373116	1357605	220552	10260

Author's calculations using data from Medicare Provider of Services, the American Hospital Association, the National Inpatient Sample, and Truven MarketScan aggregated to the Hospital Referral Region (HRR)-year.

Table 1.2: First Stage Results

	Psych Beds per 100,000	
Bed Reduc (t=0,1,2,3)	-4.31 (1.84)	
Bed Add (t=0,1,2,3)	3.42 (0.84)	
Bed Red (t=1,2,3,4)		-4.12 (1.72)
Bed Add (t=1,2,3,4)		3.47 (0.84)
Pop Quintile 2	-13.44 (8.32)	-13.42 (8.31)
Pop Quintile 3	-13.05 (8.83)	-12.70 (8.84)
Pop Quintile 4	-13.86 (8.99)	-13.38 (9.04)
Pop Quintile 5	-19.75 (10.35)	-19.38 (10.37)
Year Dummy 1989	-0.01 (1.50)	0.01 (1.50)
1990	1.66 (1.67)	1.62 (1.67)
1991	23.08 (4.47)	23.08 (4.47)
1992	23.43 (5.10)	23.35 (5.09)
1993	20.41 (4.58)	20.37 (4.58)
1994	17.26 (4.48)	17.13 (4.51)
1995	19.69 (4.98)	19.56 (4.99)
1996	19.44 (4.87)	19.30 (4.88)
1997	17.14 (4.53)	17.01 (4.54)
1998	15.45 (4.37)	15.41 (4.36)
1999	14.84 (4.04)	14.66 (4.05)
2000	13.93 (4.00)	13.95 (4.00)
2001	14.07 (3.70)	14.05 (3.71)
2002	12.68 (3.82)	12.53 (3.84)
2003	11.98 (4.05)	12.04 (4.05)
2004	10.87 (3.69)	10.83 (3.70)
2005	12.08 (3.84)	12.04 (3.85)
2006	10.48 (3.88)	10.44 (3.90)
2007	11.31 (3.62)	11.29 (3.64)
2008	11.60 (3.68)	11.45 (3.72)
2009	11.47 (3.63)	11.25 (3.68)
2010	11.84 (3.56)	11.73 (3.58)
2011	10.50 (3.38)	10.30 (3.42)
F-stat Bed Reduc	13.38	10.96
F-stat Bed Add	6.76	6.95

Author's calculations. Outcome is Psych Beds per 100,000 in HRR. N=3025. Standard deviation in parentheses. Period affected by discontinuous change defined as up to three years following observed discontinuity.

Table 1.3: Main Results: Hospitals without Psych Beds

	All Hospitals				Hospitals Without Psych Beds			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Psych Beds per	-0.001	0.005	0.006*	0.006	-0.0001	-0.0167	-0.0168	-0.0199**
100,000	(0.001)	(0.004)	(0.004)	(0.004)	(0.0001)	(0.0107)	(0.0108)	(0.0101)
Pop Quintile 2	-0.130*		-0.030	-0.030	-0.0537**		-0.2753	-0.3168
	(0.065)		(0.106)	(0.111)	(0.0234)		(0.2070)	(0.2259)
Pop Quintile 3	-0.096		-0.004	-0.004	-0.1049**		-0.3104	-0.3489
	(0.075)		(0.118)	(0.122)	(0.0380)		(0.2037)	(0.2221)
Pop Quintile 4	-0.137		-0.038	-0.038	-0.1320**		-0.3524*	-0.3936*
	(0.092)		(0.140)	(0.150)	(0.0472)		(0.1981)	(0.2145)
Pop Quintile 5	0.0940		0.245	0.245	-0.0704		-0.4044*	-0.4668*
	(0.196)		(0.233)	(0.244)	(0.0760)		(0.2401)	(0.2458)
Specification	OLS	IV	IV	IV	OLS	IV	IV	IV
IV		Add & Red	Add & Red	Red		Add & Red	Add & Red	Red
HRR FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y

Author's calculations. Outcome is the difference in psychiatric and control index. N=3023. Standard deviation in parentheses. Coefficients derived from 2SLS specification performed at the HRR-level. The instrumental variable is either the discrete, identified local-area reductions in psychiatric beds only ("Red"), or a combination of discrete, identified local-area additions and reductions in psychiatric beds ("Add & Red"). Period affected by discontinuous change defined as up to three years following observed discontinuity

Table 1.4: Main Results: Public Hospitals and Private Insurance

	Public Hospitals				Private Insurance			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Psych Beds per	0.0024**	-0.0176**	-0.0171**	-0.0167**	0.0073	0.0261*	0.0259*	0.0278*
100,000	(0.0012)	(0.0050)	(0.0051)	(0.0054)	(0.0067)	(0.0144)	(0.0143)	(0.0153)
Pop Quantile 2	0.0058		0.0346	0.0340	-0.1449			-0.4454
	(0.0599)		(0.1622)	(0.1575)	(0.2138)			(0.4065)
Pop Quantile 3	-0.0597		0.0406	0.0384	-0.0875		0.0274	-0.4210
	(0.0704)		(0.1985)	(0.1891)	(0.2101)		(0.0365)	(0.4114)
Pop Quantile 4	-0.0653		0.0888	0.0854	-0.0590		0.0381	-0.4120
	(0.0980)		(0.2836)	(0.2693)	(0.1937)		(0.0781)	(0.4019)
Pop Quantile 5	0.1557		0.2335	0.2318			0.4184	
	(0.1775)		(0.4074)	(0.3962)			(0.3873)	
Specification	OLS	IV	IV	IV	OLS	IV	IV	IV
IV		Add & Red	Add & Red	Red		Add & Red	Add & Red	Red
HRR FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y

Author's calculations. Outcome is the difference in psychiatric and control index. N=3023. Standard deviation in parentheses. Coefficients derived from 2SLS specification performed at the HRR-level. The instrumental variable is either the discrete, identified local-area reductions in psychiatric beds only ("Red"), or a combination of discrete, identified local-area additions and reductions in psychiatric beds ("Add & Red"). Period affected by discontinuous change defined as up to three years following observed discontinuity.

Table 1.5: Jail Outcomes

	Std Inmates Number				Has Alc/Drug Treatment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Psych Beds per	-.0003	-.0029	-.0029	-.0020	-.0013	.0007	.0005	.0002
100,000	(.0002)	(.0021)	(.0021)	(.0032)	(.0011)	(.0054)	(.0055)	(.0061)
Pop Quantile 2	.0675**		.0349	.0464	.0065		.0160	.0144
	(.0335)		(.0379)	(.0444)	(.0216)		(.0375)	(.0399)
Pop Quantile 3	.1180**		.0869	.0980*	.0440		.0533	.0517
	(.0564)		(.0566)	(.0583)	(.0401)		(.0504)	(.0529)
Pop Quantile 4	.1675**		.1319	.1446*	.0300		.0414	.0394
	(.0827)		(.0811)	(.0824)	(.0833)		(.0905)	(.0941)
Pop Quantile 5	.2286*		.1692	.1903	-.0664		-.0089	-.0188
	(.1253)		(.1183)	(.1200)	(.1147)		(.2150)	(.2409)
Specification	OLS	IV	IV	IV	OLS	IV	IV	IV
IV		Add & Red	Add & Red	Red		Add & Red	Add & Red	Red
HRR FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
N	2756	2756	2756	2756	1806	1806	1806	1806

Author's calculations. Standard deviation in parentheses. Coefficients derived from 2SLS specification performed at the HRR-level. The instrumental variable is either the discrete, identified local-area reductions in psychiatric beds only ("Red"), or a combination of discrete, identified local-area additions and reductions in psychiatric beds ("Add & Red"). Period affected by discontinuous change defined as up to three years following observed discontinuity.

CHAPTER 2
SPECIALIST INTERESTS AND MEDICARE REIMBURSEMENT: AN EXAMINATION OF THE RESOURCE-BASED RELATIVE VALUE SCALE

2.1 Introduction

Since 1992, the Resource Based Relative Value Scale (RBRVS) has been the basis fee schedule to determine reimbursement amounts paid for physician and clinical services under Medicare Part B. The RBRVS was intended to generate more unbiased assessments of physician work compared to the fee-for-service system it replaced [89]. However, the RBRVS process has faced criticism for the implicit favor it gives to specialists over primary care physicians [90, 19, 16].

Reimbursement levels of medical codes are defined by the RBRVS. Throughout, I will use the term "code" to refer to a service or procedure designated by an individual current procedure terminology (CPT) code. More specifically, the RBRVS divides compensation for service or procedure codes into three components: physician work, practice expenses and malpractice liability insurance. The base unit of the scale is called a relative value unit (RVU) and Medicare payments are calculated by multiplying the sum of the RVUs for each component of a code, given in accordance with the scale, against a multiplier determined by the Centers for Medicare and Medicaid Services (CMS). Reimbursements are then further adjusted for geography by use of a geographic multiplier [1]. RVU values corresponding to individual components of this scale are termed work RVUs (wRVU), practice expense RVUs, malpractice expense RVUs, and their sum is denominated as total RVUs.

Physician wRVUs are subject to a process of physician survey and subsequent committee review. wRVUs are meant to incorporate measures of the time required to perform a service, the mental effort required, the physical effort required and the psychological stress of the

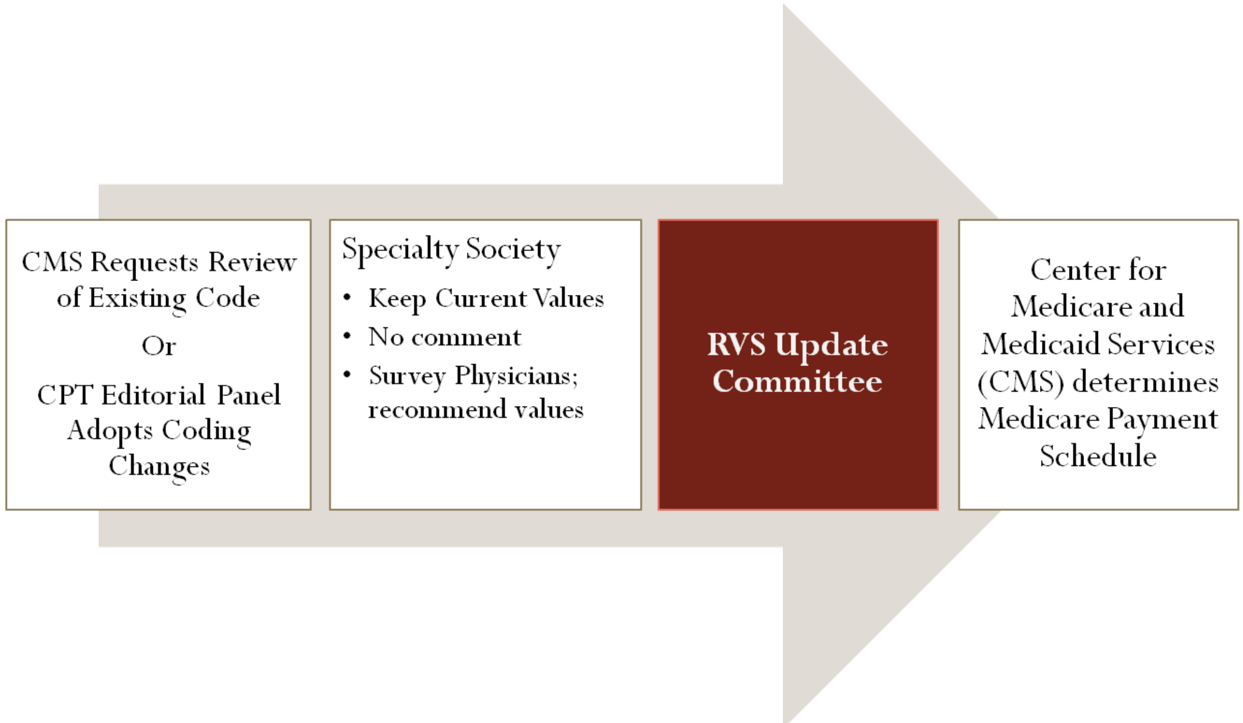


Figure 2.1: Resource-Based Relative Value Scale Revision Process Summary.

service as measured using a national survey. The process by which CMS annually updates work wRVUs for the RBRVS is as follows (Figure 2.1).

First, a code is nominated for review by either the CPT editorial panel or CMS. Nominated codes will receive input from to the RVS Update Committee (RUC), a committee composed of members of the American Medical Association (AMA). The RUC contracts the specialty society to which the code best pertains. The specialty society then performs a physician survey to determine the amount of work required by a given service or procedure and provides recommendations. The RUC can then decide whether to adopt the recommendations, refer the study back to the specialty society, or modify the recommendations before submission to CMS. In order to be submitted to CMS, a recommendation must be approved by a minimum of two-thirds of sitting RUC members [4].

Although the majority of RUC seats have been allocated to representatives of national medical specialty societies, a minority of RUC seats have corresponded to primary care specialties. Following February 2012, the RUC consisted of 31 members, of which 24 were members were appointed by major national medical specialty societies. Of the seats allocated to 24 medical specialties, 20 seats were allocated to specialties on a permanent basis and 4 were rotating seats. That is, one permanent seat apiece was allocated to internal medicine, radiology, cardiology, anesthesia, orthopedic surgery, psychiatry, otolaryngology, etc. and 4 seats are allocated to representatives of four distinct specialties not otherwise represented. Of these four rotating seats, two are reserved for internal medicine specialties such as allergy/immunology, endocrinology, gastroenterology, hematology, infectious disease, nephrology, oncology, pulmonary medicine and rheumatology and one is reserved for primary care. Prior to February 2012, the RUC was comprised of 29 total seats, 23 of which were appointed by major national medical societies. Of the 23 seats open to medical specialty society members, three of the seats rotated on a 2-year basis.

Historically, the RUC recommendations have had a high rate of adoption by CMS. In a report from the AMA, the RUC has submitted recommendations annually since 1993 in addition to the four five-year reviews. Of these recommendations between 80-100 percent of the recommendations were accepted annually, with an average of 87.4 percent of recommendations accepted in any given year [65].

Some primary care physicians and physicians groups have suggested that such disproportionate representation has contributed to a widening specialist and non-specialist income gap [20]. The Department of Health and Human Services acknowledged this possibility in 2008, when it requested that the RUC renew its efforts to identify overvalued codes rather than undervalued codes as they had in the past. However, the AMA has maintained that service on the RUC is an individual choice and that committee members "exercise their independent judgment" and "are not advocates for their specialty" [4].

The literature has offered different sources of evidence for establishing a link between RUC membership and reimbursement from individual accounts to aggregated reports with the majority focusing on time series analysis or differential growth rates across code types (e.g. evaluation and management codes versus imaging or procedures) [75, 74, 20, 65, 64].

Given the attention the issue has been given throughout the years and a setting in which variation exists in committee membership, the effects of rotating membership would seem a natural setting in which to study potential impacts of the RUC. A precedent for this type of study has been set in the political economy literature: for instance, Kuziemko and Werker use rotating seats on the U.N. security council to identify correlations between security council membership and receipt of U.N. aid [61]. Indeed, the AMA states that in 2007, an internal investigation was conducted into whether rotating seat allocation impacted final RBRVS values and determined that no such effect existed. The study results have not been publicly available.

In this paper, I investigate the relationship between rotating seat representation on RUC and Medicare reimbursement with respect to the number of physician specialties observed to bill a code. I refer to the number of physician specialties observed to be bill a code as an "inverse level of specialization", where I term codes performed by a small number of physician specialties as "highly specialized" and codes performed by a large number of physician specialties as "less specialized". This represents a deviation from the literature, which has hitherto focused on code type (e.g. evaluation and management, surgical or imaging) rather than degree of specialization. I do so because I wish focus on the statutory cap placed on the total amount that RVU changes may affect projected billing. The goal is to demonstrate how a relatively simple regulation may generate perverse incentives, which then may contribute to the differential increases observed in reimbursements across levels of specialization.

The Social Security Act requires that increases or decreases in RVUs may not cause the projected amount in expenditures under Part B for the subsequent year to differ more than 20 million dollars from what it would have been in the absence of the changes [1]. The implication of such a cap may be made clear by example. Imagine that each member of the 25 member committee represents a specialty that performs 1 million office visits annually. Imagine that in addition, the specialty also performs 1 million specialty procedures annually, which are unique that that specialty. Suppose that the cap on Medicare expenditure changes due to the fee schedule is 1 million dollars annually. If both office visits and specialty procedure codes are nominated for review, it is in the specialists interest to advocate that the 1 million dollars is dedicated solely to the specialty procedure code: if the increase is applied to office visits, the reimbursement for the office visit increases only by 4 cents, but applied to the specialty procedure, it amounts to a 1 dollar increase per procedure, which is completely internalized by the specialty. In total, the increase applied to office visits would be worth only 40,000 dollars to the specialty, whereas the same increase applied to the specialty procedure would yield 1 million dollars to the specialty. We may then anticipate that RUC membership should correlate with disproportionate increases in reimbursement to codes that are performed by a small number of physician specialties.

2.2 Methods

2.2.1 Data

Annual RBRVS values were compiled from the CMS website for the years 2003-2013. Additional RBRVS data was retrieved as revisions from the Federal Register for the years 1994-2003. Each code in the dataset is uniquely identified by its common procedure terminology (CPT) code and modifier. Because multiple revisions of the RBRVS were available from CMS for any given year of data, I use only latest revision of RBRVS values for each year 1994-2013.

RUC members are listed on the acknowledgment page of Medicare RBRVS: The Physician's Guide, an annual publication of the AMA. Due to the 2007 AMA study, rotating seats for the RUC are available upon request from the AMA. I use the 2007 AMA study rotating specialties from years 1991-1998 and the RUC members listed in *Medicare RBRVS: The Physician's Guide* for years 1999 and onward [3]. There are very few discrepancies between the two sources in the overlapping years.

Table 2.1: RVS Update Committee Rotating Specialty Seat Occupation by Year.

1991-1994	**Gastroenterology***, Nuclear Medicine
1995-1996	**Pediatric Surgery, Geriatrics**
1997-1999	Child Psychiatry, **Rheumatology, Geriatrics**, Physician Assistants
1999-2001	**Pulmonary Medicine, Oncology/Hematology, Vascular Surgery**
2002-2003	**Geriatrics, Rheumatology**, Radiation Oncology
2004	**Geriatrics, Gastroenterology, Vascular Surgery**
2005	**Gastroenterology, Vascular Surgery, Pulmonary Medicine**
2006	**Pulmonary Medicine, Oncology/Hematology**, Spine Care
2007	**Geriatrics, Oncology/Hematology**, Spine Care
2008	**Geriatrics, Pediatric Surgery, Gastroenterology**, Podiatry
2009	Podiatry, **Pediatric Surgery, Gastroenterology, Infectious Disease**
2010	Podiatry, **Infectious Disease**, Nephrology, Colon and Rectal Surgery
2011	Podiatry, Nephrology, Colon and Rectal Surgery
2012	Chiropractics, **Pulmonary Medicine, Rheumatology, Vascular Surgery**
2013	**Geriatrics, Infectious Disease, Rheumatology, Vascular Surgery**, Primary Care Rotating

*Specialties which appear more than once on non-consecutive years are shown in bold.

The Physician/Supplier Procedure Summary Master File (PSPS) is a 100 percent summary of all Medicare Part B Carrier and Durable Medical Equipment Regional Carrier Claims. It includes carrier, pricing locality, healthcare common procedure coding (HCPC) or common procedural terminology (CPT) designations, totally submitted and allowed services and charges by specialty. I have access only the PSPS file for the year 2007.

Part B National Summary Data File (BESS) is a publicly available file from the CMS website. It summarizes allowed services, charges, and payments by HCPC/CPT group. During the time of study, years 2000-2011 were available.

2.2.2 Independent Variables and Outcomes

Medicare designates 93 distinct specialist codes in the PSPS 2007 data, all of which were matched to at least one CPT code. I mark a specialty and a code as matched if the specialty is observed to bill for the code at least one time in the PSPS 2007. Due to constraints in training between specialties it is reasonable to assume that the relative level of specialization for a given code should not vary substantially across time. If a specialty represented on the RUC did not have a corresponding Medicare billing code, no codes were designated as pertaining to that specialty. The inverse level of specialization for a given code is proxyed by the number of specialties observed to bill for that code in 2007. For example, the CPT code 27703 for reconstruction of the ankle joint is billed only by orthopedic surgeons and has an inverse level of specialization of 1 and is highly specialized. By contrast the CPT code 99211 Office/outpatient visit established patient has an inverse level of specialization of 90 and is less highly specialized.

Subsequently, I use committee seat data to designate a dummy variable by code by year. Here the dummy is marked as 1 if the code is matched to a specialty that was observed to have a RUC rotating seat for that year and 0 otherwise. Codes that could not be matched to any specialty were dropped from the data. This resulted in the loss of 814,297 observations.

The outcomes variable of interest in the main analysis is the number of wRVUs associated with each given code, denominated in relative value units. To approximate the aggregate effects of RUC rotating committee memberships on total Medicare payments, I use three dependent variables. The first, projected 2012 payment, is the product of total RVUs, conversion factor, service volume, and the inflation rate given by the Bureau of Labor Statistics. The second, the actual 2012 payments, is the product of observed payments by code, and the inflation rate. The third, aggregated wRVUs billed, is the product of wRVUs, and service volume for a given code. Because level values for each of these variables differ substantially, I present outcome variables as logs to facilitate cross-comparison.

Within robustness checks, I use facility practice expense, non-facility practice expense, malpractice RVUs as placebo specifications. Although the RUC has had input into valuation for each of these components, historically its influence has been relatively limited in comparison to its role in the establishment of wRVUs: practice expenses values were historically pegged to a geographically varying index of relative costs and malpractice expenses were pegged to malpractice insurance premiums. The number of observations for practice expense RVUs differ from wRVU values because the initial RBRVS did not distinguish between facility and non-facility practice expense RVUs.

In order to determine whether any effect is due to the code nomination process or to the

RUC, I use an indicator for whether the work RVU value for any given year differs from that of the previous year. Of note, this is not necessarily equivalent to the codes reviewed in any given year, which is unobserved. Rather changes in RVU value from year to year may be due in part to payment schedule changes or changes in date of data acquisition as is the case for the years 1996, 2002, 2006 and 2009. As a result, I run a robustness check to validate study findings are not the pure result of such data abnormalities.

2.2.3 Empirical Specification

I construct independent variables for the number of specialties billing by code and RUC rotating seat representation by code by year as specified in the section above describing independent variables.

The main specification for this study is:

$$wRVU_{it} = \alpha + RotSeat_{it}\beta + \left[RotSeat_{it} \cdot f(count_i)\right]\gamma + f(count_i)\xi + \eta_t + \mu_i + \epsilon_{it} \quad (2.1)$$

where $wRVU_{it}$ is the number of work RVUs assigned to code i in year t, $RotSeat_{it}$ is an indicator for whether code i was represented by a rotating specialty on the RUC in year t, $count_i$ is the inverse level of specialization for code i as identified by the number of specialties observed to bill for code i in the 2007 Medicare data, $f(.)$ is a function, η_t is the year-level fixed effect for year t, μ_i is the code-level fixed effect for code i and ϵ_{it} is a presumed zero-mean error term.

In Figure 2.2, let $f(x) = \sum_{n=1}^{N} 1[x = n]$ where each $n \in [0, N]$ represents an inverse level of specialization and N represents the lowest level of specialization observable as codes with $count_i = N$ are shared by all observed specialties. Then:

$$wRVU_{it} = \alpha + RotSeat_{it}\beta + \left[RotSeat_{it} \cdot \sum_{n=1}^{N} 1[count_i = n]\gamma_n\right] + f(count_i)\xi + \eta_t + \mu_i + \epsilon_{it} \quad (2.2)$$

I plot the estimates obtained from the above regression, $\hat{\gamma_n}$, against a series of dummy variables indicative for each number of specialties billing from 1 to 59, and the interaction at each level of specialization with the RUC rotating seat variable, again with both year and code-level fixed effects. Due to collinearity, it was not possible to evaluate rotating seat-associated effects for codes with a number of specialties billing greater than 59 as the majority of such codes do not exhibit substantial variation by rotating seat. Figure 2.2 plots regression coefficients from these interaction terms of RUC rotating seat occupation and number of specialties billing on the y axis against number of specialties billing on the x-axis. Bars indicate estimated standard errors.

The series observed in Figure 2.2 is downward sloping, thus there is a positive interaction effect between rotating seat membership and higher specialization. In each regression in which no estimate for interacted rotating seat effect and level of specialization is shown, $f(x) = 0$, thus yielding a form:

$$wRVU_{it} = \alpha + RotSeat_{it}\beta + \eta_t + \mu_i + \epsilon_{it} \tag{2.3}$$

Notice here that because no variation in $f(count_i)\xi$ exists, we distinguish mean effects by level of code specialization from μ_i as both effects vary only at the code level i, thus no coefficient $\hat{\xi}$ is reported. I report the coefficients $\hat{\beta}, \hat{\eta_t}$ for $t \in [1995, 2012]$ with the year 1994 acting as the excluded year.

To obtain the aggregated results displayed in the subsequent columns of Table 2.3, in which estimation by level of specialization is shown, I let $f(x) = 1[count_i > \bar{n}]$ where $\bar{n}$ is the threshold for which the groups higher and lower specialization are drawn. To wit, codes i with $count_i \leq \bar{n}$ are termed as highly specialized and codes i with $count_i > \bar{n}$ I term to be less specialized. I choose the cutpoint of number of specialties billing, $\bar{n}$, equal to 6 as this was the median for my sample.

I regressed work RVUs onto the dummy variable indicating specialty occupation of a

RUC rotating seat, level of specialization group (i.e. highly specialized or less specialized), the interaction of the RUC rotating seat dummy and level of specialization, an indicator for each year of data in a fixed effects framework. In this model, work RVUs are demeaned by code in order to generate within-code level estimates of effects. Degrees of freedom and statistical testing are adjusted accordingly. I present aggregated results by regressing log of projected 2012 payment and log of actual 2012 payments, and log aggregated work RVUs onto RUC rotating seat representation, year fixed effects, and code-level fixed effects.

I conduct robustness checks accounting for data inconsistencies, as well as exercises in mechanistic confirmation using the same specification as for the main analysis. These robustness checks are displayed in Table 2.4. Robustness checks are as follows: main specification excluding data from abnormal years (classified as above), demonstration of robustness to grouping specification using cutoffs at levels of specialization at the 25th and 75th percentiles rather than the median, clustered standard errors at the code-level and code-type level, removal of code-level fixed effects using first-differencing rather than fixed effects, placebo regressions using practice expense and malpractice RVUs as outcomes of regression, and falsification test of predicting any change in wRVU value. Code-type is the 7-level CPT classification of anesthesia, evaluation and management, medicine, pathology/lab, radiology, or surgery assigned to the code by CMS.

In addition to the aforementioned approach, I use an event study to confirm that some of the timing for increases corresponds to years of committee membership rather than general increases to specialty-associated codes over time. Here I regress wRVUs on a series of lagged dependent variables ranging from four years prior to the first year of rotating seat occupation to four years after the first year of rotating seat occupation separately for both specialized and less specialized codes where levels of specialization groups are defined as for the main analysis. Regressions contain code-level and year-level fixed effects.

$$wRVU_{it} = \alpha + \sum_{-J}^{J} \phi_j L^j RotSeat_{it} + \eta_t + \mu_i + \epsilon_{it} \tag{2.4}$$

where $L^j RotSeat_{it} = RotSeat_{i,t-j}$. Specialties often repeat membership (shown in Table 2.1), thus a given year may be double marked. For example, if geriatrics occupies a RUC seat in years t and $t+1$, and then again in years $t+4$, $t+5$, then both year t and $t+4$ will be marked as the first year of membership, $t+2$ will be marked both as two years following membership and two years before membership. In Figure 2.3, the choice $J = 4$ allows for the display of some pre- and post- trends. Figure 2.3 plots estimates $\hat{\phi}_j$ versus j.

2.3 Results

Descriptive statistics at the code-year level are summarized in Table 2.2 for the variables wRVUs, number of specialties billing, and rotating seat representation including means, and standard deviation. Codes that could not be matched to specialty were dropped, thus this table displays only the average values within remaining codes in the dataset. As anticipated, there is significant variation in wRVUs between different codes, with less variation within any given code over time. The average code in the sample has 12.7 specialties billing. Billing by a rotating seat specialty is observed for 28 percent of the sample. For comparison, I also report the percent of the sample identified as being billed for a permanent seat specialty.

Because we cannot expect that the effect of rotating seat occupancy will be the same for highly specialized and less specialized codes, I display the effect size by number of specialties billing in Figure 2.2. I find that the effect of rotating seat representation on average reimbursements seems to decline with increasing numbers of specialties billing. That is, the more specialized a code, the higher associated change corresponding to RUC representation. Moreover, I find that for highly specialized codes, the association between RUC rotating seat representation and wRVU is positive. Conversely, for less specialized codes the association between RUC rotating seat representation and wRVU is negative.

Table 2.3 displays the average effects of rotating seat occupancy on corresponding code wRVUs as well as back-of-the-envelope projected payments. In a code-level fixed-effects specification where wRVUs are regressed onto RUC rotating seat representation and year fixed effects, we can see a small overall increase associated with RUC rotating seat membership 0.028 (se=0.010). When I decompose this measure into that derived from highly specialized versus less specialized codes, one may see that the result in the former column is driven disproportionately by increases to codes highly specialized codes 0.177 (se=0.024) and controlling for average year on year increases, less specialized codes actually decline in value with associated RUC rotating seat representation -0.183 (se=0.026).

The following three columns in Table 2.3 should be interpreted as back-of-the-envelope calculations giving aggregated RUC rotating seat effects on spending. Because outcomes are given in log values, regression coefficients may be interpreted as percentages. RUC rotating seat membership corresponds to an average 5.0 percent (se=0.6 percent) increase in naively computed projected payments, 3.0 percent (se=0.7 percent) increase in observed Medicare payments and a 4.9 percent (se=0.5 percent) increase in the aggregate number of wRVUs billed.

In Figure 2.3, I display the results of the event study. For highly specialized codes, the pre-RUC membership trend is downwards, however, a discontinuous increase can be observed with a timing corresponding to the beginning in the first year of RUC membership. This pattern is not observed in the sample restricted to less specialized codes.

Robustness checks are displayed in Table 2.4. Results were robust to numerous specifications. Placebo regressions using facility practice RVUs, non-facility practice RVUs, and malpractice RVUs are available. Facility practice and non-facility practice RVUs do not exhibit the same pattern of increases reflected in wRVUs. For facility practice RVUs, RUC representation for highly specialized codes was associated with -.170 (se=.036) RVUs and for non-facility practice RVUs, RUC representation for highly specialized codes was associated with -.060 (se=.048) RVUs. Malpractice RVUs exhibit a similar pattern of increases in RVUs with estimated correlation .137 (se=.023) and -.142 (se=.025) for highly and less specialized codes, respectively. Regression with an outcome of any observed change in wRVU value was conducted and found that there is a negative association between RUC rotating seat membership and observing a change in the wRVU value for highly specialized codes -.039 (se=.004) and positive association between RUC rotating seat membership and observing a change in the wRVU value for less specialized codes .0183 (se=.004).

2.4 Discussion

Since the adoption of the RBRVS, CMS have faced the problem of how to value physician work. As such, they have relied on an advisory committee of which the majority of members are medical specialists. This analysis provides both quantification of these effects and their distributional impact using a plausibly exogenous source of variation. This paper is able to suggest that some proportion of changes are specifically correlated to committee membership rather than due solely to global payment schedule changes related, for instance, to technological changes alone.

I find a positive association between code-level reimbursement values and specialty committee representation for highly specialized codes, which occur at the expense of reimbursement for less specialized codes. Increases correlate with years of committee membership and are not consistently observed in other reimbursement components for which historically there have been caps on the extent of committee action i.e. practice expense components. Malpractice RVUs, which are computed from malpractice insurance rates, do exhibit an associated change in correspondence to RUC rotating membership, however, to the extent that insurance rates could be tiered by RVUs generated, malpractice RVUs may respond to wRVUs changes.

There does not seem to be a correspondence between likelihood of code valuation change and committee membership. This suggests that the increases in specialty-relevant code reimbursements associated with rotating seat occupancy are a consequence of the RUC rather than as a submission bias on the part of CMS or the CPT editorial panel.

In terms of discussion of the event study, it is necessary to note that due to repeated recycling of the same rotating subspecialties on the RUC, it is difficult to independently estimate the particular effects for each year in the cycle. For instance, the reader should interpret the large average increase in wRVUs observed in the third year post-membership as implicitly combined with the first year pre-membership results as, due to the predominance of two-year terms, these years are often tagged together. Nevertheless, the event study results are informative. Firstly, they demonstrate that over time there is an increase in the overall level of reimbursement. Secondly, they demonstrate that this effect is concentrated in highly specialized codes, thus decreases in the value of less specialized codes in the main specification may be occurring as a consequence of failure to increase the value of these codes in correspondence with year-over-year changes rather than the initiation of purposeful decreases. Thirdly, it demonstrates that, for codes associated with RUC rotating seats, positive changes in the valuation of associated codes are timed with RUC membership and that the effect is sustained. This means that after a rotating member loses his seat, the work value for specialty-related codes remains high and other rotating seat members do not actively lower wRVUs for non-specialty associated codes in representative absence.

Overall, these changes represent on average a 3-5 percent increase in Medicare specialty reimbursement per year of committee membership. As an example of implicit scale, according to the PSPS in 2007, Medicare paid 1.5 billion dollars to gastroenterology. Thus, a three percent annual increase would represent a 45 million dollar increase in Medicare annual payments to gastroenterologists as a result of a single year of RUC membership.

Several additional limitations of this paper are as follows. I am limited by restricted data access. During the 2007 internal review of specialty participation on the RUC, the AMA noted that no significant distributive outcomes were found from voting records. I am able to roughly replicate the results of this review in my initial analysis using publicly available data, however, the inability to observe votes somewhat limits my ability to discriminate the mechanism for any RUC-associated increases. Additionally, it is not possible to provide an estimate for the larger part of the RUC composition because it is composed of permanent representation. Because this paper looks only at within variation for codes matched to specialties in the year 2007, it is not able to comment on the extent to which committee composition may have an effect between codes or in the introduction of new codes.

Nevertheless, these findings are important for health care regulation. The short-term marginal increases reported here cannot explain the growth in the income gap between general practitioners and specialists, nor should they be expected to. Physician income is driven by a variety of factors related or unrelated to within code price effects. However, these findings suggest that the current procedural evaluation process may be biased toward the perpetuation of higher specialty reimbursement shares of overall health expenditures. As such, the current committee dynamic has the possibility to exacerbate existing disparities, which occur to the detriment of aims to increase the supply ratio of primary care physicians to specialists.

Changes in RUC composition have already been made. However, given the current cap in overall spending, it is difficult to imagine a context in which committee incentives would align to give generalists a bureaucratic advantage. One possibility is to consider less specialized codes, or codes of particular concern by some other metric (e.g. evaluation and management codes), as separate from highly specialized codes and then subject such codes to a segregated pool of spending. This would enable more targeted control of reimbursement growth in given categories while acknowledging the essential differences that exist between types of services and procedures.

A larger issue at stake is how to assess physician work overall. Like fee-for-service, the RBRVS implies reimbursement should be coupled with marginal costs. In the context of an increasing emphasis on accountable care, future physician payments may be tied instead to marginal patient benefit. Systemic reorganization of physician reimbursement may constitute an opportunity to create an improved balance between specialty expenditure shares, however, it is no guarantee. For one, specialty and non-specialty care often have different aims and different measurable outcomes for patients within the short, medium, and long-terms. While treatment of an illness may constitute a discrete measure, prevention of one is worthwhile, but often more difficult to quantify. It may be necessary to recognize that specialists and non-specialists function differently within the context of modern medical care and it may similarly be reasonable to formulate accountable reimbursements separately with these differences in mind.

Table 2.2: Distributional summary of codes by RVU component, payments, RUC representation and code type.

	Mean	Overall Std Dev.	Std Dev Between Codes	Std Dev Within Code	Num Obs
Distribution of RVU Components					
Number of work RVUs	4.75	8.36	7.53	1.37	234027
facility practice RVUs	3.68	5.8	5.28	1.97	223085
non-facility practice RVUs	3.97	6.94	8.73	2.62	223372
malpractice RVUs	0.84	2.13	1.4	1.51	234025
Distribution of Payments Projections					
Projected payments by code from total RVUs	8,844,740	1.01×10^8	-	-	89468
Total Medicare payments by code	7,434,888	8.40×10^7	-	-	93555
Number of total work RVUs billed	114,694	1.64×10^6	-	-	90942
Distribution of RUC Representation					
Number of specialties billing a code in 2007	12.66	16.21	-	-	234027
Pct of codes rep by a RUC rotating seat	28%	45%	-	-	234027
Pct codes rep by a RUC permanent seat	79%	41%	-	-	234027
Pct codes work RVU change from last year	15%	36%	-	-	234027
Distribution of Code Types					
Percent of codes type anesthesia	1%	9%	-	-	185070
Percent of codes type eval and managem't	1%	12%	-	-	185070
Percent codes type medicine	13%	33%	-	-	185070
Percent codes type path or lab	12%	32%	-	-	185070
Percent codes type radiology	20%	40%	-	-	185070
Percent codes type surgery	53%	50%	-	-	185070

Table 2.3: Average Effect of Rotating RUC Seat by Level of Specialization.

Variable	Work RVUs	Work RVUs	log(Projected Payments)	log(Actual Payments)	log(Aggregated wRVUs)
Code represented by RUC rotating seat	0.028 (0.010)**	0.177 (0.024)**	0.050 (.006)**	0.030 (.007)**	0.049 (.005)**
Int with less specialized code		-0.183 (0.026)**			
Year 1995	-0.023 (0.020)	-0.024 (0.020)			
1996	0.086 (0.019)**	0.084 (0.019)**			
1997	0.398 (0.019)**	0.398 (0.019)**			
1998	0.336 (0.021)**	0.337 (0.021)**			
1999	0.379 (0.019)**	0.382 (0.019)**			
2000	0.421 (0.019)**	0.423 (0.019)**			
2001	0.592 (0.019)**	0.594 (0.019)**	0.009 (0.009)	0.051 (0.011)**	0.053 (0.009)**
2002	0.595 (0.020)**	0.597 (0.020)**	-0.032 (0.009)**	0.087 (0.012)**	0.094 (0.009)**
2003	0.633 (0.019)**	0.634 (0.019)**	0.037 (0.009)**	0.092 (0.011)**	0.144 (0.008)**
2004	0.635 (0.019)**	0.636 (0.019)**	0.014 (0.008)	0.092 (0.011)**	0.156 (0.008)**
2005	0.643 (0.019)**	0.644 (0.019)**	0.010 (0.008)	0.093 (0.011)**	0.165 (0.008)**
2006	0.659 (0.018)**	0.660 (0.018)**	-0.044 (0.008)**	0.050 (0.011)**	0.142 (0.008)**
2007	1.062 (0.018)**	1.062 (0.018)**	-0.080 (0.009)**	-0.020 (0.011)	0.176 (0.008)**
2008	1.071 (0.018)**	1.070 (0.018)**	-0.081 (0.008)**	-0.060 (0.011)**	0.205 (0.008)**
2009	1.069 (0.018)**	1.069 (0.018)**	-0.092 (0.009)**	-0.008 (0.011)	0.236 (0.008)**
2010	1.122 (0.018)**	1.122 (0.018)**	-0.062 (0.009)**	0.001 (0.011)	0.240 (0.008)**
2011	1.127 (0.018)**	1.127 (0.018)**	-0.079 (0.009)**	0.002 (0.011)	0.240 (0.008)**
2012	1.159 (0.018)**	1.161 (0.018)**			

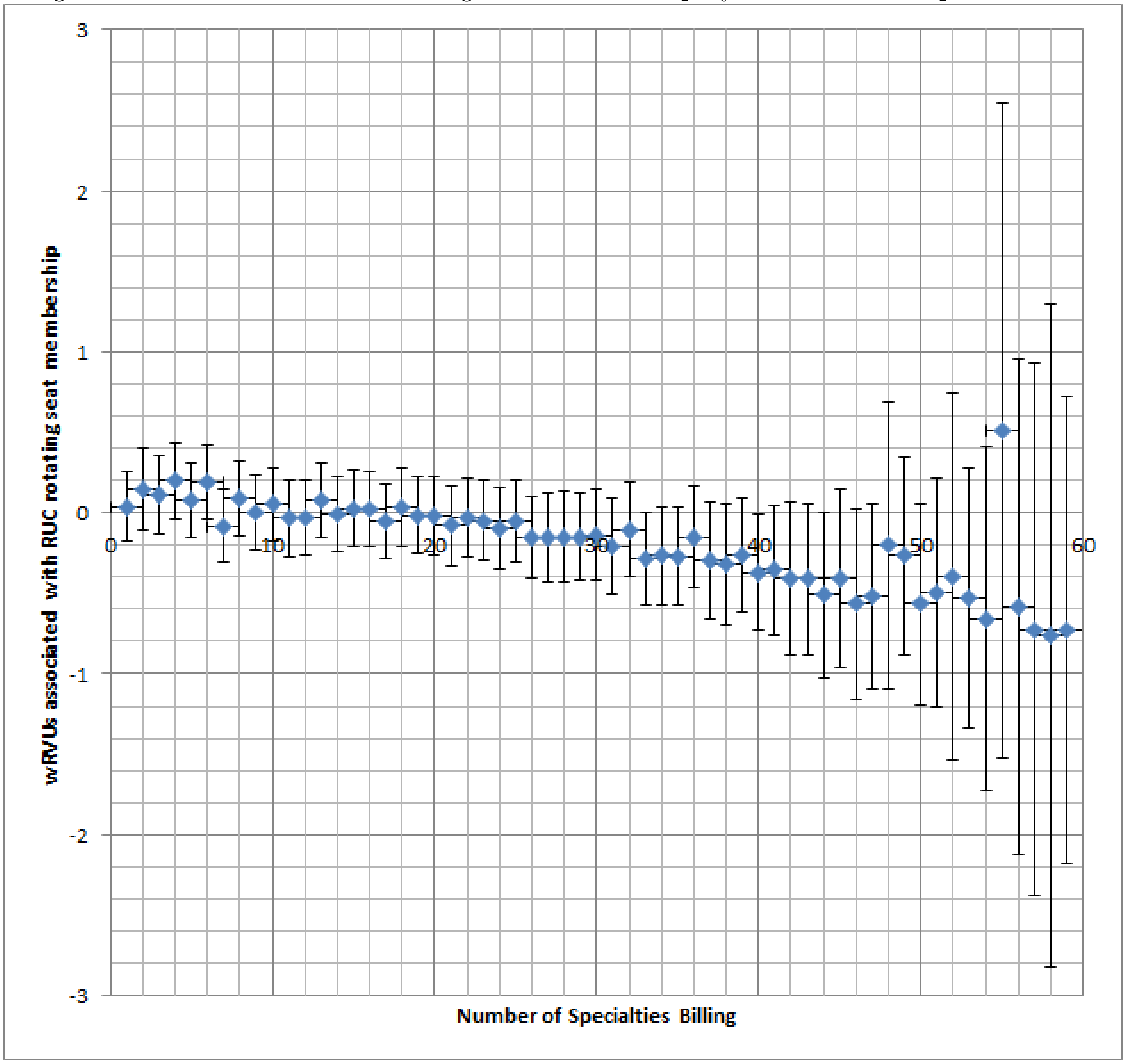
Columns headers give the dependent variable of each regression. Projected payments is the product of total RVUs, conversion factor, service volume and the inflation rate given by the Bureau of Labor Statistics. Actual payment is the product of observed code payments and the inflation rate. Aggregated work RVUs is the product of wRVUs and Service Volume for a given code. Interaction with less specialized code indicates the interaction effect of RUC rotating seat membership and a dummy variable indicating the code is billed by more than 6 specialties in 2007 Medicare. All regressions are linear and include year indicators (shown) and code-level fixed effects. Standard errors displayed in parentheses. ** Indicates Wald test was significant at the at the 5% level.

Table 2.4: Robustness Checks

Outcome variable	Code represented by RUC rotating seat	Interaction with less specialized code	Model	N
(1) wRVUs, excluding abnormal years	.209 (.027)**	-.180 (.030)**	FE	186108
(2) wRVUs, less specialized code defined as # specialties billing > 2 (25th p'ile)	.139 (.060)**	-.115 (.061)*	FE	234027
(3) wRVUs, less specialized code defined as as # specialties billing > 18 (75th p'ile)	.083 (.012)**	-.173 (.021)**	FE	234027
(4) wRVUs, clustered stand errors by CPT	.177 (.039)**	-.183 (.011)**	FE	234027
(5) wRVUs, clustered stand errors by type	.184 (.024)**	-.181 (.005)**	FE	185070
(6) wRVUs, first differences	.106 (.023)**	-.135 (.023)**	FD	209449
(7) facility practice RVUs	-.170 (.036)**	.250 (.040)**	FE	223085
(8) non-facility practice RVUs	-.060 (.048)	.194 (.053)**	FE	223372
(9) malpractice RVUs	.137 (.023)**	-.142 (.025)**	FE	234025
(10) observed wRVU value differs , linear from previous year value	-.039 (.004)**	.0183 (.004)**	FE	234027

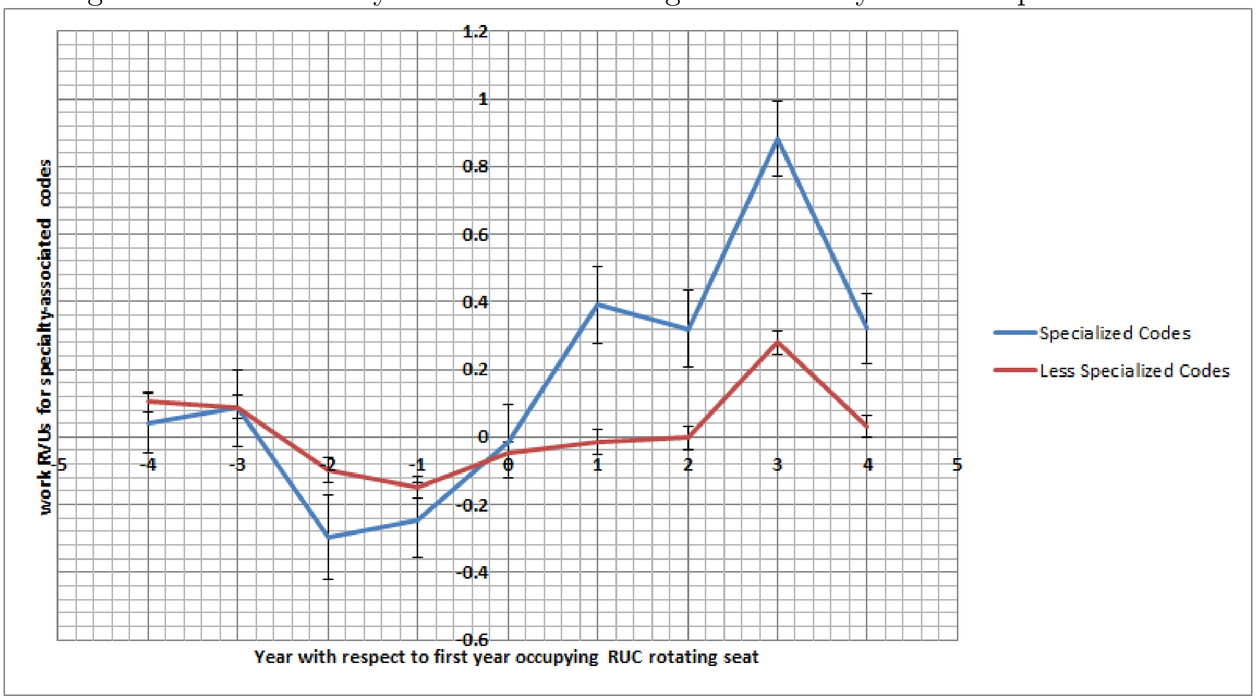
All regressions include year dummies. Interaction with less specialized codes denotes the interaction of rotating seat representation and an indicator for codes with the number of specialties billing greater than 6 unless otherwise specified. Observed wRVU value differs from previous year value is a binary outcome variable for this outcome. Abnormal years defined as 1996, 2002, 2006, 2009. Clustered standard errors by CPT is the same as HCPC and modifier. Clustered standard errors by type are the Medicare category classifications of CPT code. The classification categories are anesthesia, evaluation and management, medicine, pathology/lab, radiology, and surgery. ** Indicates Wald test was significant at the at the 5% level. * Indicates Wald test was significant at the 10% level.

Figure 2.2: Effects of RUC Rotating Seat Membership by Code-Level of Specialization.



Plotted circles represent the average effect size on wRVUs of RUC rotating seat membership among codes with a given level of specialization. They are plotted against the corresponding levels of specialization on the x-axis. Because little rotating seat variation exists for codes billed by large numbers of specialties, estimates of rotating seat effects by code specialization could not be obtained for these codes. Thus, only effects for codes shared by a maximum of 59 specialties are reported. Bars indicate 95% confidence interval.

Figure 2.3: Event Study of Effect of Rotating RUC Seat by Level of Specialization.



The blue line displays the effect on work RVUs for codes for which 6 or fewer specialties bill. The red line displays the effect on work RVUs for codes for which greater than 6 specialties bill. The zero line on the x-axis represents the first year of that specialtys term in a RUC rotating seat. Points -4, -3, -2, -1 on the x-axis denote the years 4, 3, 2, and 1 year before RUC membership begins respectively. 1, 2, 3, 4 are equal to one in the first, second, third and fourth years following the first year term in a RUC rotating seat. 95% confidence intervals are given by the error bars.

CHAPTER 3

MENTAL HEALTH MEASUREMENT IN PROGRAM EVALUATION: AN EXAMPLE FROM THE OREGON HEALTH INSURANCE EXPERIMENT.

3.1 Introduction

Program evaluators commonly appropriate clinical mental health questionnaire when the outcome of interest in a study is mental health. For example, both the Johns Hopkins Longitudinal Effects of Housing on Health and Social Adjustment of the 1950s and the Moving To Opportunity (MTO) examined the mental health consequences of neighborhoods and housing by this method [57]. Several papers in the development literature have also examined mental health using clinical screening questionnaire outcomes, as has the landmark Oregon Health Insurance Experiment (OHIE), in which [10] make use of scores from the eight question version of the Patient Health Questionnaire (PHQ-8), a common self-report questionnaire used to screen patients for clinical depression [17, 47, 10, 60]. However, clinical screening questionnaires have not been designed for use in such program evaluations, and their use comes with drawbacks. For example, the score obtained from the PHQ-9 may be difficult to compare to a score produced by the PHQ-2, the two-item version of the same questionnaire. What is more, it is difficult to relate a score from the PHQ-9, a scale for depression, to other scales for anxiety, health-related quality of life, or other reasonable candidates for an outcome variable related to mental health. In the past, this incompatibility for comparison led researchers to use dichotomized outcomes, e.g. the participant screened positive for depression or not, but this dichotomization comes at the cost of the efficiency of estimation [25].

In this paper, I borrow two tools from psychometrics: item-response theory and the bifactor model. Both these tools are potentially valuable for program evaluators interested in mental health. I model physical and mental health jointly and demonstrate that there is a stable relationship between different items from several health-related questionnaires. Subsequently, I derive a score from the model for mental health-related functional status scores. I then use mental health-related functional status (I will call it functional status, for short) as the outcome variable in my analysis using data from a well-known setting with exogeneous treatment variation: the Oregon Health Insurance Experiment (OHIE). I demonstrate several advantages to using functional status scores including an improvement in the power to detect changes in mental health due to program participation, an improvement in the power to detect treatment heterogeneity, and, a philosophical improvement, which better aligns the measured quantity in the study with the entity of primary interest for many program evaluators.

In the next section, I discuss the some background on psychiatric instruments, as well as, the modeling approach used. In Section 3, I describe my dataset and the methods by which I design my measure of mental health-related functional status and the context of OHIE, which I have chosen as the application for this method. In Section 4, I will present the results. In the final section, I conclude.

3.2 Background

3.2.1 Mental Health as an Outcome in Program Evaluation

Although the goal of mental health measurement may differ by program, most economic interest stems from poor socioeconomic outcomes or disability associated with poor mental health [80, 51, 83]. However, disability or socioeconomic outcome attributable to mental illness is not directly observable, thus, economists frequently employ clinical screening questionnaire scores as a proxy for poor mental health [17, 47, 10, 57]. Within economics, the attenuation bias produced by the use of such proxy variables for mental health has been widely recognized [21]. However, the clinical literature reveals that there is only a weak correspondence between clinical screening questionnaire scores and general functional impairment [8]. Thus, it is possible that some of the most commonly used proxy variables are in fact poorer proxies for mental health-related functionality than previously appreciated.

While the weak correspondence between symptom scores and functionality may appear unintuitive, this weak correspondence is sensible when we consider the goal of designing a screening questionnaire for a specific diagnosis. Many psychiatric diagnoses, for example major depression, are substantially heterogeneous. Screening questionnaires, however, must be brief. As a result, assessment items in many clinical screening questionnaires may be restricted to items that test close to the diagnostic threshold. For example, items such as, "Over the last two weeks, I have felt tired or had little energy more than half the days," are more likely to be incorporated into a screening instrument for major depression than others such as, "Over the past month, I have attempted suicide." Because the role of a screening questionnaire is to flag a patient for referral, the severity of symptoms need not be reflected directly by the score on the questionnaire. This task is reflected in studies of clinical screening questionnaires, which report the questionnaire's percent correspondence with clinician's binary diagnosis (i.e. depression or not) rather than a correspondence with a clinician's assessment for severity [60, 99].

While a questionnaire may excel at the screening task by incorporating many questions reflecting only moderate severity, this same feature becomes a bug when used for the program evaluation task. After all, a questionnaire score reflects the patient's risk for a particular psychiatric diagnosis, it does not monotonically reflect underlying mental health for very ill patients. As an example, consider a screening questionnaire for clinical depression. One person may concurrently endorse mild fatigue, sleep difficulties, sadness, poor concentration, and guilt and yet may continue to function normally, while another may report amotivation, but amotivation enough to impair the activities of daily living. This severity, is not captured by the questionnaire score. The result is clinical screening questionnaire scores may be non-monotonic to latent mental health in the following way. As mental health worsens toward the diagnostic threshold, the questionnaire will produce a corresponding increase in the score. However, beyond the diagnostic threshold, the score is not guaranteed to be monotonic to mental health in any way. This problem is of particular concern for those using scores as the outcome of interest or a linear transformation of score, as is done by [57].

Recognizing this problem, many investigators have corrected the non-monotonic score problem by specifying it as a binary outcome. However, this approach is not without disadvantages. Because the binary outcomes used typically identify only those who screen positive for a given mental illness, there may be many false positives within this sample, particularly for rare diagnoses. Again, this is because clinical screening questionnaires are by nature conservative and refer many patients who, ultimately, will not be diagnosed with a psychiatric illness. What is more, by condensing the data to a single binary outcome, the investigator also throws away much of the information about ordering we could have obtained from the items [25].

One simple solution is to incorporate direct general functional status instruments to study mental health. Several general functional status items abound have already appeared in such large national surveys such as the Medical Expenditure Panel Survey and the National Longitudinal Survey of Youth, such as the 12- and 36-item short-form health survey (SF-12, SF-36). By comparison to diagnosis-specific questionnaires, these health surveys can be administered at a similar cost, many of them come in a version with 12-items or fewer, but are not used as frequently for mental health. One possible reason they have not been more widely used for mental health has to do with perceived interpretability. This is because functionality surveys also measure dysfunction related to poor physical health. Investigators who draw a sharp distinction between physical and mental health-related functional status, may define the former and the latter as disjoint, or at least amenable to independent modeling. For instance, SF-12 may be decomposed into its physical and mental health components and analyzed as two separate outcomes. Also related to interpretability, a diagnosis-specific questionnaire produces a reportable of patients screening positive for a given diagnosis whereas relative functionality is more difficult summarize. Researchers may therefore demonstrate a preference for the more "tangible" outcome.

Despite the infrequency of its use for mental health in policy studies, functional status is an important concept in psychiatry. A view of diagnosis-specific functional status is not generally supported by the psychiatric literature, whether the diagnosis is mental or physical. Both the Diagnostic and Statistical Manual-V (DSM-V), the cornerstone manual of psychiatric diagnosis, and a considerable clinical literature describe physical and somatic manifestations of psychiatric illness, as well as considerable heterogeneity across demographic dimensions [101, 11, 42, 22, 67]. As a result, the clinical definition of functionality is purposefully general and is not specific to any specific diagnosis. Recently, the DSM-V adopted World Health Organization Disability Assessment Schedule 2.0 over the previous Global Assessment of Functioning Scale (GAF), the former being a scale metric for disability which had its origin in physical health [41]. These clinical changes are in accordance with findings from neuroscience, psychology, and anthropological accounts of mental illness, which have yielded the theory of constructed emotion [13, 12]. According to this theory, some baseline somatic experience is common across sufferers within diagnoses. Thus, while item responses based on the culturally-embedded interpretation of those symptoms vary, some physical component should remain common across cases. What is more, unlike diagnosis, functionality is defined by the presence of some poor outcome or disability. This consideration makes it perhaps the most relevant outcome for policy investigators.

3.2.2 A Bifactor Modeling Approach

The bifactor model belongs to a class of latent response factor models first introduced by [49], although the particular item-response theory version of this model I borrow from [37]. As in other latent response model, we can imagine that N participants, indexed by i, may be asked to respond to a sequence of J items indexed by j. Their responses X_{ij} are coded as affirmative if their latent response propensity, X^*_{ij} exceeds a given threshold, γ_j. Latent response propensity is modeled in accordance with a known parametric distribution, in this case logistic, and as a function of individual-level latent factors θ_{ik} across K subdimensions with $K < J$. More explicitly,

$$\begin{aligned} \Pr[X_{ij} = 1|\theta_{i1}, ..., \theta_{ik}] &= \Pr[X^*_{ij} > \gamma_j|\theta_{i1}, ..., \theta_{ik}] \\ &= \Phi\left(\frac{\gamma_j - X^*_{ij}(\theta_{i1}, ..., \theta_{ik})}{\sigma_j}\right) \end{aligned} \tag{3.1}$$

where

$$X^*_{ij}(\theta_{i1}, ..., \theta_{ik}) = \sum_{k=1}^{K} \alpha_{jk}\theta_{ik} + \epsilon_{ij} \tag{3.2}$$

and ϵ_{ij} is a standard normal error term and σ_j is the standard deviation for item j.

The bifactor restriction we impose is that, in addition to $\alpha_{j1} \neq 0$, only one of α_{jk} can be non-zero for $k = 2, ..., K$. Put another way, the matrix α of discrimination parameters takes the form

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} & 0 \\ \alpha_{21} & \alpha_{22} & 0 \\ \alpha_{31} & 0 & \alpha_{33} \\ \alpha_{41} & 0 & \alpha_{43} \end{bmatrix}. \tag{3.3}$$

Notice that as an alternative to the full unrestricted multivariate probit, the bifactor restrictions permit the investigator to model multidimensional data while maintaining both identification and computational tractability [18]. Investigators may find that this model is particularly well-suited to situations where the outcome of interest is best represented as an underlying component to a multidimensional entity, for instance depression may present across a range of items within different domains such as sleep, appetite, amotivation, or hopelessness.

For each item, the parameters estimated by the model can be estimated via maximum likelihood. In particular, we will estimate the threshold, $\hat{\gamma}_j$, and the factor loadings, $\hat{\alpha}_{jk}$, for each item j across each subdimension k. From these estimates, we can derive the predicted values for each individual's latent ability score, $\hat{\theta}_{ik}$, for each individual i over each subdimension k. In the subsequent sections, I will call $\hat{\theta}_{i1}$ the "common dimension score" or, simply the "score".

The parameters of this latent-factor model have interpretations. γ_j, the threshold, reflects the quantity of the average latent parameter required in order to elicit an endorsement of the item j. For example, imagine a unidimensional model where latent mental health is ordered such that lower θ_{ik} is worse mental health. Imagine our questionnaire contains two items: $j = 1$, "Over the last two weeks, I have felt tired or had little energy more than half the days," and $j = 2$ "In the past month, I have attempted suicide." We might expect that item thresholds, γ_1 and γ_2, to possess the feature that $\gamma_1 > \gamma_2$ because a lower latent level of mental health is required in order to positively endorse item 2. Making this distinction between item thresholds is useful. At the mean level of population mental health, very few respondents will positively endorse item 2, however, item 2 may be better able to discriminate between two individuals with low versus very low levels of mental health. Thus, although only item 1 is a viable candidate as an item on a standard clinical depression screening questionnaire, item 2 is useful for the purposes of program evaluation because it provides an ordering over $\hat{\theta}_i$ for low mental health patients that would have been otherwise unavailable.

The parameter α_{jk} represents how good an item j is at discriminating respondents on the basis of subdimension k. That is, imagine that our items fall into one of two subdimensions, θ_{i1} depression and θ_{i2} anxiety. Now imagine we have three items: $j = 1$ "Over the past two weeks, I felt down, depressed, or hopeless most days," $j = 2$ "Over the past two weeks, I have had trouble falling or staying asleep most days," and $j = 3$ "Over the past two weeks, I have not been able to stop or control worrying most days." In accordance with DSM-V criteria, item 1 and 2 correspond to depressive symptoms, while items 2 and 3 correspond to anxiety symptoms [7]. We would expect the factor loadings α_{11}, α_{21}, α_{22} and α_{32} to be significant, that is, closer to one. On the other hand, we would expect that α_{12} is close to zero because we would not expect this item to closely map with other items pertaining specifically to anxiety. In general, a factor loading is considered insignificant at any value below 0.3 [45]. [27] suggest that any factor loading with a value below 0.32 to be poor, factor loadings above 0.45 to be fair, factor loadings above 0.55 to be good, and above 0.71 to be excellent.

The use of factor loadings to find items that more meaningfully correlate with underlying factors provides an improvement over the use of the simple sum of item scores, as is the current common practice. Use of the simple sum implicitly imposes an assumption that all items contribute equally to our understanding of the latent factor regardless of the item's specificity for that factor. Taking the example of the three items from the paragraph above, item 2, which pertains to sleep, may be less informative with regard to patient's underlying level of depression in comparison to the more specific item, item 1, because item 2 may also be endorsed by participants due to underlying anxiety. Furthermore, the possibility of low factor loadings may guard against the effects of minimally significant items that may have been erroneously included into the questionnaire.

The model limitations arise primarily from the bifactor restrictions imposed. That is, because each item must only load onto one sub-dimension in addition to the common dimension, it may not be appropriate for data in which the sub-dimensional membership of items is not straightforward [70]. [86] suggests that one way to obviate this difficulty is perform an exploratory bifactor analysis in advance of the confirmatory investigation to confirm that cross-loadings are approximately zero. Another possible difficulty with use of the bifactor model, which, in this case, is general to most IRT approaches, is the number of observations must be reasonably large. [53] report that for short tests (fewer than 240-items) scored using a multidimensional graded response model, appropriate parameter estimates can be obtained from a sample size of greater than 500.

In spite of the model limitations, the existing range of applications for bifactor modeling is reasonably broad. Current applications include the exploration of psychological constructs such as depression or self-esteem in psychology as well as medical applications such as the development of computer adaptive tests to screen for mental illness [100, 88, 87, 39, 38].

In this paper, I roughly follow a framework provided by [86]. First, I use a primary exploratory factor analysis to describe groups of variance found in the variables released with [10]. Subsequently, I perform a confirmatory bifactor analysis using an IRT framework, describe the estimated thresholds and factor loadings of the associated items, and compute the estimated scores on common dimension for experiment participants, which I will interpret as mental health-related functional status. Finally, I reevaluate the study findings to describe the effect that health insurance receipt had on mental health-related functional status in this setting.

3.3 Methods

3.3.1 Data

In 2008, the state of Oregon wished to expand its Medicaid program, but did not have to funds to enroll 100% of the interested population. Therefore, Oregon drew names by lottery for its Medicaid program for low-income, uninsured adults generating a large-scale randomized control trial on the effects of health insurance [34]. Among the results of the first year of results, the study authors reported a statistically significant -9.15 percentage point decrease in positive screening rates for depression. In addition to depression screening, the study collected responses to questions pertaining to health-related quality of life, happiness, a battery of health-related diagnoses, and healthcare utilization [10].

The public use data for the OHIE were acquired from the National Bureau of Economic Research. These data include a study population of 20,745 participants living in the Portland, Oregon metropolitan area. 10,405 participants were selected by the lottery to receive Oregon Health Plan Standard public insurance and 10,340 control group participants were selected; the latter were those who had entered the lottery, but who were not selected to receive insurance. At the time of acquisition, outcomes were available from September 2006 to December 2010, or up to 12 months subsequent to randomization. Item-level responses for the PHQ-9 were not yet available.

3.3.2 Measurement Design

Selection of included variables were defined as those factors considered in [10] as relating to depression, health-related quality of life, happiness, with the inclusion of several indicators of high levels of medical service utilization. Variables containing more than 6 categories such as PHQ-9 score were discretized into 6 quantiles to facilitate model convergence.

In order to insure that the restrictions imposed on factor loadings by the bifactor model would not be significantly violated, an exploratory bifactor model was fitted, the results of which are shown in Table B.1. This was done via a traditional factor analysis, followed by bifactor rotation was performed following [52].

The OHIE public use data included data collected from treatment and control groups across four time periods: 0-month mail survey (June-November 2008), 6-month mail survey (January-May 2009), 12-month mail survey (July 2009-March 2010), and the in-person survey (August 2009-October 2011). Because item responses across each of these periods were not identical, I speculate that the relationship between latent functional status and item responses is stable across time. I verify this assumption of temporal stability by first fitting the bifactor model to each time period separately, and present the results in Table B.2.

The use of the 0-month through 12-month data is a novel analysis of this data. By comparison, the main analysis presented in [10] relies only on the in-person survey data along with participant retrospective responses in order to obtain difference-in-difference estimates. It is possible these retrospective responses contain recall errors as they do not necessarily correspond to the responses given by participants during the 0-month period.

Having verified temporal stability of latent functional status with item parameters, I append all four periods into a single dataset of item responses in which repeated observations by the same individual are treated as independent observations. The final bifactor model was fitted to this data using the BIFACTOR program publicly available from the Center for Health Statistics [37]. Estimated factor loadings ($\hat{\alpha}_{jk}$), thresholds ($\hat{\gamma}_j$), and individual-level scores ($\hat{\theta}_{i1t}$) were obtained where i is the index for the individual, $k = 1$ indicates that this value was obtained over the common dimension, and $t \in \{0, 1, 2, 3\}$ is the time period from which that estimate was obtained.

3.3.3 Application

After obtaining individual-time scores for functional status ($\hat{\theta}_{it}$), the analysis can be collapsed into a program evaluation context in which longitudinal data are available for a group of study participants. I consider a simple specification as follows:

$$\hat{\theta}_{i1t} = \beta_1 + \beta_2 Treat_i + \beta_3 Time_t + \beta_4 Treat_i \times Time_t + \upsilon_{it} + \epsilon_{it} \tag{3.4}$$

where $\hat{\theta}_{i1t}$ is the mental health-related functional status score for person i in time t derived from the bifactor fitting, $Treat_i$ is an indicator equal to one if person i was randomized to receive insurance, $Time_t$ is a variable denominated as $\{0, 1, 2, 3\}$ as corresponding to 0-month, 6-month, 12-month, and in-person surveys respectively, and $Treat_i \times Time_t$ is an interaction term. Due to the randomized nature of this intervention, $cov(Treat_i, \epsilon_{it})$, is considered to be zero. However, I also consider fully-interacted specification using a person's baseline level of functional status, where the baseline level of functional status is defined as the estimated 0-month score $\hat{\theta}_{i10}$.

$$\begin{aligned}\hat{\theta}_{i1t} =& \beta_1 + \beta_2 Treat_i + \beta_3 Time_t + \beta_4 \hat{\theta}_{i10} + \beta_5 Treat_i \times Time_t + \beta_6 Treat_i \times \hat{\theta}_{i10} \\ &+ \beta_7 Time_t \times \hat{\theta}_{i10} + \beta_8 \left[Treat_i \times Time_t \times \hat{\theta}_{i10} \right] + \upsilon_{it} + \epsilon_{it}\end{aligned} \tag{3.5}$$

I model υ_{it} using both a mixed model with person-specific time trends and a fixed effects formulation.

An individual must have completed a sufficient number of items across multiple time periods in order to estimate the functional status scores $\hat{\theta}_{i1t}$. A total of 14,402 observations contained repeated observations for that individual, but did not include baseline scores. They are dropped from the model fully-interacted with the baseline.

3.4 Results

3.4.1 Findings from Bifactor Modeling

Descriptive statistics for the variables used in the model are presented in Table 3.1. The score mean value is $\hat{\theta}_{i1t}$. Score variance represents the variance estimate reported by BIFACTOR of $\hat{\theta}_{i1t}$, treating repeated observations of the same individual as independent observations. The number of observations differs across variables due to the use of different items in each time point during follow-up. The item correspondences with the periods in which they were surveyed can be found in Table B.2. As we can see, the experiment administrators were much more consistent in their collection of medical utilization data across periods compared to psychological assessment items. What is more, the particular psychological assessment items are variably collected, with diagnosis of depression collected at higher frequency in comparison to the SF-8 score, which was collected only at the in-person interview.

Table 3.2 presents the estimated item thresholds ($\hat{\gamma}_j$) and factor loadings ($\hat{\alpha}_{jk}$) for each item available at each time point. Factor loadings that are set to zero by assumption are shown as blank. Recall that items with common dimension factor loadings less than 0.3 have poor correlation between item response and underlying latent quantity [45]. Positive subdimension factor loadings reflect the positive correlation of items within each subdomain. For example, column (6) shows the estimated subdimension for hospitalization is 0.93, meaning it is significantly positively correlated with other metrics of utilization such as emergency room visits. Item thresholds reflect the average mental health-related functional status associated with an item endorsement. For example, the estimated item threshold for self-rated general health across all samples, reported in column (1), is -1.155 versus -0.514 for self-rated happiness indicating that a lower level of underlying functional status is required to positively endorse general in comparison to happiness.

Across periods, self-rated mental health items (items with non-zero loading onto subdimension 2 in column (4)) load most strongly onto the common dimension with factor loadings between 0.557 and 0.857. Thus, we confirm that the latent common dimension may have the interpretation of mental health-related functional status rather than general functional status. Although items related to physical health are traditionally thought of as separate from mental health, self-rated physical health items (items with non-zero loading onto subdimension 1 in column (3)) also loaded signficantly onto the common dimension with loadings between 0.511 and 0.653. Previous studies have analyzed each of these outcome separately, without recognizing that they are substantially correlated and reliably ordered. As such, the extent to which changes in self-rated general health may be derived from underlying mental health rather than changes in medical diagnostic test results may have been under-appreciated.

In general, medical utilization variables such as emergency room utilization and hospitalization loaded poorly onto the common dimension with common dimension factor loadings between 0.265 and 0.332. One possible explanation for this findings is that the intervention itself may have decoupled a relationship between utilization and functional status. In other words, Medicaid significantly increased medical service utilization among the treatment group. Since treatment was randomized, we might expect that utilization would be less indicative of functional status than treatment status. This would lead to a violation of the assumption of temporal stability. From Table B.2, this does not appear to be the case. Rather the factor loadings for utilization items onto the common dimension appear to be stably low from the 0-month period onwards. If anything, they increase across time, perhaps indicating that such variables were minimally predictive of functional status at baseline because participants without insurance were reticent to seek care. The stable, insignificant relationship between utilization and mental health may also imply that any changes in mental health derived from an intervention cannot be attributed to utility derived from medical utilization alone, a finding from this study which was previously not well-described.

One thing we glean from examining relative item thresholds is that we are able to distinguish items that have sensitivity at high versus low functional status. Sometimes, the results of such study yields surprising results. A key example of this phenomenon are the items for diagnosis of depression versus self-rated measures of mental health status. In Table 3.2, column (1), we see that the threshold for a diagnosis of depression is -0.149, which is higher than the thresholds for self-rated mental health status, which range from -1.1 to -0.514. Given a traditional prior for the sequence of care, we might expect that in order to be diagnosed with depression, many patients may endorse depressive symptoms, some subset of which will present for treatment and some further subset of which will be diagnosed. Under this paradigm, fewer patients should be diagnosed with depression than endorse symptoms of depression. However, from the bifactor examination, we find the opposite. Because the threshold for receiving a prescription for depression medication is -0.706, lower than the self-rated mental health items, the finding cannot be driven by treated depression. One possible interpretation of this finding is that the appropriate clinical variable to use is that of prescribing as the "previous diagnosis of depression" item may be broader than previously recognized. That is, perhaps the previous diagnosis question captures patients who were diagnosed many years in the past and may flag patients who are more prone toward depression, however, in itself, previous diagnosis is not necessarily useful for program evaluations interested primarily in the mental health determinants of poor functional status.

A final findings from the bifactor analysis portion of the study is, in comparison to individual item responses reflected by the mail survey responses, the aggregations of mental health and physical component items contain significantly less ordinal information regarding functional status by comparison. We find in column (1) the SF-8 physical component threshold ($\hat{\gamma}_j = -0.972$), SF-8 mental component ($\hat{\gamma}_j = -1.051$), PHQ-9 ($\hat{\gamma}_j = -0.978$), and SF-4 pain severity scales ($\hat{\gamma}_j = -0.961$) all measure at approximately the same threshold of functional status. Thus, regression analysis using each of these outcomes should yield similar results as these measures are also substantially correlated. This finding highlights the importance of analyzing item-level outcomes rather than aggregated scoring only, a point made also by [54].

3.4.2 Program Evaluation Results

Figure 3.1 compares the treatment versus the control groups at each of the 0-month (baseline), 6-month, and 12-month time points. Here we see that mental health related functional status is on average better at baseline for the treatment group. This functional status difference is perpetuated to the 12-month mark and to in-person interviews, although the exact program effect is difficult to evaluate from visual inspection alone. Previously, the method of analysis was limited by retrospective accounts in order to determine participant baselines. As such, the baseline difference between groups was not obvious. If an investigator wished to assess the baseline characteristics of the patient population using the 0-month data, she might find the task non-trivial, as the questions assessed at 0-months differed substantially from the set assessed at the final period. Any difference in baseline mental health could potentially impact final study findings in a large way. Mean regression is a well described phenomenon in psychiatric data, and clinical trials for psychoactive agents commonly exhibit high placebo response rates [58, 81].

To examine potentially heterogeneous treatment effects, Figure 3.2 shows fixed effect regression coefficients for the treatment by time effect stratified by baseline score quantile with period dummies. The pooled sample is shown in the dotted line above and below. The solid line in the panel above corresponds to the treatment effect by period for those with below median baseline functional status scores. The solid line in the panel below corresponds to the treatment effect by period for those with above median baseline functional status scores. Here we can see that in the pooled sample, there is a large effect of program receipt at the 6-month mark, but this effect decreases over time. The 6-month effect is being driven by those with baseline functional status scores below the median. For these below median baseline individuals, they experienced a large immediate treatment effect, but by the time of the in-person interview, the treatment effect is not distinguishable from zero. By contrast, the individuals who began the study at above-median functional status did not see a significant treatment effect. The ability to subsegment the population by baseline, as well as to conduct a cross-time analysis of the treatment effect, would have been previously limited due to item-incompatibility across period.

Table 3.3 records regression results using functional status scores ($\hat{\theta}_{i1t}$) as the outcome of interest. In columns (1) and (2), I present regression results, unadjusted for participant baseline ($\hat{\theta}_{i10}$). In columns (3) and (4), I present the unadjusted regression results restricted to the participants with non-missing baseline scores. In columns (5) and (6) I report the baseline-adjusted treatment results. In the initial four columns, we find that the estimated treatment effect across time are significantly attenuated in comparison to the magnitude found in columns (5) and (6). Specifically, the mixed-effects treatment effect in columns (1) and (3) were comparable, 0.002 (se=0.004) and 0.006 (se=0.005) respectively, in comparison to the mixed-effects treatment effect after adjusting for baseline found in column (5), 0.019 (se=0.004). Similarly, the fixed-effects treatment effect in columns (2) and (4) were comparable, 0.005 (se=0.005) and 0.005 (se=0.005) respectively, in comparison to the fixed-effects treatment effect after adjusting for baseline found in column (6), 0.017 (se=0.004). The baseline-time interaction recapitulates the significance of mean regression, it is -0.145 (se=0.004) and -0.140 (se=0.004) in column (5) and column (6) respectively.

3.4.3 Costs to Dichotomization

From Fig. 3.2, we know to expect some degree of treatment effect heterogeneity, but we would like to know what our interpretation would have been if we had used the thresholding metric alone. Fig. 3.3 displays the coefficient for treatment by time for both the model excluding baseline interaction (above) and including baseline interaction (below). Regression coefficient for treatment by time is plotted by variation of the threshold level. That is, the outcome here is whether the treatment reduced the number of people whose scores could be found below a given threshold. The adjustment for baseline score makes a large difference to the interpretation. Above, when no baseline adjustment is made, we would likely conclude that the treatment was largely ineffective. We would have concluded that the treatment was successful at reducing depression levels if we had pre-specified a score threshold at -2.1 or on [-1.8, -1.3], we could have measured a significant treatment effect, however, we would have been unlikely to specify these regions as our thresholds of interest. The vertical dotted line here indicates the score threshold of -0.978, or the threshold corresponding to the PHQ-9 score. The solid vertical lines correspond to -1.031 and -0.673, the questions relating to the endorsement of sadness and the endorsement of lack of interest respectively. As these instruments are designed to measure at the diagnostic screening threshold for depression, they have been interpreted as the clinically-relevant diagnostic threshold. However, all of these measure lie to the right of the prespecified area for which an effect is measurable. Had we specified an outcome at this threshold, we would have erroneously concluded the treatment had no effect.

By contrast, the lower panel in Fig. 3.3 shows the result of the baseline-adjusted metric. Again, we see that the thresholding outcome is somewhat sensitive to the particular threshold selected, however, now, in the majority of cases, we see that the treatment is significant. In general, a model without baseline adjustment may be misspecified because it cannot allow for the baseline-dependent mean-regression over time.

3.5 Discussion

Researchers in public policy are often interested in studying the potential mental health impacts of a variety of interventions, however, many of the common practices for studying mental health are unnecessarily restrictive. In a setting of a randomized experiment, in which we know that treatment is exogeneously assigned, I show that a variety of items can be used to assess mental-health related functional status in order to obtain a more complete picture of relative mental health among participants. Further, I show that having these orderings can change the way we design our program evaluation analysis, which can meaningfully change the results of the study.

The bifactor method supplies a matrix of estimated thresholds and factor loadings which provide item-level detail with regard to the behavior of each item. As I show in the case of depression diagnosis versus depression symptom endorsement, this item-level behavior can provide important interpretive clues about the context of the measurement, which would not be otherwise available when using a borrowed clinical questionnaire. What an item measures in actuality and what we might believe it measure a priori can differ significantly. For example, I find in this study that self-rated general health, pain severity, and SF-8 physical component score are actually quite closely correlated to self-rated mental health both in terms of their item thresholds and factor loadings. Furthermore, the threshold of the item, previous diagnosis of depression, is in fact quite high, meaning that it may be an inadequate control for de facto depression. Such a relationship has implications for a literature which has sought to explain the long-observed correlation between self-rated general health and mortality [30].

In terms of a more practical benefit, short psychological health items are often included on population surveys such as the National Health Interview Survey, the Medical Expenditure Panel Survey and the National Longitudinal Survey of Youth. However, between both time periods on the same survey, or across surveys, the variation in the particular items used can limit analysis. By recognizing that many of the items included on such survey are related and designed to measurement identical latent constructs, a bifactor approach can improve cross-comparability between datasets or time periods within a dataset. As such, using this approach could significantly broaden the types of questions that can be asked regarding mental health using observational data.

By far, however, the greatest benefit to the use of latent response modeling for mental health evaluations is an improvement in power. Current estimates of program effects on mental health are limited by a large degree of measurement error (attenuation bias) and by reducing mental health into a binary outcome (cost of dichotomization). The result is that several large-scale public programs that have been well powered to examine other outcomes, have nevertheless reported equivocal findings on mental health [71, 57, 79]. In fact, I show that, even given the strong initial results reported in [10], a more conventional empirical specification has the potential to eliminate the significance of the findings.

In spite of this long track of null results, this proposal to address to the attenuation problem is reasonably straightforward.

1. Choose a measure that corresponds to an outcome that is meaningful, rather than a diagnostic proxy that is merely convenient.

2. Compute the full range of scores for individuals, rather than a binary outcome. Item-response theory can help. In the case of multidimensionality, a bifactor structure can also be useful.

3. Mental health scores commonly exhibit mean-regression. Control for baseline-time interactions in any longitudinal study on mental health.

Table 3.1: Descriptive Statistics

Variable	Mean	Std. Dev.	Obs
Treatment	.421	.494	105,264
Score Mean	.000	.901	68,781
Score Variance	.422	.117	68,781
Self-Rated Health	.475	.637	39,490
SF-4 Pain Severity	2.37	.949	12,229
SF-8 Physical Component Score	3.49	1.71	12,229
SF-8 Mental Component Score	3.49	1.71	12,229
Quantile PHQ-9 Score	3.60	1.77	12,229
Self-Rated Happiness	1.94	.669	12,229
Quantile Bad Days Physical	1.72	1.25	34,843
Quantile Bad Days Mental	1.28	.907	26,131
Diagnosis of Depression	.561	.496	39,443
Prescriptions for Depression	.756	.430	38,896
Any Hospitalization	.920	.272	68,286
Any Doctor's Office Visit	.398	.315	68,261
Any Emergency Room Visit	.711	.453	68,241

Table 3.2: Estimated Thresholds and Factor Loadings

Item Description	(1) Item Threshold	(2) Common Dim Fac Loading	(3) Subdim Load 1	(4) Subdim Load 2	(5) Subdim Load 3	(6) Subdim Load 4
Self-Rated General Health	-1.155	0.542	0.399			
Self-Rated Pain Severity	-0.961	0.56	0.641			
SF-8 Physical Component Score	-0.972	0.653	0.569			
# Bad Days Due to Physical Health	-0.965	0.511	0.849			
SF-8 Mental Component Score	-1.051	0.817		0.566		
# Bad Days Due to Mental Health	-1.1	0.81		0.354		
Endorse Sadness	-0.673	0.557		0.421		
Endorse Lack of Interest	-1.031	0.844		0.328		
PHQ-9 Score	-0.978	0.716		0.458		
Self-Rated Happiness	-0.514	0.857		0.219		
Diagnosis of Depression	-0.149	0.675			0.64	
Prescriptions for Depression	-0.706	0.565			0.767	
Any Hospitalizations	-1.403	0.296				0.93
Any Doctor's Visits	0.259	0.265				0.336
Any Emergency Room Visits	-0.555	0.332				0.632

Common Dim Threshold column reports the estimated item threshold on the common dimension. Common Dim Fac Loading reports the mean estimated factor loading onto the common dimension. Subdim Load 1-4 reports the mean estimated factor loading onto the first to fourth sub-dimensions respectively.

Table 3.3: Mental Health-Related Functional Status Across Time in Treatment and Control from the Oregon Health Insurance Experiment.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.067**		0.073**		-.013**	
	(0.013)		(.013)		(.004)	
Time	-0.019**	-0.027**	-0.029**	-.028**	-.033**	-.030**
	(0.003)	(0.003)	(.005)	(.003)	(.003)	(.003)
Treatment x Time	0.002	0.005	.006	0.005	.019**	.017**
	(0.004)	(0.005)	(.005)	(0.005)	(.004)	(.004)
Baseline Score					1.12**	
					(.004)	
Baseline x Time					-.145**	-.140**
					(.004)	(0.004)
Baseline x Treat					.006	
					(.005)	
Baseline x Treat x Time					-.002	-.001
					(.005)	(.005)
Specification	ME	FE	ME	FE	ME	FE
N	68781	68781	54379	54379	54379	54379

Author's calculations using public data from the Oregon Health Insurance Experiment. Outcome variable is mental health related functional status score. ME indicates a mixed effects model including person-time random effects. FE denotes a person-level fixed effects model. All regressions clustered at the household level. Standard errors shown in parentheses below coefficient. ** Indicates Wald statistic is significant at the 95% level.

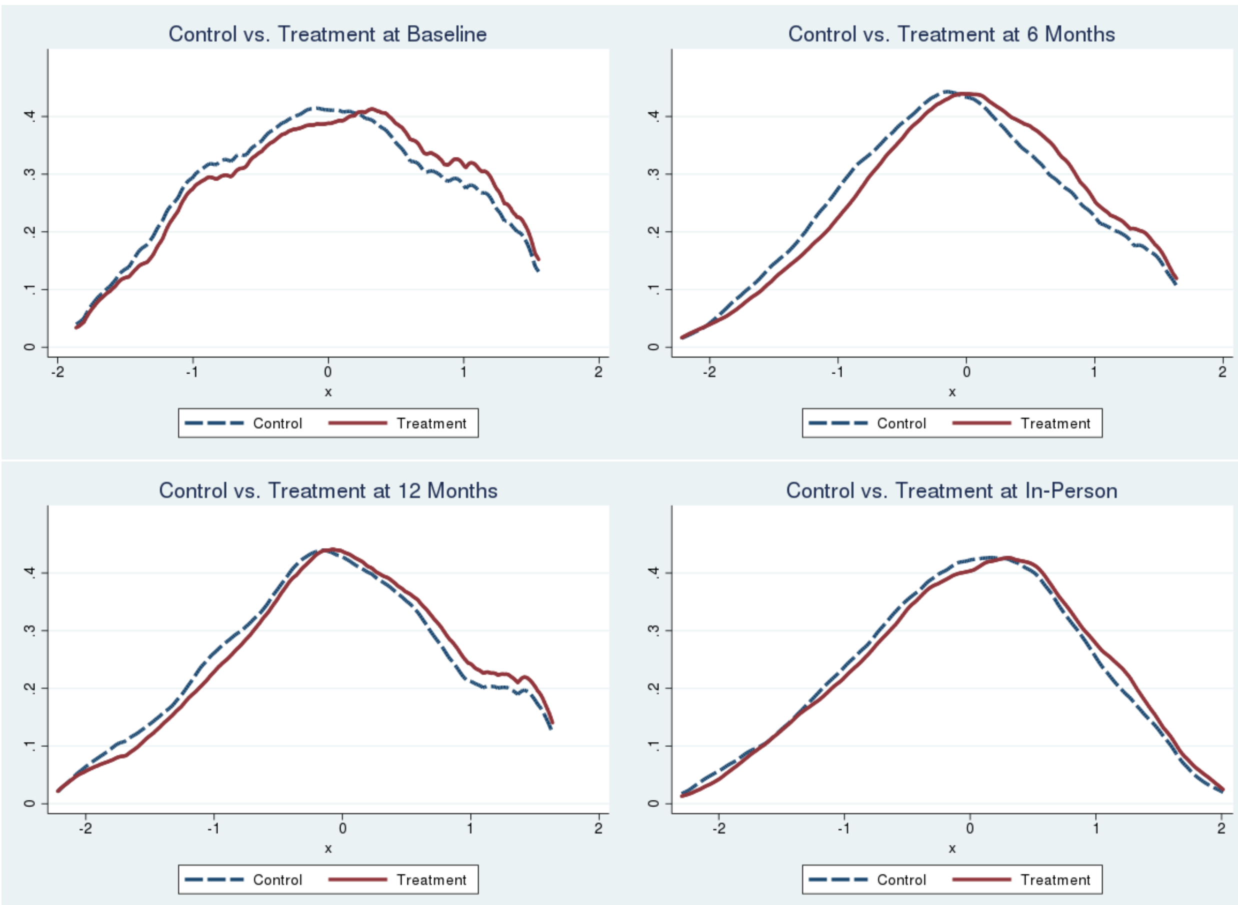
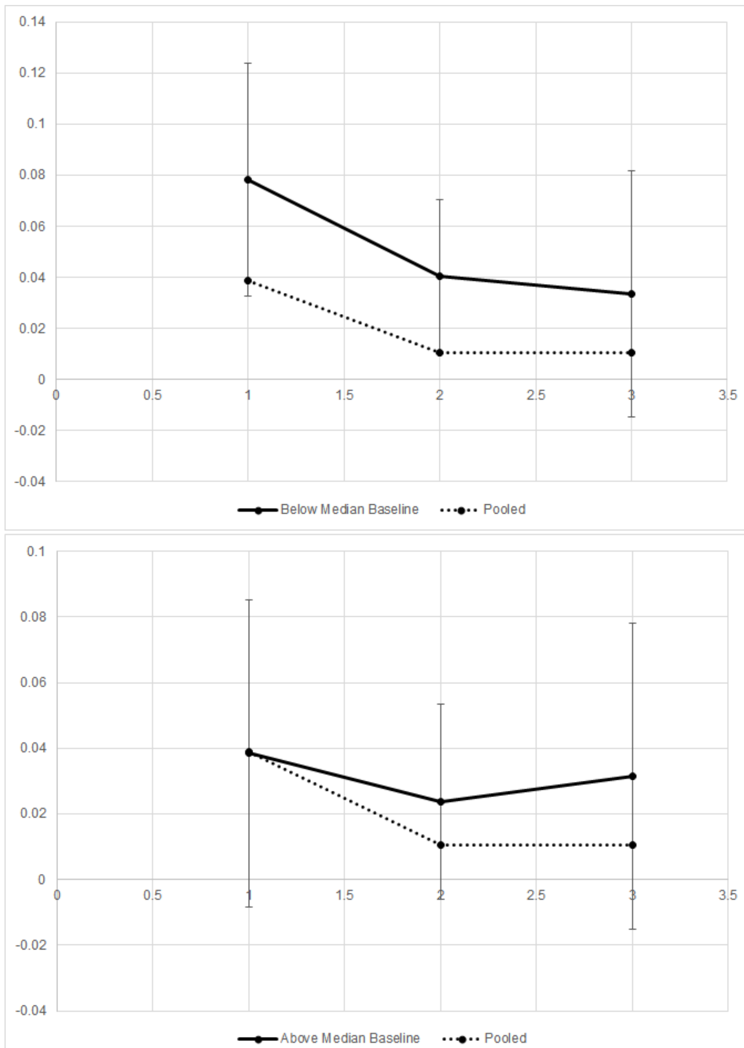


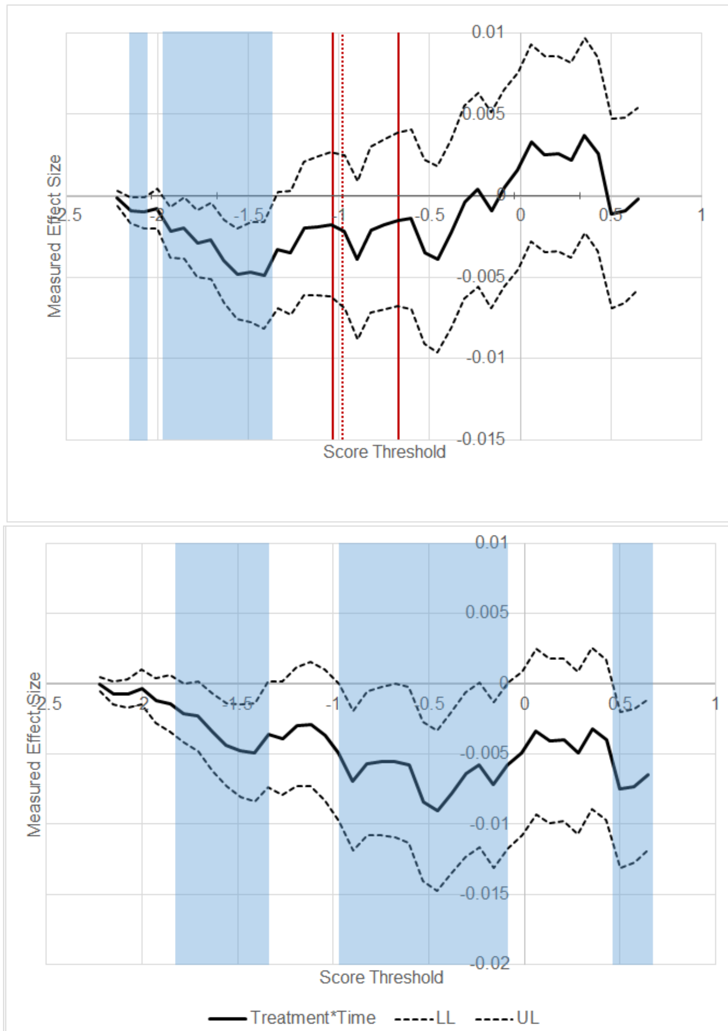
Figure 3.1: Kernel Density Plots by Time Point.

Figure 3.2: Stratified Treatment Effects by Baseline Quintile and Period Dummies.



The population was stratified into those with above median and those below median functional status at the baseline period. Score was then modeled as a function of treatment status, time period dummies and the interaction. Above in the solid line, the coefficients for treatment and time dummy interaction are plotted for those with below median baseline functional status. Below in the solid line, coefficients for treatment and time dummy interaction are plotted for those with above median baseline functional status. In both panels, the dotted line represents the coefficients for treatment and time dummy interaction pooling all baseline score groups.

Figure 3.3: Estimated Treatment Effect by Score Threshold Cutoff.



Regression results for fixed effects model. Above: Main treatment effects by score cutoff, model with no baseline-time interaction. Below: Main treatment effects for fully-interacted model with baseline inclusion. Blue boxes indicated regressions for which Wald estimator for main treatment effects were estimated to be significant at the 95% confidence level. The vertical dotted line indicates the score threshold for the PHQ-9 score. The solid vertical lines indicate the score thresholds for the items corresponding to the endorsement of sadness and the endorsement of lack of interest.

REFERENCES

[1] Social Security Act: Payment for Physician Services.

[2] Vidhya Alakeson, Nalini Pande, and Michael Ludwig. A Plan To Reduce Emergency Room Boarding Of Psychiatric Patients. *Health Affairs*, 29(9):1637–1642, September 2010.

[3] {American Medical Association}. *Medicare RBRVS : the physicians' guide.* American Medical Association, Chicago, IL, 1992.

[4] {American Medical Association}. *RVS Update Process 2012.* American Medical Association., 2012.

[5] {American Psychiatric Association}. Practice guideline for the treatment of patients with substance use disorders: alcohol, cocaine, opioids. American Psychiatric Association. *The American Journal of Psychiatry*, 152(11 Suppl):1–59, November 1995.

[6] {American Psychiatric Association}. Practice Guideline for the Treatment of Patients With Substance Use Disorders. Technical report, August 2006.

[7] {American Psychiatric Association}. *Diagnostic and statistical manual of mental disorders: DSM-5.* American Psychiatric Association, Arlington, VA, 5th ed. edition, 2013.

[8] W. A. Anthony and M. A. Jansen. Predicting the vocational capacity of the chronically mentally ill. Research and policy implications. *The American Psychologist*, 39(5):537–544, May 1984.

[9] Paul S. Appelbaum. Boarding Psychiatric Patients in Emergency Rooms: One Court Says No More. *Psychiatric Services*, 66(7):668–670, July 2015.

[10] Katherine Baicker, Sarah L. Taubman, Heidi L. Allen, Mira Bernstein, Jonathan H. Gruber, Joseph P. Newhouse, Eric C. Schneider, Bill J. Wright, Alan M. Zaslavsky, and Amy N. Finkelstein. The Oregon Experiment Effects of Medicaid on Clinical Outcomes. *New England Journal of Medicine*, 368(18):1713–1722, May 2013.

[11] J. C. Ballenger, J. R. Davidson, Y. Lecrubier, D. J. Nutt, L. J. Kirmayer, J. P. Lpine, K. M. Lin, O. Tajima, and Y. Ono. Consensus statement on transcultural issues in depression and anxiety from the International Consensus Group on Depression and Anxiety. *The Journal of Clinical Psychiatry*, 62 Suppl 13:47–55, 2001.

[12] Lisa Feldman Barrett. The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, page nsw154, October 2016.

[13] Lisa Feldman Barrett, Michael Lewis, and Jeannette M. Haviland-Jones. *Handbook of emotions.* Guilford Press, New York, fourth edition. edition, 2016.

[14] Colleen L Barry and Emma Elizabeth McGinty. Stigma and Public Support for Insurance Parity and Government Spending on Mental Health: Findings from a 2013 National Opinion Survey. *Psychiatric services (Washington, D.C.)*, 65(10):1265–1268, October 2014.

[15] Gary S. Becker and Kevin M. Murphy. A theory of rational addiction. *Journal of political Economy*, 96(4):675–700, 1988.

[16] Robert A. Berenson and John D. Goodson. Finding Value in Unexpected Places Fixing the Medicare Physician Fee Schedule. *New England Journal of Medicine*, 374(14):1306–1309, April 2016.

[17] Christopher Blattman, Eric P. Green, Julian Jamison, M. Christian Lehmann, and Jeannie Annan. The Returns to Microenterprise Support among the Ultrapoor: A Field Experiment in Postwar Uganda. *American Economic Journal: Applied Economics*, 8(2):35–64, April 2016.

[18] R. Darrell Bock and Murray Aitkin. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4):443–459, December 1981.

[19] Thomas Bodenheimer, Robert A. Berenson, and Paul Rudolf. The primary care-specialty income gap: why it matters. *Annals of Internal Medicine*, 146(4):301–306, February 2007.

[20] Thomas Bodenheimer and Hoangmai H. Pham. Primary care: current problems and proposed solutions. *Health Affairs (Project Hope)*, 29(5):799–805, May 2010.

[21] John Bound, Charles Brown, and Nancy Mathiowetz. Chapter 59 - Measurement Error in Survey Data. volume 5 of *Handbook of Econometrics*, pages 3705 – 3843. Elsevier, 2001.

[22] Diane R. Brown, Faroz Ahmed, Lawrence E. Gary, and Norweeta G. Milburn. Major depression in a community sample of African Americans. *American Journal of Psychiatry*, 152(3):373–378, March 1995.

[23] {Bureau of Justice Statistics}. Drug Use, Dependence, and Abuse Among State Prisoners and Jail Inmates, 2007-2009. Special Report NCJ 250546, U.S. Department of Justice, June 2017.

[24] {Bureau of Justice Statistics}. Indicators of Mental Health Problems Reported by Prisoners and Jail Inmates, 2011-12. Special Report NCJ 250612, U.S. Department of Justice, June 2017.

[25] Jacob Cohen. The cost of dichotomization. *Applied psychological measurement*, 7(3):249–253, 1983.

[26] Michael T. Compton, Jason Craw, and Bruce E. Rudisch. Determinants of inpatient psychiatric length of stay in an urban county hospital. *The Psychiatric Quarterly*, 77(2):173–188, 2006.

[27] Andrew L. Comrey and Howard B. Lee. *A First Course in Factor Analysis*. Psychology Press, 2 edition edition, January 2016.

[28] G. E. Davis, W. E. Lowell, and G. L. Davis. Determining the number of state psychiatric hospital beds by measuring quality of care with artificial neural networks. *American Journal of Medical Quality: The Official Journal of the American College of Medical Quality*, 13(1):13–24, 1998.

[29] Jonathan De Quidt and Johannes Haushofer. Depression for economists. Technical report, National Bureau of Economic Research, 2016.

[30] Karen B. DeSalvo, Nicole Bloser, Kristi Reynolds, Jiang He, and Paul Muntner. Mortality Prediction with a Single General SelfRated Health Question. *Journal of General Internal Medicine*, 21(3):267–275, March 2006.

[31] David Dranove, Craig Garthwaite, and Christopher Ody. How do nonprofits respond to negative wealth shocks? The impact of the 2008 stock market collapse on hospitals. *The Rand Journal of Economics*, 48(2):485–525, 2017.

[32] David Dranove and Mark A. Satterthwaite. The industrial organization of health care markets. *Handbook of health economics*, 1:1093–1139, 2000.

[33] Richard J. Durbin. Medicaid Coverage for Addiction Recovery Expansion Act, May 2017.

[34] Amy Finkelstein, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, and Katherine Baicker. The Oregon Health Insurance Experiment: Evidence from the First Year*. *The Quarterly Journal of Economics*, 127(3):1057–1106, August 2012.

[35] Matt Ford. America's Largest Mental Hospital Is a Jail. *The Atlantic*, June 2015.

[36] Michel Foucault. *Madness and civilization; a history of insanity in the age of reason.* Pantheon Books, New York, 1965. OCLC: 676127.

[37] Robert D. Gibbons and Donald R. Hedeker. Full-information item bi-factor analysis. *Psychometrika*, 57(3):423–436, September 1992.

[38] Robert D. Gibbons, David Kupfer, Ellen Frank, Tara Moore, David G. Beiser, and Edwin D. Boudreaux. Development of a Computerized Adaptive Test Suicide Scale-The CAT-SS. *The Journal of Clinical Psychiatry*, May 2017.

[39] Robert D. Gibbons, David J. Weiss, Paul A. Pilkonis, Ellen Frank, Tara Moore, Jong Bae Kim, and David J. Kupfer. Development of a Computerized Adaptive Test for Depression. *Archives of General Psychiatry*, 69(11):1104–1112, November 2012.

[40] Jenny Gold. A Dearth Of Psychiatric Hospital Beds For California Patients In Crisis, April 2016.

[41] Liza H. Gold. DSM-5 and the Assessment of Functioning: The World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). *Journal of the American Academy of Psychiatry and the Law Online*, 42(2):173–181, June 2014.

[42] Jacqueline M Golding, Marvin Karno, and Carolyn M. Rutter. Symptoms of major depression among Mexican-Americans and non-Hispanic whites. *American Journal of Psychiatry*, 147(7):861–866, July 1990.

[43] Jan Ellen Goldstein. *Console and classify: the French psychiatric profession in the nineteenth century*. 1987.

[44] Gerald N. Grob. *From asylum to community : mental health policy in modern America*. Princeton University Press, Princeton, N.J., 1991.

[45] Joseph F. Hair, Ronald L. Tatham, Rolph E. Anderson, and William Black. *Multivariate Data Analysis*. Prentice Hall, Upper Saddle River, N.J, 5th edition edition, March 1998.

[46] Bernard E. Harcourt. An Institutionalization Effect: The Impact of Mental Hospitalization and Imprisonment on Homicide in the United States, 19342001. *The Journal of Legal Studies*, 40(1):39–83, January 2011.

[47] Johannes Haushofer and Jeremy Shapiro. The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya. *The Quarterly Journal of Economics*, 131(4):1973–2042, November 2016.

[48] {HCUP Fast Stats}. Healthcare Cost and Utilization Project (HCUP)., November 2017.

[49] Karl J. Holzinger and Frances Swineford. The Bi-factor method. *Psychometrika*, 2(1):41–54, March 1937.

[50] Jill R. Horwitz. Making Profits And Providing Care: Comparing Nonprofit, For-Profit, And Government Hospitals. *Health Affairs*, 24(3):790–801, May 2005.

[51] Christopher G. Hudson. Socioeconomic Status and mental Illness: Tests of the Social Causation and Selection Hypotheses. *American Journal of Orthopsychiatry*, 75(1):3–18, 2005.

[52] Robert I. Jennrich and Peter M. Bentler. Exploratory Bi-factor Analysis. *Psychometrika*, 76(4):537–549, October 2011.

[53] Shengyu Jiang, Chun Wang, and David J Weiss. Sample Size Requirements for Estimation of Item Parameters in the Multidimensional Graded Response Model. *Frontiers in Psychology*, 7, February 2016.

[54] Brian Junker, Lynne Steuerle Schofield, and Lowell J. Taylor. The use of cognitive ability measures as explanatory variables in regression analysis. *IZA Journal of Labor Economics*, 1:4, October 2012.

[55] Daniel P. Kessler and Mark B. McClellan. Is hospital competition socially wasteful? *The Quarterly Journal of Economics*, 115(2):577–615, 2000.

[56] Dae-Young Kim. Psychiatric Deinstitutionalization and Prison Population Growth: A Critical Literature Review and Its Implications. *Criminal Justice Policy Review*, 27(1):3–21, February 2016.

[57] J.R. Kling, J.B. Liebman, and L.F. Katz. Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119, 2007.

[58] Helena Chmura Kraemer, G. Terence Wilson, Christopher G. Fairburn, and W. Stewart Agras. Mediators and Moderators of Treatment Effects in Randomized Clinical Trials. *Archives of General Psychiatry*, 59(10):877–883, October 2002.

[59] Despo Kritsotaki, Vicky Long, and Matthew Smith, editors. *Deinstitutionalisation and After: Post-War Psychiatry in the Western World*. Mental Health in Historical Perspective. Palgrave Macmillan, 2016.

[60] Kurt Kroenke, Robert L Spitzer, and Janet B W Williams. The PHQ-9. *Journal of General Internal Medicine*, 16(9):606–613, September 2001.

[61] Ilyana Kuziemko and Eric Werker. How Much Is a Seat on the Security Council Worth? Foreign Aid and Bribery at the United Nations. *Journal of Political Economy*, 114(5):905–930, 2006.

[62] Elizabeth M. La, Kristen Hassmiller Lich, Rebecca Wells, Alan R. Ellis, Marvin S. Swartz, Ruoqing Zhu, and Joseph P. Morrissey. Increasing Access to State Psychiatric Hospital Beds: Exploring Supply-Side Solutions. *Psychiatric Services (Washington, D.C.)*, 67(5):523–528, May 2016.

[63] H. R. Lamb and L. E. Weinberger. Persons with severe mental illness in jails and prisons: a review. *Psychiatric Services (Washington, D.C.)*, 49(4):483–492, April 1998.

[64] Miriam Laugesen. *Fixing Medical Prices: How Physicians are Paid*. 2016.

[65] Miriam J. Laugesen, Roy Wada, and Eric M. Chen. In Setting Doctors Medicare Fees, CMS Almost Always Accepts The Relative Value Update Panels Advice On Work Values. *Health Affairs*, 31(5):965–972, May 2012.

[66] Douglas L. Leslie and Robert A. Rosenheck. Comparing Quality of Mental Health Care for Public-Sector and Privately Insured Populations. *Psychiatric Services*, 51(5):650–655, May 2000.

[67] Keh-Ming Lin and Freda Cheung. Mental Health Issues for Asian Americans. *Psychiatric Services*, 50(6):774–780, June 1999.

[68] T. M. Luhrmann. *Of two minds: the growing disorder in American psychiatry*. Knopf, New York, 1st ed. edition, 2000.

[69] Shefali Luthra. How Gaps In Mental Health Care Play Out In Emergency Rooms, October 2016.

[70] Robert C. MacCallum and Ledyard R. Tucker. Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin*, 109(3):502–511, May 1991.

[71] Willard G. Manning, Kenneth B. Wells, Joan L. Buchanan, Emmett B. Keeler, R. Burciaga Valdez, and Joseph P. Newhouse. Effects of Mental Health Insurance. Product Page R-3815-NIMH/HCFA, RAND Corporation, Santa Monica, CA, October 1989.

[72] {Maryland Department of Health and Mental Hygiene}. Maryland HealthChoice Program Section 1115 Waiver Amendment., June 2016.

[73] {Mathematica Policy Research}. Medicaid Emergency Psychiatric Services Demonstration Evaluation: Final Report. Final Report 40123.444, Centers for Medicare and Medicaid Services, Baltimore, MD, August 2016.

[74] Stephanie Maxwell, Stephen Zuckerman, and Robert A. Berenson. Use of Physicians' Services under Medicare's Resource-Based Payments. *New England Journal of Medicine*, 356(18):1853–1861, May 2007.

[75] Nancy McCall, Jerry Cromwell, and Peter Braun. Validation of Physician Survey Estimates of Surgical Time Using Operating Room Logs. *Medical Care Research and Review*, 63(6):764–777, December 2006.

[76] Jonathan Metzl. *The protest psychosis: how schizophrenia became a black disease.* Beacon, Boston, Mass, 2011. OCLC: 809699885.

[77] Adrian P. Mundt, Winnie S. Chow, Margarita Arduino, Hugo Barrionuevo, Rosemarie Fritsch, Nestor Girala, Alberto Minoletti, Flvia Mitkiewicz, Guillermo Rivera, Mara Tavares, and Stefan Priebe. Psychiatric hospital beds and prison populations in South America since 1990: does the Penrose hypothesis apply? *JAMA psychiatry*, 72(2):112–118, February 2015.

[78] Carles Muntaner, William W. Eaton, Richard Miech, and Patricia OCampo. Socioeconomic Position and Major Mental Disorders. *Epidemiologic Reviews*, 26(1):53–62, July 2004.

[79] Edgar O. Olsen and Jeffrey E. Zabel. Chapter 14 - US Housing Policy. In J. Vernon Henderson and William C. Strange Gilles Duranton, editor, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 887–986. Elsevier, 2015.

[80] Johan Ormel, Michael VonKorff, T. Bedirhan Ustun, Stefano Pini, Ailsa Korten, and Tineke Oldehinkel. Common Mental Disorders and Disability Across Cultures: Results From the WHO Collaborative Study on Psychological Problems in General Health Care. *JAMA*, 272(22):1741–1748, December 1994.

[81] George I. Papakostas and Maurizio Fava. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *European Neuropsychopharmacology*, 19(1):34–40, January 2009.

[82] L. S. Penrose. Mental Disease and Crime: Outline of a Comparative Study of European Statistics. *British Journal of Medical Psychology*, 18(1):1–15, March 1939.

[83] Kate E. Pickett and Richard G. Wilkinson. Income inequality and health: A causal review. *Social Science & Medicine*, 128:316–326, March 2015.

[84] Harold Pollack. What happened to U.S. mental health care after deinstitutionalization? *Washington Post*, June 2013.

[85] Steven Raphael and Michael A. Stoll. Assessing the Contribution of the Deinstitutionalization of the Mentally Ill to Growth in the U.S. Incarceration Rate. *The Journal of Legal Studies*, 42(1):187–222, January 2013.

[86] Steven P. Reise. Invited Paper: The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, 47(5):667–696, September 2012.

[87] Steven P. Reise, Dale S. Kim, Maxwell Mansolf, and Keith F. Widaman. Is the Bifactor Model a Better Model or Is It Just Better at Modeling Implausible Responses? Application of Iteratively Reweighted Least Squares to the Rosenberg Self-Esteem Scale. *Multivariate Behavioral Research*, 51(6):818–838, December 2016.

[88] Steven P. Reise, Joseph Ventura, Richard S. E. Keefe, Lyle E. Baade, James M. Gold, Michael F. Green, Robert S. Kern, Raquelle Mesholam-Gately, Keith H. Nuechterlein, Larry J. Seidman, and Robert Bilder. Bifactor and Item Response Theory Analyses of Interviewer Report Scales of Cognitive Impairment in Schizophrenia. *Psychological assessment*, 23(1):245–261, March 2011.

[89] James C. Robinson. Theory and Practice in the Design of Physician Payment Incentives. *The Milbank Quarterly*, 79(2):149–177, June 2001.

[90] S A Schroeder and L G Sandy. Specialty Distribution of U.S. Physicians The Invisible Driver of Health Care Costs. *New England Journal of Medicine*, 328(13):961–962, 1993.

[91] Edward Shorter. *A History of Psychiatry: From the Era of the Asylum to the Age of Prozac*. Wiley, New York, 1 edition edition, March 1998.

[92] Edward Shorter. A history of psychiatry from the era of the asylum to the age of Prozac, 2011. OCLC: 940169125.

[93] Martha Shumway, Jennifer Alvidrez, Mark Leary, Deborah Sherwood, Eric Woodard, Emily K. Lee, Heather Hall, Ralph A. Catalano, and James W. Dilley. Impact of capacity reductions in acute public-sector inpatient psychiatric services. *Psychiatric Services (Washington, D.C.)*, 63(2):135–141, February 2012.

[94] Frank A. Sloan. Not-for-profit ownership and hospital behavior. *Handbook of health economics*, 1:1141–1174, 2000.

[95] Estina E. Thompson, Harold W. Neighbors, Cheryl Munday, and Steve Trierweiler. Length of Stay, Referral to Aftercare, and Rehospitalization Among Psychiatric Inpatients. *Psychiatric Services*, 54(9):1271–1276, September 2003.

[96] E F Torrey. Jails and prisons–America's new mental hospitals. *American Journal of Public Health*, 85(12):1611–1613, December 1995.

[97] E. Fuller Torrey. A Dearth of Psychiatric Beds. *Psychiatric Times*, February 2016.

[98] Alex D. Tulloch, Paul Fearon, and Anthony S. David. Length of stay of general psychiatric inpatients in the United States: systematic review. *Administration and Policy in Mental Health*, 38(3):155–168, May 2011.

[99] Nicholas T. Van Dam and Mitch Earleywine. Validation of the Center for Epidemiologic Studies Depression ScaleRevised (CESD-R): Pragmatic depression assessment in the general population. *Psychiatry Research*, 186(1):128–132, March 2011.

[100] David Watson, Michael W. O'Hara, Leonard J. Simms, Roman Kotov, Michael Chmielewski, Elizabeth A. McDade-Montez, Wakiza Gamez, and Scott Stuart. Development and validation of the Inventory of Depression and Anxiety Symptoms (IDAS). *Psychological Assessment*, 19(3):253–268, September 2007.

[101] M. M. Weissman, R. C. Bland, G. J. Canino, C. Faravelli, S. Greenwald, H. G. Hwu, P. R. Joyce, E. G. Karam, C. K. Lee, J. Lellouch, J. P. Lpine, S. C. Newman, M. Rubio-Stipec, J. E. Wells, P. J. Wickramaratne, H. Wittchen, and E. K. Yeh. Cross-national epidemiology of major depression and bipolar disorder. *JAMA*, 276(4):293–299, July 1996.

[102] Robert D. Wilton. More responsibility, less control: psychiatric survivors and Welfare State restructuring. *Disability & Society*, 19(4):371–385, June 2004.

[103] Jangho Yoon and Jeff Luck. Intersystem return on investment in public mental health: Positive externality of public mental health expenditure for the jail system in the U.S. *Social Science & Medicine (1982)*, 170:133–142, December 2016.

APPENDIX A

APPENDIX TO PSYCHIATRIC SPILLOVERS

A.1 Long Run Bed Supply

Economists have devoted some attention to medical providers, but generally have not discussed psychiatric care specifically [55, 32, 31]. Much of the economics literature specific to mental health has focused on individual-level decision-making rather than providers [15, 29]. As a result, the institutional context of psychiatric care may be under-appreciated by economists. By contrast, sociologists, anthropologists, and historians have written widely about psychiatric care, detailing some differences between psychiatric care provision and medical care, such as the historically large public role in care provision and the large oscillations in public attitudes that affect this public provision. However, authors in history, anthropology and sociology do not focus their discussions on the provider objectives, provider costs, or how these factors might affect how we interpret changes in long-run psychiatric bed supply [36, 43, 44, 91, 68]. To bridge the gap, I provide a brief discussion of long-run provider decisions here.

Regardless of the provider ownership, inpatient psychiatric providers face cost and revenue curves by ward with a standard shape. Specifically, providers face large fixed costs in the determination of the number of beds provided, which dominate variable costs [84]. On a per-ward basis, for each additional psychiatric ward added to a hospital, the provider must pay a fixed cost for the associated facilities, security, 24-hour staffing requirements, as well as psychiatric-care-specific capital such as recreation rooms with observation windows and furniture that does not promote self-injury. Thus while there are a small number of beds per ward, each additional psychiatric bed decreases the average cost per patient. As the number of beds per ward increases, however, the average cost per bed also begins to increase. For example, patients on large or crowded wards (wards with many beds) may be inadequately monitored, self-injure at higher rates, or engage in other behaviors that are

Table A.1: Psychiatric Beds by Hospital Type Over Time.

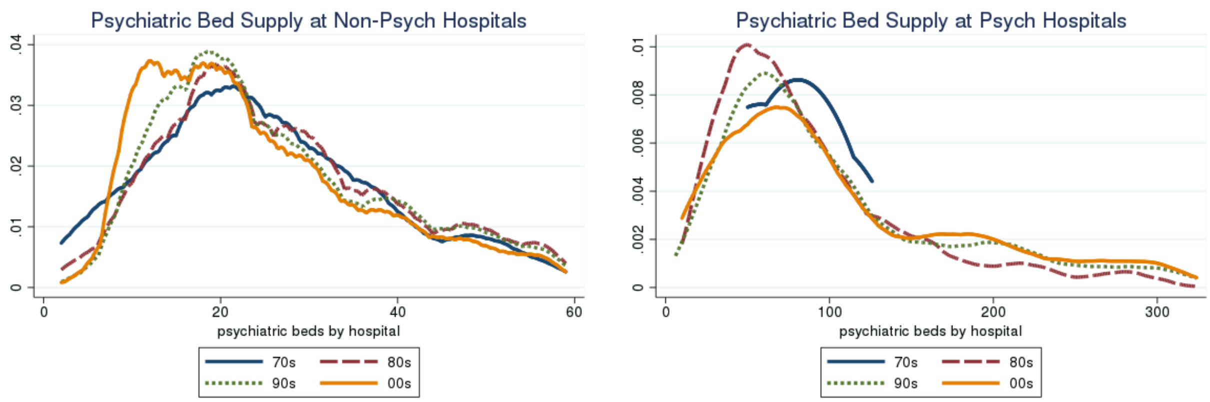
	General Hospitals					
	Years	Mean	SD	Median	Mode	N
Private	1970-1979	12.3	23.2	25.0	20.0	2974
	1980-1989	19.7	28.0	27.0	20.0	11082
	1990-1999	25.2	30.0	27.0	20.0	13417
	2000-2014	22.9	29.0	25.0	20.0	16314
Public	1970-1979	22.6	72.3	30.0	20.0	1580
	1980-1989	43.3	99.0	38.0	30.0	5118
	1990-1999	54.7	108.2	34.0	20.0	4461
	2000-2014	28.5	54.8	23.0	10.0	4723
	Psychiatric Hospitals					
	Years	Mean	SD	Median	Mode	N
Private	1970-1979	300.5	440.9	96.0	.	4
	1980-1989	80.8	54.4	70.0	60.0	1110
	1990-1999	79.8	47.4	69.0	60.0	1377
	2000-2014	84.6	53.8	75.0	100.0	1387
Public	1970-1979	.	.	.	.	0
	1980-1989	260.3	393.7	146.0	60.0	440
	1990-1999	272.9	246.4	210.0	28.0	1236
	2000-2014	211.8	211.5	175.0	50.0	1338

Author's calculations at a hospital-level from American Hospital Association Data 1970-2014. Hospitals were tagged as psychiatric hospitals if the number of reported psychiatric beds exceeded 95% of the total bed number. Hospitals reporting fewer than two psychiatric beds excluded from sample.

costly for providers. In contrast to the large fixed costs of inpatient care provision, however, revenue is often received on the margin per-patient-per-diem.

In general, we may expect efficient providers to supply beds stepwise in multiples of their efficient ward size. Providers would like to choose a number of beds such that each ward functions at minimum cost while each ward operates as close to full capacity as possible. A constraint on the provider's decision is the the shape of the cost curve, which dictates an efficient ward size. Given the efficient ward size, however, a provider may choose the total number of wards freely. Anecdotally, some providers will report an efficient ward size at close to 20 beds per ward. Consistent with these provider accounts, the data from the American Hospital Association (AHA) shows a modal ward size of 20 beds across multiple decades of data. I show these data in Fig. A.1 and Table A.1.

Figure A.1: Distributions of Psychiatric Bed Supply by Decade by Hospital Type.



Source: Author's calculations from the American Hospital Association Survey. Hospitals were tagged as psychiatric hospitals if the number of reported psychiatric beds exceeded 95% of the total bed number. The kernel density by decade is indicated by the legend. Of note, the decade "00s" corresponds to the years 2000-2014.

While both all psychiatric providers face similar cost curves in terms of the number of beds per ward, provider objectives may differ substantially by whether the hospital is public or private. For example, a public provider may care about the quantity of care delivered versus a private provider who may wish strictly to maximize profit. In the case of a public hospital, consider a provider with utility over the quantity of psychiatric care provided [94]. However, the hospital also values quantities of other care delivered, say from an additional medical ward. One possible formulation for a public provider's optimization problem is as follows:

$$\max_{Q_o, Q_\psi} V(Q_o, Q_\psi) \quad s.t. \quad R(Q_\psi) + T \geq C_{Q_o} + \bar{C}_\psi Q_\psi \tag{A.1}$$

where Q_ψ is the number of psychiatric wards, Q_o is the quantity of all other public goods, $R(.)$ is the revenue function for the psychiatric wards, $C(.)$ is the cost function for all other public goods, $\bar{C}_\psi$ is the fixed per-ward costs for each psychiatric ward provided at the efficient number of beds per ward, and T is the total tax revenue, which defines the budget constraint. Under the standard assumptions, $MV_o > 0, MV_\psi > 0$ and $\frac{\partial MV_o}{\partial Q_o} < 0, \frac{\partial MV_\psi}{\partial Q_\psi} < 0$, public general hospitals will choose Q_o, Q_ψ such that

$$\frac{MV_o}{MV_\psi} = \frac{MC_o}{\bar{C}_\psi - MR_\psi}. \tag{A.2}$$

Thus, for a public provider, the quantity of wards chosen depends on the relative preference for psychiatric care, as well as the relative cost net of revenues for psychiatric care.

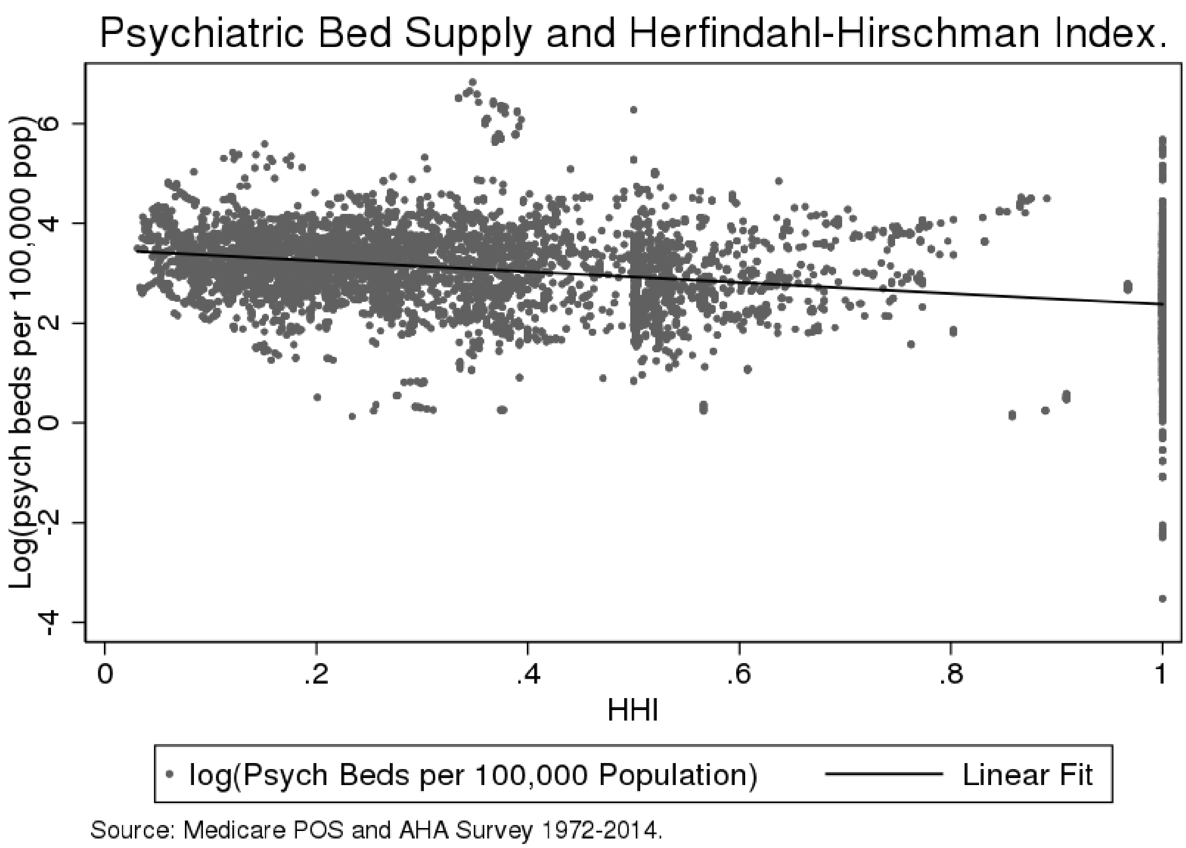
By contrast, a private hospitals may be strictly profit-maximizing. Under this condition, a private hospital will choose a non-zero number of wards only if the total revenue earned at optimality equals or exceeds the total costs at optimality, otherwise it will exit. Among the providers who choose a positive number of psychiatric wards, the total number of psychiatric beds supplied privately depends on the provider's local market power. In a perfectly competitive setting, the private provider will choose a number of wards such that the marginal revenue per ward is equal to the ward-level marginal cost, the sum of fixed and variables costs at the efficient ward size. However, given the inpatient psychiatric provider market has some barriers to entry, perfectly competitive private providers are likely less common that private providers with some market power. As such, a private provider with market power may undersupply the number of beds relative to the perfectly competitive equilibrium.[1]

Some suggestive evidence of total psychiatric undersupply secondary to market power can be found in the data. Figure A.2 shows market concentration for HRR-years present in the merged AHA and Medicare POS data. From this figure, we see first that the majority of HRR-years in sample were observed to have an HHI greater than 0.18 and, second, we see that psychiatric bed supply per population decreases with increasing market concentration.

Given the above framework for long-run psychiatric bed supply, I make several speculations regarding the way that changes in reimbursement and policy have influenced long-run psychiatric bed supply. First, provider cost is a meaningful way of interpreting the declines

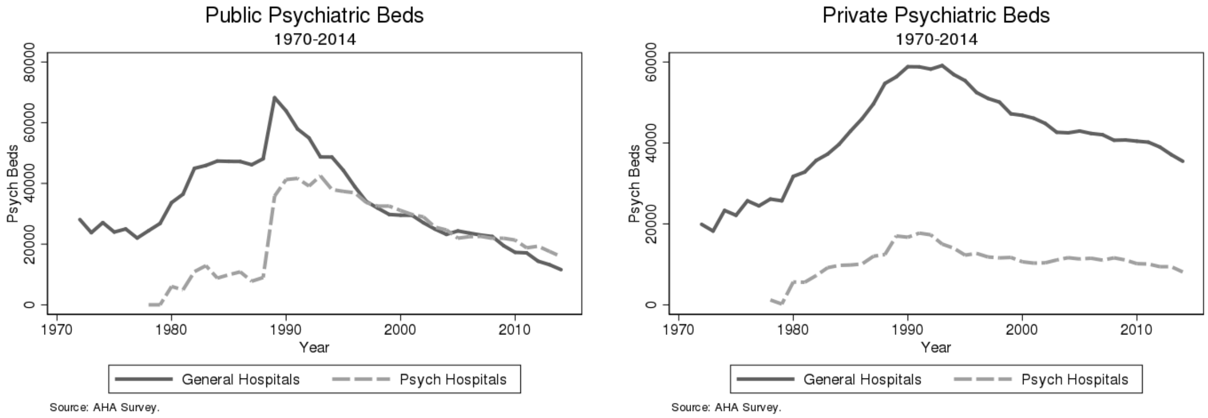
1. Although price equilibrium is reached in this market through bargaining with insurers, private providers with market power are price setters through their ability to negotiate higher payment rates with insurers due to plan-holder demand.

Figure A.2: Psychiatric Bed Supply and Market Concentration.



Source: Author's calculations from the American Hospital Association Survey 1972-2014 and Medicare Provider of Services Data 1991-2014. Market concentration computed at the hospital referral region level as the Herfindahl-Hirschman Index (HHI). Black line indicates linear fit to the scatter plot shown given by the equation y=3.47-1.09*HHI. R-squared 0.162.

Figure A.3: Psychiatric Bed Supply Across Time.



Source: Author's calculations from the American Hospital Association Survey 1970-2014. The line corresponds to the total number of psychiatric beds in sample over the total number of hospital beds in sample. Left: the ratio for psychiatric beds provided by public hospitals. Right: the ratio for psychiatric beds provided by private psychiatric hospitals.

in psychiatric bed supply in the period 1990-2014, and, second, individual providers act in accordance with local optimal conditions rather than en bloc, as has been suggested by previous national-level studies. First, with regard to the subject of cost, Figure A.3 shows the time series for 1980-2014, during which time neither public nor private providers wholly dominated inpatient psychiatric supply. From 1990-2014, the number of psychiatric beds supplied by all provider types fell. As I mentioned in the previous section, this decline in the number of beds supplied likely occurred secondary to changes in psychiatric reimbursement under managed care, which became widely disseminated during the 1990s. While it is possible that these declines could have been due to changes in public preferences, the concurrent declines in public and private supply support the speculation that these changes were a response to changes in costs net of revenue.[2]

2. Preference changes have historically accounted for some of the large increases and then declines in psychiatric bed numbers as described in the U.S. by [44], or in the international setting by [59]. From these historical studies, we have examples that attitudes toward mental health entitlements have run parallel to attitudes toward entitlement programs more generally. The enduring correlation in attitudes toward mental health care provision and entitlements is evidenced in more recent studies by [14]. From a historical perspective, the 1990s witnessed a popularity shift away from entitlement programs toward "welfare-to-work," as evidenced by the passage of the Personal Responsibility and Work Opportunity Act of 1996 and others. There has been some speculation that such policies have had large effects on the lives of patients

Costs structures also elucidate the relative prevalence of general hospital-provided psychiatric care. General hospitals have been cited as having higher ward-level costs in comparison to free-standing psychiatric hospitals [72]. If higher general hospital costs is true, one may expect general hospital psychiatric beds to be disproportionately affected by decreases to reimbursements. However, in comparison to the relative public supply, which may have responded to cost increases by preferentially reducing general hospital beds, many private general hospitals retained their psychiatric wards. There are two likely reasons for this trend. One possible explanation is private general hospitals retained their costly psychiatric wards because demand was insufficient in many areas to offset the hospital-level fixed costs incurred by opening a free-standing psychiatric facility. In the absence of such free-standing psychiatric facilities, a general hospital could provide beds at a higher cost and still make a profit. Another possible explanation is the Medicaid IMD exclusion. Care delivered at general hospitals is favored by the IMD Exclusion over care provided at free-standing psychiatric hospitals. Given the large volume of psychiatric inpatients covered by Medicaid, general private hospitals could be compensated for a much larger volume of care in comparison to free-standing hospitals. The difference in volume and, therefore revenues, too, could have offset higher costs of the general hospital setting.

Second, returning to my assertion that individual providers act in accordance with local optimal conditions rather than en bloc, a second trend to notice in Figure A.3 is the gradual change in the number of psychiatric beds across time. This gradual decline is, in part, the result of the composition of the market. Table A.1 describes the composition of psychiatric supply as the summation of multiple, small, local providers. In contrast to the rapid declines in the rates of institutionalization post-1965, the rate of institutionalization halved in just five years, the 1990s witnessed a very different picture for deinstitutionalization [46]. That is, while bed supply declined steadily from 1990 onwards, the decline was not driven by the

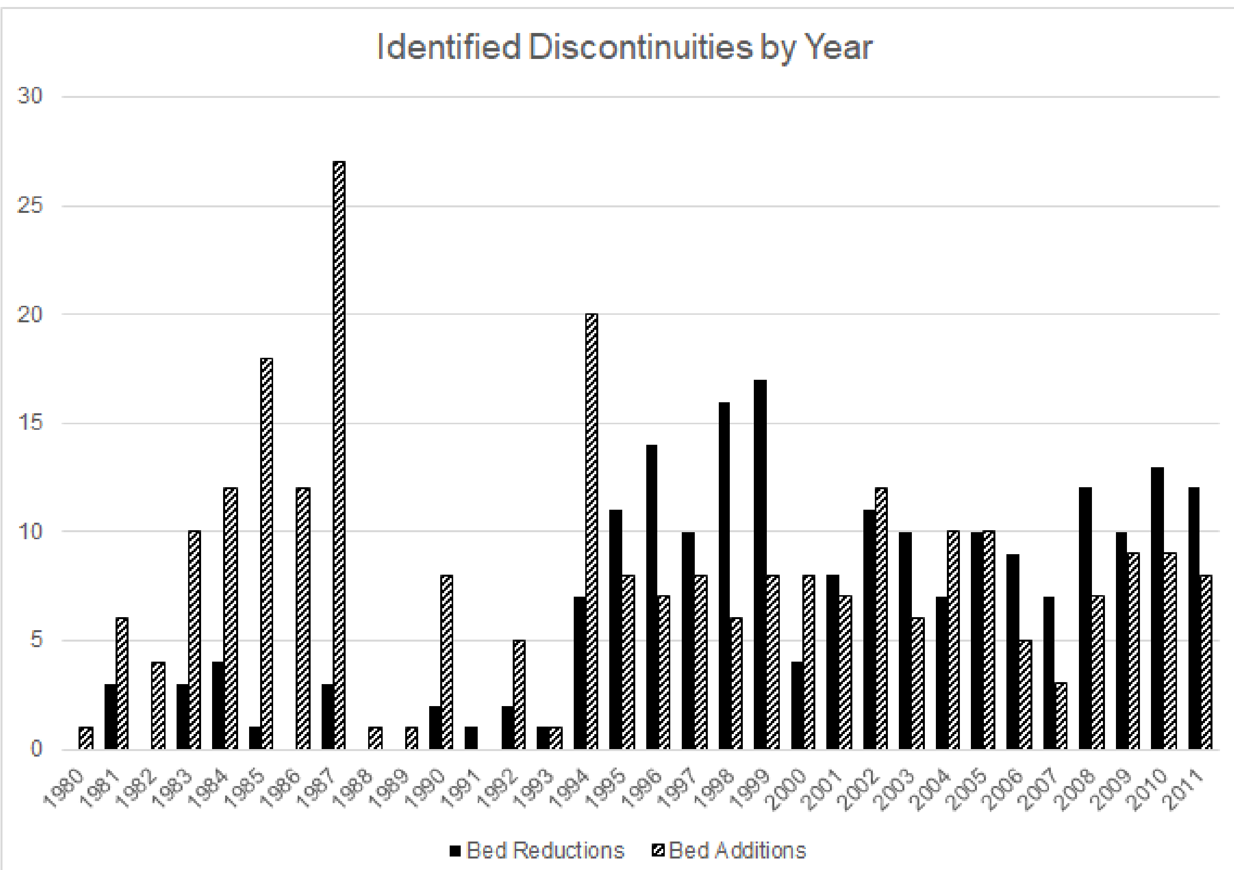
with psychiatric illness, although it is not clear whether that mechanism occurred through primarily through defunding long-run financial assistance programs or psychiatric bed supply itself [102].

concurrent movement of a few large, state-level providers responding to a single large change in federal policy. Rather, provider exit occurred gradually, in an ad hoc fashion, throughout the following two decades. Thus, while the gradual changes in the psychiatric market present a difficulty for analysts interested in long-run decisions, the local area entry and exit of small providers may nevertheless be useful for short-run investigation.

A.2 Summary of Discontinuities

The discontinuities found by the kernel methods from the Medicare Provider-of-Services 1991-2014 and American Hospital Association data 1980-2014 are summarized in Figure A.4. The pattern of bed additions and reductions are consistent with the description of psychiatric history. During the 1980s, elective inpatient substance use treatment and adolescent psychiatric care expanded, a trend which is reflected by the bed additions found in the data. Although the 1990s are sometimes cited as an era of expanded managed care models in psychiatry, which curtailed inpatient psychiatric profitability, the bed reductions found in this data occur predominantly from 1994 onwards.

Figure A.4: Number of Identified Discontinuities by Year.

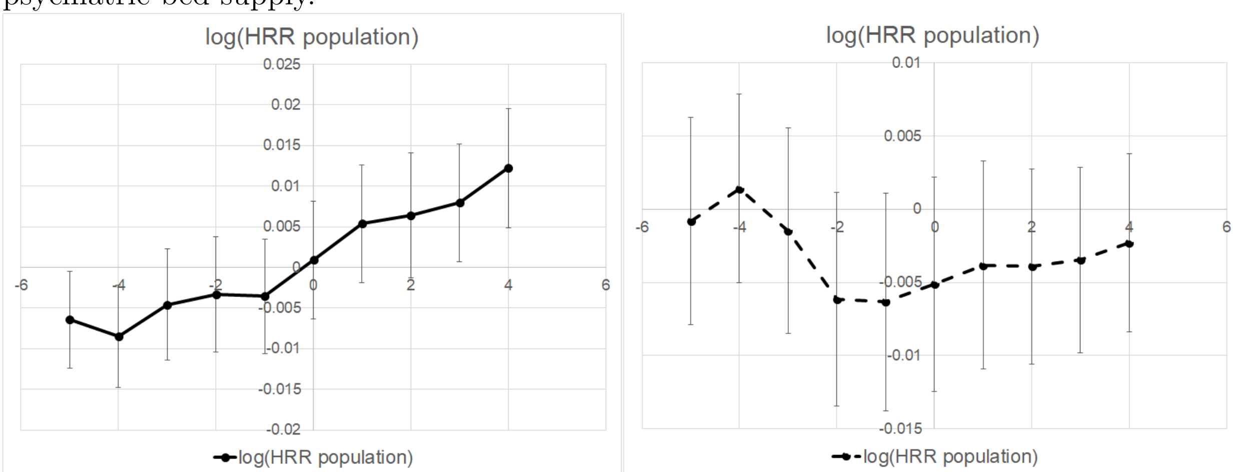


Black bars show the number of identified HRR-level bed reductions per year. Gray bars show the number of identified HRR-level bed additions per year.

A.3 Checks on the First Stage

One of the limitations of this analysis is the inability to identify the reason for psychiatric bed additions or reductions. With this limitation in mind, I wished to investigate possible correlates to psychiatric additions and reductions using the data at hand. Figure A.5 shows the event study conducted around the timing of the identified discontinuities for the population at the HRR-level. Here we see that the discontinuities, either additions of reductions do not appear to be substantially correlated with any discontinuous change in the HRR population.

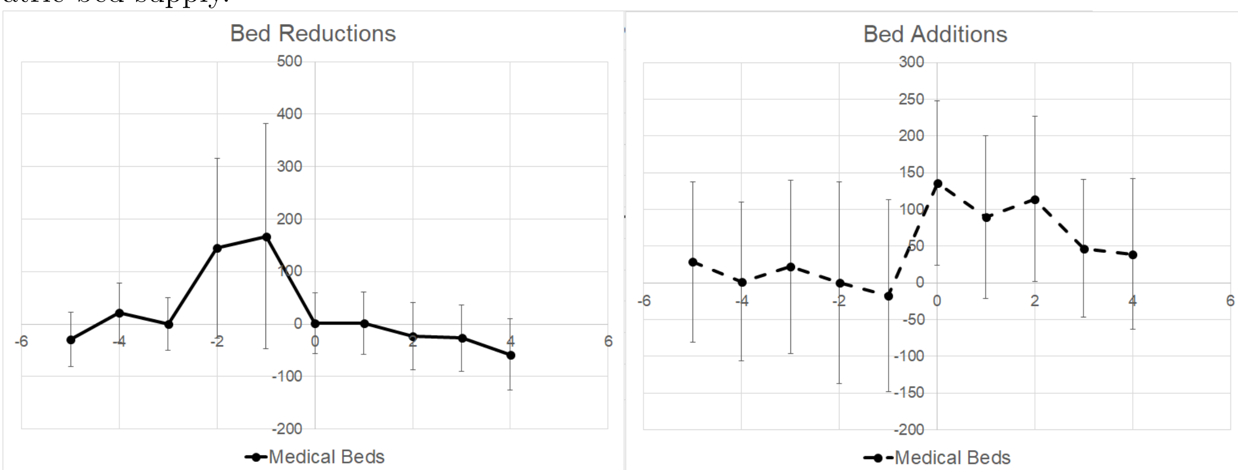
Figure A.5: Hospital referral region population with respect to identified discontinuities in psychiatric bed supply.



Author's calculations. HRR-level population versus years from identified discontinuities. Left: Discontinuous bed reductions. Right: Discontinuous bed additions. The zero point in both graphs is the year of the identified discontinuity. The plotted line is the average number of psychiatric beds within each HRR. The x-axis points -5, -4, -3, -2, -1 correspond to the averages 5, 4, 3, 2, and 1 years before the discontinuity respectively. X-axis points 1, 2, 3, 4 correspond to the years 1, 2, 3 and 4 years following the discontinuity respectively. Standard errors given by the error bars. All 95% confidence intervals contain zero.

In Figure A.6, I show the event study conducted around the timing of the identified discontinuities for medical beds at the HRR-level. Here medical beds is defined as the total number of hospital beds reported in the data minus the number of psychiatric beds. Here we find some discontinuity coinciding with the timing of the reduction and addition of psychiatric beds, although in neither case is the size of this discontinuity statistically significant. In particular, it appears that psychiatric bed additions may coincide with medical bed additions along with a baseline shift. This may indicate that psychiatric bed additions may occur in the form of opening an individual psychiatric ward, but also in the form of opening an additional medical facility. To contend with this shifting baseline, I correct for the change in the number of medical admissions, I difference the baseline shift in medical admissions from the shift in psychiatric admissions in my main analysis.

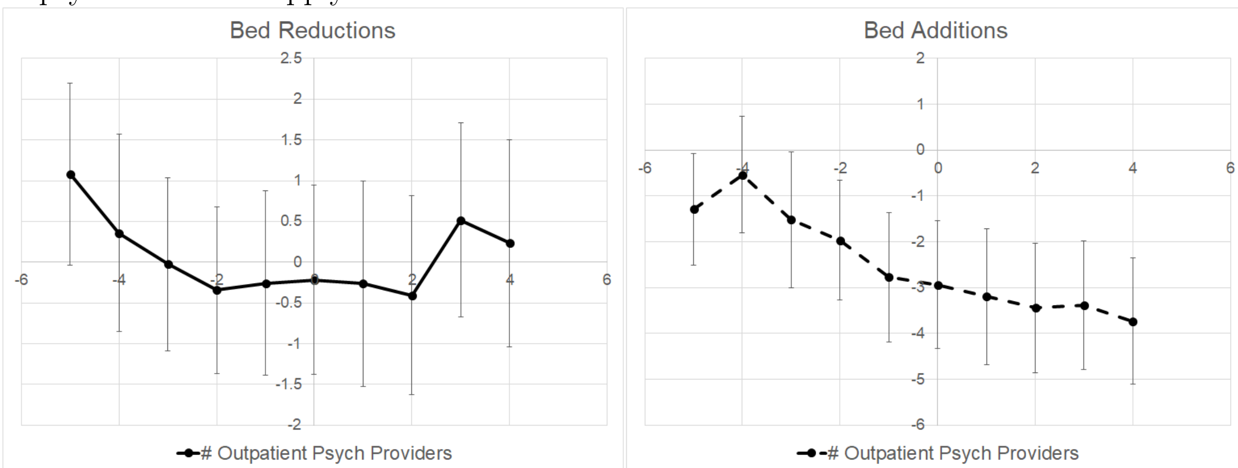
Figure A.6: Non-psychiatric bed supply with respect to identified discontinuities in psychiatric bed supply.



Author's calculations. Number of total beds minus the number of psychiatric beds by HRR versus years from identified discontinuities. Left: Discontinuous bed reductions. Right: Discontinuous bed additions. The zero point in both graphs is the year of the identified discontinuity. The plotted line is the average number of psychiatric beds within each HRR. The x-axis points -5, -4, -3, -2, -1 correspond to the averages 5, 4, 3, 2, and 1 years before the discontinuity respectively. X-axis points 1, 2, 3, 4 correspond to the years 1, 2, 3 and 4 years following the discontinuity respectively. Standard errors given by the error bars. All 95% confidence intervals contain zero.

An additional potential for confounding can be due to discrete changes in the level of outpatient service provision around the time of discontinuous change. If hospitals are responding to a decline in inpatient demand due to some exogenous shock, say introduction of a medication, then we should expect to find that psychiatric outpatient service provision is inversely correlated with inpatient service provision. On the other hand, we could also have a prior hypothesis that since inpatient bed reductions are responding to some shock that globally decreasing the profitability or desirability of providing psychiatric care, for example, if a state has withdrawn funding from mental health services writ large.

Figure A.7: Outpatient psychiatric service provision with respect to identified discontinuities in psychiatric bed supply.



Author's calculations. Number of psychiatric outpatient providers by HRR versus years from identified discontinuities. Left: Discontinuous bed reductions. Right: Discontinuous bed additions. The zero point in both graphs is the year of the identified discontinuity. The plotted line is the average number of psychiatric beds within each HRR. The x-axis points -5, -4, -3, -2, -1 correspond to the averages 5, 4, 3, 2, and 1 years before the discontinuity respectively. X-axis points 1, 2, 3, 4 correspond to the years 1, 2, 3 and 4 years following the discontinuity respectively. Standard errors given by the error bars. All 95% confidence intervals contain zero.

To check that these large-scale changes are not driving the observed changes in inpatient bed supply, I use Medicare POS and the AHA again, this time looking at the number of providers that state they provide outpatient psychiatric services with respect to the timing of discontinuities. The results, shown in Figure A.7 show again, that there is no significant outpatient shock, which would potentially confound the interpretation of the main results.

Table A.2: Placebo Variable Summaries

	Mean	Median	Std Dev.	Min	Max
F-Values	1.03	0.32	1.60	0.00	8.97
Beta All Hospital	0.04	0.00	0.59	-0.82	8.21
Beta Non Psych Hosp	-0.13	0.00	1.83	-25.82	0.95
Beta Public Hospitals	0.04	0.00	0.32	-0.67	2.97
Beta Private Insurance	0.00	0.02	0.31	-2.48	2.06
Beta Std Inmates per 100,000	0.05	-0.00	1.45	-9.91	17.51
Beta Has Alc/Drug Treat	0.03	-0.01	0.72	-3.03	5.43

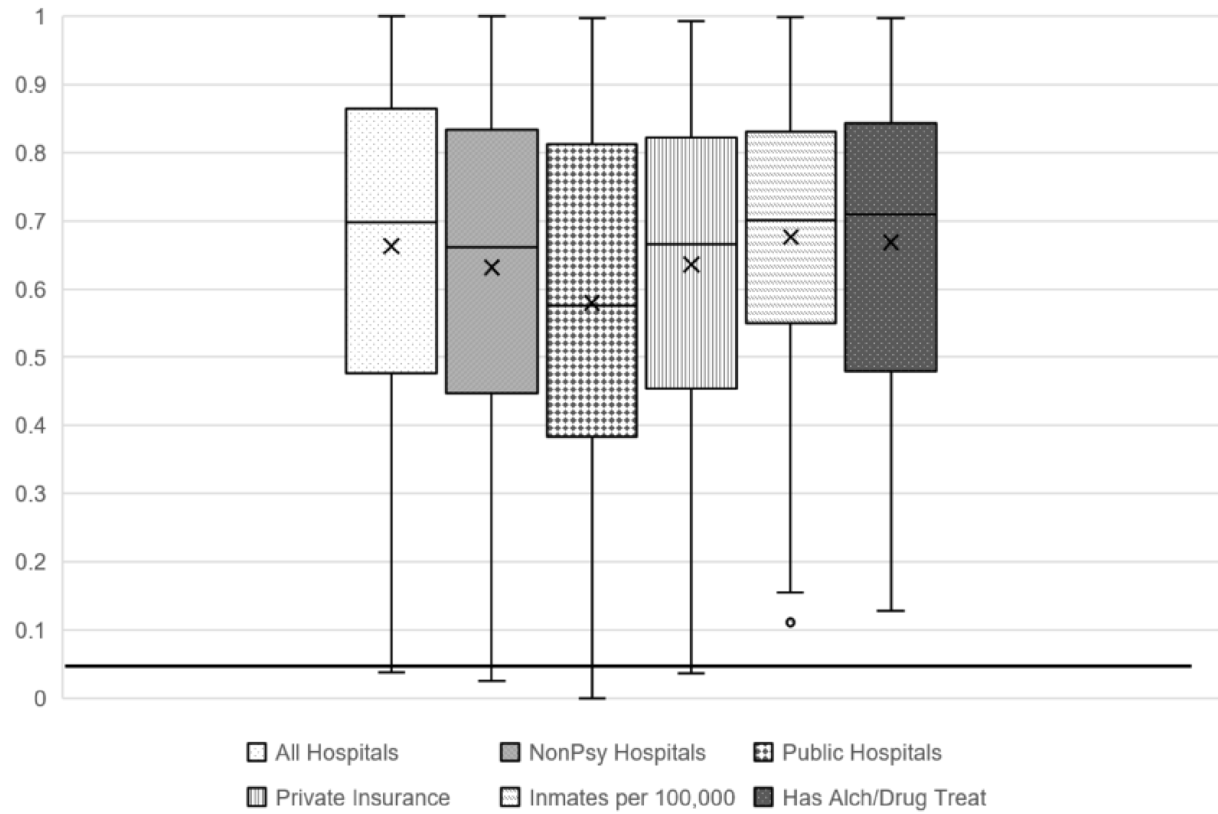
Author's calculations. 4 percent of the HRR-year observations were tagged as discontinuities at random. Subsequently, 2SLS regression was performed using the randomly specified discontinuities as the instrument. Table summarizes results of 200 iterations these random discontinuities using 2SLS specification with outcomes listed.

A.4 Placebo Regressions

To check that a strong underlying time trend in the data was driving the results of the analysis, I performed 200 placebo regressions, in which I tagged discontinuous years at random. Table A.2 summarizes the results of those regressions. In short, the kernel-identified regressions were a substantial improvement over the placebo in the first stage. On average, the size of the estimates across all regressions found a spillover effect size close to zero.

Figure A.8 summarizes the P-values from these placebo regressions with statistically significant findings tagged as those with a p-value of less than 0.05, as marked by the horizonal line. In general, finding a significant effect was a rare outcome for all of the outcomes examined. In the case of the jail outcomes, no significant effects at all were detected indicating that the analysis is likely underpowered to detect these effects.

Figure A.8: P-Values Derived from 200 Placebo Regressions.



Author's calculations. 4 percent of the HRR-year observations were tagged as discontinuities at random. Subsequently, 2SLS regression was performed using the randomly specified discontinuities as the instrument. The figure summarizes the p-values obtained on each of the main outcomes used in this analysis. For each of the hospital outcomes, the outcome of interest was the difference between psychiatric and medical admissions index.

A.5 Supplemental Tables

Table A.3: Event Study of Bed Reductions

	(1) t=-5	(2) t=-4	(3) t=-3	(4) t=-2	(5) t=-1	(6) t=0	(7) t=1	(8) t=2	(9) t=3	(10) t=4
Psych Ind - Ctrl Ind (All Hosp, HCUP)	0.011	-0.026	0.035	0.089	-0.042	-0.095	-0.130	-0.103	-0.152	-0.076
	(0.054)	(0.054)	(0.056)	(0.065)	(0.076)	(0.073)	(0.086)	(0.066)	(0.053)	(0.060)
Psych Ind - Ctrl Ind (All Hosp, MarketScan)	0.016	0.061	-0.135	-0.104	-0.086	-0.228	-0.207	-0.203	-0.178	-0.036
	(0.044)	(0.061)	(0.153)	(0.100)	(0.077)	(0.100)	(0.080)	(0.095)	(0.098)	(0.086)
Psych Ind - Ctrl Ind (Hosp w/o Psych Beds)	-0.030	-0.013	-0.055	0.014	0.081	0.102	0.092	0.114	0.008	0.015
	(0.025)	(0.025)	(0.042)	(0.032)	(0.087)	(0.098)	(0.102)	(0.111)	(0.038)	(0.032)
Psych Ind - Ctrl Ind (Public Hosp)	-0.071	-0.079	0.039	0.293	0.184	0.210	0.186	0.227	-0.141	-0.195
	(0.066)	(0.047)	(0.048)	(0.257)	(0.148)	(0.173)	(0.148)	(0.178)	(0.122)	(0.164)
Std Admits for Oth Psych Dx (All Hosp)	0.090	0.004	-0.045	-0.086	-0.294	-0.302	-0.304	-0.386	-0.326	-0.156
	(0.104)	(0.105)	(0.113)	(0.162)	(0.225)	(0.231)	(0.177)	(0.205)	(0.128)	(0.103)
Std Admits for Psychosis (All Hosp)	-0.003	-0.075	-0.122	-0.149	-0.352	-0.359	-0.291	-0.344	-0.310	-0.162
	(0.060)	(0.071)	(0.076)	(0.116)	(0.219)	(0.235)	(0.187)	(0.198)	(0.125)	(0.114)
Std Admits with Psych Eval (All Hosp)	0.025	-0.047	-0.137	-0.038	-0.228	-0.139	-0.187	-0.258	-0.274	-0.110
	(0.133)	(0.145)	(0.133)	(0.150)	(0.179)	(0.171)	(0.141)	(0.132)	(0.114)	(0.126)
Std Admits for Weird Urine (All Hosp)	-0.008	-0.046	-0.132	0.022	-0.249	-0.432	-0.495	-0.335	-0.302	-0.153
	(0.070)	(0.085)	(0.069)	(0.146)	(0.379)	(0.371)	(0.417)	(0.211)	(0.186)	(0.126)
Std Admits Medical Seq of Psych Dx (All Hosp)	0.014	-0.063	-0.139	-0.210	-0.305	-0.278	-0.209	-0.253	-0.212	-0.122
	(0.055)	(0.062)	(0.078)	(0.136)	(0.192)	(0.209)	(0.166)	(0.169)	(0.121)	(0.110)
Std Admits for V-Codes (All Hosp)	-0.021	-0.198	-0.006	-0.155	-0.414	-0.437	-0.247	-0.304	-0.484	-0.313
	(0.132)	(0.098)	(0.139)	(0.143)	(0.238)	(0.267)	(0.213)	(0.237)	(0.142)	(0.205)
Index of Psych Outcomes (HCUP)	0.016	-0.071	-0.097	-0.102	-0.307	-0.324	-0.289	-0.313	-0.318	-0.169
	(0.065)	(0.073)	(0.077)	(0.117)	(0.230)	(0.239)	(0.203)	(0.177)	(0.119)	(0.116)
Index of Control Outcomes (HCUP)	0.006	-0.045	-0.131	-0.191	-0.265	-0.230	-0.159	-0.210	-0.166	-0.093
	(0.051)	(0.057)	(0.071)	(0.123)	(0.178)	(0.187)	(0.141)	(0.145)	(0.104)	(0.092)
Std Admits for Oth Psych Dx (Hosp w/o Psych Beds)	0.019	0.065	0.180	0.063	-0.258	-0.272	-0.356	-0.435	-0.324	-0.186
	(0.053)	(0.076)	(0.218)	(0.075)	(0.278)	(0.327)	(0.364)	(0.406)	(0.264)	(0.207)
Std Admits for Psychosis (Hosp w/o Psych Beds)	0.004	0.054	0.117	0.009	-0.237	-0.258	-0.286	-0.342	-0.270	-0.164
	(0.047)	(0.070)	(0.165)	(0.042)	(0.244)	(0.280)	(0.293)	(0.325)	(0.221)	(0.174)
Std Admits with Psych Eval (Hosp w/o Psych Beds)	-0.082	0.022	-0.094	-0.077	-0.130	-0.041	-0.149	-0.204	-0.237	-0.058
	(0.060)	(0.109)	(0.113)	(0.063)	(0.199)	(0.227)	(0.140)	(0.142)	(0.107)	(0.104)
Std Admits with Weird Urine (Hosp w/o Psych Beds)	0.032	0.033	0.145	-0.034	-0.212	-0.260	-0.328	-0.423	-0.360	-0.268
	(0.061)	(0.056)	(0.174)	(0.045)	(0.260)	(0.298)	(0.337)	(0.372)	(0.312)	(0.259)
Std Admits Med Seq of Psych Dx (Hosp w/o Psych Beds)	0.020	0.058	0.137	0.014	-0.274	-0.284	-0.338	-0.411	-0.305	-0.208
	(0.047)	(0.068)	(0.179)	(0.045)	(0.274)	(0.318)	(0.342)	(0.375)	(0.261)	(0.214)
Std Admits for V-Code (Hosp w/o Psych Beds)	-0.023	-0.001	0.036	0.111	-0.061	-0.058	-0.085	-0.098	-0.105	-0.095
	(0.013)	(0.018)	(0.040)	(0.122)	(0.045)	(0.057)	(0.072)	(0.084)	(0.071)	(0.078)
Index of Psych Dx Hosp w/o Psych Beds)	-0.005	0.038	0.087	0.014	-0.195	-0.196	-0.257	-0.319	-0.267	-0.163
	(0.037)	(0.051)	(0.144)	(0.047)	(0.211)	(0.243)	(0.255)	(0.282)	(0.202)	(0.168)
Index of Control Dx (Hosp w/o Psych Beds)	0.025	0.051	0.142	0.001	-0.276	-0.298	-0.348	-0.433	-0.275	-0.178
	(0.054)	(0.052)	(0.181)	(0.039)	(0.293)	(0.337)	(0.356)	(0.392)	(0.235)	(0.194)
Std Admits for Oth Psych Dx (Public Hosp)	-0.058	-0.127	0.019	-0.038	-0.266	-0.459	-0.408	-0.519	-0.179	-0.087
	(0.093)	(0.095)	(0.120)	(0.163)	(0.352)	(0.399)	(0.420)	(0.456)	(0.120)	(0.109)
Std Admits for Psychosis (Public Hosp)	-0.061	-0.172	-0.101	-0.031	-0.218	-0.361	-0.280	-0.332	-0.113	-0.085
	(0.087)	(0.084)	(0.115)	(0.118)	(0.265)	(0.288)	(0.276)	(0.293)	(0.084)	(0.072)
Std Admits with Psych Eval (Public Hosp)	-0.118	-0.243	-0.244	0.140	-0.138	0.005	0.012	0.148	-0.040	-0.132
	(0.153)	(0.182)	(0.194)	(0.227)	(0.298)	(0.267)	(0.177)	(0.218)	(0.185)	(0.114)
Std Admits with Weird Urine (Public Hosp)	-0.094	-0.175	-0.197	-0.319	-0.198	-0.514	-0.428	-0.539	-0.483	-0.269
	(0.146)	(0.169)	(0.159)	(0.368)	(0.481)	(0.466)	(0.495)	(0.526)	(0.325)	(0.222)
Std Admits Med Seq of Psych Dx (Public Hosp)	-0.022	-0.144	-0.099	-0.247	-0.319	-0.511	-0.368	-0.474	-0.111	-0.072
	(0.064)	(0.079)	(0.113)	(0.250)	(0.334)	(0.395)	(0.378)	(0.400)	(0.070)	(0.073)
Std Admits for V-Code (Public Hosp)	-0.132	-0.092	0.100	0.202	-0.064	-0.077	-0.024	-0.035	-0.134	-0.095
	(0.069)	(0.051)	(0.093)	(0.281)	(0.091)	(0.070)	(0.118)	(0.155)	(0.083)	(0.111)
Index of Psych Dx Public Hosp)	-0.081	-0.159	-0.087	-0.049	-0.201	-0.319	-0.249	-0.292	-0.177	-0.123
	(0.084)	(0.075)	(0.091)	(0.121)	(0.252)	(0.282)	(0.276)	(0.300)	(0.092)	(0.071)
Index of Control Dx (Public Hosp)	-0.010	-0.080	-0.126	-0.342	-0.385	-0.529	-0.435	-0.519	-0.036	0.072
	(0.050)	(0.065)	(0.098)	(0.326)	(0.365)	(0.434)	(0.402)	(0.444)	(0.106)	(0.144)

Author's calculations. Standard deviation in parentheses. Coefficients derived from 2SLS specification performed at the HRR-level. Discontinuous psychiatric bed additions or reductions as IV as specified. Period affected by discontinuous change defined as up to three years following observed discontinuity.

Table A.4: Event Study Discrete Bed Addition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	t=-5	t=-4	t=-3	t=-2	t=-1	t=0	t=1	t=2	t=3	t=4
Psych Ind - Ctrl Ind (All Hosp, HCUP)	-0.027	-0.016	0.007	-0.027	-0.016	-0.044	0.032	0.049	0.009	0.016
	(0.079)	(0.094)	(0.110)	(0.076)	(0.062)	(0.067)	(0.071)	(0.074)	(0.044)	(0.067)
Psych Ind - Ctrl Ind (All Hosp, MarketScan)	0.045	0.063	-0.141	-0.046	-0.022	-0.069	0.235	0.193	0.080	-0.065
	(0.095)	(0.112)	(0.133)	(0.134)	(0.123)	(0.164)	(0.154)	(0.131)	(0.104)	(0.266)
Psych Ind - Ctrl Ind (Hosp w/o Psych Beds)	0.097	0.067	0.055	-0.052	-0.041	-0.033	-0.019	-0.009	-0.058	-0.090
	(0.085)	(0.081)	(0.080)	(0.044)	(0.029)	(0.042)	(0.049)	(0.019)	(0.060)	(0.067)
Psych Ind - Ctrl Ind (Public Hosp)	0.109	0.064	0.081	-0.071	-0.103	-0.252	-0.103	-0.094	-0.053	-0.244
	(0.123)	(0.091)	(0.088)	(0.150)	(0.195)	(0.178)	(0.110)	(0.098)	(0.082)	(0.144)
Std Admits for Oth Psych Dx (All Hosp)	-0.114	0.029	0.061	0.108	0.162	0.214	0.162	0.150	0.133	0.154
	(0.181)	(0.168)	(0.228)	(0.167)	(0.157)	(0.163)	(0.139)	(0.147)	(0.133)	(0.132)
Std Admits for Psychosis (All Hosp)	-0.177	-0.065	-0.118	-0.042	0.016	0.039	0.077	0.088	0.064	0.075
	(0.204)	(0.175)	(0.165)	(0.147)	(0.149)	(0.137)	(0.126)	(0.141)	(0.129)	(0.134)
Std Admits with Psych Eval (All Hosp)	-0.077	0.020	-0.044	-0.112	-0.076	-0.087	-0.007	-0.009	0.095	0.159
	(0.169)	(0.252)	(0.278)	(0.222)	(0.168)	(0.192)	(0.173)	(0.150)	(0.120)	(0.214)
Std Admits for Weird Urine (All Hosp)	-0.260	-0.347	-0.156	-0.146	0.087	-0.026	0.171	0.081	0.038	0.088
	(0.353)	(0.307)	(0.110)	(0.118)	(0.186)	(0.152)	(0.220)	(0.160)	(0.159)	(0.149)
Std Admits Medical Seq of Psych Dx (All Hosp)	-0.166	-0.078	-0.094	-0.077	0.040	0.062	0.100	0.079	0.077	0.092
	(0.174)	(0.138)	(0.117)	(0.117)	(0.137)	(0.142)	(0.124)	(0.132)	(0.119)	(0.121)
Std Admits for V-Codes (All Hosp)	-0.196	0.061	0.003	-0.148	0.015	-0.043	0.205	0.380	0.085	0.046
	(0.265)	(0.220)	(0.312)	(0.178)	(0.225)	(0.210)	(0.225)	(0.308)	(0.193)	(0.207)
Index of Psych Outcomes (HCUP)	-0.165	-0.063	-0.058	-0.069	0.041	0.026	0.118	0.128	0.082	0.102
	(0.201)	(0.176)	(0.164)	(0.118)	(0.128)	(0.121)	(0.130)	(0.142)	(0.123)	(0.141)
Index of Control Outcomes (HCUP)	-0.138	-0.048	-0.065	-0.042	0.056	0.071	0.086	0.080	0.073	0.087
	(0.148)	(0.112)	(0.101)	(0.108)	(0.115)	(0.122)	(0.113)	(0.123)	(0.110)	(0.112)
Std Admits for Oth Psych Dx (Hosp w/o Psych Beds)	-0.292	-0.301	-0.314	-0.196	-0.140	-0.179	-0.197	-0.119	-0.021	0.015
	(0.330)	(0.319)	(0.306)	(0.168)	(0.136)	(0.161)	(0.135)	(0.089)	(0.050)	(0.071)
Std Admits for Psychosis (Hosp w/o Psych Beds)	-0.240	-0.244	-0.263	-0.163	-0.113	-0.163	-0.158	-0.071	0.008	0.016
	(0.270)	(0.244)	(0.232)	(0.129)	(0.104)	(0.123)	(0.099)	(0.059)	(0.058)	(0.077)
Std Admits with Psych Eval (Hosp w/o Psych Beds)	-0.090	-0.093	-0.185	-0.320	-0.196	-0.245	-0.321	-0.097	-0.114	-0.138
	(0.196)	(0.222)	(0.249)	(0.225)	(0.167)	(0.239)	(0.259)	(0.116)	(0.229)	(0.219)
Std Admits with Weird Urine (Hosp w/o Psych Beds)	-0.246	-0.256	-0.254	-0.206	-0.154	-0.094	-0.106	-0.059	-0.064	0.026
	(0.270)	(0.272)	(0.261)	(0.183)	(0.145)	(0.072)	(0.057)	(0.042)	(0.043)	(0.112)
Std Admits Med Seq of Psych Dx (Hosp w/o Psych Beds)	-0.280	-0.262	-0.265	-0.169	-0.109	-0.129	-0.129	-0.057	0.031	0.047
	(0.305)	(0.285)	(0.267)	(0.153)	(0.128)	(0.138)	(0.104)	(0.049)	(0.056)	(0.076)
Std Admits for V-Code (Hosp w/o Psych Beds)	-0.067	-0.088	-0.090	-0.081	-0.056	-0.067	-0.040	-0.020	-0.000	-0.060
	(0.077)	(0.077)	(0.080)	(0.068)	(0.066)	(0.069)	(0.061)	(0.054)	(0.068)	(0.040)
Index of Psych Dx Hosp w/o Psych Beds)	-0.203	-0.207	-0.228	-0.189	-0.128	-0.146	-0.158	-0.071	-0.026	-0.016
	(0.234)	(0.217)	(0.207)	(0.126)	(0.104)	(0.110)	(0.098)	(0.049)	(0.064)	(0.076)
Index of Control Dx (Hosp w/o Psych Beds)	-0.300	-0.274	-0.284	-0.137	-0.087	-0.114	-0.139	-0.062	0.031	0.075
	(0.314)	(0.286)	(0.269)	(0.115)	(0.096)	(0.108)	(0.089)	(0.043)	(0.066)	(0.106)
Std Admits for Oth Psych Dx (Public Hosp)	-0.487	-0.395	-0.313	0.041	0.116	-0.029	-0.159	0.004	-0.074	0.038
	(0.381)	(0.306)	(0.306)	(0.133)	(0.167)	(0.102)	(0.137)	(0.137)	(0.198)	(0.161)
Std Admits for Psychosis (Public Hosp)	-0.329	-0.299	-0.247	0.026	0.131	-0.023	-0.054	0.068	-0.074	0.025
	(0.259)	(0.204)	(0.185)	(0.073)	(0.128)	(0.091)	(0.105)	(0.157)	(0.163)	(0.097)
Std Admits with Psych Eval (Public Hosp)	-0.105	-0.063	-0.002	0.193	0.152	-0.152	0.097	-0.008	-0.183	0.079
	(0.093)	(0.204)	(0.224)	(0.201)	(0.157)	(0.228)	(0.126)	(0.264)	(0.281)	(0.196)
Std Admits with Weird Urine (Public Hosp)	-0.379	-0.286	-0.310	-0.086	-0.092	0.261	-0.033	-0.177	0.423	-0.026
	(0.354)	(0.280)	(0.280)	(0.167)	(0.124)	(0.285)	(0.309)	(0.395)	(0.608)	(0.385)
Std Admits Med Seq of Psych Dx (Public Hosp)	-0.365	-0.283	-0.188	0.074	0.084	0.093	-0.010	0.169	0.209	0.112
	(0.332)	(0.244)	(0.227)	(0.115)	(0.138)	(0.099)	(0.087)	(0.193)	(0.324)	(0.185)
Std Admits for V-Code (Public Hosp)	-0.151	-0.074	-0.072	-0.011	0.045	-0.110	-0.172	0.057	-0.219	-0.353
	(0.128)	(0.121)	(0.135)	(0.125)	(0.167)	(0.139)	(0.146)	(0.226)	(0.124)	(0.152)
Index of Psych Dx Public Hosp)	-0.303	-0.233	-0.189	0.039	0.073	0.007	-0.055	0.019	0.014	-0.021
	(0.249)	(0.182)	(0.185)	(0.093)	(0.097)	(0.089)	(0.094)	(0.113)	(0.207)	(0.111)
Index of Control Dx (Public Hosp)	-0.411	-0.298	-0.270	0.110	0.176	0.259	0.048	0.113	0.067	0.223
	(0.365)	(0.244)	(0.230)	(0.155)	(0.227)	(0.215)	(0.110)	(0.153)	(0.184)	(0.229)

Author's calculations. Standard deviation in parentheses. Coefficients derived from 2SLS specification performed at the HRR-level. Discontinuous psychiatric bed additions or reductions as IV as specified. Period affected by discontinuous change defined as up to three years following observed discontinuity.

APPENDIX B

APPENDIX TO BIFACTOR MODELING

B.1 Exploratory Bifactor Analysis

See Table B.1.

B.2 Individual Period Bifactor Loading Estimates

See Table B.2.

Table B.1: Exploratory Bifactor Analysis

Item Description	(1) Common Dim	(2) Subdim 1	(3) Subdim 2	(4) Subdim 3	(5) Subdim 4
0-Month Mail (June-Nov 2008)					
Self-Rated General Health	-0.4	-0.3	0.0		
# Bad Days Due to Physical Health	-0.6	-0.3	0.0		
# Bad Days Due to Mental Health	-0.7	-0.1	-0.1		
Diagnosis of Depression	-0.5	0.1	-0.1		
Any Hospitalizations	-0.2	0.0	0.5		
Any Doctor's Visits	-0.3	0.1	0.2		
Any Emergency Room Visits	-0.3	0.0	0.4		
6-Month Mail (Jan-May 2009)					
Self-Rated General Health	0.6	0.0	0.2	0.1	
# Bad Days Due to Physical Health	0.6	-0.1	0.2	0.0	
# Bad Days Due to Mental Health	0.7	0.0	0.1	0.2	
Endorse Sadness	0.6	0.0	0.0	0.5	
Endorse Lack of Interest	0.6	0.0	0.0	0.5	
Diagnosis of Depression	0.5	0.0	-0.4	0.1	
Prescriptions for Depression	0.4	0.0	-0.5	0.0	
Any Hospitalizations	0.2	-0.5	0.0	0.0	
Any Doctor's Visits	0.3	-0.3	-0.2	-0.1	
Any Emergency Room Visits	0.3	-0.5	0.0	0.0	
12-Month Mail (July 2009-Mar 2010)					
Self-Rated General Health	0.7	0.0	0.0	0.0	
# Bad Days Due to Physical Health	0.7	-0.1	0.1	0.1	
# Bad Days Due to Mental Health	0.6	-0.1	0.0	-0.3	
Endorse Sadness	0.6	0.0	0.0	-0.5	
Endorse Lack of Interest	0.6	0.0	0.0	-0.5	
Diagnosis of Depression	0.3	0.8	0.0	0.0	
Prescriptions for Depression	0.2	0.8	0.0	0.0	
Any Hospitalizations	0.2	0.0	0.4	0.0	
Any Doctor's Visits	0.2	0.1	0.3	0.0	
Any Emergency Room Visits	0.3	0.0	0.4	0.0	
In-Person (Aug 2009-Oct 2011)					
Self-Rated General Health	0.6	0.3	-0.1	0.0	-0.1
Self-Rated Pain Severity	0.5	0.6	0.1	0.0	0.0
SF-8 Physical Component Score	0.5	0.7	0.0	0.0	0.0
Self-Rated Happiness	0.6	0.0	0.0	-0.2	-0.1
PHQ-9 Score	0.8	0.0	0.0	-0.2	0.0
SF-8 Mental Component Score	0.7	-0.1	0.0	-0.3	0.0
Diagnosis of Depression	0.4	0.0	0.0	-0.1	0.4
Prescriptions for Depression	0.4	0.0	0.0	0.0	0.4
Any Hospitalizations	0.3	0.0	0.0	0.4	0.0
Any Doctor's Visits	0.3	0.1	0.0	0.2	0.3
Any Emergency Room Visits	0.4	0.0	0.0	0.4	0.0

Table B.2: Consistency in Item Loading and Threshold Across Period.

Item Description	(1) Item Threshold	(2) Common Dim Fac Loading	(3) Subdim Load 1	(4) Subdim Load 2	(5) Subdim Load 3	(6) Subdim Load 4
0-Month Mail (June - Nov 2008)						
Self-Rated General Health	-1.182	0.478	0.260			
# Bad Days Due to Physical Health	-0.961	0.806	0.262			
# Bad Days Due to Mental Health	-0.505	0.706		0.485		
Diagnosis of Depression	-0.184	0.475		0.646		
Any Hospitalizations	-1.478	0.361			0.913	
Any Doctor's Visits	0.192	0.274			0.258	
Any Emergency Room Visits	-0.577	0.349			0.640	
6-Month Mail (July 2009-Mar 2010)						
Self-Rated General Health	-1.137	0.602	0.485			
# Bad Days Due to Physical Health	-0.980	0.652	0.521			
# Bad Days Due to Mental Health	-0.477	0.839		0.270		
Endorse Sadness	-0.965	0.794		0.598		
Endorse Lack of Interest	-1.148	0.799		0.402		
Diagnosis of Depression	-0.196	0.640			0.679	
Prescriptions for Depression	-0.801	0.513			0.777	
Any Hospitalizations	-1.544	0.292				0.920
Any Doctor's Visits	0.178	0.259				0.407
Any Emergency Room Visits	-0.758	0.365				0.645
12-Month Mail (July 2009-Mar 2010)						
Self-Rated General Health	-1.160	0.632	0.450			
# Bad Days Due to Physical Health	-0.999	0.665	0.507			
# Bad Days Due to Mental Health	-0.528	0.819		0.271		
Endorse Sadness	-1.061	0.800		0.588		
Endorse Lack of Interest	-1.071	0.805		0.363		
Diagnosis of Depression	-0.030	0.665			0.691	
Prescriptions for Depression	-0.533	0.584			0.759	
Any Hospitalizations	-1.464	0.310				0.904
Any Doctor's Visits	0.265	0.257				0.348
Any Emergency Room Visits	-0.659	0.343				0.620
In-Person (Aug 2009-Oct 2011)						
Self-Rated General Health	-1.117	0.610	0.355			
Self-Rated Pain Severity	-0.964	0.585	0.618			
SF-8 Physical Component Score	-0.968	0.547	0.831			
Self-Rated Happiness	-0.676	0.549		0.428		
PHQ-9 Score	-1.033	0.823		0.369		
SF-8 Mental Component Score	-0.981	0.699		0.488		
Diagnosis of Depression	-0.255	0.673			0.584	
Prescriptions for Depression	-0.935	0.556			0.686	
Any Hospitalizations	-1.124	0.318				0.922
Any Doctor's Visits	0.440	0.349				0.327
Any Emergency Room Visits	-0.236	0.376				0.569

Here a bifactor model is fit to each period separately to demonstrate consistency in item loading and thresholds across measurement periods. Common Dim Threshold column reports the estimated item threshold on the common dimension. Common Dim Fac Loading reports the mean estimated factor loading onto the common dimension. Subdim Load 1-4 reports the mean estimated factor loading onto the first to fourth sub-dimensions respectively.