

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Essays in Health and Behavioral Economics

Permalink

<https://escholarship.org/uc/item/8kj9f0q6>

Author

Boone, Claire

Publication Date

2022

Peer reviewed|Thesis/dissertation

Essays in Health and Behavioral Economics

by

Claire E. Boone

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Health Policy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Paul J. Gertler, Chair

Professor Jonathan Kolstad

Professor Ulrike Malmendier

Professor Ziad Obermeyer

Spring 2022

Essays in Health and Behavioral Economics

Copyright 2022
by
Claire E. Boone

Abstract

Essays in Health and Behavioral Economics

by

Claire E. Boone

Doctor of Philosophy in Health Policy

University of California, Berkeley

Professor Paul J. Gertler, Chair

I examine medical adherence among patients and health care providers. Non-adherence is common throughout health care: providers frequently do not follow clinical practice guidelines, and patients often do not take their prescribed treatments. These behaviors are widespread and highly consequential: in a given year in the United States alone, it is estimated that diagnostic errors lead to 40,000 to 80,000 deaths, and patient non-adherence to medication leads to \$300 billion in preventable health care costs. For these reasons, the World Health Organization has stated that “increasing adherence may have a greater effect on health than improvements in specific medical therapy”. Using data from Chile’s public, primary health care system I first show that sending reminders to patients with chronic conditions can improve health-seeking behavior. Through nudging patients at the beginning of the cascade of care, preventative care visit reminders increase not only visits but also testing, medication adherence, and tertiary care utilization. In the second paper I show that healthcare providers deviate from clinical practice guidelines, using discretion to make diagnostic and treatment decisions about hypertension. This behavior leads to a more accurate sorting of patients on either side of the diagnostic threshold, and ultimately benefits patients through avoiding unnecessary diagnoses.

To Philippe Garneau-Halliday, and Charlie, Erica, Hannah, and Bunny Boone.

Contents

Contents	ii
List of Figures	iv
List of Tables	v
1 Introduction	1
2 Nudging Patients at Scale: Evidence from Text-message Appointment Reminders	6
2.1 Introduction	7
2.2 Setting	10
2.2.1 Health and Healthcare in Chile	11
2.2.2 SMS Reminder Program	12
2.3 Data	12
2.3.1 Dates of SMS Reminder Implementation	12
2.3.2 Phone Records	13
2.3.3 Electronic Health Records	13
2.3.4 Medication Data	14
2.3.5 Hospitalization Records	14
2.3.6 Mortality Records	14
2.3.7 Descriptive Statistics	15
2.4 Empirical Approach	15
2.4.1 Event Studies	16
2.4.2 Difference-in-Differences	16
2.5 Results	17
2.5.1 Program Compliance	17
2.5.2 Impact of Nudges on Visits and Tests	17
2.5.3 Impact of Nudges on Medication Refill Behavior	19
2.5.4 Impact of Nudges on Hospitalizations	20
2.5.5 Impact of Nudges on Mortality	21
2.6 Conclusion	22

3	The Role of Discretion in Clinical Decision Making: Evidence from Thresholds	38
3.1	Introduction	39
3.2	Setting and Data	42
3.2.1	Health and Healthcare in Chile	43
3.2.2	Medical Context	44
3.2.3	Data	45
3.3	Empirical Strategy	46
3.3.1	Measuring Provider Discretion With Bunching	47
3.3.2	Variable Construction	50
3.3.3	Measuring the Impact of Discretion	51
3.4	Results	52
3.4.1	Evidence of Discretion in Blood Pressure Measurement	52
3.4.2	Impact of Discretion on Clinical Decisions	53
3.4.3	Discretion and Selection by Cardiovascular Risk	53
3.4.4	Discretion and Selection by Patient Characteristics	55
3.5	Conclusion	56
4	Conclusion	69
	Bibliography	71
A	Chapter 2 – Supplementary Tables and Figures	86
B	Chapter 3 – Supplementary Tables and Figures	100

List of Figures

2.1	Compliance with SMS reminder intervention, by clinic-year	24
2.2	Event studies: Visits and tests	25
2.3	Event studies: Medication adherence among patients with type 2 diabetes . . .	26
2.4	Event studies: Medication adherence among patients with hypertension	27
2.5	Event studies: Hospitalization per 1000 population	28
2.6	Event studies: Mortality per 1000 population	29
3.1	Expected vs. observed distribution of systolic blood pressure	58
3.2	Examples of the distribution of systolic blood pressure at 12 clinics	59
3.3	Example of counterfactual bunching estimation	60
3.4	Estimated magnitude of discretion of blood pressure	61
3.5	Balance of patient and clinic characteristics by magnitude of discretion	62
B.1	Expected vs. observed distribution of diastolic blood pressure	102
B.2	Histogram of systolic blood pressure by high vs. low discretion clinics	103
B.3	Comparison of results under different parameterizations of the magnitude of discretion: adherence to clinical practice guidelines	110
B.4	Comparison of results under different parameterizations of the magnitude of discretion: cardiovascular hospitalizations	111
B.5	Comparison of results under different parameterizations of the magnitude of discretion: patient characteristics	112

List of Tables

2.1	Take up of SMS reminders by public primary care clinics	30
2.2	Take up of SMS reminders by county	30
2.3	Baseline patient characteristics: electronic health record data	31
2.4	Baseline patient characteristics: mortality records	32
2.5	Baseline county characteristics	33
2.6	Impact of SMS reminders on primary care visits and tests	34
2.7	Impact of SMS reminders on medication adherence	35
2.8	Impact of SMS reminders on hospitalization (per 1000 population) among all patients	36
2.9	Impact of SMS reminders on mortality, among all patients)	37
3.1	Patient reasons for admission to cardiovascular health program	63
3.2	Summary statistics, clinic and municipality characteristics	63
3.3	Summary statistics, patient characteristics	64
3.4	Impact of discretion on hypertension diagnosis and prescription	65
3.5	Discretion and cardiovascular hospitalization 3, 6, and 12 months after primary care visit	66
3.6	Discretion and non-cardiovascular hospitalization 3, 6, and 12 months after primary care visit	67
3.7	Selection by patient characteristics representative of high cardiovascular risk	68
A.1	ICD-10 Codes included in cardiovascular hospitalization and mortality outcomes	87
A.2	Event study: ITT estimates of the effect of SMS reminders on visits and tests	88
A.3	Event study: ATE estimates of the effect of SMS reminders on visits and tests	89
A.4	Event study: ITT estimates of the effect of SMS reminders on medication adherence among patients with type 2 diabetes	90
A.5	Event study: ATE estimates of the effect of SMS reminders on medication adherence among patients with type 2 diabetes	91
A.6	Event study: ITT estimates of the effect of SMS reminders on medication adherence among patients with hypertension	92
A.7	Event study: ATE estimates of the effect of SMS reminders on medication adherence among patients with type 2 diabetes	93

A.8	Event study: ITT estimates of the effect of SMS reminders on hospitalizations per 1000 population	94
A.9	Event study: ATE estimates of the effect of SMS reminders on hospitalizations per 1000 population	95
A.10	Event study: ITT estimates of the effect of SMS reminders on mortality per 1000 population	96
A.11	Event study: ATE estimates of the effect of SMS reminders on mortality per 1000 population	97
A.12	Event study: ITT estimates of the effect of SMS reminders on visits (TWFE model)	98
A.13	Difference-in-differences estimates of the effect of SMS reminders on visits (TWFE model)	99
B.1	Inclusion criteria for primary care electronic health records data	101
B.2	Chile's clinical practice guidelines: Frequent errors that result in an inaccurate measurement of blood pressure	101
B.3	Construction of hospitalization variables	104
B.4	Sample of primary care 3, 6, and 12 months after initial visit	104
B.5	Impact of discretion on patients' retention in the data	105
B.6	Discretion and all-cause hospitalization 3, 6, and 12 months after primary care visit among retained patients	106
B.7	Discretion and cardiovascular hospitalization 3, 6, and 12 months after primary care visit among retained Patients	107
B.8	Selection by patient characteristics - health history questions	108
B.9	Selection by patient age	109

Acknowledgments

I am grateful to my dissertation chair and cheerleader from day one, Paul Gertler. Without you I would not be here. Thank you to my dissertation committee Jon Kolstad, Ziad Obermeyer, and Ulrike Malmendier for the invaluable insights, feedback, and great conversations. Thank you to my co-authors and mentors Tadeja Gracner and Pablo Celhay for taking a chance on me five years ago. Thank you to my academic family, Nicole Perales, Chris Lowenstein, Nick Otis, Calvin Chiu, and Austin Land. The lower part of the bimodal distribution would have been lonely without you. Thank you to Dominique Glassman for the decade+ of support. Thank you to Hannah, Erica, and Charlie Boone for teaching me how to be more creative, for the brainwashing, and most importantly, for the genes. Last, merci à Phil for sticking with me from math camp through the job market and distracting me from work just the right amount.

Chapter 1

Introduction

Medicine and medical technology are rapidly advancing. The past decade has brought us breakthroughs in gene editing technology, mRNA vaccines, and the proliferation of wearable devices. In parallel, many existing treatments are becoming easier to access. It is now possible to, for example, receive cognitive behavioral therapy to treat mental illness on a smartphone application, to speak with a primary care provider over video, and to have medications delivered right to your door. On the health care provider side, clinicians have access to more data than ever about individual patients, helping them to form more accurate predictions and conduct precision medicine.

New and existing medicine most benefits society if it is used appropriately, which often not the case. Misuse is common throughout health care: providers frequently do not follow clinical practice guidelines, and patients often do not adhere to their prescribed treatment (Finkelstein et al., 2021; Osterberg & Blaschke, 2005). These behaviors are widespread and highly consequential: in a given year in the United States alone, it is estimated that diagnostic errors lead to 40,000 to 80,000 deaths (Graber et al., 2012; Restrepo et al., 2020), and patient non-adherence to medication leads to \$300 billion in preventable health care costs (Osterberg & Blaschke, 2005). For these reasons, the World Health Organization (WHO) has stated that “increasing adherence may have a greater effect on health than improvements in specific medical therapy” (Burkhart & Sabaté, 2003).

In this dissertation I examine the question of medical adherence. First from the demand side in the context of treatment adherence among patients with chronic conditions, and second from the supply side in the context of health care providers’ diagnostic decisions and the use of clinical practice guidelines. The appropriate use of medicine and medical knowledge by patients and providers alike is a fundamentally behavioral question, and in both papers I explore how adherence behaviors can be understood, and ultimately improved, using insights from behavioral economics.

Treatment for chronic conditions such as type 2 diabetes and hypertension is more effective than ever: it is well known that both type 2 diabetes and hypertension can be controlled through weight loss, a healthy diet, regular physical activity, and low-cost medications (Clark & Royer, 2013). As a result, patients who get diagnosed with chronic conditions are living

longer, healthier lives than ever: in the past ten years in the United States, cardiovascular-related mortality rates have fallen overall, for patients living with type 2 diabetes (L. Chen et al., 2020), with hypertension (Rethy et al., 2020), and among those with a previous acute myocardial infarction (Wang et al., 2022). These gains are in spite of the dramatic rise in the prevalence of obesity, poor diet, and sedentary behaviors in both high and middle-income countries alike (Malik et al., 2020; Malik et al., 2013).

Still, more patients could be benefiting from these advances through improving adherence to treatment. As many as 80% of patients treated for hypertension or high cholesterol do not take their medication as prescribed, and more than half fail at maintaining behaviors like exercise, not smoking and eating a healthy diet (Krass et al., 2015; Lee et al., 2006). Similarly, an estimated 45% of patients with type 2 diabetes fail to achieve glycemic control (Ali et al., 2012; Brown & Bussell, 2011; Ford, 2011; Vonbank et al., 2013). This lack of adherence is costly: it has been linked to nearly 25% of avoidable hospitalizations, an estimated 125,000 premature deaths and up to \$300 billion in preventable health care costs annually in the United States alone (Desai et al., 2019; Osterberg & Blaschke, 2005).

Why is there a disconnect between effective, available treatment and its use by patients? While treatment for hypertension and type 2 diabetes is often simple from a medical standpoint, adherence and behavior change require substantial effort to both initiate and maintain (Oster, 2018). Other barriers to behavior change include financial and time costs, as well as psychological barriers such as inertia, inattention, present bias, self-control issues, and a lack of salience (Della Vigna & Malmendier, 2006; DellaVigna, 2009; Gabaix, 2019; Kessler & Zhang, 2014; Roberto & Kawachi, 2015).

Medication for chronic conditions can be costly, either with or without insurance. In the United States, where health insurance coverage is highly variable, patients are price elastic with respect to medications, even when those medications are necessary for their survival (S.-Y. Chen et al., 2014; Dusetzina et al., 2022; Einav et al., 2018; Einav et al., 2015). Other research shows that patients cut back on both high and low value care when costs increase (Baicker et al., 2015).

Even if a patient can afford necessary medications, adhering to medication and other treatment is challenging for many. Unlike acute conditions where treatment lasts a short duration, chronic diseases must be actively managed over many years, and for some patients, for the remainder of their lifetime. Because of this long time horizon and the incremental nature of health improvements, management of chronic conditions involves a series of inter-temporal choice problems where the patient trades off current pleasures, such as eating junk food, or not taking a medication with painful side effects, against the future costs and benefits of these behaviors (Della Vigna & Malmendier, 2006; Roberto & Kawachi, 2015). Here, a patient who suffers from present bias is less likely to pay the cost in the present day, such as taking a medication, to achieve long term health.

The repetitive and long term nature of chronic conditions also means their treatment is unlikely to be salient to patients. When a preventative care appointment is scheduled for every three months, visiting the doctor is no longer salient. In addition, for many conditions, once a patient has achieved control, that is their blood pressure or blood glucose

has returned to normal levels through treatment, their condition is no longer salient. This commonly happens in the case of blood pressure, which is referred to as the ‘silent killer’ as it has no symptoms (WHO, 2013).

For these reasons, encouraging individuals to adopt and sustain healthy lifestyles and to adhere to medical treatment is a major challenge in public health and health policy, and one that lies at the intersection of health and behavioral economics. In chapter 2, written with coauthors Pablo Celhay, Paul Gertler, and Tadeja Gracner, we evaluate the impact of sending text-message (SMS) reminders to patients diagnosed with hypertension and/or type 2 diabetes. These reminders work through increasing the salience of the patients’ upcoming medical appointment, while also providing information about the date, time and location, as forgetfulness is a major contributor to no-shows (Dantas et al., 2018; Free et al., 2013; Geraghty et al., 2008; Lacy et al., 2004; Murdock et al., 2002). Making the preventative care appointment top of mind could have spillovers into other health behaviors, such as taking medication, exercising, or eating a healthy diet. We study this question in the setting of Chile’s public health care system, which offers three key advantages. First, SMS reminders were scaled-up nationally, allowing us to measure their effectiveness using first intent to treat models and second models that incorporate clinic-level compliance with the intervention. Second, Chile’s public healthcare system is used by 80% of the population, enabling us to study the impact of additional primary care visits on downstream health behaviors and outcomes. Because we observe such outcomes for 364,426 patients, followed on average for 3.5 years, a third advantage is that we have statistical power to detect small effects on medication adherence, hospitalization and mortality, magnitudes we would expect from a light touch nudge such as this one.

We find that nudging patients to attend their preventative care appointments increases visits by 5 to 7%, results that are about half as large as published smaller-scale or efficacy trials. Our results increase to 9 to 11% after taking clinic-level compliance into account, where on average 53% of eligible patients were sent reminders. This increase in primary care utilization translates into an 11% increase in health screening. Reminders also lead to a 42% improvement in anti-hypertensive medication adherence, which is largely attributed to patients having any active prescription. Last, SMS reminders lead to a 17% increase in cardiovascular hospitalizations coupled with a reduction in in-hospital mortality, suggesting an increase in referrals through primary care or timely care-seeking behavior of emergency and specialized health services.

Adherence is imperfect not only on the demand side of healthcare, but also on the supply side: widespread non-adherence to clinical practice guidelines has been documented in high- and middle-income countries (Fischer et al., 2016; Jeffery et al., 2015; Murad, 2017; Shanbhag et al., 2018). At their core, clinical practice guidelines are rule-based decision aids for determining whether a patient should be diagnosed and/or treated for a particular condition, given their current state of health. These guidelines aim to standardize medicine using evidence-based recommendations, yet large inefficiencies in care delivery remain (Abaluck et al., 2020; Abaluck et al., 2016; Chandra & Staiger, 2014; Mullainathan & Obermeyer, 2021). Given the large non-adherence rates found in medicine, a natural question is *how* are

providers making these decisions?

I explore three categories of non-medical factors that could be influencing medical decision-making when providers deviate from clinical practice guidelines: patient characteristics, behavioral biases, and discretion. First, ample research has shown that patient characteristics such as race/ethnicity, gender, and weight can influence medical decision making (Borkhoff et al., 2009; Greenwood et al., 2018; Mehok et al., 2019). In many cases, these decisions contribute to healthcare disparities, where females and non-white patients receive a lower volume or quality of care (Alsan, 2015; Borkhoff et al., 2009; Obermeyer et al., 2019). Similarly, patient-provider characteristic match can also influence medical decision-making (Alsan, 2015; Sandhu et al., 2009).

Second, like all humans, healthcare providers sometimes rely on heuristics when making decisions. Heuristics are mental shortcuts or rules of thumb that can help us to make complex decisions quickly (Tversky & Kahneman, 1974). For example, Singh, 2021 shows that physicians who recently experienced an adverse event in the delivery room are subsequently more likely to switch delivery methods; a recency heuristic that increases adverse outcomes for patients. In other cases, heuristic thinking and patient characteristics may be correlated with health or risk. Olenski et al., 2020 find that physicians perform significantly more coronary-artery bypass grafting on older patients, which is logical on the surface as risk is positively correlated with age. However, the underlying cause of this behavior is a bias: clinicians perceive patients who are 80.0 to be substantially older, and therefore higher risk than patients who are 79.9 years old leading to a discontinuity in testing rates where there is no discontinuity in risk (Olenski et al., 2020). Similarly, Mullainathan and Obermeyer, 2021 show that physicians' testing behavior changes based on whether a patient's symptoms are representative of a stereotypical heart attack.

Last, providers may be using their discretion when deviating from clinical practice guidelines, that is, relying on their own judgement and expertise to make decisions instead of prescriptive rules (Abaluck et al., 2020). Indeed, critiques of guidelines, which often come from physicians, include that diagnostic rules are written for the average patient and not necessarily the individual they are treating, guidelines may be too simple and ignore some clinically relevant information, and they do not allow for role of expertise (Costantini et al., 1999; Knaapen, 2014; Timmermans & Mauck, 2005). Clinical decisions made using discretion instead of guidelines is another form of non-adherence. In this case, however, nonadherence to guidelines is an active choice by the healthcare provider, as opposed to a passive decision or one influenced by factors such as fatigue, a busy schedule, or behavioral biases. For these reasons it is possible that this type of non-adherence to clinical practice guidelines could lead to improvements in the quality of medical decisions or patient health.

Discretionary decision-making is likely important in medicine's many areas of ambiguity, where the correct clinical decision is not obvious. These ambiguous areas might arise due to sparse or incomplete clinical practice guidelines, or uncertain and incomplete information about patient risk. In chapter 3, I study how health care providers behave when diagnosing and treating hypertension, defined as blood pressure $\geq 140/90$ mmHg. Because blood pressure is an important but noisy measure of risk, this is one example of the many decisions

providers must make with poor or limited information.

In chapter 3 using bunching estimation and electronic health records from over 600,000 patients I find that providers round some patients' test results below the diagnostic threshold, using their discretion to turn a positive test for hypertension into a negative. This behavioral response varies at the primary care clinic level, and can be large: up to 62% of patients expected just above the diagnostic threshold are instead recorded just below it. This behavior leads to a more accurate sorting of patients with respect to the threshold: 9% and 45% fewer hospitalizations for heart attack and stroke are seen in patients moved below the threshold at high discretion clinics. I explore heuristic thinking as one way providers might be making this decision. I find that among patients with identical test results, those with characteristics representative of high cardiovascular risk are less likely to be sorted below the diagnostic threshold. These results suggest that in the case of limited information, heuristic thinking and discretionary diagnosis can lead to more accurate decisions and to better patient outcomes.

Like all humans, patients and providers make mistakes in medical and healthcare settings. While medicine is rapidly advancing, a key challenge in health policy is promoting adherence to medical treatment, on the side of patients, and adherence to clinical practice guidelines on the side of providers. This challenge is behavioral in nature, and in the remainder of this dissertation I show how insights from behavioral economics can improve our understanding decision-making and treatment adherence in medicine.

Chapter 2

Nudging Patients at Scale: Evidence from Text-message Appointment Reminders

ABSTRACT

We study how reminding high-risk patients of their upcoming preventative, primary care appointments impacts health behaviors. We use a natural experiment in Chile, where text-message appointment reminders were scaled up nation-wide over three years at public, primary care clinics. Our data come from national medical, medication, hospitalization, and mortality records and include over 300,000 patients, all recently diagnosed with type 2 diabetes and/or hypertension. We find that nudging patients to attend their preventative care appointments increases visits by 5 to 7%, results that are about half as large as published smaller-scale or efficacy trials. Our results increase to 9 to 11% after taking clinic-level compliance into account, where on average 53% of eligible patients were sent text-message reminders. This increase in primary care utilization translates into an 11% increase in health screening. Reminders also lead to a 42% improvement in anti-hypertensive medication adherence, which is largely attributed to patients having any active prescription. Last, SMS reminders lead to a 17% increase in cardiovascular hospitalizations coupled with a reduction in in-hospital mortality, suggesting an increase in referrals through primary care or timely care-seeking behavior of emergency and specialized health services. Our findings are particularly important for settings with a gate-keeping healthcare model, which is common in OECD countries. In this type of system, patients must first visit their primary care provider before being approved for tests, prescribed medication, or referred to speciality and hospital care. This paper shows that through intervening at the first step in the cascade of care, even a light touch intervention such as text-message reminders can have large and meaningful downstream impacts.¹

¹This chapter presents results from an analysis conducted in collaboration with Paul Gertler, Pablo Celhay, and Tadeja Gracner. We gratefully acknowledge financial support from Center on the Economics

2.1 Introduction

Chronic conditions such as hypertension and type 2 diabetes are among the highest contributors to both excess mortality and health care spending in high- and middle-income countries (Egan et al., 2019; Piper et al., 2015). Health monitoring through screenings, and treatment through medication and behavior change can improve the control of chronic conditions. However, many patients do not receive this care because they do not attend medical appointments: half of all patients miss their scheduled appointments (Ciechanowski et al., 2006; Hardy et al., 2001; Karter et al., 2004; Schechtman et al., 2008), and between 12% and 50% drop out of care in their first year of treatment (Gucciardi, 2008; Yamaguchi et al., 2017). Patients failing to keep just one primary care appointment are 70% more likely not to return to medical care within the next 18 months (Fullerton et al., 2012; Gucciardi, 2008). Reducing missed appointments is therefore a promising strategy to improve the health of patients living with chronic conditions.

Patients’ reasons for missing appointments include behavioral biases such as inattention, present bias, self-control issues, and a lack of salience (Della Vigna & Malmendier, 2006; Della Vigna, 2009; Gabaix, 2019; Kessler & Zhang, 2014; Roberto & Kawachi, 2015). For these reasons, light-touch nudges such as appointment reminders sent by text message (or short message service, SMS) are a promising strategy to reduce no-shows and promote timely preventative care visits (da Costa et al., 2010; Ellanti et al., 2011; Hamine et al., 2015; Jongh et al., 2012; Leong et al., 2006; Liew et al., 2009). Existing studies largely consist of small to medium sized controlled clinical trials with limited to no integration with existing health information systems (WHO Global Observatory for eHealth, 2011), and pilots with short follow-ups (e.g. Boksmati et al., 2016; da Costa et al., 2010; Hallsworth et al., 2015; Hamine et al., 2015). While this type of intervention is frequently described as “scalable”, to our knowledge the effectiveness of SMS appointment reminders at scale is unknown (Tomlinson et al., 2013).

We study the impact of sending SMS appointment reminders on both preventative care attendance, and on downstream outcomes such as medication prescriptions and adherence, hospitalization, and mortality. We study this question among high risk, high-cost patients: those recently diagnosed with type 2 diabetes and/or hypertension (defined here as blood pressure $\geq 140/90$ mmHg). The setting of this study is Chile’s public healthcare system, which is used by over 15 million individuals, or 80% of the population (FONASA, 2018).

We leverage the variation in implementation of SMS reminders across primary care clinics to measure the causal effect of nudges using a generalized difference-in-differences approach. For each outcome we use an event study design to compare pre-implementation trends between treated and control clinics, and then estimate the impact of reminders using first an

and Demography of Aging - NIH 2P30AG012839. This project does not meet the threshold definition of “human subjects” research set forth in Federal Regulations at 45 CFR 46.102(f) as it only includes publicly available data, and coded private information data that was not collected for the currently proposed project, and the investigators cannot readily ascertain the identity of the individuals to whom the coded private information pertains.

intent to treat (ITT) approach that assumes all patient-visits after a clinic implemented SMS reminders were sent an SMS. Second, we recover the average treatment effect (ATE) by accounting for variation in clinic-level compliance over time. We do this using national data from 364,426 patients at 315 primary care clinics from the point of each patient’s diagnosis with hypertension (328,037 patients), or type 2 diabetes (75,899 patients), onward. Patients are observed for on average 3.5 years after their diagnosis. We link these data at the individual level to three datasets, each collected over 2013-2018: electronic health records (EHR) that include information on visits and screenings received in primary care (2,608,221 patient-visits), prescription and refill data from pharmacies nationwide, and the universe of hospitalizations. Last we link these data at the county level to county-year mortality estimates, constructed using and the universe of deaths in Chile from 2012-2020.

First, we show that SMS reminders - a low-cost, light-touch nudge - can increase health seeking behavior among high-risk patients when implemented at scale by a national health-care system. In this setting SMS reminders increase the probability that patients with type 2 diabetes visit primary care by 5.4%, and the probability that patients with hypertension visit primary care by 6.7%, per semester. We find no change in screenings among patients with type 2 diabetes, but patients with hypertension sent reminders were 7.6% more likely to have their blood pressure measured.

SMS reminders cause plausibly exogenous variation in the use of primary healthcare, allowing us to also measure how nudges change downstream health behaviors. We find that SMS reminders improve medication adherence: hypertensive patients who receive reminders were 22% more likely to have adequate medication adherence (measured by proportion of days covered of at least 80%), or 42% after accounting for compliance. Patients with type 2 diabetes were 28.6% more likely to ever visit the pharmacy in a given semester, but we found no impact on adequate medication adherence for this group.

Last, we study the impact of SMS reminders on hospitalization and mortality. Among patients sent SMS reminders, we observe an additional 1.4 cardiovascular-related hospitalizations per 1000 population, a relative increase of 12% increase. At the same time, we find a modest reduction in in-hospital mortality, and no change in cardiovascular related mortality attributable to SMS reminders.

This paper makes three contributions: first, to our understanding of the effectiveness of sending SMS appointment reminders to patients diagnosed with chronic conditions, and second, to the differences between such programs when they are implemented in controlled, experimental environments compared to within national health systems. Third, this study contributes to our understanding of the impact of primary healthcare utilization on patient behaviors and health.

Our estimates of the impact of SMS reminders on primary care visits are smaller than many in this literature: Boksmati et al., 2016 systematically review SMS appointment reminder trials and find changes of 4 - 123%, and their meta-analysis estimate from randomized controlled trials (RCTs) is approximately 20%. This is likely due to two reasons: first, previous work largely comes from small-scale trials, often at one medical center or clinic (Boksmati et al., 2016; Ellanti et al., 2011; Hamine et al., 2015; Jongh et al., 2012; Leong et al., 2006;

Liew et al., 2009). Under these controlled conditions, implementation fidelity is likely high yielding an estimate of efficacy, rather than effectiveness. The discrepancy in effect size between our at-scale results and the broader literature on SMS reminders is in line with findings from DellaVigna and Linos, 2022, who compare the effects of various nudge interventions when implemented at scale by governments to results found in the academic literature. The authors report an average effect of 1.4 percentage points (pp) among government-run and at-scale nudges, compared to 8.7pp in published literature. DellaVigna and Linos, 2022 also document substantial publication bias in the nudging literature, where large and statistically significant effects are reported, and null effects are unlikely to be published.

Using phone records we provide a unique look into the implementation fidelity of this program at the clinic level and over time. We find that on average 53% of eligible patients at treated clinics were sent SMS reminders. Compliance increased over time, from a low of 46% in 2016, to 56% in 2018. After taking clinic-semester level compliance into account, we estimate that SMS reminders increased the probability of a visit by 8.5% for patients with type 2 diabetes, or 10.8% for patients with hypertension. This pattern holds for tests, medication adherence, and hospitalizations: estimated effect sizes are over a third larger after accounting for compliance. This paper therefore contributes to our understanding of how programs operate at scale, where variation in compliance fidelity may occur (Heckman et al., 2010; King et al., 2009). Understanding whether the magnitude of programs and policies will differ from existing literature after they are scaled is important, especially as governments increasingly use experimentation in policy-making².

Last, these findings contribute to our understanding of the impact of primary care utilization on hospital admissions and other healthcare spending. Identifying the causal impact of healthcare utilization on health and behaviors is challenging because healthcare seeking behavior is usually endogenous to past utilization or diagnoses (Levy & Meltzer, 2008). Here, we solve this endogeneity problem by leveraging plausibly exogenous variation in primary care visit behavior caused by SMS reminders which, importantly, allows us to also examine downstream outcomes. Other research has solved the endogeneity problem by examining the impact of diagnosis using regression discontinuity design (Alalouf et al., 2019; Moscoe et al., 2015; O’Keeffe et al., 2014; Venkataramani et al., 2016) or through leveraging an exogenous change in insurance availability or copayments, as these often lead to an increase in primary care utilization (Adams et al., 2021; Aron-Dine et al., 2013; Baicker et al., 2013; Card et al., 2008, 2009; Finkelstein et al., 2012; King et al., 2009; Taubman et al., 2014).

While health insurance is an important financial tool to avoid catastrophic spending, when it comes to its effect on health and mortality results from experimental and quasi-experimental studies are mixed (Aron-Dine et al., 2013; Finkelstein et al., 2012; Levy & Meltzer, 2008; Manning, 1987; Taubman et al., 2014). A major reason is that in these settings, health is challenging to measure. Like our study, many are unable to provide an

²Currently, there are more than 200 government nudge units, a movement that began with the UK’s Behavioral Insights Team launched in 2010, in the United States the Office of Evaluation Sciences launched in 2015

unbiased estimate of utilization on health as measured by the results of screenings tests (Manning, 1987; Taubman et al., 2014; Zhang et al., 2010), and others find null effects on test results (Baicker et al., 2013; King et al., 2009). Several papers measure the effects of utilization on self-reported health, which has been shown to be a good proxy for true health status, however issues such as Hawthorne effect, poor recall, or conflating mental and physical health still remain (Baicker et al., 2013; Finkelstein et al., 2012). These same studies have all either found no effect of increased utilization on mortality or are under-powered because death is a rare event. A notable exception here is Card et al., 2009, who estimate that Medicare eligibility increases health care utilization and leads to a 1 percentage point, or 20% reduction in 7-day mortality.

The direction and magnitude of our results are similar to these studies. Alalouf et al., 2019 use regression discontinuity design to understand the impact of a new diagnosis with type 2 diabetes on patient health care spending. They find that while health spending increases by over USD \$1000 the year following diagnosis, there is no measurable change in patients' health. In addition, new diagnosis with related diseases such as diabetic retinopathy, neuropathy, and kidney disease each increase after diagnosis with type 2 diabetes due to more intensive screening (Alalouf et al., 2019).

An increase in hospitalization after an increase in primary care utilization is also found in the Oregon health insurance experiment (Baicker et al., 2013; Finkelstein et al., 2012; Taubman et al., 2014). In Oregon, low-income residents became eligible for Medicare through a lottery, which led to a 37% increase in outpatient care, a 15% increase in blood sugar testing, and a subsequent 40% increase in hospitalizations (Finkelstein et al., 2012; Taubman et al., 2014).

The remainder of the paper is as follows: in section 2 we describe health and healthcare in Chile and the SMS reminder program we evaluate, in section 3 we describe our data sources, in section 4 we outline the empirical approach, section 5 describes results, and section 6 concludes.

2.2 Setting

We study the impact of appointment reminders, sent by text message (SMS) to patients recently diagnosed with chronic conditions at public, primary care clinics in Chile. This setting offers several key advantages. In Chile, public healthcare is low cost and guaranteed for all residents. Primary care is received at local clinics or low-complexity hospitals (if the community only has one health center), centrally organized by the Ministry of Health. Going forward, we refer to all health facilities where a patient might seek primary care as 'clinics'. The public health care system does not operate like a market, and instead patients using public health insurance are assigned to one primary care clinic based on their home address. This makes contamination between treated and control clinics unlikely. Chile's national health care system allows us to link data at the individual level country-wide from primary care, hospitals, and pharmacies. This analysis therefore includes a large cross section of

patients in terms of location, socio-economic status, wealth, and health from a population with few access to care issues.

2.2.1 Health and Healthcare in Chile

Chile has two health care systems: the public system, used by nearly 80% of the population that is funded by a 7% tax on earnings, and the private system, used by the remainder of the population (FONASA, 2018; Goic, 2015). The public system, the setting of this study, is called Fondo Nacional de Salud or FONASA. It is organized in a gatekeeping model where patients are required to visit a general practitioner at a primary care clinic before receiving prescriptions or referrals to specialists or more advanced care units (Brekke et al., 2007; Rotar et al., 2018).

Patients covered by FONASA are administratively assigned exclusively to use on one clinic based on their place of residence and are not allowed to obtain primary care at any other public clinic. Therefore, while the public healthcare system is characterized by excess demand, seeking better access to care by trying another clinic is unlikely (MINSAL, 2021). Similarly, it is unlikely that patients will seek care at private clinics because their out of pocket costs are high, and public insurance is unlikely to cover care at private facilities.

Chile's burden of chronic, non-communicable diseases is high: approximately 57% of the population is living with at least one chronic condition (Margozzini & Passi, 2018). These patients consume 84% of healthcare resources (MINSAL, 2008) and contribute to the high national rate of missed appointments: in 2019, 16.7% of scheduled appointments were missed (Boone et al., 2022). High utilization is partly because patients with chronic conditions are prioritized for care: in 2005 Chile passed a healthcare reform that aimed to reduce wait times and guarantee access to primary care for patients with 85 priority conditions, including hypertension and type 2 diabetes (FONASA, 2018; Martinez et al., 2019; MINSAL, 2017a; Vargas & Poblete, 2008). For these patients, primary care appointments are scheduled in advance, while for others, appointments are on a first come, first served basis. Patients living with chronic conditions receive primary care at no cost, and co-payments for their medications are between 0% and 20% of cost, depending on their household income³ (A. Aguilera et al., 2015).

The share of patients with controlled hypertension and/or type 2 diabetes has risen since the 2005 health reform, however there is still room for improvement. Chile's 2016-2017 National Health Survey found that 27.6% of the adult population had hypertension, defined as blood pressure $\geq 140/90$ mmHg, and 9.5% had type 2 diabetes, both similar prevalences to other high income countries such as the United States (Lanas et al., 2020; Ostchega & Nguyen, 2020). Among Chileans with hypertension, 68.7% were aware of their condition,

³For 2020, the FONASA copayment groups (in 2020 US Dollars) were: group A, extreme poverty and/or unhoused, group B monthly income $< \$320$ or on a government pension, group C, monthly income $\$320$ - 465 , and group D monthly income $> \$465$ (de Salud - Gobierno de Chile, 2020). In 2020 Chile's mean and median monthly income per capita were USD\$784.45 and USD\$518.74 respectively (de Estadísticas, 2020).

and 33.3% had controlled blood pressure (blood pressure < 140/90 mmHg) (Lanas et al., 2020).

2.2.2 SMS Reminder Program

In response to low levels of disease control, and high rates of missed appointments, beginning in January 2015, the Chilean Ministry of Health offered the option for public, primary care clinics to opt in to an SMS reminder program. Clinics were eligible for this program if they had an electronic health records system. Within clinics that took up the program, patients were eligible for SMS reminders if they had been diagnosed with type 2 diabetes and/or hypertension (MINSAL, 2017b).

The program, which is a software integrated into the clinic’s electronic medical record, automatically send a text-message, or SMS, to patients 24-48 hours before their appointment:

“Dear [Patient Name], this is a reminder that you have a medical appointment on the day [Date of appointment] at [Time] hours at [Facility Name] with the doctor [Name of the doctor]. Do you confirm your time? Yes/No”

Patients’ appointment slot was kept if they responded yes, or if they did not respond at all.

2.3 Data

To evaluate the impact of SMS reminders we use two datasets that measure the policy implementation, and four datasets that contain patient behaviors and health outcomes. We include patients newly diagnosed with type 2 diabetes and/or hypertension, and measure the effect of SMS reminders on all subsequent visits to primary care, medication prescriptions and pick up, hospitalizations, and mortality.

2.3.1 Dates of SMS Reminder Implementation

We obtained a list of public primary care clinics that had implemented the SMS reminder program from the Chilean Ministry of Health, together the with date of first implementation. Clinics were eligible for the program if they had an electronic health record system. These data were available for 2013-2019. The SMS reminder program was first offered in 2015, and 172 clinics implemented the program that year.

Out of 877 public primary care clinics that were eligible for the program, 757 (86%) were in non-extreme geographic locations⁴, and had a non zero number of patients diagnosed with chronic conditions (Boone et al., 2022). Among these clinics, 435 (57%) used the

⁴We exclude primary care clinics in De Arica Parinacota, Aisén del General Carlos Ibáñez del Campo, the Chilean Antarctic, and Rapa Nui (Easter Island).

electronic health record provider from which we could access health record data. Finally, in this analysis we include only treated clinics for which phone record data was available, as well as all control clinics, resulting in a sample of 315 clinics (73%). These 315 clinics represent 275 counties; 79% of all counties in Chile.

By the end of our study period, in 2018, 208 clinics had implemented the program, and 107 clinics remained as controls. Table 2.1 describes clinic-level program implementation. In this setting patients are assigned a primary care clinic based on their home address, allowing us to assign exposure to SMS reminders to patients using their home clinic.

We merged the clinic data with socio-economic controls, such as mean age, sex, income per capita, share of rural population, and share of population below the poverty line at municipality-level from the National Socioeconomic Survey (MDS, 2013).

2.3.2 Phone Records

At clinics that took up the SMS reminder program, all patients diagnosed with type 2 diabetes and/or hypertension were eligible to receive SMS reminders about their upcoming primary care visits. However, as this program was implemented at-scale nationwide, we hypothesized there would be differences in implementation fidelity, or compliance. We measure clinic-level compliance with the intervention using phone records, collected in 2016, 2017, and 2018. Many clinics first implemented the program in 2015, however phone records are not available for this year. These data contain a variable indicating whether an SMS was sent to an individual with an appointment. We construct a variable at the clinic-semester level that measures the share of patients who were sent an SMS reminder, among all patients who were eligible for a reminder. In our analyses we assign the clinic's 2016 semester 1 compliance level to 2015.

2.3.3 Electronic Health Records

To evaluate the impact of SMS reminders on patient health seeking behaviors and disease management, we use electronic health records (EHR) provided by the Division of Primary Care at Chile's Ministry of Health. The EHR dataset contains all visits from patients attending clinics where Chile's most popular EHR software is used. These data contain information from each patient-visit to primary care, from January 1 2013 - December 31 2018. In particular, these data only contain visit information from patients once they have been diagnosed with type 2 diabetes and/or hypertension.

SMS reminders are only sent to patients diagnosed with type 2 diabetes and/or hypertension, so we limit our analysis to visits from patients newly diagnosed with one of these conditions. Our decision is based on the literature on visit adherence and medication adherence, where these outcomes are defined among patients with incident disease only; mostly to address concerns related to left censoring (Dusetzina et al., 2022; Hernán, 2010; Schisterman et al., 2013; Schneeweiss, 2010). Because we do not observe date of diagnosis for all patients, we examine the distribution of number of days between a patients' first and second visit,

among patients who were in the dataset on January 1 2013. We found that 100% of patients who were in the data on January 1 2013 and who had any subsequent visit had presented for their second visit within 361 days of January 1 2013. Therefore, if a patient appears in the data 362 days after January 1 2013, they are very likely to be a new patient, who is presenting at their first post-diagnosis visit. For this reason, we exclude all patients who had a visit that occurred before January 1 2014. Second, we also exclude visits to providers that are not necessary for management of type 2 diabetes and/or hypertension, such as psychologists.

2.3.4 Medication Data

To understand whether SMS reminders improve medication adherence we use administrative records from pharmacies during 2013-2019. These records contain information on pharmacy name and unique identifier, prescription date, prescribed medication name, number of units, active ingredient of the drug, and date of medication withdrawal. Each prescription contains a patient ID that can be linked to the EHR dataset. That is, all patients in our EHR dataset who have been prescribed a medication for their chronic condition(s) also appear in this pharmacy dataset: during 2013-2019, that translates into 276,964 patients (76% of total). These data are automatically generated at each pharmacy and stored in an electronic record, allowing us to calculate proxies of medication adherence based on pharmacy refill data such as proportion of days covered (PDC).

2.3.5 Hospitalization Records

We also examine the impact of SMS reminders on hospitalization, using the universe of hospital admissions in Chile from 2013-2019. Importantly, these records include both public and private hospitals nationwide, and can be linked to both the EHR and medication datasets with the patient identifier. These data also contain hospital name and its unique identifier, as well as date of admission or hospital visit, length of stay in hospital, ICD-10 diagnoses recording primary and secondary cause of admission, and death in the hospital.

2.3.6 Mortality Records

We use mortality records to understand the impact of SMS reminders on mortality and cause of death. We use the universe of deaths in Chile from 2012-2020, which include the individual's sex, date of birth, date of death, and ICD-10 code for primary cause of death. We are not able to match these data at the individual level to the dataset including EHR records, medication history, and hospitalizations because a different type of ID encryption was used. Instead, we first match them, at the individual level, to annual records containing universe of residents in Chile from 2012-2020. The residents records contain, importantly, the date an individual was diagnosed with key diseases, including hypertension and/or type 2 diabetes, as well as date of birth, type of insurance (public vs. private), and county of residence. In our analyses we include patient-years after an individual was diagnosed with

type 2 diabetes and/or hypertension, from patients using public health insurance. We are not able to obtain information on the primary care clinic these patients are assigned to. Instead, we assign treatment status at the county level, and include only counties where all or no primary clinics took up the program ($N=275/347$). If a county contains clinics with different implementation years we use the earliest year of program implementation.

2.3.7 Descriptive Statistics

A total of 315 primary care clinics were analyzed; 208 of them implemented the program between 2015 and 2019 (table 2.1), and 107 clinics never implemented the program. At both ever and never treated clinics, 41% of patients were male, and mean age was 62 years (table 2.3). At treated clinics 91% and 20% of SMS eligible patients had been diagnosed with hypertension and type 2 diabetes, respectively, compared to 89% and 23% at control clinics (table 2.3). The health of patients at their time of diagnosis was similar across treated and control clinics, as measured by blood pressure, hemoglobin A1c, blood glucose, and body mass index (BMI) (table 2.3). Patients at treated clinics were observed for 0.2 semesters (approximately 35 days) longer than patients at control clinics, which may have been impacted by SMS reminders. The probability of a medication prescription at the time of diagnosis was similar across treated and control clinics, as was the probability of key tests, indicating that clinics with similar qualities of care did and did not take up the program (table 2.3).

Using mortality records data a total of 275 counties were analyzed: 97 counties where all public, primary care clinics eventually took up the program by 2020, and 178 counties where no clinics took up the program (table 2.2). At both ever and never treated counties, 44% of patients were male, mean age was 54 years, and the average age of hypertension and type 2 diabetes diagnosis was 55 and 53 years, respectively (table 2.4). At treated counties 78% and 49% of SMS eligible patients had been diagnosed with hypertension and type 2 diabetes, respectively, compared to 80% and 45% at control counties (table 2.4). The probability and number of existing comorbidities at the time of diagnosis with type 2 diabetes and/or hypertension was also similar across county type (table 2.4). Treated and control counties were also similar on socioeconomic characteristics including population, mean income per capita, employment rate, and education (table 2.5).

2.4 Empirical Approach

The SMS reminder program was implemented by clinics and counties over several years, allowing us to use a staggered adoption, difference-in-differences approach (DID) to estimate the effects of the program. For each outcome we estimate flexible event study models, which show that the effects of the program increased over time for most outcomes. Because of this heterogeneity over time, the standard two-way fixed effects estimator is likely a biased estimate of the difference-in-differences estimand (Callaway & Sant’Anna, 2020; de

Chaisemartin & D’Haultfoeulle, 2020; Goodman-Bacon, 2018). For this reason, we use Borusyak, 2021’s difference-in-differences estimator for our intent to treat (ITT) models, where the treatment variable is either 0 or 1. For each outcome we also estimate average treatment effect (ATE) models, that account for imperfect compliance at the clinic or county level. Here, the treatment variable is a number between 0 and 1. Currently, there is no DID software available that allows for a continuous independent variable, so we use two-way fixed effects models for ATE models. Equations included here are represent the specification for both the ITT and ATE models, as the same control variables are included in each.

2.4.1 Event Studies

We start by estimating a non-parametric event-study design to study the link between SMS program adoption and change in patient-level outcomes over time.

$$Y_{ijt} = \sum_{\tau=-3}^6 \beta_{\tau} Q_{\tau} + X'_{it} \delta + \lambda_t + \gamma_i + \epsilon_{ijt} \quad (2.1)$$

Y_{ijt} is an outcome for patient i at clinic j in semester t . β_{τ} are coefficients on semester indicators (Q_{τ}) for time relative to SMS program adoption (at $\tau = 0$) at a clinic j . The key coefficients of interest are the β_{τ} ’s that estimate the difference in outcomes at a clinic j at a given τ relative to the omitted category, Q_{-1} . Each model is adjusted for seasonality and common temporary shocks with semester indicators (λ_t) and clinic-level fixed effects (γ_i). We also include a vector of patient-level controls (X'_{it}): all models include fixed effects for semesters since the patient was diagnosed, and hospitalization and mortality models also include male, and 2-year age fixed effects. ϵ_{ijt} is an error term correlated within clinics across time. We calculate robust standard errors, clustered at the level of the intervention, that is, at the clinic level for primary care, medication, and hospitalization outcomes, and county level for mortality (Abadie et al., 2020).

2.4.2 Difference-in-Differences

We then use a parametric, DID approach to obtain the average effect of SMS reminders on patient-level outcomes. We estimate the following regression model:

$$Y_{ijt} = \alpha + \beta SMS_{jt} + X'_{it} \delta + \lambda_t + \gamma_i + \epsilon_{ijt} \quad (2.2)$$

Y_{ijt} is an outcome for patient i at clinic j in semester t . In ITT models, SMS_{jt} is an indicator variable that takes value one for all semesters t in each clinic j after SMS reminders were implemented. In ATE models, SMS_{jt} is a number between 0-1 that measures clinic-semester compliance in all semesters t in each clinic j after SMS reminders were implemented. For clinics that did not implement the SMS reminder program by the end of the last semester of 2016, this variable is always zero. The main coefficient of interest is β ; a DID estimate

which measures the impact of SMS reminders on the outcome of interest. As above, each model was adjusted for seasonality, common temporary shocks with semester indicators (λ_t), and clinic-level fixed effects (γ_i). Each model is adjusted for seasonality and common temporary shocks with semester indicators (λ_t). We also include a vector of patient-level controls (X'_{it}): all models include fixed effects for semesters since the patient was diagnosed, and hospitalization and mortality models also include male, and 2-year age fixed effects. We again calculated robust standard errors, clustered at the clinic or county level.

The goal of this analysis is two-fold: first, to understand whether sending SMS appointment reminders increases the probability that patients living with chronic conditions attend preventative, primary care visits. We create a panel dataset with a row for each patient-semester, and an indicator for if they had a visit in that semester. Second, to understand how these additional visits might improve patient health through screenings, referrals, and improved medication adherence; outcomes that might occur if a patient visits primary care, and do not occur if a patient drops out of care. To avoid introducing bias, we only construct outcomes at the patient-semester level for which we can assign patients a zero if they did not visit primary care. For example, we estimate the impact of SMS reminders on the probability a patient with type 2 diabetes received a blood glucose test; if the patient's EHR data shows they visited and received this test, this outcome is equal to one. If they visited and did not receive a test, or did not visit, this outcome is equal to zero. We are therefore unable to analyze patient health, as measured by the results of such tests, because to do this we would have to condition on the patient having a test, which would introduce bias.

2.5 Results

2.5.1 Program Compliance

We measure compliance at the clinic-semester level using phone records from 2016-2018. Figure 2.1 shows the distribution of compliance across primary care clinics by year. In each year, there is large variation in compliance: among treated clinics between 0 and 100% of eligible patients were sent SMS reminders. Compliance improves over time: average compliance in 2016 was 47%, in 2017 was 55%, and in 2018 was 56% (table 2.1).

2.5.2 Impact of Nudges on Visits and Tests

For each outcome we first report results from a flexible event study model, which allows us to test for pre-trends, and understand whether the effect of SMS reminders changes over time (equation 2.1). We do this using the Borusyak, 2021 estimator (ITT), and also estimate two-way fixed effect models that account for compliance (ATE). Next, we report results from a difference-in-differences analysis of both intent-to-treat models, which assume that all patients with appointments at treated clinics after SMS implementation were sent SMS reminders (equation 2.2). Last, we compare these estimates to the average treatment effect

(ATE) where we replace the treatment indicator with the share of patients actually sent SMS reminders, among eligible patients in each clinic-semester (equation 2.2).

Results from the event study analysis on visits and tests are presented in figure 2.2 (see tables A.2 and A.3 for corresponding regression analyses). The key identifying assumption that allows us to interpret coefficients on indicators for time after program adoption as causal is that, conditional on its adoption and included controls, potential outcomes are uncorrelated with the timing of the program adoption. Although this is not testable, figure 2.2 shows that there are no trends in outcomes in periods leading to the program adoption, yet the differences in outcome patterns change sharply after the event. For all outcomes, we also conduct an F-test to test whether the pre-program coefficients are jointly different from zero. The p -value from this test is included at the bottom of all event study ITT tables (for visits and tests see table A.2).

Figure 2.2, panel A, illustrates an increase in the total number of visits among patients with type 2 diabetes soon after the program implementation, and that number increases in the first two semesters after adoption before stabilizing at approximately 5 percentage points. Figure 2.2, panel B shows a similar pattern for patients with hypertension.

The magnitude of the effect of SMS reminders increases with time since program implementation, a pattern that is also seen in our compliance data (figure 2.1). In event studies E-H we take clinic-level compliance into account, and observe two key findings 1) the magnitude of the coefficients increase up to two-fold in all cases, and 2) the pattern over time is further exaggerated, that is, after two years of program implementation the effect of reminders is much larger than in the first year (figure 2.2).

Table 2.6 describes the average effect of SMS reminders on visits and tests using the difference-in-differences approach. We find that on average, adopting the SMS program increases the probability of a primary care visit in a given semester by 3.5 percentage points (pp) among patients with type 2 diabetes, and 4.3 pp among patients with hypertension ($p < 0.05$, table 2.6 columns 1 and 3 respectively).

Chile's clinical practice guidelines state that after diagnosis with type 2 diabetes, patients should visit primary care every three months, irrespective of their health status (MINSAL, 2017b). In reality, visits are much lower: 64.4% of patients visit each semester (where a semester is six months, or two three-month periods). SMS reminders increase the likelihood of a visit by 5.4%. This result is smaller than what is found in the literature: Boksmati et al., 2016 systematically review SMS appointment reminders and find changes of 4 - 123%, and their meta-analysis estimate from RCTs is approximately 20%. However, none of these studies were of at-scale programs. All were clinical or randomized controlled trials with close monitoring. We argue these trials measure the efficacy of SMS reminders, and here, we measure the effectiveness of an at-scale program, where, as is common in the real world, implementation issues arose and compliance at the clinic level was varied.

For this reason, we also present ATE estimates. After taking time-varying, clinic-level compliance into account, our estimate is closer to the effect of a *patient being sent an SMS reminder*, as opposed to being assigned to a clinic that adopted SMS reminders. In panel B of table A.3 we find that on average, SMS reminders increase the probability of a primary care

visit in a given semester by 5.5 percentage points (pp) among patients with type 2 diabetes (compared to the ITT estimate of 3.5pp), and 6.9 pp among patients with hypertension (compared to the ITT estimate of 4.3) ($p < 0.05$, table 2.6 columns 1 and 3 respectively).

We find that at primary care visits for patients with hypertension, 96% of patients receive a blood pressure test in both treated and control clinics (table 2.3). This pattern is also reflected in the ITT and ATE estimates of the impact of SMS reminders on the probability of a blood pressure test, which are very similar to the impact of reminders on visits: 4.7pp, or 6.7pp if compliance is taken into account (table 2.6, column 4).

For patients with type 2 diabetes, at baseline there is an 80% chance that a blood glucose test was performed at a treatment clinic, and an 85% chance at control clinics (table 2.3). We find no statistically significant effect of SMS reminders on the probability of a blood glucose or hemoglobin A1c test for patients with type 2 diabetes (table 2.6, column 2). Hemoglobin A1c is a long-term measure of glucose in the blood. If a patient visits primary care more frequently due to SMS reminders, it is therefore possible that they would not need their glucose re-measured, which could explain this null result. This is in contrast to blood pressure, which is recommended to be measured at every primary care visit, regardless of patient health status (MINSAL, 2017b).

2.5.3 Impact of Nudges on Medication Refill Behavior

Next, we estimate the effect of SMS reminders on medication prescriptions and adherence. There are three main pathways through which more visits to primary care might improve patient health. First, through regular health screenings, which we showed increased for patients with hypertension, and may have increased for patients with type 2 diabetes. Second, through information and learning: increased interaction with healthcare providers may reinforce the behavior changes needed to improve control of chronic diseases, such as weight loss, physical activity, maintaining a healthy diet, and medication adherence. These interactions might also provide information to patients about when to seek care at a hospital, if needed. Last, visiting primary care enables patients to maintain active prescriptions for anti-hypertensive and anti-diabetic medications.

In this section we test for whether SMS reminders increased medication adherence. Here, we assume that a patient who has been diagnosed with type 2 diabetes and/or hypertension and who is prescribed a medication for that condition should have an active prescription going forward. This allows us to assign a medication adherence value of zero for patients who did not fill any prescriptions in a given semester⁵.

Medication adherence can only be calculated among patients with a prescription. 44.7% of patients with type 2 diabetes, and 47.3% of patients with hypertension were prescribed a medication for their disease at their time of diagnosis. We find no impact of SMS reminders

⁵It is possible for patients to be de-prescribed medication for hypertension or type 2 diabetes. However, this is only done if the patient is regularly visiting primary care, and has made substantial lifestyle changes such as consistently eating a healthy diet and exercising, which research shows few patients do successfully (Oster, 2018).

on this outcome (table 2.7 column 1), indicating that clinicians at clinics that did and did not take up the program prescribe similarly. Not only is there no effect on average, but event studies A and D in figure 2.3 and A and D in figure 2.4 show this estimate is close to zero at every post-implementation time-point.

Given this null effect, we next estimate medication adherence outcomes among patients with any prescription at the time of their diagnosis. Here, we observe large differences between the ITT and ATE models: for patients with type 2 diabetes, there is no statistically significant effect of clinic assignment to SMS reminders on the probability of any medication pick up (table 2.7, column 2, panel A). However, once compliance is taken into account, we find a 7.9pp impact of reminders on pick up, or a relative increase of 28.6%. Similarly, among patients with hypertension, we find a 7.2pp increase in pick up, or a relative increase of 25.9% (table 2.7, column 2, panel D).

In addition to any pick up, which is a relatively low adherence level, we also find that SMS reminders increased the probability that hypertensive patients had adequate medication adherence, defined here as proportion of days covered $\geq 80\%$ (2.7, column 3, panels c and d). We find no statistically significant effect among patients with type 2 diabetes. Notably, the share of patients in the control group with adequate medication adherence is less than 4% for both patient groups. Similarly, in the control group only 28% of patients pick up *any* medication in a given semester, among those who were prescribed at their first visit. A large share of patients, therefore, never pick up their prescription in a given semester, and may not have an active prescription at all. For these patients medication adherence is zero. This coding - which is necessary to estimate unbiased outcomes - likely contributes to the very low adherence and pick up rates seen here. Recent work has shown that medication non-initiation, that is, when a patient never fills their first prescription for a condition and is subsequently likely to never be medicated, is much higher than previously thought. A recent paper found that on average 39.4% of all new prescriptions among Medicare Part D patients were never initiated (Dusetzina et al., 2022). These results were true even for highly effective, necessary medications: non-initiation was 28.3% for cancer drugs, 20.9% for hepatitis C drugs, 54.2% for immune system drugs, and 66.8% for hyper-cholesterolemia drugs (Dusetzina et al., 2022).

Importantly, in Chile's public health care system, patients' medication prescriptions are valid for a short duration; between 1-3 months, and this information is visible in our pharmacy database. This might encourage patients to visit primary care more regularly, to have their prescriptions updated, or, for some patients it might instead lead to reduced medication adherence if their valid refills run out.

2.5.4 Impact of Nudges on Hospitalizations

We have seen that more visits to primary care can increase health screenings, medication fills and adherence. We now examine the last pathway through which more visits to primary care might improve patient health: information and referrals. We hypothesize that through more frequent interaction with clinicians, patients may have more information or a better

understanding of how to manage their chronic condition. While we do not have survey data, we can test this hypothesis using hospitalizations. In particular, given that all patients studied here have at least one chronic condition, we hypothesize that more visits to primary care would lead to *more* hospitalizations, due to increased testing and referrals from their primary care clinician, and patient knowledge of when to seek care during an episode.

Here, to improve statistical power we analyze patients with hypertension and type 2 diabetes together. We separate hospitalizations into those that are cardiovascular-related and non-cardiovascular related, based on ICD-10 codes for primary and secondary diagnosis (for list of ICD-10 codes see table B.3). We expect a change in cardiovascular hospitalizations only, as these are the types of conditions that could be affected by increased information, screenings, or referrals through primary care. In both ITT and ATE models we find that SMS reminders increased the probability that a patient was hospitalized for a cardiovascular-related condition in a given semester. Specifically, at SMS clinics there was an additional 1.4 hospitalizations per 1000 patients, or 2.0 per 1000 patients after taking compliance into account, representing a 12-17% increase (table 2.8, column 1). At the same time, we document a small reduction in the probability that patients died in-hospital for cardiovascular-related reasons (table 2.8, column 2). Together these findings indicate that while patients receiving SMS reminders were more likely to visit the hospital, they were healthier upon arrival, due to referral, better medication adherence, or information from their provider about when to seek care.

This interpretation is further supported by the fact that we find no impact of SMS reminders on non-cardiovascular hospitalization, or in-hospital mortality for reasons not related to cardiovascular disease (table 2.8, columns 3, 4).

2.5.5 Impact of Nudges on Mortality

Last, we examine the impact of SMS reminders on mortality by cause of death: cardiovascular vs. non-cardiovascular (for list of ICD-10 codes see table B.3). Here, as described in section 2.3.6, we begin with the universe of deaths in Chile from 2012-2020, then restrict our analysis to counties where either all or none of the public, primary care clinics implemented the SMS reminder program. We do this because we are not able to assign individuals in this dataset to a primary care clinic because this variable is missing in all pre-years.

We find no statistically significant effect of SMS reminders on cardiovascular or non-cardiovascular mortality. However, we strongly reject the null hypothesis of no pre-trends (see figure 2.6, and table A.10, pre-trend F test p-value 0.017 and 0.000), and so we interpret this finding with caution. We can use these findings to rule out that patients receiving SMS reminders had significantly poorer health - which is one interpretation of a higher hospitalization rate - and a higher mortality rate.

2.6 Conclusion

This study uses a difference-in-differences design to measure the impact of sending SMS appointment reminders on healthcare utilization. We studied this question among high risk, high spending patients: those recently diagnosed with type 2 diabetes and/or hypertension. We do this in the setting of Chile’s public health care system, which offers three key advantages. First, SMS reminders were scaled-up nationally, allowing us to measure the effectiveness of SMS reminders with and without accounting for variation in clinic-level compliance with the intervention. Second, Chile’s public healthcare system is used by 80% of the population, allowing us to examine the impact of additional primary care visits on downstream health behaviors and outcomes. Because we observe such outcomes for 364,426 patients, followed on average for 3.5 years, a third advantage is that we have statistical power to detect small effects on medication adherence, hospitalization and mortality; magnitudes we would expect from a light touch nudge such as this one.

We find that patients who received reminders attended 8.5 to 10.7% more primary care visits. At those additional visits, patients with hypertension received more blood pressure screenings, but the probability of blood glucose screening among patients with type 2 diabetes did not change.

We next study three pathways through which primary care visits might lead to improved health: increased screenings, valid prescriptions and medication adherence, hospital referrals and information. First, we find that at additional visits patients with hypertension received more blood pressure screenings. Second, SMS reminders also improve medication refill behaviors, as measured both by the probability a patient *ever* visited a pharmacy during a given semester, and the probability they had adequate medication adherence (proportion of days covered of at least 80%). Third, we find that SMS reminders increased cardiovascular hospitalizations, which may be due to referrals from primary care visits. Because we also find a reduction in in-hospital mortality, we conclude that patients who received SMS visited a hospital earlier in a disease episode; information they likely learned in primary care.

Finally, we have shown that in a setting with free or low-cost primary care visits and medication, and with SMS reminders, many patients still do not regularly visit primary care or pick up their medications, even after diagnosis with a highly consequential disease such as hypertension or type 2 diabetes. For these patients, a more intensive intervention is likely need.

Our analysis has limitations, the first being that we cannot produce an unbiased estimate of the effect of SMS reminders on intermediate health outcomes, such as blood pressure and hemoglobin A1c. This is because we find that SMS reminders increase the probability of health screening, meaning estimates of the impact of reminders on the results of those screening tests would be biased. This is a common challenge faced by studies with exogenous variation in healthcare utilization (Adams et al., 2021; Fan et al., 2019; King et al., 2009; Manning, 1987; Pan et al., 2016). A second limitation is that we are also unable to definitively provide an estimate of the impact of SMS reminders on mortality due to pre-trends in this outcome, however, we do rule out a large positive effect on mortality in the compliance

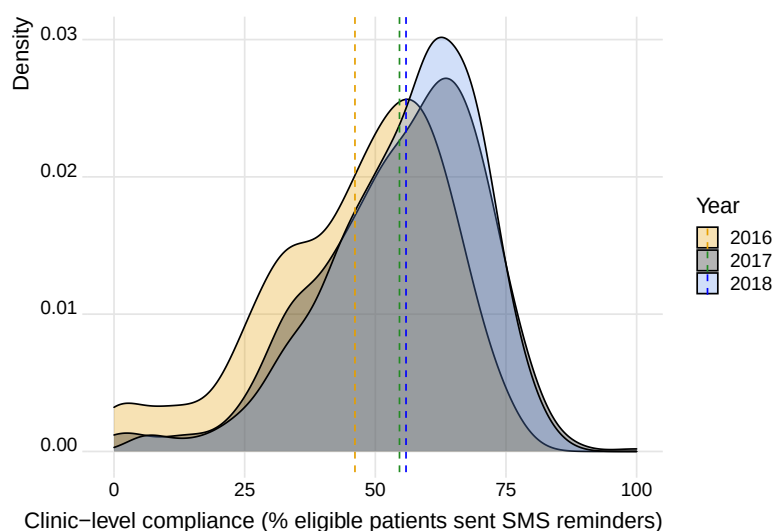
analyses where pre-trends are less severe. A final limitation is that we do not observe medication ingestion, and instead must rely on refill behavior.

Our findings have several implications for policy. Most importantly, we show that SMS reminders are effective at scale, even if implemented with variation in compliance. We show that in the context of public, government run health care clinics in Chile, implementation fidelity improved over time suggesting clinic-level learning. As a result, we see that the impacts of the program on visits, medication adherence, and hospitalizations also increase over time.

Control of chronic conditions like type 2 diabetes and hypertension is a global issue. Our findings are particularly important for settings with a gate-keeping healthcare model – where patients must first visit their primary care provider or general practitioner before being referred to speciality care, approved for tests, or prescribed new medication. This model is common in other OECD countries with public health care systems, such as Canada, the United Kingdom, Spain, and Italy (Blöndal & Ásgeirsdóttir, 2019; Brekke et al., 2007; Kringos et al., 2015; Reibling & Wendt, 2012; Rotar et al., 2018; Watt, 1987). In these settings, a light touch intervention such as nudging patients to attend primary care can have potentially large and meaningful impacts because it is intervening on the first step in the cascade of care. Nudges such as text-message reminders are increasingly used by governments to encourage constituents to take-up programs. Here, we have shown how these nudges can be used to also study the downstream effects of program take up, and that they are a promising strategy to promote treatment adherence among high-risk patients living with chronic conditions.

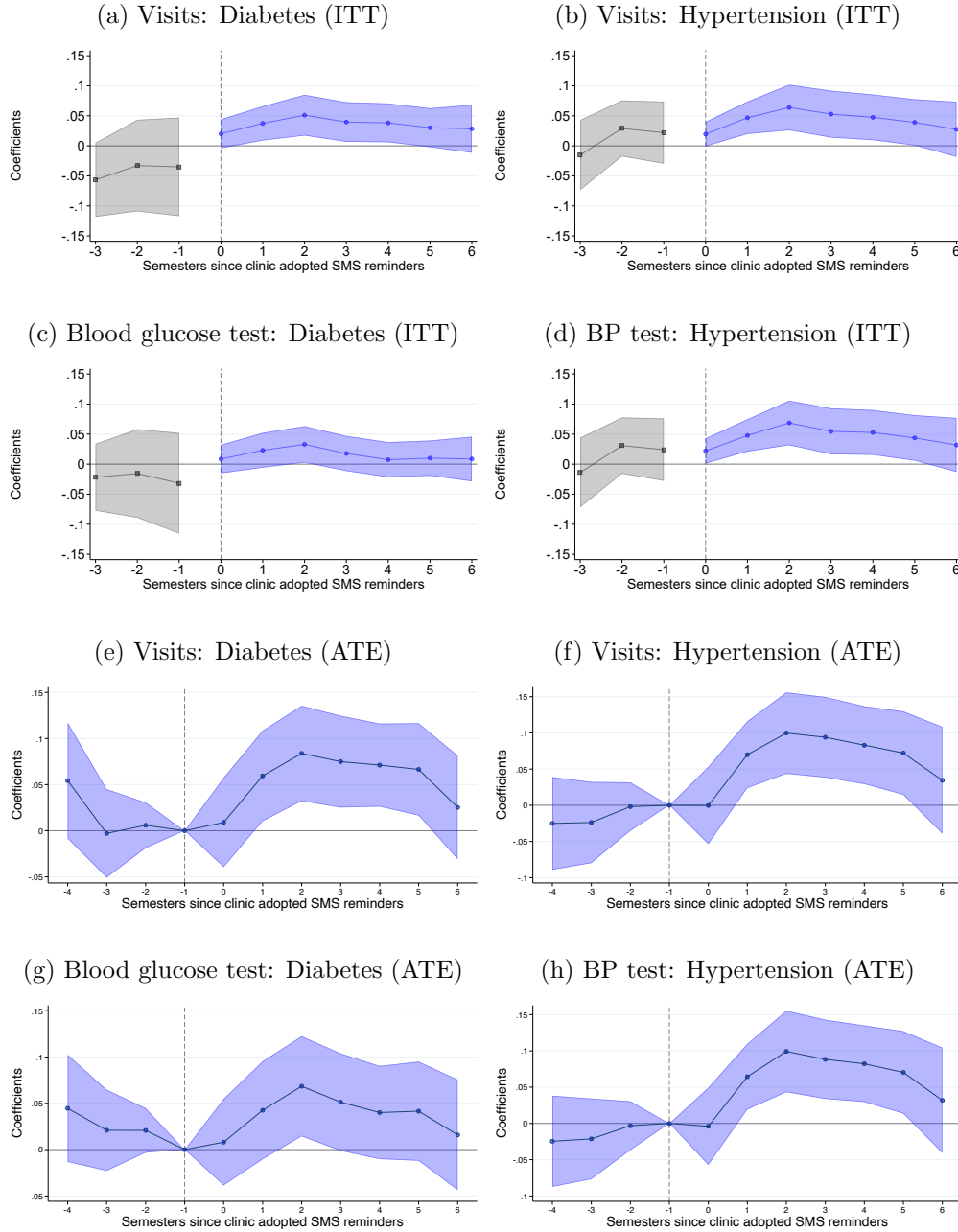
Chapter 2: Tables and Figures

Figure 2.1: Compliance with SMS reminder intervention, by clinic-year



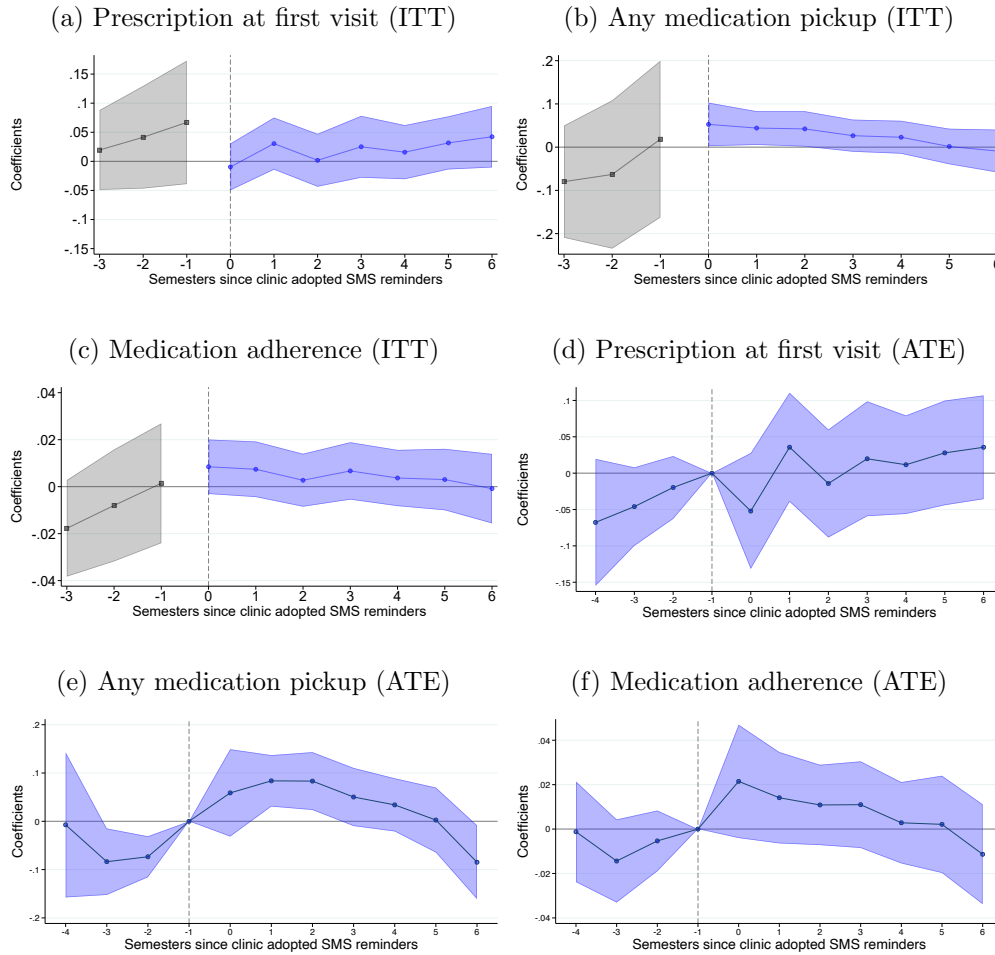
Note: Compliance was estimated at the clinic-year level using phone records. Records displayed, for each scheduled appointment, whether that patient was sent an SMS reminder. Compliance was calculated as the number of eligible appointments sent SMS reminders, divided by the total scheduled appointments for eligible patients.

Figure 2.2: Event studies: Visits and tests



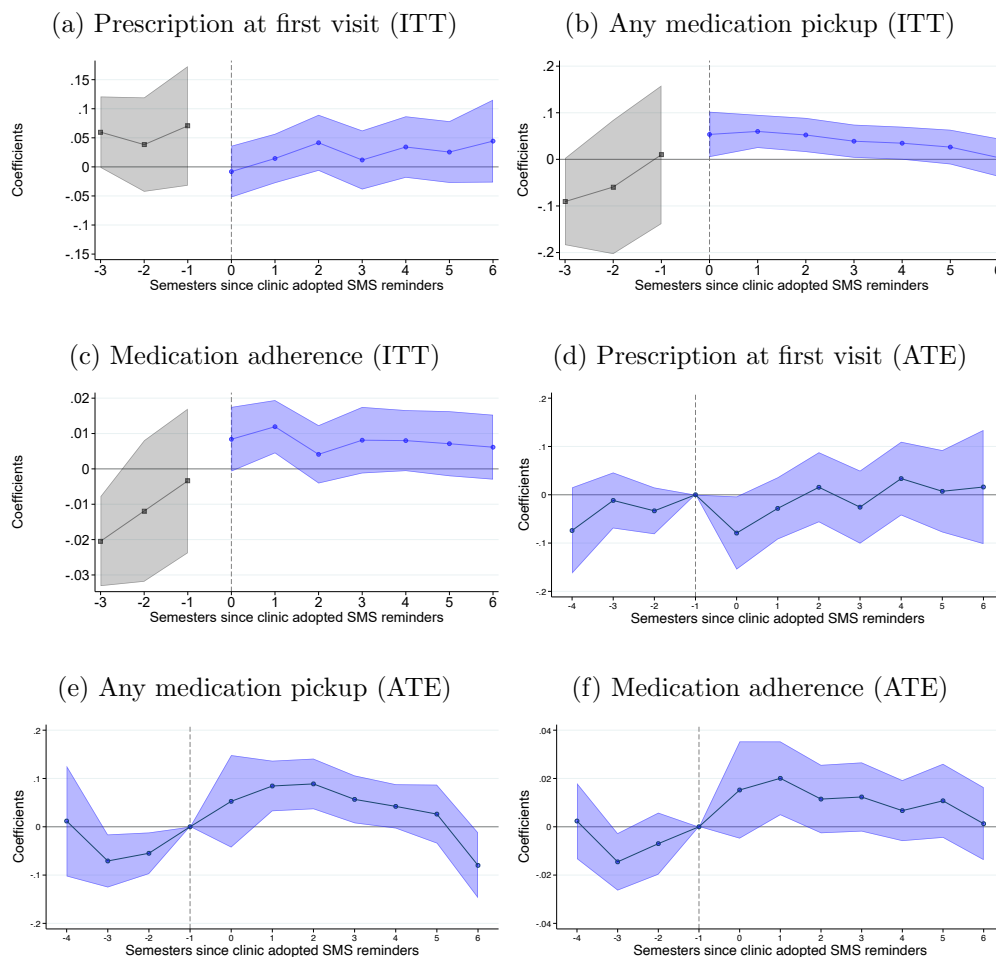
Note: Figures show regression estimates based on equation 2.1 for ITT estimates (panels a-d) and ATE estimates (panels e-h). The shaded areas denote 95% confidence intervals. The x-axis is number of semesters until or since the clinic implemented SMS reminders. DM2 stands for type 2 diabetes.

Figure 2.3: Event studies: Medication adherence among patients with type 2 diabetes



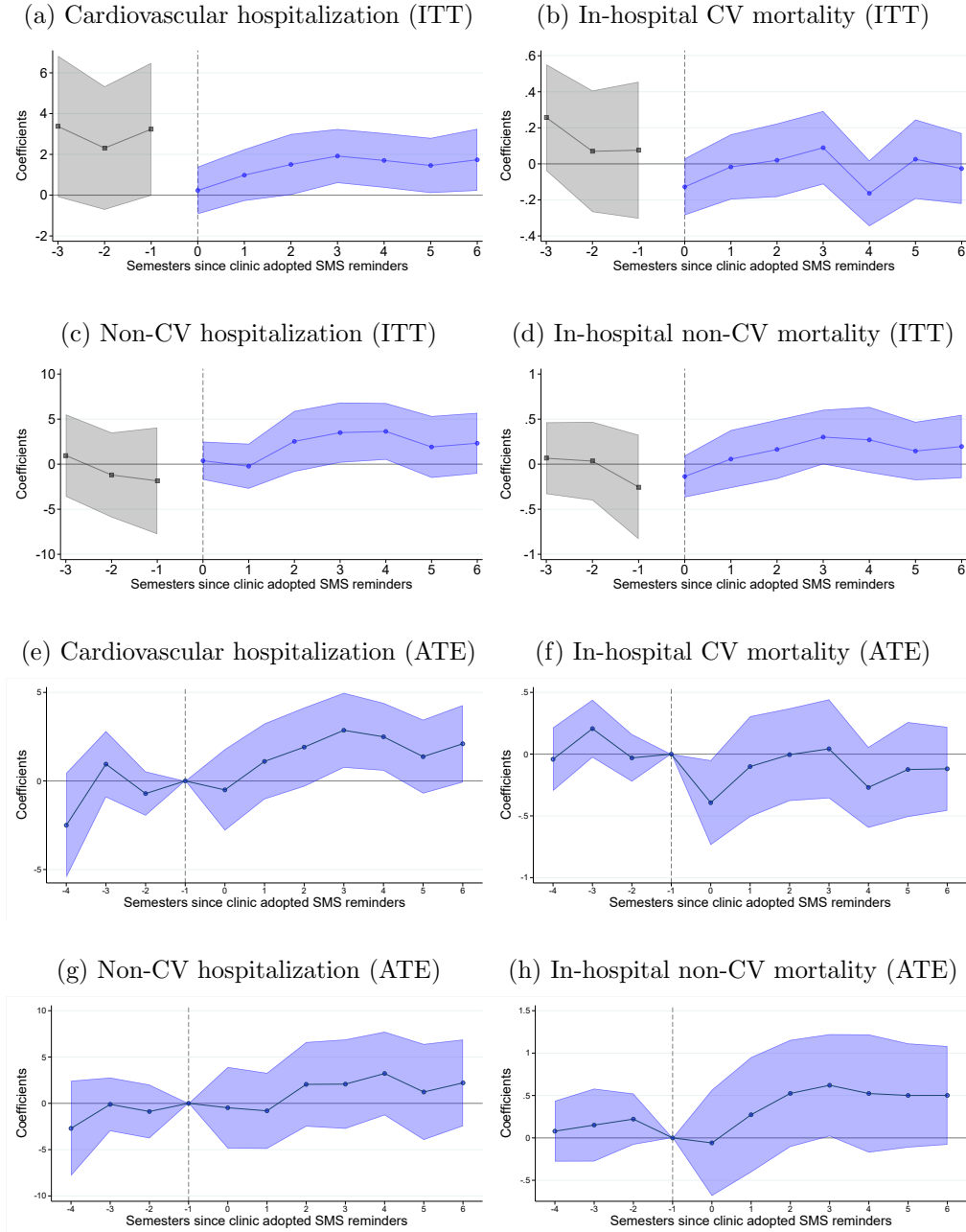
Note: Figures show regression estimates based on equation 2.1 for ITT estimates (panels a-c) and ATE estimates (panels d-f). The shaded areas denote 95% confidence intervals. The x-axis is number of semesters until or since the clinic implemented SMS reminders.

Figure 2.4: Event studies: Medication adherence among patients with hypertension



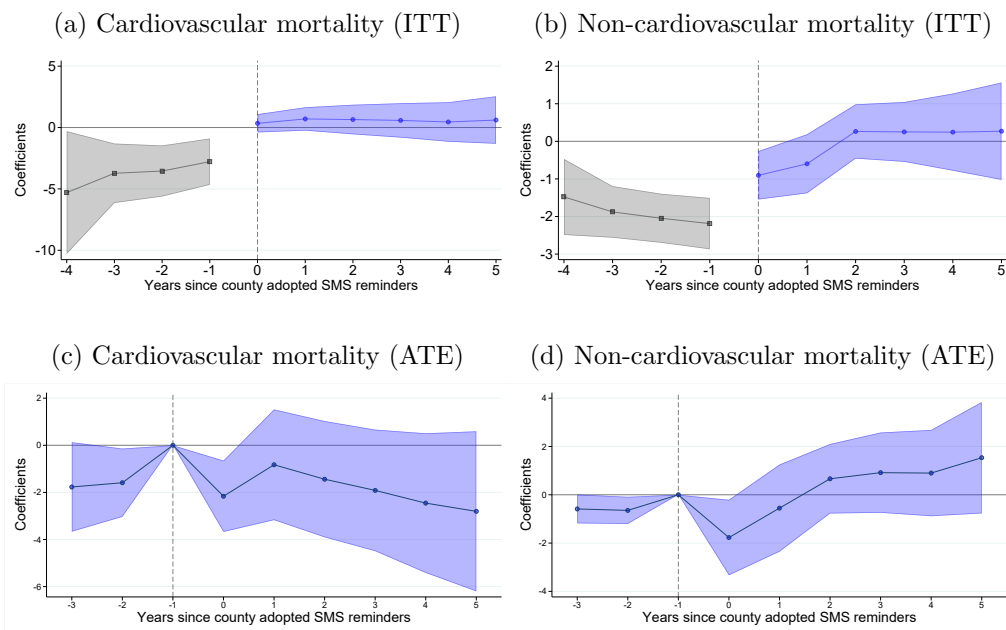
Note: Figures show regression estimates based on equation 2.1 for ITT estimates (panels a-c) and ATE estimates (panels d-f). The shaded areas denote 95% confidence intervals. The x-axis is number of semesters until or since the clinic implemented SMS reminders.

Figure 2.5: Event studies: Hospitalization per 1000 population



Note: Figures show regression estimates based on equation 2.1 for ITT estimates (panels a-d) and ATE estimates (panels e-h). The shaded areas denote 95% confidence intervals. The x-axis is number of semesters until or since the clinic implemented SMS reminders. CV stands for cardiovascular.

Figure 2.6: Event studies: Mortality per 1000 population



Note: Figures show regression estimates based on equation 2.1 for ITT estimates (panels a-b) and ATE estimates (panels c-d). The shaded areas denote 95% confidence intervals. The x-axis is number of years until or since the county implemented SMS reminders.

Table 2.1: Take up of SMS reminders by public primary care clinics

Date	N Treated	N Control	Compliance		
	Clinics	Clinics	Mean	Min.	Max.
S1 2014	0	315	NA	NA	NA
S2 2014	0	315	NA	NA	NA
S1 2015	168	147	NA	NA	NA
S2 2015	172	143	NA	NA	NA
S1 2016	203	112	46%	0%	76%
S2 2016	208	107	49%	10%	76%
S1 2017	208	107	56%	14%	76%
S2 2017	208	107	53%	16%	81%
S1 2018	208	107	58%	18%	85%
S2 2018	208	107	54%	20%	78%
Total	208	107	53%	0%	85%

Note: Number of clinics that did and did not take up SMS reminder program, among those in the EHR dataset. Compliance was measured among treated clinics only using phone records. It is the share of patients sent an SMS reminder, among eligible patients in a clinic-semester cell.

Table 2.2: Take up of SMS reminders by county

Year	N Treated Counties	N Control Counties
2013	0	275
2014	0	275
2015	86	189
2016	97	178
2017	97	178
2018	97	178
2019	97	178
2020	97	178
Total	97	178

Note: Number of counties that did and did not take up SMS reminder program, among those where all clinics either ever or never took up the SMS reminder program (275/346).

Table 2.3: Baseline patient characteristics: electronic health record data

	Ever Treated		Control	
	Mean	SD	Mean	SD
Male	0.41	(0.49)	0.41	(0.49)
Mean age (years)	62.11	(13.78)	61.69	(14.40)
Diagnosed with type 2 diabetes	0.20	(0.40)	0.23	(0.42)
Diagnosed with hypertension	0.91	(0.29)	0.89	(0.32)
Systolic blood pressure	135.96	(20.19)	135.36	(19.87)
Diastolic blood pressure	79.55	(12.15)	80.22	(12.03)
Hemoglobin A1c	7.98	(2.51)	7.97	(2.44)
Blood glucose	123.26	(57.53)	123.42	(58.21)
Weight (kg)	76.45	(15.69)	76.63	(15.92)
Body mass index	30.62	(5.63)	30.60	(5.67)
Normal BMI	0.14	(0.35)	0.15	(0.35)
Overweight BMI	0.37	(0.48)	0.37	(0.48)
Obese 1 BMI	0.30	(0.46)	0.29	(0.46)
Obese 2 BMI	0.12	(0.33)	0.12	(0.33)
Obese 3 BMI	0.07	(0.26)	0.07	(0.26)
Semesters observed	7.22	(2.45)	7.03	(2.57)
N primary care visits	4.46	(2.53)	4.23	(2.51)
Glucose test, at DM2 primary care visit	0.80	(0.40)	0.85	(0.36)
DM2 prescription at time of diagnosis	0.50	(0.50)	0.51	(0.50)
Blood pressure test, at HTN primary care visit	0.96	(0.19)	0.96	(0.19)
Hypertension prescription at time of diagnosis	0.55	(0.50)	0.57	(0.50)
Patients	241,899		122,527	
Total patients	364,426			
Visits	1,746,445		861,776	
Total visits	2,608,221			

Note: Patient health and characteristics from patient's first visit to primary care, where diagnosis with type 2 diabetes or hypertension occurred. BMI stands for body mass index. Glucose test and blood pressure test are conditional on a primary care visit.

Table 2.4: Baseline patient characteristics: mortality records

	Ever Treated		Control	
	Mean	SD	Mean	SD
Male	0.44	(0.50)	0.44	(0.50)
Mean age (years)	53.93	(14.95)	53.32	(15.18)
Diagnosed with type 2 diabetes	0.49	(0.50)	0.45	(0.50)
Diagnosed with hypertension	0.78	(0.42)	0.80	(0.40)
Age diagnosed with type 2 diabetes	53.05	(14.61)	52.88	(14.79)
Age diagnosed with hypertension	55.97	(14.46)	55.07	(14.82)
Years observed	5.49	(1.93)	5.59	(1.91)
Previous heart attack	0.09	(0.29)	0.10	(0.30)
Previous stroke	0.04	(0.19)	0.04	(0.19)
Previous COPD	0.04	(0.19)	0.04	(0.19)
N comorbidities	0.24	(0.51)	0.25	(0.52)
Patients	279,355		551,137	
Total patients	830,492			
Patient-years	1,574,319		3,051,479	
Total patient-years	4,625,798			

Note: Patient characteristics from year of when first diagnosed with type 2 diabetes or hypertension, among individuals diagnosed for the first time between 2013 and 2018. COPD stands for chronic obstructive pulmonary disease. N comorbidities is a count of diagnoses the patient had before their first diagnosis or type 2 diabetes and/or hypertension out of 14 conditions: kidney disease, heart attack, lymphoma, HIV/AIDS, stroke, COPD, asthma, leukemia, rheumatoid arthritis, juvenile arthritis, lupus, lung cancer, multiple myeloma, and Alzheimer's Disease and related dementias.

Table 2.5: Baseline county characteristics

	Ever Treated		Control	
	Mean	SD	Mean	SD
Population (2014)	1042.99	(1405.60)	1178.46	(1933.80)
Mean age	43.18	(1.81)	42.55	(1.87)
Mean income pc (USD)	1801.60	(350.85)	1621.44	(575.55)
% population under poverty line	0.36	(0.07)	0.36	(0.08)
Employment rate	0.65	(0.03)	0.64	(0.05)
Rural	0.75	(0.17)	0.78	(0.18)
Primary education	0.48	(0.11)	0.49	(0.09)
Secondary education	0.41	(0.07)	0.42	(0.06)
Tertiary education	0.11	(0.10)	0.09	(0.07)
Counties	97		178	

Note: County characteristics are from Chile's National Socioeconomic Survey (CASEN 2013).

Table 2.6: Impact of SMS reminders on primary care visits and tests

Patients:	Diabetes		Hypertension	
	Visit (1)	Blood sugar test (2)	Visit (3)	Blood pressure test (4)
Panel A: ITT				
Post SMS	0.035** (0.015)	0.015 (0.014)	0.043** (0.017)	0.047*** (0.017)
Panel B: ATE				
Post SMS	0.055** (0.022)	0.027 (0.023)	0.069*** (0.025)	0.067*** (0.024)
Mean dep. var.	0.644	0.553	0.643	0.618
Observations	497,949	497,949	2,411,117	2,411,117
Clinics	314	314	310	310

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents difference-in-differences estimates of the effect of SMS reminders. ITT stands for intent to treat, and ATE stands for average treatment effect. Coefficients in panel A are estimated using the Borusyak, Jaravel, and Spiess (2021) estimator for staggered policy adoption. The treatment variable is an indicator equal to one in the semester of, and after a clinic implemented SMS reminders and zero otherwise. Coefficients in panel B are estimated using two-way fixed effect models because the treatment variable is a continuous measure of compliance. Here, the treatment variable is the share of visits sent a reminder, among visits eligible to receive a reminder, zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. Visit is an indicator for if the patient had at least one visit to primary care in a given semester, zero otherwise. Blood sugar test is an indicator for if the patient had a hemoglobin A1c test or a blood glucose test in a given semester, zero otherwise.

Table 2.7: Impact of SMS reminders on medication adherence

	Prescription at diagnosis	Any prescription pick-up	Medication adherence
Patients:	Diabetes		
	(1)	(2)	(3)
Panel A: ITT			
Post SMS	0.016 (0.019)	0.023 (0.020)	0.004 (0.005)
Panel B: ATE			
Post SMS	0.022 (0.030)	0.079*** (0.030)	0.011 (0.009)
Mean dep. var.	0.447	0.276	0.032
Observations	75,899	242,194	242,194
Clinics	314	313	313
Patients:	Hypertension		
	(1)	(2)	(3)
Panel C: ITT			
Post SMS	0.017 (0.022)	0.036** (0.017)	0.008** (0.004)
Panel D: ATE			
Post SMS	0.003 (0.036)	0.072*** (0.027)	0.015** (0.006)
Mean dep. var.	0.473	0.278	0.036
Observations	328,037	1,213,287	1,213,287
Clinics	310	310	310

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents difference-in-differences estimates of the effect of SMS reminders. ITT stands for intent to treat, and ATE stands for average treatment effect. Coefficients in panels A and C are estimated using the Borusyak, Jaravel, and Spiess (2021) estimator for staggered policy adoption. The treatment variable is an indicator equal to one in the semester of, and after a clinic implemented SMS reminders and zero otherwise. Coefficients in panels B and D are estimated using two-way fixed effect models because the treatment variable is a continuous measure of compliance. Here, the treatment variable is the share of visits sent a reminder, among visits eligible to receive a reminder, zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. Models in column 1 are estimated using each patient's first visit after diagnosis only. Models in columns 2 and 3 are estimated among patients with diabetes and hypertension, who had any prescription at their first visit for diabetes or hypertension, respectively. Medication adherence is an indicator for if the patient's proportion of days covered 80% or greater in a given semester, zero otherwise.

Table 2.8: Impact of SMS reminders on hospitalization (per 1000 population) among all patients

Patients:	All			
	Cardiovascular		Non-cardiovascular	
	Hospital- ization (1)	In-hospital mortality (2)	Hospital- ization (3)	In-hospital mortality (4)
Panel A: ITT				
Post SMS	1.428** (0.595)	-0.025 (0.063)	2.157 (1.383)	0.159 (0.100)
Panel B: ATE				
Post SMS	2.037** (0.850)	-0.141 (0.128)	0.291 (0.202)	1.957 (1.962)
Mean dep. var.	11.900	0.527	1.330	39.700
Observations	2,607,151	2,607,151	2,607,151	2,607,151
Clinics	315	315	315	315

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents difference-in-differences estimates of the effect of SMS reminders on hospitalizations per 1000 population. ITT stands for intent to treat, and ATE stands for average treatment effect. Coefficients in panel A are estimated using the Borusyak, Jaravel, and Spiess (2021) estimator for staggered policy adoption. The treatment variable is an indicator equal to one in the semester of, and after a clinic implemented SMS reminders and zero otherwise. Coefficients in panel B are estimated using two-way fixed effect models because the treatment variable is a continuous measure of compliance. Here, the treatment variable is the share of visits sent a reminder, among visits eligible to receive a reminder, zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. All models also include fixed effects for male and 2-year age. List of ICD-10 codes included in cardiovascular related hospitalization and in-hospital mortality is in table B.3.

Table 2.9: Impact of SMS reminders on mortality, among all patients)

Patients:	All	
	Cardiovascular mortality (1)	Non-cardiovascular mortality (2)
Panel A: ITT		
Post SMS	0.562 (0.665)	0.005 (0.389)
Panel B: ATE		
Post SMS	-1.053 (1.413)	1.234 (0.787)
Mean dep. var.	7.639	7.378
Observations	4,186,391	4,186,391
Counties	275	275

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents difference-in-differences estimates of the effect of SMS reminders on deaths per 1000 population. ITT stands for intent to treat, and ATE stands for average treatment effect. Coefficients in panel A are estimated using the Borusyak, Jaravel, and Spiess (2021) estimator for staggered policy adoption. The treatment variable is an indicator equal to one in the year of, and after a clinic implemented SMS reminders and zero otherwise. Coefficients in panel B are estimated using two-way fixed effect models because the treatment variable is a continuous measure of compliance. Here, the treatment variable is the share of visits sent a reminder, among visits eligible to receive a reminder, zero otherwise. All models include fixed effects for year, county, and years since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. All models also include fixed effects for male and 2-year age. List of ICD-10 codes included in cardiovascular related mortality is in table B.3.

Chapter 3

The Role of Discretion in Clinical Decision Making: Evidence from Thresholds

ABSTRACT

When a decision maker has discretion, such as a worker reporting their taxable income, threshold-based rules or policies can distort behavior. This has not been studied in the context of medicine where thresholds guide many important diagnosis and treatment decisions. I study the decision to diagnose and treat hypertension, defined as blood pressure $\geq 140/90$ mmHg. Hypertension is the single most important risk factor for cardiovascular disease, but diagnosis can be challenging because blood pressure is a noisy measure of risk and often overstated in clinic. Using bunching estimation and electronic health records from over 600,000 patients in Chile I find that providers round the blood pressure of up to 62% of patients who test near the threshold, using their discretion to turn a positive test for hypertension into a negative. This behavioral response leads to a more accurate classification of patients in terms of their cardiovascular risk, as measured by future hospitalizations. I find that discretionary decisions are consistent with heuristic thinking: among patients with identical test results, those with characteristics representative of high cardiovascular risk are less likely to be sorted below the diagnostic threshold. These results suggest that in the case of limited information, heuristic thinking and discretionary diagnosis can lead to more accurate decisions and to better patient outcomes, underscoring the importance of clinical skill in achieving an efficient and equitable allocation of health care.¹

¹This project does not meet the threshold definition of “human subjects” research set forth in Federal Regulations at 45 CFR 46.102(f) as it only includes publicly available data, and coded private information data that was not collected for the currently proposed project, and the investigators cannot readily ascertain the identity of the individuals to whom the coded private information pertains.

3.1 Introduction

Clinical practice guidelines aim to standardize medicine using evidence-based recommendations, yet large inefficiencies in care delivery remain (Abaluck et al., 2020; Abaluck et al., 2016; Chandra & Staiger, 2014; Mullainathan & Obermeyer, 2021). Policymakers argue that improving adherence to guidelines can both reduce costs and improve quality by getting care to the patients who would benefit most (Runnacles et al., 2018). Critics of guidelines, a group largely made up of physicians, argue that diagnostic rules can ignore other clinically relevant information and the role of expertise (Knaapen, 2014; Timmermans & Mauck, 2005). In reality, medical decisions must often be made with limited or poor information, even when guidelines exist. In these ‘grey areas’, discretionary decision-making – where a provider uses their judgement to decide whether to test, treat, or act on the results of a positive test – is highly important. This is one reason clinical skill is becoming increasingly recognized as an important input to quality health care (Chan et al., 2022; Currie & Zhang, 2021).

I study health care providers’ decision to diagnose and treat hypertension, a highly prevalent disease with a simple but noisy test where false positives are common (Baron, 2018). While the clinical practice guidelines for hypertension are very simple - diagnose if blood pressure is at least 140/90 mmHg - both missed- and over-diagnosis are common because accurately measuring blood pressure is challenging (Baron, 2018). Instantaneous blood pressure fluctuates around the patient’s long-term or true blood pressure, which is a strong predictor of risk for cardiovascular events (Muntner Paul et al., 2019). Seemingly innocuous actions can change blood pressure within seconds: drinking coffee, sitting with one leg crossed over the other, and for some patients, speaking with a physician (Adiyaman et al., 2007; Muntner Paul et al., 2019; Myers et al., 2011). Substantial effort is given to developing systems to improve measurement precision such as averaging multiple measures, making sure the patient is sitting correctly, and waiting a few minutes after the doctor walks in. Still, in a busy primary care clinic where providers spend just a few minutes with each patient, knowing whether a patient testing over 140/90 mmHg is a true positive or a false positive requires substantial clinical skill.

The noise in blood pressure allows me to identify instances of discretionary diagnosis. If a provider believes a patient’s test result is a false positive, it is likely they will retest until they get a number they believe. Or, they might simply record a different number in the medical record than the one actually measured. With enough data, these behaviors are observable as bunching in the distribution of blood pressure under the 140 threshold, which would otherwise be smooth.² I estimate this behavior with a bunching estimator and interpret it as a behavioral response to the diagnostic threshold for hypertension. This allows for an investigation into when and how health care providers behave in one of medicine’s many ‘grey areas’.

²I expect excess mass below the diagnostic threshold and not above it because blood pressure measured in a medical setting, such as primary care, is well-known to be higher on average than if the same patient’s blood pressure was measured elsewhere (Muntner Paul et al., 2019; Myers et al., 2011).

Specifically, I use estimate the magnitude of bunching in blood pressure - my measure of discretion - separately at 257 public primary care clinics in Chile, using visits from 619,907 unique patients. This yields a measure of the amount of discretionary decision-making at each primary care clinic. Taking advantage of the fact that in Chile's public health care system patients are assigned a primary care clinic based on their home address, I use a difference-in-differences approach to estimate the impact of discretion. Specifically, I compare high vs. low discretion clinics, and patients below vs. above the diagnostic threshold. This allows for estimation of the impact of discretion on clinical decisions, selection by patient health as measured by future cardiovascular-related hospitalizations, and selection by patient characteristics representative of high cardiovascular risk.

The context of this study is Chile's public health care system. Chile is an OECD country with a GDP per capita comparable to Greece, Hungary, and Poland. Chile's burden of disease is similar to that of other high-income countries: non-communicable diseases account for 82% of years of life lost (PAHO, 2020). The prevalence of hypertension is 27.8% in Chile, comparable to the prevalence of 30.7% in the United States. The data used come from 257 primary care clinics that all use the same electronic health records software, and had enough visits per year to estimate a bunching parameter. I restrict the analysis to 619,907 patients who were not previously diagnosed with hypertension.

Blood pressure is an ideal candidate for bunching estimation for several reasons. First, because hypertension both lacks symptoms and can be deadly, blood pressure measurement is recommended at *all* primary care visits, alleviating concerns of selection into testing. Second, because it is a biological parameter its counterfactual population distribution is well-defined. Third, blood pressure measurement and hypertension diagnosis is a setting with few frictions. That is, providers measure and record blood pressure themselves and then decide the clinical action taken, if any. Thus, providers can precisely manipulate the measurement and outcome according to their beliefs about the patient.

Measuring discretion separately at each primary care clinic reveals that up to 62% of patients expected to have tested just above the diagnostic threshold are instead recorded just below it. These results indicate that at some clinics, the propensity for discretionary diagnosis is large. These results also provide evidence that thresholds, which are common in medicine, can distort provider behavior.

After quantifying discretion using bunching estimation, I find that clinical practice guidelines are adhered to similarly at high and low discretion clinics. That is, conditional on recorded blood pressure below the diagnostic threshold there is no difference in the probability of new hypertension diagnosis or prescription at high vs. low discretion clinics. This indicates that the value of blood pressure recorded in the electronic health record does guide decision-making, whether it was manipulated by the provider or not.

Untreated high blood pressure raises a patient's risk of adverse cardiac events, in particular stroke, heart attack, and heart failure. If discretionary decision-making meant providers were incorrectly reclassifying true positives as false positives we would expect to see more of these hospitalizations among patients at high discretion clinics who were classified below the diagnostic threshold. Instead we see the opposite. Patients at high discretion clinics are

62.1% and 44.5% less likely to be hospitalized for stroke within 3 and 12 months respectively. They are also 25.7% and 9.3% less likely to be hospitalized for heart attack within 3 and 12 months respectively. No effects are found on heart failure. These results suggest not that discretion itself reduces cardiovascular hospitalization, but that at high discretion there is more, and more accurate selection by to cardiovascular risk. In particular, providers at high discretion are more likely to use their discretion to move healthy patients under the threshold and keeping high-risk patients above it.

How do providers identify low-risk patients who happened to have tested positive? I investigate heuristic thinking as one possible mechanism. At the heart of diagnosis is the question: what is the probability that patient A belongs to the class of people with disease B? When answering this type of question it is likely that providers, like all humans, rely on the representativeness heuristic (Tversky & Kahneman, 1974). If so, instead of mentally calculating conditional probabilities they judge the similarity of the patient in front of them to a patient that is representative of one with high cardiovascular risk.

Indeed, I find that patients with characteristics representative of high cardiovascular risk - older males - are less likely to be sorted under the threshold, behavior that is consistent with the use of heuristics. These results are small and likely do not explain all providers' decision-making process, but they do suggest that providers use the representativeness heuristic to make decisions. Compared to the clinical practice guidelines, where blood pressure is the only determinant of hypertension status, by using the representativeness heuristic providers add information to the decision process.

This paper contributes to our understanding of how health care providers make decisions, in particular their use of cognitive shortcuts and rules of thumb. In recent work in this areas, Singh, 2021 documents heuristic thinking in the delivery room, showing that clinical-decisions are affected by recent adverse events experienced by the provider. Olenski et al., 2020 and Coussens, 2018 both examine the influence of patient age on test decisions in emergency departments, and find providers are subject to end digit bias. This behavior leads patients who are slightly younger than round number ages to be systematically tested less than patients who are slightly older.

Heuristic thinking has been shown to lead to biased decisions and poor outcomes for some patient groups (Coussens, 2018; Ly, 2021; Olenski et al., 2020; Singh, 2021). In contrast, in the setting of hypertension diagnosis and treatment I find that heuristic thinking contributes to better outcomes for patients. There are several reasons for this difference. First, I compare decisions at clinics that are more or less likely to use heuristics, whereas other work has compared the use of heuristics to an optimal decision rule, which will by definition be better. The policy prescription in the latter might be the inclusion of a decision-tool or algorithm to aid providers, whereas in the former education of low-discretion providers by high-discretion providers could be beneficial. Second, each clinical scenario is different, and in the case of diagnosing hypertension, adding more information to the decision process through the use of heuristics appears to be beneficial. This is likely due to the too-simple guideline given the noisy biomarker, which may not be the case for other clinical scenarios studied.

This paper aims to make two additional contributions. First, to our understanding of the

allocation of clinical resources. Focusing on the fact that health care providers determine how many and which patients receive medical tests and treatments, this literature uses observational data to examine the volume and cost of care, as well as the quality of clinical decisions. Abaluck et al., 2016, Abaluck et al., 2020, and Chan and Gruber, 2020 find that providers' testing decisions are inefficient based on patient risk. Chandra and Staiger, 2014 and Mullainathan and Obermeyer, 2021 show heart attack tests and treatment are misallocated in emergency departments compared to an optimal rule. Clinical skill, including the ability to test the 'right' patients is increasingly being recognized as an important and modifiable input to quality health care (Chan et al., 2022; Currie & Zhang, 2021).

It is well studied that when incentives or policies change discontinuously at a threshold, behavior can be distorted (Kleven, 2016). While bunching methods have been used to study patients' response to health insurance (Chandra et al., 2021; Einav et al., 2015, 2017), providers' response to health insurance (Hernández-Pizarro et al., 2020), and hospital wait times (Gruber et al., 2018), this is the first paper to my knowledge to investigate bunching in the context of diagnostic decisions. Given the vast number of medical decisions made using thresholds that categorize continuous biological measurements, there are likely many opportunities to use bunching to understand how and why clinical decisions are made. Relatedly, a small but growing literature uses regression discontinuity designs to examine the impact of diagnosis and treatment (Alalouf et al., 2019; Moscoe et al., 2015; O'Keeffe et al., 2014; Venkataramani et al., 2016). A common feature of these papers is the relatively small change in the probability of diagnosis or treatment at the clinical threshold, which is expected to be near one.³ This could be non-adherence to clinical practice guidelines, or, discretionary decision-making by health care providers.

The remainder of this paper is as follows: section 2 describes the country and medical setting as well as the data used. Section 3 describes what we can learn from bunching estimation and the empirical strategy to measure discretion and its impact. Section 4 describes results, and section 5 concludes.

3.2 Setting and Data

I study providers' decision to diagnose hypertension among public healthcare users in Chile. This setting offers several key advantages. In Chile, public healthcare is low cost and guaranteed for all. Primary care is received at local clinics or hospitals, centrally organized by the Ministry of Health. Because of this organizational structure data from primary care, hospitals, and pharmacies is available and able to be linked country-wide. This analysis therefore includes a large cross section of patients in terms of location, socio-economic status, wealth,

³Alalouf et al., 2019 find the probability of diagnosis with type 2 diabetes increases by 11 percentage points at the diagnostic threshold. Venkataramani et al., 2016 find the probability of treatment with antiretrovirals increases by 38% at the CD4 cell count threshold, Geneletti et al., 2015 find the probability of treatment with statins increases by 17% at the prescription threshold.

and health from a population with few access to care issues. These patients also receive care from clinics that face similar payment structures, incentives, and organization.

3.2.1 Health and Healthcare in Chile

Chile is an OECD country with a GDP per capita comparable to Greece, Hungary, and Poland. Chile's burden of disease is similar to that of other high-income countries: non-communicable diseases account for 82% of years of life lost (PAHO, 2020). The prevalence of hypertension is 27.8% in Chile, 30.7% in the United States⁴, and 24.2% in Canada (Garies et al., 2019; Lanas et al., 2020; Ostchega & Nguyen, 2020).

Chile has both public and private healthcare systems. All residents are defaulted in to the public system, where insurance is provided by the federal government and funded through a universal 7% tax on earnings (*Fondo Nacional de Salud*, or FONASA). Approximately 19% of the population opts-out of the public system by purchasing private insurance. Those who remain in the public system are assigned a primary care clinic based on their home address. Primary care services are considered to be of relatively high quality and organized, delivering free medical, dental, nursing, and midwifery services at local clinics (Bastías et al., 2008; Bitrán et al., 2010; Bossert & Leisewitz, 2016). Primary care is obtained either at a clinic, or a low-complexity hospital. Low-complexity hospitals are typically present in rural areas that have a single medical center. For this reason low-complexity hospitals provide a combination of primary care and some in-patient services. Hereafter, I refer to both types of facilities as clinics.

Most care in the public system is low-cost, and copayments are further reduced for chronic diseases, including hypertension, as they are a priority policy area for the federal government (I. Aguilera et al., 2015). This reduction applies to primary care visits, hospital visits, and medication. FONASA, the public insurer, assigns patients a copayment level based on their income.⁵ The lowest two income groups have a copayment of 0%, which increases to 10% and 20% for the second highest and highest income groups, respectively.

Supply-side payment and incentives are also standardized nation-wide. Payment from the public insurer to primary care clinics is capitated and determined by the number of registered patients and their risk (JLN, 2017a, 2017b). Healthcare providers in the public system are employed by the government either on a annual, salaried basis, or a fixed-term contract basis. Some primary care providers also face performance based financing - incentives to improve quality of care in key areas. Bonuses can be as large as 20% of pay, however the

⁴The prevalence of hypertension is 30.7% in the US if the threshold of 140/90 mmHg is used. In 2017 the American College of Cardiology and American Heart Association recommended the diagnostic guideline be reduced to 130/80 mmHg (Carey & Whelton, 2018). If this US-specific threshold is used over 45% of Americans are estimated to have hypertension (Ostchega & Nguyen, 2020)

⁵For 2020, the FONASA copayment groups (in 2020 US Dollars) were: group A, extreme poverty and/or unhoused, group B monthly income <\$320 or on a government pension, group C, monthly income \$320-465, and group D monthly income >\$465 (de Salud - Gobierno de Chile, 2020). In 2020 Chile's mean and median monthly income per capita were USD\$784.45 and USD\$518.74 respectively (de Estadísticas, 2020).

goals and metrics are not clear and change frequently (JLN, 2017a, 2017b). Importantly for this project, the bundle of incentives faced by providers does not indicate that there is any reason providers would avoid diagnosing or treating a patient with a (marginally) positive test for hypertension. If anything, they might be more incentivized to diagnose a relatively healthier patient, as this group is then more likely to achieve controlled hypertension.

3.2.2 Medical Context

An estimated 1.4 billion people have high blood pressure (BP) making it the largest contributor to both all-cause and cardiovascular mortality globally (Egan et al., 2019; Piper et al., 2015). Despite being a strong predictor of cardiovascular risk, hypertension lacks symptoms. For this reason it is common practice to measure patients' blood pressure at *every* primary care clinic visit along with vital signs, as this is the only way to know if BP is high. Chile's clinical practice guidelines define hypertension as blood pressure $\geq 140/90$ mmHg.⁶ Hypertension can be prevented and treated with medication and behavior change including regular exercise and improving diet.

Accurately measuring blood pressure is challenging. Instantaneous or short-term blood pressure fluctuates around the patient's long-term or true blood pressure. Blood pressure is influenced in the very short term by stimuli such as caffeine intake, sleep, nervousness, the measurement method, and for some patients, the presence of a physician (Muntner Paul et al., 2019; Myers et al., 2011).

When making a diagnosis the provider - who is typically measuring BP while interacting with the patient - observes instantaneous BP and not true, long-term BP. Substantial effort is devoted to creating training materials (e.g. Stergiou et al., 2020; Williams et al., 2009, technology (e.g. Miao et al., 2020; Socrates et al., 2021, and research (e.g. Baron, 2018; Wohlfahrt et al., 2020 to improve the accuracy of BP measured in clinic. Still, it is known that blood pressure is usually measured with error, and both missed and over-diagnoses are common. In this paper I argue that uncertainty in whether BP measured in clinic accurately reflects the patient's true, underlying health status makes this diagnostic decision one where a providers' clinical skill and use of discretion are crucial.

Noisy measurement contributes to anomalies in recorded blood pressure, including rounding down and the over-representation of numbers that end in zero,⁷ which have long been documented in the medical literature (Alcocer et al., 1997; Wilcox, 1961). An example of end digit bias in blood pressure measurement is shown in figure 3.1. These numerical anomalies have not been documented in other biomarkers, either those measured in primary care (such as pulse), or in laboratories (such as hemoglobin A1c).

⁶In the US, clinical practice guidelines were updated in 2017 to recommend diagnosis of hypertension at blood pressure $\geq 130/80$ mmHg (Carey & Whelton, 2018) However in most countries, including Chile, this recommendation had not been adapted as of 2021.

⁷The over-representation of numbers that end in zero (and sometimes 5) compared to an expected distribution is called end digit bias, round number preference, or left-digit bias. I will use end digit bias here.

Broadly, research has found that end digit bias in blood pressure measurement is associated with a reduced likelihood of diagnosis and treatment,⁸ and that physicians are more likely to round down than nurses (Burnier & Gasser, 2008; Greiver et al., 2019; Kim et al., 2007). This literature conceptualizes end digit bias as an indicator of low-quality care and has recently documented an association between this bias and increased mortality from stroke, angina, and myocardial infarction (Greiver et al., 2019).

3.2.3 Data

3.2.3.1 Electronic Health Records

I primarily use electronic health records (EHRs) from Chile’s public health care system from 2013-2017. These data record routine encounters between providers and patients in primary care and include detailed diagnoses, biomarkers (e.g., blood pressure, blood glucose, weight etc.), demographics (age, sex), comorbidities, and some self-reported health behaviors (e.g., exercise, smoking, alcohol use). Visit data was linked at the patient level to medication prescriptions (including date, drug name, and quantity) and to the universe of hospitalizations (including ICD-10 codes for diagnoses, date, and length of stay) in Chile over the same time period.

Several EHR software providers are used by public clinics, and these data come from the largest provider. Among clinics that use said software, visits in the analysis dataset come from patients in Chile’s cardiovascular control program (*Programa de Salud Cardiovascular* or PSCV), which has been operating since 2002. A patient is considered enrolled in the PSCV once they are diagnosed with one or more diseases of interest: type 2 diabetes, hypertension, dyslipidemia,⁹ or have risk factors: a family history of cardiovascular disease (CVD), or smoke and are at least 55 years old. PSCV patients continue to seek care at their assigned primary care clinic, and receive additional benefits, namely counselling about how to manage their chronic disease and when to seek urgent care, text-message reminders about upcoming visits, additional tests such as for diabetic foot, and are asked to attend preventative care more frequently than non-PSCV patients (MINSAL, 2017a). For this reason, my EHR data does not represent all patients seen by a provider on a given day, but instead, the interactions providers had with PSCV patients.

EHR data come from 257 clinics representing 113 municipalities (out of 345 in the country). I link EHR at the municipality level to Chile’s National Socioeconomic Characterization Survey (CASEN, 2015 wave), which is representative at the municipality level, to obtain socio-economic characteristics.

To compare the distribution of blood pressure in the EHR to a nationally representative distribution I use Chile’s National Health Survey (ENS, 2016-2017 wave), which is also

⁸Only one study, to my knowledge, finds that end digit bias is associated with an increased likelihood of treatment (Broad et al., 2007).

⁹Dyslipidemia refers to unhealthy cholesterol levels in the blood: too much ‘bad’ LDL cholesterol and/or too little ‘good’ HDL cholesterol.

representative of the population at the municipality level.

3.2.3.2 Sample

To measure discretion and its impact, I include visits where blood pressure could be diagnosed, from patients aged 18-80 at ‘large’ clinics. Last, I restrict the sample to patients who visited primary care at least 1 year before the end of the study observation period, so that I can measure post-visit hospitalization over a sufficiently long period.

To do this, I first exclude visits from patients previously diagnosed with hypertension in order to analyze interactions where a diagnostic decision *could be made*. This leaves patients with type 2 diabetes (29%; categories are not mutually exclusive), dyslipidemia (46%), or cardiovascular disease risk factors (63%), who are attending regular care at primary care clinics (table 3.1). I use each patient’s first observed visit to avoid over-representing blood pressure measures that are under 140/90 mmHg, which could impact bunching estimates.

I also exclude visits where blood pressure was not recorded (2.4% of visits), and patients under 18 or over 80 (9.8%). Patients over 80 are subject to a different diagnostic threshold and goal for controlled blood pressure. Last, I exclude ‘small’ clinics with too few visits for bunching estimation. I define small as clinics with below the median number of visits ($N=1165$). This causes 52% of clinics to be excluded, but only 12.5% of patients. This leaves $N=619,907$ patient-visits at $N=257$ clinics. See table B.1 for full details on sample selection.

Tables 3.2 and 3.3 describe the study population. The average patient is 58 years old with a body mass index of 30.7, the low end of the obese category. Many are low-income, making less than \$320 per month and over a quarter of patients live below the federal poverty line. Compared to the total population of Chile, or even to the population that uses public, primary care, this is a high risk sample. Because of the criteria to be in the PSCV we expect adverse cardiac events to be higher in this group than the general public. For this reason, it is also likely that providers are more attentive to blood pressure during visits with this patient group, because BP is an important predictor of cardiovascular risk. Thus, the behavioral response estimated here is likely an under-estimate of the magnitude we would observe in a less risky patient population.

3.3 Empirical Strategy

I first describe the intuition behind bunching estimation and what we can learn from it. Second, I describe construction of the independent variable, the magnitude of bunching (discretion) at the clinic level. Third, I outline the difference-in-differences approach used to understand the impact of discretion on clinical decisions, and selection by cardiovascular risk and patient characteristics.

3.3.1 Measuring Provider Discretion With Bunching

3.3.1.1 What Can We Learn From Bunching?

The bunching approach uses discontinuities in incentives to measure behavior (Kleven, 2016). Similar to regression discontinuity designs, bunching methods can be applied in settings with a running variable and a threshold that determines treatment. In the case of bunching the running variable is manipulated (compared to a known counterfactual distribution) and selection with respect to the threshold is present. This manipulation can be interpreted as a behavioral response to the threshold.

The majority of work using the bunching approach is in public finance and studies a behavioral response to taxation (e.g. (Chetty et al., 2011; Kleven & Waseem, 2013; Saez, 2010). Broadly, these papers show that some individuals and firms under-report earnings systematically to pay less in taxes, which is interpreted as a behavioral response to the tax threshold. Bunching has also been used to show that patients are price elastic with respect to drug prices when the generosity of their health insurance coverage changes ((Chandra et al., 2021; Einav et al., 2018; Einav et al., 2015, 2017). Only one paper, to my knowledge, has used bunching to examine provider behavior. Gruber et al., 2018 show that emergency department doctors respond to an incentive to see waiting patients within 4 hours by selectively reducing wait times for some patients, which increases the cost and intensity of care delivered and reduces mortality.

I study bunching of blood pressure that is recorded in electronic medical records, and interpret it as a behavioral response by healthcare providers to the diagnostic threshold for hypertension. Importantly, what is present here is manipulation of blood pressure that is *recorded by the provider* in the electronic health record, and not precise manipulation of patients' own health. This is the first paper to my knowledge to investigate providers' behavioral response to a diagnostic threshold, a type of decision rule that is common in medicine.

Using blood pressure as a running variable in which to study bunching offers several advantages. First, because hypertension lacks symptoms and can be deadly blood pressure measurement is recommended at all primary care visits, alleviating any concerns of selection into testing. Second, as a biological parameter its counterfactual population distribution is known. Third, it is a setting with few frictions. That is, providers measure and record BP themselves and then decide the clinical action taken, if any. Thus, providers can precisely manipulate the measurement and outcome according to their beliefs about the patient.

3.3.1.2 Bunching Estimation

Examining raw distributions of blood pressure at the clinic level reveals that clinics vary not only in the amount of discretion, but also in the extent to which providers round BP to numbers ending in zero (figure 3.3). This often occurs when the policy or reference point is at a salient round number, which is the case with the hypertension diagnostic threshold (Kleven & Waseem, 2013). In blood pressure measurement, over-representation of zeros is

common and thought to be driven by use of imprecise measurement tools (Greiver et al., 2019). This behavior is easily incorporated into bunching estimation to avoid overstating the behavioral response.

To quantify discretion at each clinic I compare the clinic’s observed distribution to the distribution that would exist in the absence of bunching. This counterfactual is recovered by fitting a polynomial regression to the empirical density, excluding observations in a range around the threshold $[p_l, p_u]$, where missing and excess mass is observed. To account for round number bunching I add a set of round number fixed effects. Finally, I predict the counterfactual density for the whole distribution and extrapolate into the bunching region, excluding the contribution of round number fixed effects.

To improve fit of the bunching counterfactual I use a different specification for clinics with a large amount of rounding to zero and for clinics with little rounding to zero. To classify clinics I compare the expected (10%) to observed share of observations that end in zero. I define ‘high rounding’ clinics as those where end digit zeros make up at least 20% of the clinic’s total blood pressure measurements. All other clinics are ‘low rounders’. Results are not sensitive to this classification, and are similar if a 15% or 30% threshold is used.

I collapse data into bins of 2 and 5 units of blood pressure for high and low rounding clinics, respectively. For each clinic I separately estimate:

$$n_{c,i} = \sum_{k=0}^K \beta_k p_i^k + \sum_{r \in R} \rho_r * 1 \left[\frac{p_i}{r} \in N \right] + \sum_{p=p_l}^{p_u} \gamma_b * 1\{y_i = t\} + \epsilon_i \quad (3.1)$$

Where $n_{c,i}$ is the number of observations in bin i at clinic c , k is the order of the polynomial,¹⁰ p_i is the blood pressure mid point in bin i , $[p_L, p_U]$ is the excluded region of the PDF (area that is affected by bunching), t is the diagnostic threshold (140), and $1\{y_i = t\}$ are indicators for if bin i is in the excluded region around the threshold. N is the set of natural numbers, R is a vector of numbers that end in zero, omitting 140 (the threshold) $R = \{8, 9, 10, 11, 12, 13, 15, 16, \dots, 20\}$.

The estimate of the counterfactual distribution is the predicted values from (1) after omitting the contribution of the dummies around the threshold, but not omitting the contribution of other round-number dummies. (Kleven & Waseem, 2013). To quantify the magnitude of discretion, I compare the predicted density, (1) to the observed empirical density. Figure 3.3 shows two examples of bunching estimation.

Specifically, the bunching estimator returns the number of missing, if negative, or excess, if positive, observations at the diagnostic threshold bin divided by the height of the counterfactual. This normalization allows for comparison of magnitudes across clinics with different

¹⁰For estimation I use the *bunching* package in R (Mavrokonstantis, 2019). I use the package’s default polynomial of 9. Results are robust to instead using 7 or 8 but change when a lower order polynomial is used, which results in excessive smoothing and is not recommended. I also specify that the policy threshold is the minimum of its bin, because the diagnostic threshold is $BP \geq 140$. That is, for low rounding clinics bins are constructed as $\{\dots[135, 139], [140, 144], [145, 149]\dots\}$

N visits. The bunching parameter of interest for each clinic, c is:

$$b_c = \frac{N \text{ observed patients} - N \text{ patients in counterfactual}}{N \text{ patients in counterfactual}}$$

Each term is measured in the bin containing the diagnostic threshold (140 mmHg). A clinic with $b_c = -0.4$, for example, has 40% of the expected mass at the threshold is missing. Standard errors are obtained by bootstrapping the bunching estimation 100 times per clinic.

3.3.1.3 Identifying Assumptions for the Bunching Approach

To recover a valid counterfactual for bunching estimation three assumptions must be met (Blomquist et al., 2017). First, the counterfactual distribution must be bounded and well-behaved. Because blood pressure is a biological measurement this assumption is easily satisfied. See figure B.1 for an example of the expected distribution of blood pressure, in the absence of bunching or diagnostic incentives.

The second assumption is that missing mass is not caused by selection. It is recommended that blood pressure is measured at every primary care visit making selection into testing unlikely. Blood pressure is measured along with vital signs (e.g. pulse, temperature) as a general marker of patient health, and because hypertension lacks symptoms so screening is crucial. This fact is reflected in the electronic health records data, where before any restrictions, 98% of visits contain a blood pressure measurement (table B.1).

The third assumption is that manipulation is one-sided and bounded, that is, providers are only rounding blood pressure from above to below the threshold and only for individuals in some range of blood pressure. Because hypertension status is only determined by one threshold, there is no reason for providers to systematically round blood pressure when far from the threshold, providing support for the boundedness assumption.

Here, the direction of rounding is unobservable at the patient-level, however, previous work and the institutional setting provide support for one-sided manipulation. Researchers have compared mean blood pressure in clinics with high end digit bias (where noisy measurement and thus rounding are more likely) to clinics with low end digit bias and concluded that blood pressure, if rounded or measured imprecisely, is systematically rounded down (Greiver et al., 2019; Myers et al., 2011). In addition, it is well established the blood pressure is on average higher when measured in clinic compared to at the patient's home (Muntner Paul et al., 2019; Myers et al., 2012). Indeed, Chile's clinical practice guidelines for hypertension management lists errors that commonly lead to incorrect blood pressure measurement (MINSAL, 2017b). Of the 8 common errors listed, 7 lead to an overstatement of blood pressure which can be as large as 20 mmHg, and only one leads to an understatement of blood pressure, of 7 mmHg (see appendix table B.2 for more details).

The payment scheme in Chile also suggests rounding down. Several papers have shown that healthcare providers exhibit lower effort when paid a fixed salary compared to other payment models such as fee for service or performance based financing (Das et al., 2016; Gertler & Vermeersch, 2013). Thus, in a salaried setting like Chile's public health care

system, providers may avoid diagnosing a marginal patient. While some providers also receive performance bonuses the performance indicators do not include *diagnosis* of hypertension.

For these reasons it is unlikely, but not impossible that providers manipulate blood pressure upwards. If downwards manipulation was masking some upwards manipulation then this is another reason that the behavioral response estimated here using bunching would be a lower bound on the true magnitude of discretion at each primary care clinic.

3.3.2 Variable Construction

The goal of this analysis is to understand the impact of health care providers' discretion on the quality of hypertension diagnosis decisions. Due to a limited sample size, bunching, which is our measure of discretion, must be measured at the clinic level and not the provider level. After measurement, the magnitude of bunching at the clinic level can be parameterized in several ways. My preferred specification examines the intensive margin: I construct a continuous variable equal to $b_c * -1$ if bunching was statistically significant at the 5% level, zero otherwise. For easier interpretation, this variable is then normalized across clinics to mean zero, standard deviation one.

For sensitivity analyses I construct three other variables that measure discretion. First, the magnitude of bunching for all clinics (regardless of statistical significance), normalized to mean zero, standard deviation one. Second, the magnitude of bunching for clinics with negative bunching (regardless of statistical significance), normalized to mean zero, standard deviation one. Last, I also examine the extensive margin by constructing an indicator for if the clinic's bunching estimate is negative and statistically significant, zero otherwise. In each of the independent variables constructed, a higher value indicates more excess mass below the diagnostic threshold and missing mass above it. The four independent variables yield very similar estimated effects. Figures B.3, B.4, and B.5 display coefficients and 95% confidence intervals for all main outcomes estimated with each of the four independent variables.

I study three groups of outcome variables: clinical decisions, hospitalizations, and patient characteristics.

Clinical decisions studied are the diagnosis of hypertension and prescription of anti-hypertensive medication. Hypertension diagnosis appears as a column in the EHR data, and is an indicator equal to 1 for a new diagnosis, 0 otherwise. Prescriptions are retrieved from the national medications database and linked to the EHR by patient ID and date. I construct a variable equal to 1 if a new prescription for anti-hypertension medication was written on the same date as the patient's primary care visit. Importantly, patients do not need to fill the prescription for it to appear in the medications database.

The main health outcomes of interest are hospitalization for stroke, heart attack, or congestive heart failure as these are directly affected by untreated high blood pressure. Because these are measured within 3, 6, or 12 months of the patient's diagnostic visit, they are also a measure of cardiovascular risk. Other health outcomes studied include non-hypertension related hospitalization, and hospitalization for type 2 diabetes. These two variables exclude any hospitalization for stroke, heart attack, or heart failure. The cause of

hospitalization was identified with ICD-10 codes for primary and secondary diagnosis (see table B.3 for details). For each diagnosis grouping I then create indicators equal to 1 if the hospitalization occurred within 3, 6, or 12 months of the hypertension screening visit, zero otherwise.

Last, I examine patient characteristics that are available in the EHR. I focus on characteristics both observable to the provider and potentially informative of cardiovascular risk: patient sex, obesity (as measured by body mass index ≥ 35 , previous diagnosis with type 2 diabetes, and age greater than or equal to 55 and 65 years.

3.3.3 Measuring the Impact of Discretion

To understand the impact of clinical discretion I use a difference-in-differences framework: I compare outcomes in high vs. low discretion clinics and below vs. above the clinical threshold. The coefficient of interest is β_3 in the following regression:

$$Y_{i,c} = \alpha + \beta_1 Discretion_c + \beta_2 BP < 140_i + \beta_3 (Discretion_c * BP < 140_i) + \lambda_i + \gamma_c + \epsilon_{ic} \quad (3.2)$$

Where $Discretion_c$ is the normalized excess mass in clinic c , $BP < 140_i$ is an indicator for if patient i 's recorded systolic blood pressure is less than 140 mmHg, λ_i is a vector of patient characteristics, and γ_c are clinic controls.

My preferred specification includes fixed effects for calendar year, quarter, and type of clinic (i.e. urban primary care clinic, rural primary care clinic, or low-complexity hospital) in all regressions. Regressions for hospitalization outcomes additionally include fixed effects for male and 1-year age groups. Standard errors are robust and clustered at the clinic level.

This approach takes advantage of the fact that the public healthcare system in Chile does not act as a competitive market. Instead, individuals are assigned to a primary care clinic based on their address of residence. The national insurer, FONASA, keeps a list of all enrollees and their assigned clinic. Because the public health care system does not act as a competitive market, clinics have no incentive to enroll more patients. Therefore, patients are quasi-randomly assigned to be exposed to varying levels of discretion.

Figure 3.5 provides a test of the assumption that the magnitude of clinics' discretion is uncorrelated with patient, clinic, and municipality characteristics. Specifically, I regress each characteristic separately on the continuous discretion variable. We see that these characteristics are not significantly predictive of the magnitude of discretion, and that more discretion does not appear to be systematically related to patient health or risk as measured by age, previous diagnoses, and risk factors such as obesity.

3.4 Results

3.4.1 Evidence of Discretion in Blood Pressure Measurement

We begin by examining the magnitude and variation of bunching, our measure of health care providers' propensity to use their discretion in making clinical decisions for hypertension. Figure 3.3 shows that for some clinics there is clear bunching just below the diagnostic threshold (e.g. clinic B, where BP=139 is highly over-represented), while for other clinics there is little to no bunching (e.g. clinic G).

I quantify the amount of discretion separately for each clinic using equation (1). Among the 257 clinics I study, negative bunching is observed in 140 clinics, and is statistically significant in 24 of those. Negative bunching indicates missing mass at and just above the threshold, meaning providers are rounding blood pressure down. Positive bunching is observed in the remaining 117 clinics, but is statistically significant in only 2. I do not focus on positive bunching for several reasons. First, these two clinics appear to be outliers. Second, positive bunching indicates there is excess mass in the 140-144 bin, and missing mass to the right of that. Because all blood pressure readings at or above 140 indicate the same clinical action, that is, to diagnose blood pressure, bunching at 140 is not as clinically meaningful and likely reflects rounding instead of discretion in clinical decision-making.

Figure 3.4 displays the distribution of the magnitude of discretion at the clinic level for clinics with statistically significant negative bunching. The median clinic has a b value of -0.215, with values ranging from -0.06 to -0.622. We interpret this as, depending on the primary care clinic, up to 62.2% of patients who are predicted to have blood pressure at or just above 140 mmHg - and who therefore should be diagnosed with hypertension - are instead recorded as having blood pressure under the diagnostic threshold.

These results indicate a large behavioral response to the diagnostic threshold for hypertension at some primary care clinics, and no behavioral response at others. These are the first estimates, to my knowledge, of a behavioral response to a diagnostic threshold - of which there are many highly important ones in medicine.

The observed pattern of bunching is a combination of several possible behaviors. In the simplest case, the provider measures blood pressure ≥ 140 but writes down a number under 140 in the EHR. Or, the provider may have taken multiple blood pressure measurements for the same patient during the same visit. The 'final' BP measurement recorded in the EHR might be an average of multiple readings, or if the provider retests to get to a single number they are satisfied with. Averaging at least two measurements is recommended in Chile's clinical practice guidelines as a method of obtaining a more precise signal of patients' health (MINSAL, 2017b). Unfortunately, only the final decision is present in the electronic health record. Still, in all these cases providers are faced with at least one positive test result, and ultimately record a blood pressure reading below 140 for the patient, indicating they have used their discretion to change the test result. Last, the observed pattern of bunching indicates providers are behaving this way for some patients, and for some values of BP (e.g. the area of missing mass, which is approximately 140-150 mmHg).

3.4.2 Impact of Discretion on Clinical Decisions

Given providers' behavioral response to the diagnostic threshold, we next examine the impact of this behavior on clinical decisions. Table 3.4 displays the impact of discretion on the probability of a new hypertension diagnosis or medication prescription. The coefficient of interest is the interaction between the magnitude of clinics' discretion and the indicator for blood pressure recorded below the diagnostic threshold (Bunch $\times$ BP < 140). Importantly, if providers are perfectly following clinical practice guidelines, there will be no diagnoses or prescriptions below BP=140. And, if providers are following guidelines similarly at high and low discretion clinics, we expect to see a coefficient near zero on the interaction term between discretion and BP < 140.

Indeed, the coefficients on the interaction term for both diagnosis and prescription are each close to zero and not statistically significant, indicating no impact of discretion. That is, conditional on a patient's recorded blood pressure being under 140, providers appear to adhere to clinical practice guidelines similarly at high and low discretion clinics. This suggests that the location of a patient's *recorded* blood pressure (whether manipulated or not) with respect to the threshold is meaningful, and encodes the provider's belief about their risk.

3.4.3 Discretion and Selection by Cardiovascular Risk

When a provider uses their discretion to reassign a positive hypertension test result as negative, they are effectively considering the test a false positive. Untreated high blood pressure raises a patient's risk of adverse cardiac events, in particular stroke, heart attack, and heart failure. Thus, if providers are *incorrectly* reclassifying true positives as false positives we would see a higher rate of cardiac events in the months following the primary care visit in the group affected by discretion.

Instead, we see the opposite. Table 3.5 (row 1) displays a strong negative effect of discretion on the probability of stroke or heart attack. A one standard deviation increase in discretion is associated with a lower probability of stroke hospitalization of 0.02 to 0.06 percentage points, measured within 3 to 12 months respectively (p -value < 0.01). Given baseline hospitalization rates these estimates are large and translate to a relative reduction of 62.1% and 44.5% respectively. A similar pattern is seen for heart attack, where a one standard deviation increase in discretion is associated with a probability of hospitalization that is 0.028 and 0.036 percentage point lower, within 3 or 12 months, respectively (p -value < 0.05). We observe no effect of discretion on the probability of hospitalization with congestive heart failure. All models are adjusted for patient sex and age, in addition to the controls in equation (1). Results are also robust to the inclusion of controls for type 2 diabetes, dyslipidemia, and family history of cardiovascular disease.

Importantly, I do not find evidence that discretion leads to an *overall* lower number of cardiovascular hospitalizations. Instead, the reduction in the probability of hospitalization among patients classified as hypertension negative at high discretion clinics (table 3.5, row

1) is offset by an increase in the probability of hospitalization of a similar magnitude among patients classified as hypertension positive at the same clinics (table 3.5, row 2). These results are unsurprising: as providers at high discretion clinics reassign some *low* cardiovascular risk patients to the hypertension negative group, this leads to a higher concentration of *high* cardiovascular risk patients in the hypertension positive group at high discretion clinics (compared to low discretion clinics), reflected by the coefficients in row 2 of table 3.5.

These results suggest that at clinics where providers are more likely to use their discretion when diagnosing hypertension, there is a more accurate sorting of patients with respect to the diagnostic threshold: higher risk patients are more likely to be classified as hypertension positive, and low risk patients are more likely to be classified as hypertension negative, but there is no net change in the number of hospitalizations. These results raise several questions, namely, how do providers differentiate between true and false positive tests for hypertension? Is it that providers are sorting overall healthier patients under the threshold, or do they have some particular insight about cardiovascular risk? Second, are providers at high discretion clinics simply providing an overall higher quality of care?

To answer these questions I estimate the impact of discretion on: 1) the probability of non-hypertension related hospitalization (all hospitalizations excluding stroke, heart attack, and heart failure), a measure of the overall health of patients and 2) the probability of hospitalization for type 2 diabetes causes, among patients who have been previously diagnosed with type 2 diabetes.¹¹ While an indirect measure of quality of care, if discretion is an indicator of overall clinical skill we would expect patients affected by discretion to also have fewer hospitalizations for their non-hypertension related conditions.

Table 3.6 examines the impact of discretion on hospitalizations for any cause other than stroke, heart attack, and heart failure. Results fluctuate around zero and none are statistically significant. We find similar results for hospitalizations related to type 2 diabetes, among patients diagnosed with this disease (table 3.6, columns 4-6); no statistically significant estimates, and both positive and negative coefficients for hospitalization within 3, 6, and 12 months. These results do not provide conclusive evidence that discretion leads to overall more healthy patients being sorted under the threshold, or, that providers at high discretion clinics are more effective at treating diseases other than hypertension.

To definitively conclude that discretion leads to a more accurate sorting of patients on either side of the diagnostic threshold in terms of health I would ideally measure effects on patient mortality. One limitation of this analysis is that mortality is unobserved. To address this limitation I re-estimate the effect of discretion on cardiovascular hospitalizations among patients who are known to be alive for at least 3, 6, or 12 months after their initial primary care visit. For this group admission to the hospital for a cardiovascular event is considered the most severe adverse health outcome related to high blood pressure.

¹¹The group of hospitalizations included here are mutually exclusive to stroke, heart attack, and heart failure. It is not possible to do a similar exercise for the other patient groups - dyslipidemia, smokers, or patients with a family history of cardiovascular disease - because the main health concerns among these populations are also stroke, heart attack, and heart failure.

In practice, if a patient has another primary care visit or is hospitalized after the initial visit, we can conclude they were alive during the interim. 85.4% of patients are re-observed 3 months after their first visit, 82.7% within 6 months, and 77.0% within 12 months (table B.4). Patients who are not observed during this time period may have left the country, transitioned to private health care, simply not had a scheduled appointment nor fallen ill, or may have died, however we cannot distinguish between these events. Retention in the sample is unrelated to clinic level discretion B.5, and so in tables B.6 and B.7 each model is estimated conditional on the patient being re-observed during the relevant time frame.

First, I find no impact of discretion on all-cause hospitalization among the retained sample (table B.6). Second, after conditioning on retained patients, the effect of discretion on cardiovascular hospitalization is similar to when the full sample is used: a one standard deviation in discretion is associated with a lower rate of stroke hospitalization of 0.02 to 0.06 within 3 to 6 months. For heart attack the results are also similar: a one standard deviation in discretion is associated with a lower rate of stroke hospitalization of 0.02 to 0.03 within 3 to 6 months. While the magnitudes are very similar to when the full sample is used (see table 3.5), the estimates are slightly less precise which is likely driven by the smaller number of observations used here. Overall, these results confirm that discretion leads to a more accurate sorting of patients with respect to the diagnostic threshold.

3.4.4 Discretion and Selection by Patient Characteristics

So far we have seen that providers at high and low discretion clinics are equally likely to follow diagnostic guidelines for hypertension, and yet, patients with lower cardiovascular risk are more likely to be accurately classified as hypertension negative at high discretion clinics. These two facts can both be true if providers use their discretion to identify false positives - who have lower true cardiovascular risk - and sort these patients below the threshold.

How do providers distinguish between true and false positives? One potential method is the use of the representative heuristic. When deciding if a patient belongs in the group of false positives providers may, instead of calculating conditional probabilities which is mentally taxing, judge the similarity of the patient to someone who is representative of low cardiovascular risk (Tversky & Kahneman, 1974). To test for this, I use the difference-in-differences framework to examine if key patient characteristics are over-represented below the diagnostic threshold at high discretion clinics. Specifically, I examine characteristics that are 1) easily observed by the provider, and 2) that would be present in a patient representative high cardiovascular risk. These are male sex, previous diagnosis with type 2 diabetes, severe obesity (defined as body mass index ≥ 35), age 55 and older, and age 65 and older.

Table 3.7 examines the impact of discretion on patient characteristics that are correlated with cardiovascular risk, interpreted as selection with respect to the threshold. At high discretion clinics, males are 1.6% less likely to be classified as hypertension negative. Older patients are also less likely to be classified as hypertension negative at high discretion clinics: 1.2% and 1.5% for age thresholds of 55 and 65, respectively. We observe no significant effect on the probability a patient is severely obese and previously diagnosed with type 2 diabetes.

The age thresholds of 55 and 65 were used because they are likely to be salient to providers. While age is not explicitly in the clinical practice guidelines for diagnosing hypertension, the thresholds of 55 and 65 appear frequently in Chile’s guidelines for managing other cardiovascular diseases (MINSAL, 2017b). Similar effects were not seen for ages 50, 60, and 70 (table B.9), which are less salient age cut offs in this setting. These results are consistent with providers using the representative heuristic to differentiate between a true and false positive test for hypertension, however because of the small point estimates, heuristic thinking is likely only one of the ways providers make this decision.

3.5 Conclusion

This paper documents health care providers’ behavioral response to the diagnostic threshold for hypertension. This behavior varies across primary care clinics in Chile, where up to 62% of patients’ blood pressure measurements near the diagnostic threshold were manipulated. Providers selectively round some patients’ blood pressure down, turning a positive test into a negative. I argue that the uncertainty in blood pressure measurement means discretionary decision-making, where a provider does not only rely on rules such as clinical practice guidelines, is highly important.

Using the fact that primary care clinic is assigned by patient’s home address in Chile’s public health care system, I estimate the impact of discretion on clinical decisions and hospitalizations. Comparing high and low discretion clinics reveals that providers adhere to clinical practice guidelines similarly. Yet, patients at high discretion clinics who are recorded as having a negative test are less likely to be hospitalized with adverse cardiovascular events – the highest risks from uncontrolled high blood pressure. These results are large: high discretion is associated with a 9% fewer hospitalizations for heart attack, and a 44% fewer hospitalizations for stroke within a year.

The estimated difference in the rate of adverse health outcomes is likely driven by two mechanisms. Most importantly, at high discretion clinics, false positives are more likely to be (correctly) labelled as a negative test result. This leaves more true positives above the threshold. Notably, correctly classifying patients in this way would lead to the large negative effects seen, whether the *overall* level of hospitalization at high discretion clinics changed or not.

Second, because over-treatment with prescription drugs can be harmful, not treating false positive patients could also have direct benefits. Recent work has shown there can be harms to over-prescribing: in the United States 12 in 1000 people 65 and up were hospitalized for medication harms, and this was largely driven by therapeutic medicines for chronic diseases (Budnitz et al., 2021). For this reason, and because patients are more likely to adhere to simpler treatment regimes, policymakers have recently turned to treatment simplification and de-prescribing as one method to improve health, especially for cardiovascular diseases (Böhm et al., 2021; Elnaem et al., 2020; Vrijens et al., 2017).

How do health care providers distinguish between a true and a false positive? While I can't definitively answer this question, I provide support for heuristic thinking as one component of discretionary decision-making and rule out several other methods. First, providers don't appear to use cardiovascular history questions to make this decision. They also don't seem to be providing an overall better standard of care, as there is no effect of discretion on diabetes-related hospitalizations among patients previously diagnosed with type 2 diabetes. Instead, they appear to rely on heuristics to help identify false positive patients. Specifically, they are less likely to manipulate the positive test result of patients with characteristics representative of high cardiovascular risk. In this case, providers' cognitive shortcuts add information to their decision process compared to just using the clinical practice guidelines.

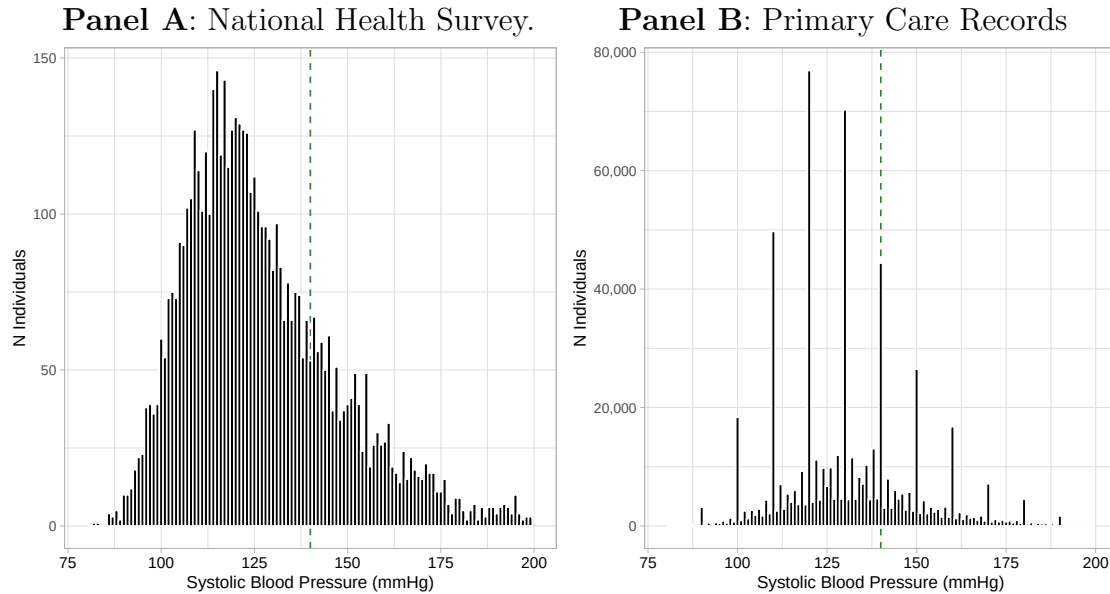
An important limitation of this analysis is the inability to understand precisely how providers identify false positive patients. It is clear that providers add more dimensions to the decision process than is suggested in the clinical practice guidelines, however the small effect sizes indicate this is not the whole story. Future work, in particular work using standardized patients which allows for an understanding of exactly what happens in a visit with a provider, could be useful here. For example, the standardized patients could record what kinds of questions the provider asks but does not record in the electronic health record, or the number of times blood pressure is re-measured.

A second limitation is the necessity of measuring discretion at the clinic level instead of the provider level. This is done for two reasons: first, because of the large data requirements for bunching estimation. Second, patients cannot choose the public, primary care clinic they attend but could choose the provider they see within the clinic. Still, by aggregating to the clinic level important heterogeneity in the use of discretion or in clinical skill might be masked. However, there are several reasons providers working at the same clinic may behave similarly, which helps to reduce concerns about heterogeneity. For example, providers working at the same clinic likely have access to the same technologies. In addition, previous work has documented that among physicians, peers' and co-workers' practice style is influenced by others in their immediate network (Arnold et al., 2021; Donohue et al., 2018; Zheng et al., 2010). As such, the measure of discretion here is a combination of clinical skill and practice style at each primary care clinic.

Many medical decisions are made in 'grey areas' where information about patient health is uncertain. For example, the decision to prescribe antibiotics is usually made before test results return from the lab, and costly test decisions, such as those for heart attack, are made largely based on subjective information from the patient. In both of these cases, research has shown that prediction algorithms could improve quality of clinical decisions (Huang & Ullrich, 2021; Mullainathan & Obermeyer, 2021). A noisy biomarker or too-simple clinical practice guidelines also create uncertainty. Here, I find that healthcare providers' mental models partially overcome this by adding in relevant information that is not in clinical practice guidelines. As decision support tools including algorithms become more common in medicine, understanding how providers are already thinking, whether this deviates from guidelines, and if that benefits or harms patients, is crucial as we work towards improving the efficiency and equity of healthcare allocation.

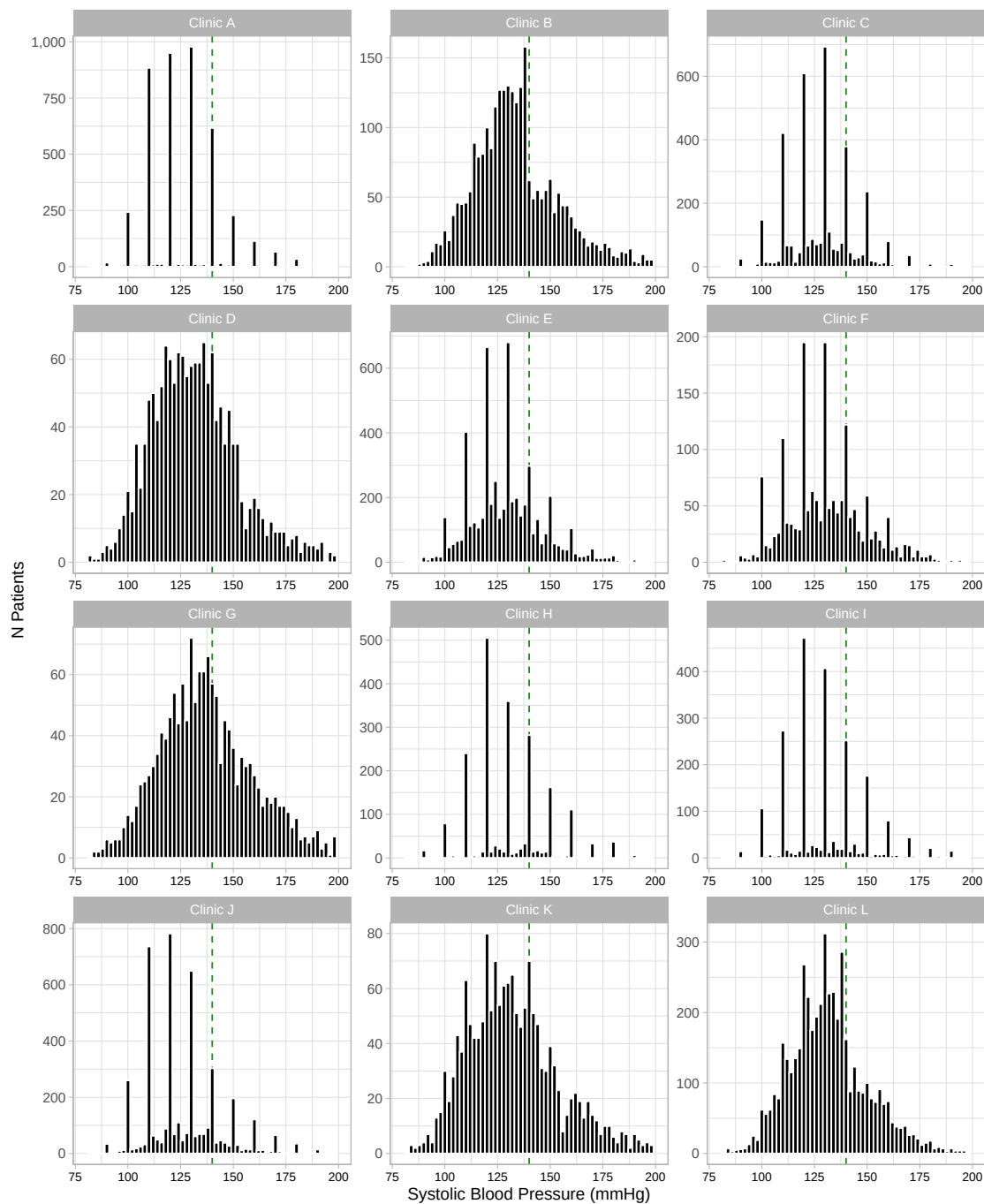
Chapter 3: Tables and Figures

Figure 3.1: Expected vs. observed distribution of systolic blood pressure



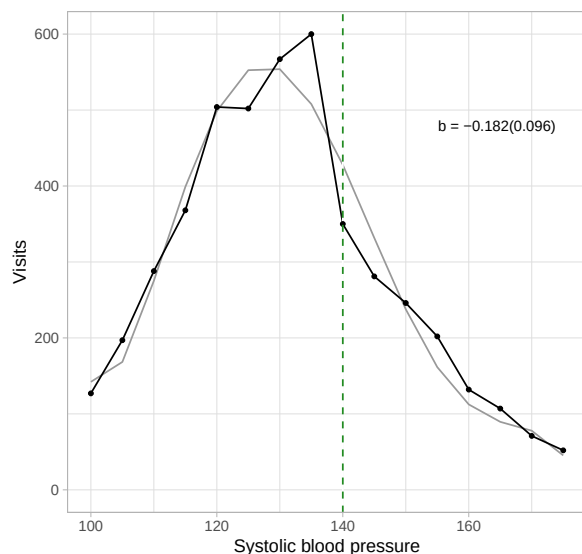
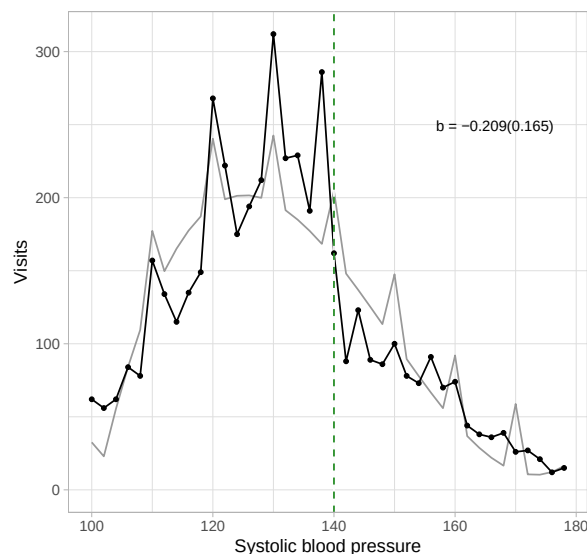
Note: Comparison of expected (panel A) to observed (panel B) distribution of systolic blood pressure in two large populations. Panel A shows the distribution of systolic blood pressure from Chile's nationally representative health survey (ENS 2016-2017). Panel B shows the distribution of systolic blood pressure from primary care electronic health records (2016-2019). Substantial end digit bias, or rounding to zero, is seen, as well as some missing mass to the right of the diagnostic threshold. Both distributions exclude individuals who have been diagnosed with hypertension before this instance of blood pressure was measured.

Figure 3.2: Examples of the distribution of systolic blood pressure at 12 clinics



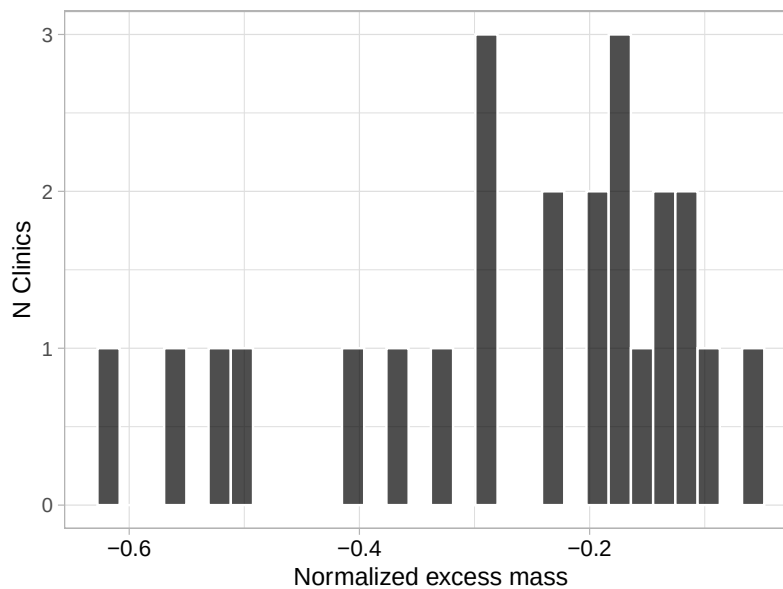
Note: Distributions of systolic blood pressure from 12/257 primary care clinics from preventative care visits among patients who were not previously diagnosed with hypertension. Clinics A, C, E, F, H, I and J show substantial end digit bias: numbers that end in zero are over-represented. Clinics A, B, E and L show substantial bunching: there is missing mass above 140 and excess mass below it.

Figure 3.3: Example of counterfactual bunching estimation

Panel A: Low rounding clinic.**Panel B:** High rounding clinic.

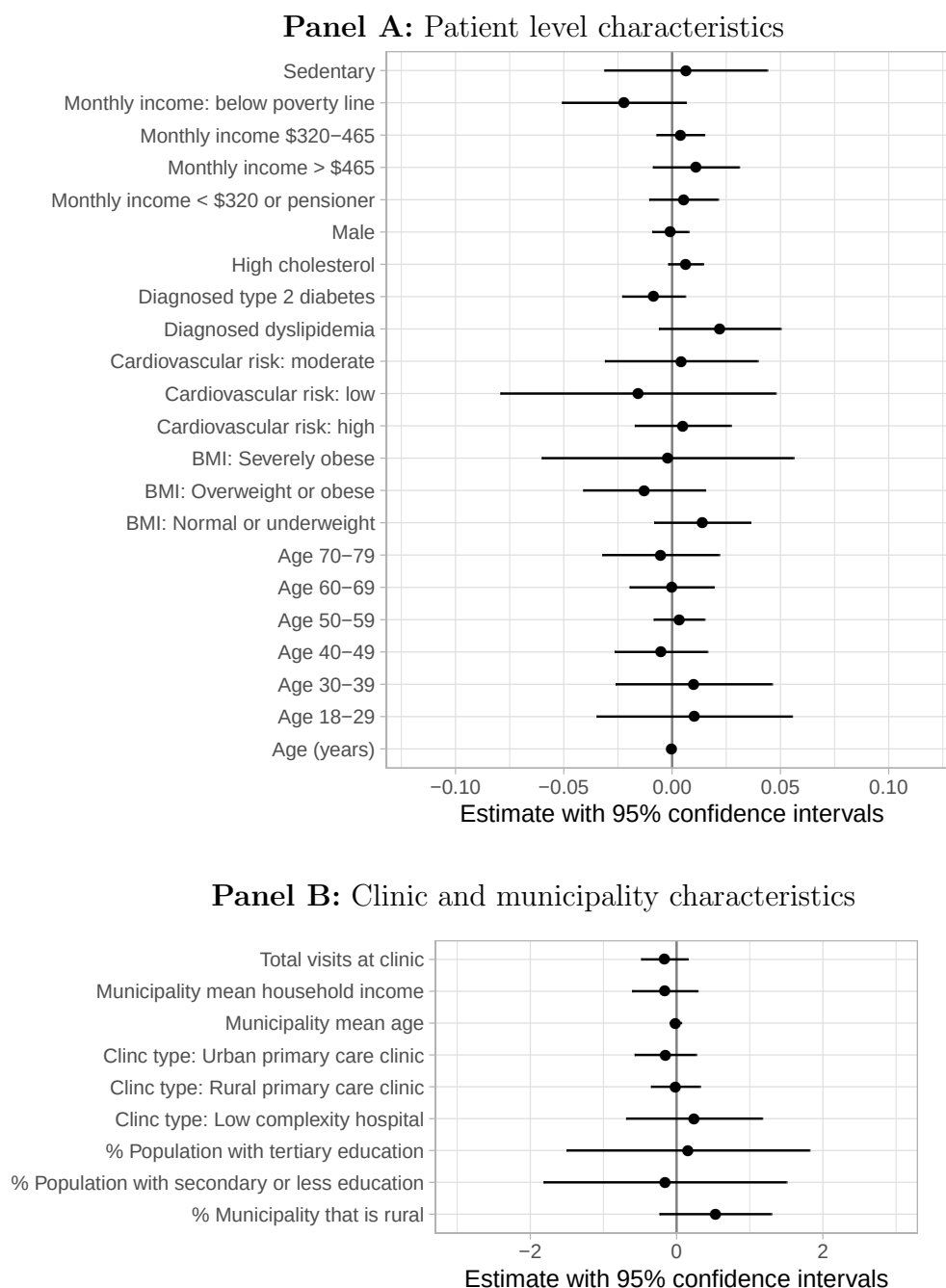
Note: Examples of bunching estimation. The grey line is the counterfactual density from equation (1). Panel A shows a clinic with little rounding to zero. Panel B shows a clinic with more rounding to zero, see as spikes at 110, 120, 130... 150. The green dashed line at 140 shows the diagnostic threshold. The magnitude of bunching (discretion) is displayed on the graph, with the bootstrap standard error in parentheses. A negative number indicates missing mass to the right of the threshold, and excess mass below it. Each b estimate is the number of missing (or excess, if positive) observations at the threshold, normalized by dividing the total excess mass by the height of the counterfactual at the threshold. This allows for comparison of magnitudes across clinics with different N visits. Estimated using the bunching package in R (Mavrokonstantis, 2019).

Figure 3.4: Estimated magnitude of discretion of blood pressure



Note: Histogram of the magnitude of discretion (estimated using bunching) for 24 of 257 clinics. Only clinics with statistically significant negative bunching are included. A negative number indicates missing mass to the right of the threshold, and excess mass below it. Each b estimate is the number of missing (or excess, if positive) observations at the threshold, normalized by dividing the total excess mass by the height of the counterfactual at the threshold. This allows for comparison of magnitudes across clinics with different N visits. Estimated using the bunching package in R (Mavrokonstantis, 2019).

Figure 3.5: Balance of patient and clinic characteristics by magnitude of discretion



Note: This figure tests for quasi-random assignment of patients to clinics by magnitude of discretion. Point estimates and 95% confidence intervals shown from the regression of clinics' continuous and standardized bunching magnitude on (A) patient characteristics, and (B) clinic and municipality characteristics. Each characteristic is from a separate regression and includes year and quarter fixed effects, with standard errors clustered at the clinic level.

Table 3.1: Patient reasons for admission to cardiovascular health program

	N	%
History of CVD	393,657	63.5
Dyslipidemia	285,304	46.0
Type 2 diabetes	177,083	28.6
Smoker and over 55	9,643	1.6
Unknown	52	0.0
Unique patients	619,907	

Note: Patients are enrolled in Chile's Cardiovascular Health Program (*Programa Salud Cardiovascular*) if they have one or more of the characteristics in the table. Categories are not mutually exclusive. Patients in the Cardiovascular Program with previously diagnosed hypertension were excluded.

Table 3.2: Summary statistics, clinic and municipality characteristics

Statistic	Mean	St. Dev.
Household income per capita	440.827	229.015
Log visits per clinic	7.938	0.460
Mean municipality age	36.521	2.343
Secondary or primary education	0.865	0.088
% of municipality that is rural	0.148	0.173
Tertiary education	0.135	0.088
Urban primary care clinic	0.825	0.381
Rural primary care clinic	0.089	0.286
Low complexity hospital	0.058	0.235
Literacy rate in municipality	0.966	0.024
Employment rate in municipality	0.525	0.056

Note: Mean and standard deviation of clinic- and municipality-level characteristics for 257 clinics. Municipality-level variables come from Chile's Socio-economic Survey, and are linked to clinics based on their location. Household income per capita is monthly a municipality mean, and in 2020 USD. St. Dev. stands for standard deviation.

Table 3.3: Summary statistics, patient characteristics

Statistic	Mean	St. Dev.
Female	0.618	0.486
Mean age	57.711	13.087
Age 18-29	0.031	0.173
Age 30-39	0.069	0.253
Age 40-49	0.172	0.377
Age 50-59	0.263	0.440
Age 60-69	0.264	0.441
Age 70-79	0.201	0.401
Body mass index	30.714	6.013
Normal or underweight BMI	0.211	0.408
Overweight BMI	0.375	0.484
Obese BMI	0.482	0.500
Severely obese BMI	0.067	0.251
Type 2 Diabetes	0.286	0.452
Dyslipidemia	0.641	0.480
High cholesterol	0.351	0.477
Cardiovascular risk: high	0.271	0.445
Cardiovascular risk: moderate	0.168	0.374
Cardiovascular risk: low	0.129	0.335
Sedentary	0.557	0.497
Monthly income: below poverty line	0.266	0.442
Monthly income <\$320 or pensioner	0.435	0.496
Monthly income \$320-465	0.122	0.327
Monthly income >\$465	0.171	0.376

Note: Mean and standard deviation of patient-level characteristics for 619,907 patients. BMI stands for body mass index. Type 2 diabetes and dyslipidemia are previously diagnosed. Cardiovascular risk is assigned by the healthcare provider using a modified Framingham 10-year cardiovascular risk algorithm. Sedentary is an indicator for if the provider noted in the electronic medical record the patient leads a sedentary (little physical activity) lifestyle. Monthly income categories come from the patient's FONASA insurance level. All variables are indicators except for age and body mass index.

Table 3.4: Impact of discretion on hypertension diagnosis and prescription

	Hypertension Diagnosis	Hypertension Prescription
	(1)	(2)
Discretion x BP<140	−0.008 (0.005)	−0.019 (0.014)
Discretion	−0.006 (0.006)	0.007 (0.010)
BP<140	−0.251*** (0.005)	−0.169*** (0.005)
Observations	619,907	619,907
Mean dep. var.	0.522	0.376
Clinics	257	257

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents difference-in-differences estimates of the impact of crossing the diagnostic threshold from above to below 140 mmHg at high vs. low discretion clinics. BP < 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of discretion at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Each dependent variable is an indicator for if the patient was hospitalized with the listed condition, within 3, 6 or 12 months of their primary care visit, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for year and quarter of primary care visit, and type of primary care clinic. Hypertension diagnosis is an indicator for if the patient was newly diagnosed. Hypertension prescription is an indicator for if the patient was prescribed a new hypertension medication.

Table 3.5: Discretion and cardiovascular hospitalization 3, 6, and 12 months after primary care visit

	Stroke			Heart Attack			Congestive Heart Failure		
	≤ 3 mo.	≤ 6 mo.	≤ 12 mo.	≤ 3 mo.	≤ 6 mo.	≤ 12 mo.	≤ 3 mo.	≤ 6 mo.	≤ 12 mo.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Disc. x BP < 140	-0.018** (0.009)	-0.039*** (0.012)	-0.061*** (0.019)	-0.028** (0.012)	-0.040*** (0.013)	-0.036* (0.019)	0.003 (0.006)	0.001 (0.010)	-0.001 (0.013)
Discretion	0.018** (0.009)	0.039*** (0.012)	0.060*** (0.017)	0.022* (0.012)	0.033** (0.014)	0.030 (0.020)	-0.005 (0.005)	-0.008 (0.008)	-0.011 (0.010)
BP < 140	-0.029*** (0.005)	-0.061*** (0.009)	-0.107*** (0.014)	-0.025*** (0.009)	-0.062*** (0.013)	-0.119*** (0.019)	0.010 (0.007)	0.013 (0.010)	0.003 (0.015)
Observations	619,907	619,907	619,907	619,907	619,907	619,907	619,907	619,907	619,907
Mean dep. var. (%)	0.029	0.066	0.137	0.109	0.205	0.385	0.062	0.124	0.232
Clinics	257	257	257	257	257	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences estimates of the impact of crossing the diagnostic threshold from above to below 140 mmHg at high vs. low discretion clinics. BP < 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of discretion at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. All dependent variables in this table are multiplied by 100 and can be interpreted as percentages. Each dependent variable is an indicator for if the patient was hospitalized with the listed condition, within 3, 6 or 12 months of their primary care visit, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit, and type of primary care clinic. Stroke includes any cerebral infarction (ICD-10 codes I63). Heart attack includes acute coronary syndrome, myocardial infarction, and any other condition associated with sudden, reduced blood flow to the heart (I20-22, I24-25). Congestive heart failure is I50.

Table 3.6: Discretion and non-cardiovascular hospitalization 3, 6, and 12 months after primary care visit

	Non-cardiovascular hospitalization			Type 2 diabetes hospitalization		
	≤ 3 mo.	≤ 6 mo.	≤ 12 mo.	≤ 3 mo.	≤ 6 mo.	≤ 12 mo.
	(1)	(2)	(3)	(4)	(5)	(6)
Discretion x BP < 140						
	-0.028 (0.051)	0.010 (0.065)	-0.027 (0.100)	-0.047 (0.105)	0.038 (0.160)	0.114 (0.216)
Discretion	-0.004 (0.047)	-0.039 (0.063)	0.017 (0.092)	0.066 (0.102)	0.007 (0.160)	-0.152 (0.230)
BP < 140	0.008 (0.048)	-0.125* (0.065)	-0.311*** (0.088)	0.003 (0.091)	-0.228* (0.124)	-0.480*** (0.163)
Observations	619,907	619,907	619,907	177,083	177,083	177,083
Mean dep. var. (%)	2.374	4.603	8.642	2.655	5.132	9.568
Clinics	257	257	257	254	254	254

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences estimates of the impact of crossing the diagnostic threshold from above to below 140 mmHg at high vs. low discretion clinics. BP < 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of discretion at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. All dependent variables in this table are multiplied by 100 and can be interpreted as percentages. Each dependent variable is an indicator for if the patient was hospitalized with the listed condition, within 3, 6 or 12 months of their primary care visit, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit, and type of primary care clinic.

Table 3.7: Selection by patient characteristics representative of high cardiovascular risk

	Male	Type 2 Diabetes	Severe Obesity	Age 55+	Age 65+
	(1)	(2)	(3)	(4)	(5)
Discretion x BP<140	−0.006*** (0.002)	−0.003 (0.005)	0.002 (0.001)	−0.007*** (0.003)	−0.005*** (0.002)
Discretion	0.008*** (0.002)	−0.005 (0.006)	−0.001 (0.003)	−0.005 (0.005)	−0.008 (0.005)
BP<140	−0.096*** (0.002)	0.007* (0.004)	−0.019*** (0.001)	−0.141*** (0.004)	−0.127*** (0.003)
Observations	619,907	619,907	582,880	619,907	619,907
Mean dep. var.	0.382	0.286	0.067	0.596	0.336
Clinics	257	257	257	257	257

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents difference-in-differences estimates of the impact of crossing the diagnostic threshold from above to below 140 mmHg at high vs. low discretion clinics. BP < 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of discretion at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Each dependent variable is an indicator for if the patient has the characteristic listed, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for year and quarter of primary care visit, and type of primary care clinic.

Chapter 4

Conclusion

In this dissertation I study the behavior of both patients and healthcare providers managing chronic conditions. While effective treatments for conditions like hypertension and type 2 diabetes exist, deciding which patients should be treated and encouraging their adherence can be challenging. Many health care providers do not follow clinical practice guidelines, and many patients do not adhere to treatment; understanding why, and how to change these behaviors are fundamental questions in health policy.

In Chapter 2, we show how nudging patients to attend their preventative care appointments increases visits by 5 to 7% per semester, suggesting effective habit formation or learning on behalf of patients. This increase in primary care utilization translates into an 11% increase in health screening. Reminders also lead to a 42% improvement in anti-hypertensive medication adherence, which is largely attributed to patients having any active prescription. Last, SMS reminders lead to a 17% increase in cardiovascular hospitalizations coupled with a reduction in in-hospital mortality, suggesting an increase in referrals through primary care or timely care-seeking behavior of emergency and specialized health services.

These findings have several implications for policy. Most importantly, we show that SMS reminders are effective at scale, even if implemented with variation in compliance. We show that in the context of public, government run health care clinics in Chile, implementation fidelity improved over time suggesting clinic-level learning. As a result, we see that the impacts of the program on visits, medication adherence, and hospitalizations also increase over time.

These findings are particularly important for settings with a gate-keeping healthcare model – where patients must first visit their primary care provider or general practitioner before being referred to speciality care, approved for tests, or prescribed new medication. This model is common in other OECD countries with public health care systems, such as Canada, the United Kingdom, Spain, and Italy (Blöndal & Ásgeirsdóttir, 2019; Brekke et al., 2007; Kringos et al., 2015; Reibling & Wendt, 2012; Rotar et al., 2018; Watt, 1987). In these settings, a light touch intervention such as nudging patients to attend primary care can have potentially large and meaningful impacts because it is intervening on the first step in the cascade of care.

In chapter 3, I study health care providers' decision to diagnose and treat hypertension. I find that providers round the blood pressure of up to 62% of patients who test near the threshold, using their discretion to turn a positive test for hypertension into a negative. This behavioral response leads to a more accurate classification of patients in terms of their cardiovascular risk, as measured by future hospitalizations. I find that discretionary decisions are consistent with heuristic thinking: among patients with identical test results, those with characteristics representative of high cardiovascular risk are less likely to be sorted below the diagnostic threshold.

Many medical decisions are made in 'grey areas' where uncertainty is high. This is partly because decisions often must be made with limited information about patient health. For example, the decision to prescribe antibiotics is usually made before test results return from the lab, and costly test decisions, such as those for heart attack, are made largely based on subjective information from the patient. In both of these cases, research has shown that prediction algorithms could improve quality of clinical decisions (Huang & Ullrich, 2021; Mullainathan & Obermeyer, 2021). A noisy biomarker or too-simple clinical practice guidelines also create uncertainty. Here, I find that healthcare providers' mental models partially overcome this by adding in relevant information that is not in clinical practice guidelines. As decision support tools including algorithms become more common in medicine, understanding how providers are already thinking, whether this deviates from guidelines, and if that benefits or harms patients, is crucial as we work towards improving the efficiency and equity of healthcare allocation.

In conclusion, it is clear that non-adherence to guidelines and to treatment is widespread in medicine. However, with large-scale and granular data, these behaviors are predictable, and in some cases can be modified through interventions like SMS reminders.

Bibliography

- Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2020). Sampling-Based versus Design-Based Uncertainty in Regression Analysis. *Econometrica*, 88(1), 265–296. <https://doi.org/10.3982/ECTA12675>
- Abaluck, J., Agha, L., Jr, D. C. C., Singer, D., & Zhu, D. (2020). Mastering the Art of Cookbook Medicine: Machine Learning, Randomized Trials, and Misallocation, 71.
- Abaluck, J., Agha, L., Kabrhel, C., Raja, A., & Venkatesh, A. (2016). The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care. *American Economic Review*, 106(12), 3730–3764. <https://doi.org/10.1257/aer.20140260>
- Adams, A. S., Kluender, R., Mahoney, N., Wang, J., Wong, F., & Yin, W. (2021). *The Impact of Financial Assistance Programs on Health Care Utilization* (Working Paper No. 29227) [Series: Working Paper Series]. National Bureau of Economic Research. <https://doi.org/10.3386/w29227>
- Adiyaman, A., Tosun, N., Elving, L. D., Deinum, J., Lenders, J. W. M., & Thien, T. (2007). The effect of crossing legs on blood pressure. *Blood Pressure Monitoring*, 12(3), 189–193. <https://doi.org/10.1097/MBP.0b013e3280b083a7>
- Aguilera, A., Schueller, S. M., & Leykin, Y. (2015). Daily mood ratings via text message as a proxy for clinic based depression assessment. *Journal of Affective Disorders*, 175, 471–474. <https://doi.org/10.1016/j.jad.2015.01.033>
- Aguilera, I., Infante, A., Ormeño, H., & Urriola, C. (2015). *Implementation of the Universal Access with Explicit Guarantees (AUGE) reform Chile case Study* (tech. rep.). World Health Organization.
- Alalouf, M., Miller, S., & Wherry, L. (2019). *What Difference Does a Diagnosis Make? Evidence from Marginal Patients* (tech. rep. w26363). National Bureau of Economic Research. Cambridge, MA. <https://doi.org/10.3386/w26363>
- Alcocer, L., Novoa, G., & Sotres, D. (1997). Digit preferences observed in the measurement of blood pressure: Repercussions on the success criteria in current treatment of hypertension. *American Journal of Therapeutics*, 4(9-10), 311–314. <https://doi.org/10.1097/00045391-199709000-00005>
- Ali, M. K., McKeever Bullard, K., Imperatore, G., Barker, L., Gregg, E. W., & Centers for Disease Control and Prevention (CDC). (2012). Characteristics associated with poor glycemic control among adults with self-reported diagnosed diabetes—National Health

- and Nutrition Examination Survey, United States, 2007-2010. *MMWR supplements*, 61(2), 32–37.
- Alsan, M. (2015). The Effect of the TseTse Fly on African Development. *American Economic Review*, 105(1), 382–410. <https://doi.org/10.1257/aer.20130604>
- Arnold, C., Koetsenruijter, J., Forstner, J., Peters-Klimm, F., & Wensing, M. (2021). Influence of physician networks on prescribing a new ingredient combination in heart failure: A longitudinal claim data-based study. *Implementation Science*, 16(1), 84. <https://doi.org/10.1186/s13012-021-01150-y>
- Aron-Dine, A., Einav, L., & Finkelstein, A. (2013). The RAND Health Insurance Experiment, Three Decades Later. *Journal of Economic Perspectives*, 27(1), 197–222. <https://doi.org/10.1257/jep.27.1.197>
- Baicker, K., Mullainathan, S., & Schwartzstein, J. (2015). Behavioral Hazard in Health Insurance. *The Quarterly Journal of Economics*, 130(4), 1623–1667. <https://doi.org/10.1093/qje/qjv029>
- Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., Schneider, E. C., Wright, B. J., Zaslavsky, A. M., & Finkelstein, A. N. (2013). The Oregon Experiment — Effects of Medicaid on Clinical Outcomes [Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMsa1212321>]. *New England Journal of Medicine*, 368(18), 1713–1722. <https://doi.org/10.1056/NEJMsa1212321>
- Baron, R. B. (2018). Treating Blood Pressure Correctly by Measuring It Correctly. *JAMA Internal Medicine*, 178(6), 860–861. <https://doi.org/10.1001/jamainternmed.2018.0311>
- Bastías, G., Pantoja, T., Leisewitz, T., & Zárata, V. (2008). Health care reform in Chile. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne*, 179(12), 1289–1292. <https://doi.org/10.1503/cmaj.071843>
- Bitrán, R., Escobar, L., & Gassibe, P. (2010). After Chile's Health Reform: Increase In Coverage And Access, Decline In Hospitalization And Death Rates. *Health Affairs*, 29(12), 2161–2170. <https://doi.org/10.1377/hlthaff.2010.0972>
- Blomquist, S., Newey, W., Kumar, A., & Liang, C.-Y. (2017). *On Bunching and Identification of the Taxable Income Elasticity* (tech. rep. w24136). National Bureau of Economic Research. <https://doi.org/10.3386/w24136>
- Blöndal, B., & Ásgeirsdóttir, T. L. (2019). Costs and efficiency of gate-keeping under varying numbers of general practitioners [_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hpm.2601>]. *The International Journal of Health Planning and Management*, 34(1), 140–156. <https://doi.org/10.1002/hpm.2601>
- Böhm, A.-K., Schneider, U., Aberle, J., & Stargardt, T. (2021). Regimen simplification and medication adherence: Fixed-dose versus loose-dose combination therapy for type 2 diabetes [Publisher: Public Library of Science]. *PLOS ONE*, 16(5), e0250993. <https://doi.org/10.1371/journal.pone.0250993>

- Boksmati, N., Butler-Henderson, K., Anderson, K., & Sahama, T. (2016). The Effectiveness of SMS Reminders on Appointment Attendance: A Meta-Analysis. *Journal of Medical Systems*, 40(4), 90. <https://doi.org/10.1007/s10916-016-0452-2>
- Boone, C. E., Celhay, P., Gertler, P., Gracner, T., & Rodriguez, J. (2022). How scheduling systems with automated appointment reminders improve health clinic efficiency. *Journal of Health Economics*, 82, 102598. <https://doi.org/10.1016/j.jhealeco.2022.102598>
- Borkhoff, C. M., Hawker, G. A., Kreder, H. J., Glazier, R. H., Mahomed, N. N., & Wright, J. G. (2009). Patients' gender affected physicians' clinical decisions when presented with standardized patients but not for matching paper patients. *Journal of Clinical Epidemiology*, 62(5), 527–541. <https://doi.org/10.1016/j.jclinepi.2008.03.009>
- Borusyak, K. (2021). Revisiting Event Study Designs: Robust and Efficient Estimation, 48.
- Bossert, T. J., & Leisewitz, T. (2016). Innovation and Change in the Chilean Health System. *New England Journal of Medicine*, 374(1), 1–5. <https://doi.org/10.1056/NEJMp1514202>
- Brekke, K. R., Nuscheler, R., & Straume, O. R. (2007). Gatekeeping in health care. *Journal of Health Economics*, 26(1), 149–170. <https://doi.org/10.1016/j.jhealeco.2006.04.004>
- Broad, J., Wells, S., Marshall, R., & Jackson, R. (2007). Zero end-digit preference in recorded blood pressure and its impact on classification of patients for pharmacologic management in primary care — PREDICT-CVD-6 [Publisher: British Journal of General Practice Section: Original Papers]. *British Journal of General Practice*, 57(544), 897–903. Retrieved May 4, 2020, from <https://bjgp.org/content/57/544/897>
- Brown, M. T., & Bussell, J. K. (2011). Medication Adherence: WHO Cares? [Publisher: Elsevier]. *Mayo Clinic Proceedings*, 86(4), 304–314. <https://doi.org/10.4065/mcp.2010.0575>
- Budnitz, D. S., Shehab, N., Lovegrove, M. C., Geller, A. I., Lind, J. N., & Pollock, D. A. (2021). US Emergency Department Visits Attributed to Medication Harms, 2017–2019. *JAMA*, 326(13), 1299. <https://doi.org/10.1001/jama.2021.13844>
- Burkhart, P. V., & Sabaté, E. (2003). Adherence to long-term therapies: Evidence for action. *Journal of Nursing Scholarship: An Official Publication of Sigma Theta Tau International Honor Society of Nursing*, 35(3), 207.
- Burnier, M., & Gasser, U. E. (2008). End-digit preference in general practice: A comparison of the conventional auscultatory and electronic oscillometric methods [Publisher: Taylor & Francis eprint: <https://doi.org/10.1080/08037050801972881>]. *Blood Pressure*, 17(2), 104–109. <https://doi.org/10.1080/08037050801972881>
- Callaway, B., & Sant'Anna, P. H. (2020). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, S0304407620303948. <https://doi.org/10.1016/j.jeconom.2020.12.001>
- Card, D., Dobkin, C., & Maestas, N. (2008). The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare. *American Economic Review*, 98(5), 2242–2258. <https://doi.org/10.1257/aer.98.5.2242>
- Card, D., Dobkin, C., & Maestas, N. (2009). Does Medicare Save Lives?*. *The Quarterly Journal of Economics*, 124(2), 597–636. <https://doi.org/10.1162/qjec.2009.124.2.597>

- Carey, R. M., & Whelton, P. K. (2018). Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Synopsis of the 2017 American College of Cardiology/American Heart Association Hypertension Guideline [Publisher: American College of Physicians]. *Annals of Internal Medicine*, 168(5), 351–358. <https://doi.org/10.7326/M17-3203>
- Chan, D. C., Gentzkow, M., & Yu, C. (2022). Selection with Variation in Diagnostic Skill: Evidence from Radiologists*. *The Quarterly Journal of Economics*, qjab048. <https://doi.org/10.1093/qje/qjab048>
- Chan, D. C., & Gruber, J. (2020). Provider Discretion and Variation in Resource Allocation: The Case of Triage Decisions. *AEA Papers and Proceedings*, 110, 279–283. <https://doi.org/10.1257/pandp.20201032>
- Chandra, A., Flack, E., & Obermeyer, Z. (2021). The Health Costs of Cost-Sharing, 66.
- Chandra, A., & Staiger, D. O. (2014). EXPERTISE, UNDERUSE, AND OVERUSE IN HEALTHCARE, 25.
- Chen, L., Islam, R. M., Wang, J., Hird, T. R., Pavkov, M. E., Gregg, E. W., Salim, A., Tabesh, M., Koye, D. N., Harding, J. L., Sacre, J. W., Barr, E. L. M., Magliano, D. J., & Shaw, J. E. (2020). A systematic review of trends in all-cause mortality among people with diabetes. *Diabetologia*, 63(9), 1718–1735. <https://doi.org/10.1007/s00125-020-05199-0>
- Chen, S.-Y., Shah, S. N., Lee, Y.-C., Boulanger, L., Mardekian, J., & Kuznik, A. (2014). Moving Branded Statins to Lowest Copay Tier Improves Patient Adherence. *Journal of Managed Care Pharmacy*, 20(1), 34–42. <https://doi.org/10.18553/jmcp.2014.20.1.34>
- Chetty, R., Friedman, J. N., Olsen, T., & Pistaferri, L. (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *The Quarterly Journal of Economics*, 126(2), 749–804. <https://doi.org/10.1093/qje/qjr013>
- Ciechanowski, P., Russo, J., Katon, W., Simon, G., Ludman, E., Von Korff, M., Young, B., & Lin, E. (2006). Where is the patient? The association of psychosocial factors and missed primary care appointments in patients with diabetes. *General Hospital Psychiatry*, 28(1), 9–17. <https://doi.org/10.1016/j.genhosppsych.2005.07.004>
- Clark, D., & Royer, H. (2013). The Effect of Education on Adult Mortality and Health: Evidence from Britain. *American Economic Review*, 103(6), 2087–2120. <https://doi.org/10.1257/aer.103.6.2087>
- Costantini, O., Papp, K. K., Como, J., Aucott, J., Carlson, M. D., & Aron, D. C. (1999). Attitudes of faculty, housestaff, and medical students toward clinical practice guidelines. *Academic medicine*, 74(10), 1138–1143. <https://doi.org/10.1097/00001888-199910000-00019>
- Coussens, S. (2018). Heuristic Thinking in the Emergency Department, 59.
- Currie, J., & Zhang, J. (2021). *Doing More with Less: Predicting Primary Care Provider Effectiveness* (Working Paper No. 28929) [Series: Working Paper Series]. National Bureau of Economic Research. <https://doi.org/10.3386/w28929>

- da Costa, T. M., Salomão, P. L., Martha, A. S., Pisa, I. T., & Sigulem, D. (2010). Hallsworth. *International Journal of Medical Informatics*, 79(1), 65–70. <https://doi.org/10.1016/j.ijmedinf.2009.09.001>
- Dantas, L. F., Fleck, J. L., Cyrino Oliveira, F. L., & Hamacher, S. (2018). No-shows in appointment scheduling - a systematic literature review. *Health Policy (Amsterdam, Netherlands)*, 122(4), 412–421. <https://doi.org/10.1016/j.healthpol.2018.02.002>
- Das, J., Holla, A., Mohpal, A., & Muralidharan, K. (2016). Quality and Accountability in Health Care Delivery: Audit-Study Evidence from Primary Care in India. *American Economic Review*, 106(12), 3765–3799. <https://doi.org/10.1257/aer.20151138>
- de Chaisemartin, C., & D’Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9), 2964–2996. <https://doi.org/10.1257/aer.20181169>
- de Estadísticas, I. N. (2020). *Síntesis de Resultados Encuesta Suplementaria de Ingresos* (tech. rep.). Retrieved September 30, 2021, from [http://www.ine.cl/prensa/2021/08/24/ingreso-laboral-promedio-mensual-en-chile-fue-de-\\$635.134-en-2020](http://www.ine.cl/prensa/2021/08/24/ingreso-laboral-promedio-mensual-en-chile-fue-de-$635.134-en-2020)
- Della Vigna, S., & Malmendier, U. (2006). Paying Not to Go to the Gym. *American Economic Review*, 96(3), 694–719. <https://doi.org/10.1257/aer.96.3.694>
- DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, 47(2), 315–372. <https://doi.org/10.1257/jel.47.2.315>
- DellaVigna, S., & Linos, E. (2022). RCTs to Scale: Comprehensive Evidence From Two Nudge Units. *Econometrica*, 90(1), 81–116. <https://doi.org/10.3982/ECTA18709>
- Desai, R., Thakkar, S., Fong, H. K., Varma, Y., Khan, M. Z. A., Itare, V. B., Raina, J. S., Savani, S., Damarlapally, N., Doshi, R. P., Gangani, K., & Sitammagari, K. (2019). Rising Trends in Medication Non-compliance and Associated Worsening Cardiovascular and Cerebrovascular Outcomes Among Hospitalized Adults Across the United States [Publisher: Cureus]. *Cureus*, 11(8). <https://doi.org/10.7759/cureus.5389>
- de Salud - Gobierno de Chile, M. (2020). FONASA. Retrieved February 3, 2021, from <https://www.fonasa.cl/sites/fonasa/documentos>
- Donohue, J. M., Guclu, H., Gellad, W. F., Chang, C.-C. H., Huskamp, H. A., Choudhry, N. K., Zhang, R., Lo-Ciganic, W.-H., Junker, S. P., Anderson, T., & Richards-Shubik, S. (2018). Influence of peer networks on physician adoption of new drugs [Publisher: Public Library of Science]. *PLOS ONE*, 13(10), e0204826. <https://doi.org/10.1371/journal.pone.0204826>
- Dusetzina, S. B., Huskamp, H. A., Rothman, R. L., Pinheiro, L. C., Roberts, A. W., Shah, N. D., Walunas, T. L., Wood, W. A., Zuckerman, A. D., Zullig, L. L., & Keating, N. L. (2022). Many Medicare Beneficiaries Do Not Fill High-Price Specialty Drug Prescriptions. *Health Affairs*, 41(4), 487–496. <https://doi.org/10.1377/hlthaff.2021.01742>
- Egan, B. M., Kjeldsen, S. E., Grassi, G., Esler, M., & Mancia, G. (2019). The global burden of hypertension exceeds 1.4 billion people: Should a systolic blood pressure target below 130 become the universal standard? *Journal of Hypertension*, 37(6), 1148–1153. <https://doi.org/10.1097/HJH.0000000000002021>

- Einav, L., Finkelstein, A., & Polyakova, M. (2018). Private Provision of Social Insurance: Drug-Specific Price Elasticities and Cost Sharing in Medicare Part D. *American Economic Journal: Economic Policy*, 10(3), 122–153. <https://doi.org/10.1257/pol.20160355>
- Einav, L., Finkelstein, A., & Schrimpf, P. (2015). The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D *. *The Quarterly Journal of Economics*, 130(2), 841–899. <https://doi.org/10.1093/qje/qjv005>
- Einav, L., Finkelstein, A., & Schrimpf, P. (2017). Bunching at the kink: Implications for spending responses to health insurance contracts. *Journal of Public Economics*, 146, 27–40. <https://doi.org/10.1016/j.jpubeco.2016.11.011>
- Ellanti, P., Manecksha, R. P., & Flynn, R. (2011). The use of text messaging to reduce non-attendance at outpatients clinic—a departmental experience. *Irish Medical Journal*, 104(1), 28–29.
- Elnaem, M. H., Irwan, N. A., Abubakar, U., Syed Sulaiman, S. A., Elrggal, M. E., & Cheema, E. (2020). Impact of Medication Regimen Simplification on Medication Adherence and Clinical Outcomes in Patients with Long-Term Medical Conditions. *Patient preference and adherence*, 14, 2135–2145. <https://doi.org/10.2147/PPA.S268499>
- Fan, H., Yan, Q., Coyte, P. C., & Yu, W. (2019). Does Public Health Insurance Coverage Lead to Better Health Outcomes? Evidence From Chinese Adults. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*, 56, 0046958019842000. <https://doi.org/10.1177/0046958019842000>
- Finkelstein, A., Persson, P., Polyakova, M., & Shapiro, J. M. (2021). *A Taste of Their Own Medicine: Guideline Adherence and Access to Expertise* (Working Paper No. 29356) [Series: Working Paper Series]. National Bureau of Economic Research. <https://doi.org/10.3386/w29356>
- Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., & Oregon Health Study Group. (2012). The Oregon Health Insurance Experiment: Evidence from the First Year*. *The Quarterly Journal of Economics*, 127(3), 1057–1106. <https://doi.org/10.1093/qje/qjs020>
- Fischer, F., Lange, K., Klose, K., Greiner, W., & Kraemer, A. (2016). Barriers and Strategies in Guideline Implementation—A Scoping Review [Number: 3 Publisher: Multidisciplinary Digital Publishing Institute]. *Healthcare*, 4(3), 36. <https://doi.org/10.3390/healthcare4030036>
- FONASA. (2018). Bases de datos, informes y documentos. Retrieved April 21, 2022, from <https://www.fonasa.cl/sites/fonasa/documentos>
- Ford, E. S. (2011). Trends in the control of risk factors for cardiovascular disease among adults with diagnosed diabetes: Findings from the National Health and Nutrition Examination Survey 1999-2008*. *Journal of Diabetes*, 3(4), 337–347. <https://doi.org/10.1111/j.1753-0407.2011.00148.x>
- Free, C., Phillips, G., Watson, L., Galli, L., Felix, L., Edwards, P., Patel, V., & Haines, A. (2013). The Effectiveness of Mobile-Health Technologies to Improve Health Care

- Service Delivery Processes: A Systematic Review and Meta-Analysis. *PLOS Medicine*, 10(1), e1001363. <https://doi.org/10.1371/journal.pmed.1001363>
- Fullerton, B., Erler, A., Pöhlmann, B., & Gerlach, F. M. (2012). Predictors of dropout in the German disease management program for type 2 diabetes. *BMC health services research*, 12, 8. <https://doi.org/10.1186/1472-6963-12-8>
- Gabaix, X. (2019). Chapter 4 - Behavioral inattention. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of Behavioral Economics: Applications and Foundations 1* (pp. 261–343). North-Holland. <https://doi.org/10.1016/bs.hesbe.2018.11.001>
- Garies, S., Hao, S., McBrien, K., Williamson, T., Peng, M., Khan, N. A., Padwal, R. S., Quan, H., Leung, A. A., & for Hypertension Canada’s Research and Evaluation Committee. (2019). Prevalence of Hypertension, Treatment, and Blood Pressure Targets in Canada Associated With the 2017 American College of Cardiology and American Heart Association Blood Pressure Guidelines. *JAMA Network Open*, 2(3), e190406. <https://doi.org/10.1001/jamanetworkopen.2019.0406>
- Geneletti, S., O’Keeffe, A. G., Sharples, L. D., Richardson, S., & Baio, G. (2015). Bayesian regression discontinuity designs: Incorporating clinical knowledge in the causal analysis of primary care data. *Statistics in Medicine*, 34(15), 2334–2352. <https://doi.org/10.1002/sim.6486>
- Geraghty, M., Glynn, F., Amin, M., & Kinsella, J. (2008). Patient mobile telephone ‘text’ reminder: A novel way to reduce non-attendance at the ENT out-patient clinic. *The Journal of Laryngology and Otology*, 122(3), 296–298. <https://doi.org/10.1017/S0022215107007906>
- Gertler, P., & Vermeersch, C. (2013). *Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes* (tech. rep. w19046). National Bureau of Economic Research. Cambridge, MA. <https://doi.org/10.3386/w19046>
- Goic, A. (2015). The Chilean Health Care System: The task ahead. 1. Retrieved April 21, 2022, from https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0034-98872015000600011&lng=en&nrm=iso&tlng=en
- Goodman-Bacon, A. (2018). *Difference-in-Differences with Variation in Treatment Timing* (tech. rep. w25018). National Bureau of Economic Research. Cambridge, MA. <https://doi.org/10.3386/w25018>
- Graber, M. L., Wachter, R. M., & Cassel, C. K. (2012). Bringing Diagnosis Into the Quality and Safety Equations. *JAMA*, 308(12), 1211. <https://doi.org/10.1001/2012.jama.11913>
- Greenwood, B. N., Carnahan, S., & Huang, L. (2018). Patient–physician gender concordance and increased mortality among female heart attack patients [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, 115(34), 8569–8574. <https://doi.org/10.1073/pnas.1800097115>
- Greiver, M., Kalia, S., Voruganti, T., Aliarzadeh, B., Moineddin, R., Hinton, W., Dawes, M., Sullivan, F., Syed, S., Williams, J., & Lusignan, S. d. (2019). Trends in end digit preference for blood pressure and associations with cardiovascular outcomes in Canadian and UK primary care: A retrospective observational study [Publisher:

- British Medical Journal Publishing Group Section: Cardiovascular medicine]. *BMJ Open*, 9(1), e024970. <https://doi.org/10.1136/bmjopen-2018-024970>
- Gruber, J., Hoe, T. P., & Stoye, G. (2018). *Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers* (tech. rep. w24445). National Bureau of Economic Research. <https://doi.org/10.3386/w24445>
- Gucciardi, E. (2008). A Systematic Review of Attrition from Diabetes Education Services: Strategies to Improve Attrition and Retention Research. *Canadian Journal of Diabetes*, 32, 53–65. [https://doi.org/10.1016/S1499-2671\(08\)21011-7](https://doi.org/10.1016/S1499-2671(08)21011-7)
- Hallsworth, M., Berry, D., Sanders, M., Sallis, A., King, D., Vlaev, I., & Darzi, A. (2015). Stating Appointment Costs in SMS Reminders Reduces Missed Hospital Appointments: Findings from Two Randomised Controlled Trials. *PLOS ONE*, 10(9), e0137306. <https://doi.org/10.1371/journal.pone.0137306>
- Hamine, S., Gerth-Guyette, E., Faulx, D., Green, B. B., & Ginsburg, A. S. (2015). Impact of mHealth Chronic Disease Management on Treatment Adherence and Patient Outcomes: A Systematic Review. *Journal of Medical Internet Research*, 17(2), e52. <https://doi.org/10.2196/jmir.3951>
- Hardy, K. J., O'Brien, S. V., & Furlong, N. J. (2001). Information given to patients before appointments and its effect on non-attendance rate. *BMJ (Clinical research ed.)*, 323(7324), 1298–1300. <https://doi.org/10.1136/bmj.323.7324.1298>
- Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., & Yavitz, A. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE8>]. *Quantitative Economics*, 1(1), 1–46. <https://doi.org/10.3982/QE8>
- Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology (Cambridge, Mass.)*, 21(1), 13–15. <https://doi.org/10.1097/EDE.0b013e3181c1ea43>
- Hernández-Pizarro, H. M., Nicodemo, C., & Casasnovas, G. L. (2020). Discontinuous system of allowances: The response of prosocial health-care professionals. *Journal of Public Economics*, 190, 104248. <https://doi.org/10.1016/j.jpubeco.2020.104248>
- Huang, S., & Ullrich, H. (2021). Physician effects in antibiotic prescribing: Evidence from physician exits, 62.
- Jeffery, R. A., To, M. J., Hayduk-Costa, G., Cameron, A., Taylor, C., Van Zoost, C., & Hayden, J. A. (2015). Interventions to improve adherence to cardiovascular disease guidelines: A systematic review. *BMC Family Practice*, 16(1), 147. <https://doi.org/10.1186/s12875-015-0341-7>
- JLN. (2017a). *Case Studies on Payment Innovation for Primary Health Care* (tech. rep.). Retrieved October 5, 2021, from https://improvingphc.org/sites/default/files/JLN-GIZ_Case_Studies_on_Payment_Innovation_for_Primary_Health_Care__0.pdf
- JLN. (2017b). *Financing and Payment Models for Primary Health Care: Six Lessons from JLN Country Implementation Experience* (tech. rep.).
- Jongh, T. d., Gurol-Urganci, I., Vodopivec-Jamsek, V., Car, J., & Atun, R. (2012). Mobile phone messaging for facilitating self-management of long-term illnesses [Publisher:

- John Wiley & Sons, Ltd]. *Cochrane Database of Systematic Reviews*, (12). <https://doi.org/10.1002/14651858.CD007459.pub2>
- Karter, A. J., Parker, M. M., Moffet, H. H., Ahmed, A. T., Ferrara, A., Liu, J. Y., & Selby, J. V. (2004). Missed Appointments and Poor Glycemic Control: An Opportunity to Identify High-Risk Diabetic Patients. *Medical Care*, 42(2), 110–115. Retrieved December 14, 2018, from <https://www.jstor.org/stable/4640707>
- Kessler, J., & Zhang, C. (2014). Behavioral Economics and Health.
- Kim, E. S. H., Samuels, T. A., Yeh, H.-C., Abuid, M., Marinopoulos, S. S., McCauley, J. M., & Brancati, F. L. (2007). End-Digit Preference and the Quality of Blood Pressure Monitoring in Diabetic Adults [Publisher: American Diabetes Association Section: Clinical Care/Education/Nutrition/Psychosocial Research]. *Diabetes Care*, 30(8), 1959–1963. <https://doi.org/10.2337/dc07-0020>
- King, G., Gakidou, E., Imai, K., Lakin, J., Moore, R. T., Nall, C., Ravishankar, N., Vargas, M., Téllez-Rojo, M. M., Avila, J. E. H., Avila, M. H., & Llamas, H. H. (2009). Public policy for the poor? A randomised assessment of the Mexican universal health insurance programme. *Lancet (London, England)*, 373(9673), 1447–1454. [https://doi.org/10.1016/S0140-6736\(09\)60239-7](https://doi.org/10.1016/S0140-6736(09)60239-7)
- Kleven, H. J., & Waseem, M. (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan*. *The Quarterly Journal of Economics*, 128(2), 669–723. <https://doi.org/10.1093/qje/qjt004>
- Kleven, H. J. (2016). Bunching [_eprint: <https://doi.org/10.1146/annurev-economics-080315-015234>]. *Annual Review of Economics*, 8(1), 435–464. <https://doi.org/10.1146/annurev-economics-080315-015234>
- Knaapen, L. (2014). Evidence-Based Medicine or Cookbook Medicine? Addressing Concerns over the Standardization of Care [_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/soc4.12184>]. *Sociology Compass*, 8(6), 823–836. <https://doi.org/10.1111/soc4.12184>
- Krass, I., Schieback, P., & Dhippayom, T. (2015). Adherence to diabetes medication: A systematic review. *Diabetic Medicine*, 32(6), 725–737. <https://doi.org/10.1111/dme.12651>
- Kringos, D. S., Boerma, W. G. W., Hutchinson, A., & Saltman, R. B. (2015). *Building primary care in a changing Europe* [OCLC: 957235404]. Retrieved April 30, 2022, from <http://www.euro.who.int/en/about-us/partners/observatory/publications/studies/building-primary-care-in-a-changing-europe>
- Lacy, N. L., Paulman, A., Reuter, M. D., & Lovejoy, B. (2004). Why we don't come: Patient perceptions on no-shows. *Annals of Family Medicine*, 2(6), 541–545. <https://doi.org/10.1370/afm.123>
- Lanas, F., Garcia, M. S., Paccot, M., Peña, J., Torres, P., Beaney, T., Ster, A. C., Poulter, N. R., & Xia, X. (2020). May Measurement Month 2018: An analysis of blood pressure screening results from Chile. *European Heart Journal Supplements*, 22(Supplement_H), H37–H39. <https://doi.org/10.1093/eurheartj/suaa023>

- Lee, W. C., Balu, S., Cobden, D., Joshi, A. V., & Pashos, C. L. (2006). Prevalence and economic consequences of medication adherence in diabetes: A systematic literature review. *Managed care interface*, 19(7), 31–41.
- Leong, K. C., Chen, W. S., Leong, K. W., Mastura, I., Mimi, O., Sheikh, M. A., Zailinawati, A. H., Ng, C. J., Phua, K. L., & Teng, C. L. (2006). The use of text messaging to improve attendance in primary care: A randomized controlled trial. *Family Practice*, 23(6), 699–705. <https://doi.org/10.1093/fampra/cml044>
- Levy, H., & Meltzer, D. (2008). The Impact of Health Insurance on Health. *Annual Review of Public Health*, 29(1), 399–409. <https://doi.org/10.1146/annurev.publhealth.28.021406.144042>
- Liew, S.-M., Tong, S. F., Lee, V. K. M., Ng, C. J., Leong, K. C., & Teng, C. L. (2009). Text messaging reminders to reduce non-attendance in chronic disease follow-up: A clinical trial. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners*, 59(569), 916–920. <https://doi.org/10.3399/bjgp09X472250>
- Ly, D. P. (2021). The Influence of the Availability Heuristic on Physicians in the Emergency Department. *Annals of Emergency Medicine*, 78(5), 650–657. <https://doi.org/10.1016/j.annemergmed.2021.06.012>
- Malik, V. S., Willet, W. C., & Hu, F. B. (2020). Nearly a decade on — trends, risk factors and policy implications in global obesity [Number: 11 Publisher: Nature Publishing Group]. *Nature Reviews Endocrinology*, 16(11), 615–616. <https://doi.org/10.1038/s41574-020-00411-y>
- Malik, V. S., Willett, W. C., & Hu, F. B. (2013). Global obesity: Trends, risk factors and policy implications [Number: 1 Publisher: Nature Publishing Group]. *Nature Reviews Endocrinology*, 9(1), 13–27. <https://doi.org/10.1038/nrendo.2012.199>
- Manning, W. G. (1987). Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. *American Economic Review*, 77(3), 251–77. https://econpapers.repec.org/article/aeaarec/v_3a77_3ay_3a1987_3ai_3a3_3ap_3a251-77.htm
- Margozzini, P., & Passi, Á. (2018). Encuesta Nacional de Salud, ENS 2016-2017: Un aporte a la planificación sanitaria y políticas públicas en Chile [Number: 1]. *ARS MEDICA Revista de Ciencias Médicas*, 43(1), 30–34. <https://doi.org/10.11565/arsmed.v43i1.1354>
- Martinez, D. A., Zhang, H., Bastias, M., Feijoo, F., Hinson, J., Martinez, R., Dunstan, J., Levin, S., & Prieto, D. (2019). Prolonged wait time is associated with increased mortality for Chilean waiting list patients with non-prioritized conditions. *BMC Public Health*, 19. <https://doi.org/10.1186/s12889-019-6526-6>
- Mavrokonstantis, P. (2019). Introduction to the bunching Package, 19.
- MDS. (2013). Encuesta Casen. Retrieved April 21, 2022, from <http://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen>
- Mehok, L. E., Miller, M. M., Trost, Z., Goubert, L., De Ruddere, L., & Hirsh, A. T. (2019). Pain Intensity And Attribution Mediate The Impact Of Patient Weight And Gender

- On Activity Recommendations For Chronic Pain. *Journal of Pain Research*, 12, 2743–2753. <https://doi.org/10.2147/JPR.S218761>
- Miao, F., Wen, B., Hu, Z., Fortino, G., Wang, X.-P., Liu, Z.-D., Tang, M., & Li, Y. (2020). Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques. *Artificial Intelligence in Medicine*, 108, 101919. <https://doi.org/10.1016/j.artmed.2020.101919>
- MINSAL. (2008). *INFORME FINAL ESTUDIO DE CARGA DE ENFERMEDAD Y CARGA ATRIBUIBLE* (tech. rep.). Retrieved April 21, 2022, from <http://www.repositoriodigital.minsal.cl/bitstream/handle/2015/602/3895.pdf?sequence=1>
- MINSAL. (2017a). DISPONE CREACIÓN DE COMISIÓN MÉDICA ASESORA PARA ANALIZAR LA SITUACIÓN DE PERSONAS QUE FALLECEN HABIENDO ESTADO EN UNA LISTA DE ESPERA O CON UNA GARANTÍA DE OPORTUNIDAD GES RETRASADA. Retrieved April 21, 2022, from <https://www.bcn.cl/leychile>
- MINSAL. (2017b). Orientación Técnica Programa de Salud Cardiovascular. Retrieved October 17, 2018, from https://www.u-cursos.cl/medicina/2017/0/DPPAEF/1/foro/r/OT-PROGRAMA-DE-SALUD-CARDIOVASCULAR_05.pdf
- MINSAL. (2021). Todo lo que debe saber para atenderse. Retrieved April 21, 2022, from <https://ssms.gob.cl/como-me-atiendo/todo-lo-que-debe-saber-para-atenderse/>
- Moscoe, E., Bor, J., & Bärnighausen, T. (2015). Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: A review of current and best practice. *Journal of Clinical Epidemiology*, 68(2), 132–143. <https://doi.org/10.1016/j.jclinepi.2014.06.021>
- Mullainathan, S., & Obermeyer, Z. (2021). Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care*. *The Quarterly Journal of Economics*, qjab046. <https://doi.org/10.1093/qje/qjab046>
- Muntner Paul, Shimbo Daichi, Carey Robert M., Charleston Jeanne B., Gaillard Trudy, Misra Sanjay, Myers Martin G., Ogedegbe Gbenga, Schwartz Joseph E., Townsend Raymond R., Urbina Elaine M., Viera Anthony J., White William B., Wright Jackson T., & null null. (2019). Measurement of Blood Pressure in Humans: A Scientific Statement From the American Heart Association [Publisher: American Heart Association]. *Hypertension*, 73(5), e35–e66. <https://doi.org/10.1161/HYP.0000000000000087>
- Murad, M. H. (2017). Clinical Practice Guidelines: A Primer on Development and Dissemination. *Mayo Clinic Proceedings*, 92(3), 423–433. <https://doi.org/10.1016/j.mayocp.2017.01.001>
- Murdock, A., Rodgers, C., Lindsay, H., & Tham, T. C. K. (2002). Why do patients not keep their appointments? Prospective study in a gastroenterology outpatient clinic. *Journal of the Royal Society of Medicine*, 95(6), 284–286. Retrieved February 6, 2020, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1279909/>
- Myers, M. G., Godwin, M., Dawes, M., Kiss, A., Tobe, S. W., Grant, F. C., & Kaczorowski, J. (2011). Conventional versus automated measurement of blood pressure in primary

- care patients with systolic hypertension: Randomised parallel design controlled trial. *The BMJ*, 342. <https://doi.org/10.1136/bmj.d286>
- Myers, M. G., Godwin, M., Dawes, M., Kiss, A., Tobe, S. W., & Kaczorowski, J. (2012). Conventional versus automated measurement of blood pressure in the office (CAMBO) trial. *Family Practice*, 29(4), 376–382. <https://doi.org/10.1093/fampra/cmr113>
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
- O’Keeffe, A. G., Geneletti, S., Baio, G., Sharples, L. D., Nazareth, I., & Petersen, I. (2014). Regression discontinuity designs: An approach to the evaluation of treatment efficacy in primary care using observational data. *BMJ*, 349(sep08 3), g5293–g5293. <https://doi.org/10.1136/bmj.g5293>
- Olenski, A. R., Zimmerman, A., Coussens, S., & Jena, A. B. (2020). Behavioral Heuristics in Coronary-Artery Bypass Graft Surgery. *New England Journal of Medicine*, 382(8), 778–779. <https://doi.org/10.1056/NEJMc1911289>
- Ostchega, Y., & Nguyen, D. T. (2020). Hypertension Prevalence Among Adults Aged 18 and Over: United States, 2017–2018. (364), 8.
- Oster, E. (2018). Diabetes and Diet: Purchasing Behavior Change in Response to Health Information. *American Economic Journal: Applied Economics*, 10(4), 308–348. <https://doi.org/10.1257/app.20160232>
- Osterberg, L., & Blaschke, T. (2005). Adherence to Medication [Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMr050100>]. *New England Journal of Medicine*, 353(5), 487–497. <https://doi.org/10.1056/NEJMr050100>
- PAHO. (2020). Chile. Retrieved April 23, 2021, from https://www.paho.org/salud-en-las-americas-2017/?page_id=105
- Pan, J., Lei, X., & Liu, G. G. (2016). Health Insurance and Health Status: Exploring the Causal Effect from a Policy Intervention [_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hec.3225>]. *Health Economics*, 25(11), 1389–1402. <https://doi.org/10.1002/hec.3225>
- Piper, M. A., Evans, C. V., Burda, B. U., Margolis, K. L., O’Connor, E., & Whitlock, E. P. (2015). Diagnostic and Predictive Accuracy of Blood Pressure Screening Methods With Consideration of Rescreening Intervals: A Systematic Review for the U.S. Preventive Services Task Force [Publisher: American College of Physicians]. *Annals of Internal Medicine*, 162(3), 192–204. <https://doi.org/10.7326/M14-1539>
- Reibling, N., & Wendt, C. (2012). Gatekeeping and provider choice in OECD healthcare systems. *Current Sociology*, 60(4), 489–505. <https://doi.org/10.1177/0011392112438333>
- Restrepo, D., Armstrong, K. A., & Metlay, J. P. (2020). Annals Clinical Decision Making: Avoiding Cognitive Errors in Clinical Decision Making [Publisher: American College of Physicians]. *Annals of Internal Medicine*, 172(11), 747–751. <https://doi.org/10.7326/M19-3692>
- Rethy, L., Shah, N. S., Paparello, J. J., Lloyd-Jones, D. M., & Khan, S. S. (2020). Trends in Hypertension-Related Cardiovascular Mortality in the United States, 2000 to 2018

- [Publisher: American Heart Association]. *Hypertension*, 76(3), e23–e25. <https://doi.org/10.1161/HYPERTENSIONAHA.120.15153>
- Roberto, C. A., & Kawachi, I. (2015). *Behavioral Economics and Public Health* [Google-Books-ID: WXQ_CgAAQBAJ]. Oxford University Press.
- Rotar, A. M., Van Den Berg, M. J., Schäfer, W., Kringos, D. S., & Klazinga, N. S. (2018). Shared decision making between patient and GP about referrals from primary care: Does gatekeeping make a difference? *PLoS ONE*, 13(6), e0198729. <https://doi.org/10.1371/journal.pone.0198729>
- Runnacles, J., Roueché, A., & Lachman, P. (2018). The right care, every time: Improving adherence to evidence-based guidelines [Publisher: Royal College of Paediatrics and Child Health Section: E&P: Quality improvement]. *Archives of Disease in Childhood - Education and Practice*, 103(1), 27–33. <https://doi.org/10.1136/archdischild-2017-312740>
- Saez, E. (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy*, 2(3), 180–212. <https://doi.org/10.1257/pol.2.3.180>
- Sandhu, H., Adams, A., Singleton, L., Clark-Carter, D., & Kidd, J. (2009). The impact of gender dyads on doctor–patient communication: A systematic review. *Patient Education and Counseling*, 76(3), 348–355. <https://doi.org/10.1016/j.pec.2009.07.010>
- Schechtman, J. M., Schorling, J. B., & Voss, J. D. (2008). Appointment Adherence and Disparities in Outcomes Among Patients with Diabetes. *Journal of General Internal Medicine*, 23(10), 1685. <https://doi.org/10.1007/s11606-008-0747-1>
- Schisterman, E. F., Cole, S. R., Ye, A., & Platt, R. W. (2013). Accuracy loss due to selection bias in cohort studies with left truncation. *Paediatric and Perinatal Epidemiology*, 27(5), 491–502. <https://doi.org/10.1111/ppe.12073>
- Schneeweiss, S. (2010). A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiology and Drug Safety*, 19(8), 858–868. <https://doi.org/10.1002/pds.1926>
- Shanbhag, D., Graham, I. D., Harlos, K., Haynes, R. B., Gabizon, I., Connolly, S. J., & Spall, H. G. C. V. (2018). Effectiveness of implementation interventions in improving physician adherence to guideline recommendations in heart failure: A systematic review [Publisher: British Medical Journal Publishing Group Section: Health services research]. *BMJ Open*, 8(3), e017765. <https://doi.org/10.1136/bmjopen-2017-017765>
- Singh, M. (2021). Heuristics in the delivery room. *Science*, 374(6565), 324–329. <https://doi.org/10.1126/science.abc9818>
- Socrates, T., Krisai, P., Vischer, A. S., Meienberg, A., Mayr, M., & Burkard, T. (2021). Improved agreement and diagnostic accuracy of a cuffless 24-h blood pressure measurement device in clinical practice [Bandiera_abtest: a Cc_license_type: cc-by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cardiology;Medical research;Risk factors Subject_term_id: cardiology;medical-research;risk-factors]. *Scientific Reports*, 11(1), 1143. <https://doi.org/10.1038/s41598-020-80905-x>

- Stergiou, G. S., O'Brien, E., Myers, M., Palatini, P., Parati, G., & Board, t. S. B. S. A. (2020). STRIDE BP: An international initiative for accurate blood pressure measurement. *Journal of Hypertension*, 38(3), 395–399. <https://doi.org/10.1097/HJH.0000000000002289>
- Taubman, S. L., Allen, H. L., Wright, B. J., Baicker, K., & Finkelstein, A. N. (2014). Medicaid Increases Emergency-Department Use: Evidence from Oregon's Health Insurance Experiment [Publisher: American Association for the Advancement of Science]. *Science*, 343(6168), 263–268. <https://doi.org/10.1126/science.1246183>
- Timmermans, S., & Mauck, A. (2005). The Promises And Pitfalls Of Evidence-Based Medicine [Publisher: Health Affairs]. *Health Affairs*, 24(1), 18–28. <https://doi.org/10.1377/hlthaff.24.1.18>
- Tomlinson, M., Rotheram-Borus, M. J., Swartz, L., & Tsai, A. C. (2013). Scaling Up mHealth: Where Is the Evidence? [Publisher: Public Library of Science]. *PLOS Medicine*, 10(2), e1001382. <https://doi.org/10.1371/journal.pmed.1001382>
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. 185, 10.
- Vargas, V., & Poblete, S. (2008). Health prioritization: The case of Chile. *Health Affairs (Project Hope)*, 27(3), 782–792. <https://doi.org/10.1377/hlthaff.27.3.782>
- Venkataramani, A. S., Bor, J., & Jena, A. B. (2016). Regression discontinuity designs in healthcare research. *BMJ*, i1216. <https://doi.org/10.1136/bmj.i1216>
- Vonbank, A., Saely, C. H., Rein, P., Sturn, D., & Drexel, H. (2013). Current cholesterol guidelines and clinical reality: A comparison of two cohorts of coronary artery disease patients. *Swiss Medical Weekly*, 143, w13828. <https://doi.org/10.4414/smw.2013.13828>
- Vrijens, B., Antoniou, S., Burnier, M., de la Sierra, A., & Volpe, M. (2017). Current Situation of Medication Adherence in Hypertension. *Frontiers in Pharmacology*, 8. <https://doi.org/10.3389/fphar.2017.00100>
- Wang, Y., Leifheit, E. C., & Krumholz, H. M. (2022). Trends in 10-Year Outcomes Among Medicare Beneficiaries Who Survived an Acute Myocardial Infarction. *JAMA Cardiology*. <https://doi.org/10.1001/jamacardio.2022.0662>
- Watt, W. D. (1987). The Family Physician: Gatekeeper to the Health-Care System. *Canadian Family Physician*, 33, 1101–1104. Retrieved April 30, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2218510/>
- WHO. (2013). *A global brief on Hypertension: Silent killer, global public health crisis* (tech. rep.). Retrieved May 4, 2022, from <https://apps.who.int/iris/bitstream/handle/10665/79059/W?sequence=1>
- WHO Global Observatory for eHealth. (2011). *mHealth: New horizons for health through mobile technologies: Second global survey on eHealth* (tech. rep.) [ISBN: 9789244564257 ISSN: 2305-0934 (Online) number-of-pages: viii, 102]. World Health Organization. Retrieved April 30, 2022, from <https://apps.who.int/iris/handle/10665/44607>
- Wilcox, J. (1961). Observer Factors in the Measurement of Blood Pressure. *Nursing Research*, 10(1), 4–17. Retrieved May 13, 2020, from <https://journals.lww.com/>

- nursingresearchonline/Citation/1961/01010/Observer_Factors_in_the_Measurement_of_Blood.2.aspx
- Williams, J. S., Brown, S. M., & Conlin, P. R. (2009). Blood-Pressure Measurement [Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMvcm0800157>]. *New England Journal of Medicine*, 360(5), e6. <https://doi.org/10.1056/NEJMvcm0800157>
- Wohlfahrt, P., Cífková, R., Krajcoviechová, A., Šulc, P., Bruthans, J., Linhart, A., Filipovský, J., Mayer, O., & Widimský, J. J. (2020). Comparison of three office blood pressure measurement techniques and their effect on hypertension prevalence in the general population. *Journal of Hypertension*, 38(4), 656–662. <https://doi.org/10.1097/HJH.0000000000002322>
- Yamaguchi, S., Waki, K., Tomizawa, N., Waki, H., Nannya, Y., Nangaku, M., Kadowaki, T., & Ohe, K. (2017). Previous dropout from diabetic care as a predictor of patients' willingness to use mobile applications for self-management: A cross-sectional study. *Journal of Diabetes Investigation*, 8(4), 542–549. <https://doi.org/10.1111/jdi.12613>
- Zhang, Y., Lave, J. R., Donohue, J. M., Fischer, M. A., Chernew, M. E., & Newhouse, J. P. (2010). The Impact of Medicare Part D on Medication Adherence Among Older Adults Enrolled in Medicare-Advantage Products. *Medical Care*, 48(5), 409–417. <https://doi.org/10.1097/MLR.0b013e3181d68978>
- Zheng, K., Padman, R., Krackhardt, D., Johnson, M. P., & Diamond, H. S. (2010). Social networks and physician adoption of electronic health records: Insights from an empirical study. *Journal of the American Medical Informatics Association*, 17(3), 328–336. <https://doi.org/10.1136/jamia.2009.000877>

Appendix A

Chapter 2 – Supplementary Tables and Figures

Chapter 2: Appendix

Table A.1: ICD-10 Codes included in cardiovascular hospitalization and mortality outcomes

Condition	ICD-10 Codes
Diabetes mellitus	E 10.0, E 10.1, E 10.2, E 10.3, E 10.4, E 10.5, E 10.6, E 10.7, E 10.8, E 10.9, E 11.0, E 11.1, E 11.2, E 11.3, E 11.4, E 11.5, E 11.6, E 11.7, E 11.8, E 11.9
Primary hypertension	I 10.X
Hypertensive heart disease	I 11.0, I 11.9
Hypertensive chronic kidney disease	I 12.0, I 12.9, I 13.0, I 13.1, I 13.2, I 13.9
Acute myocardial infarction	I 21.0, I 21.1, I 21.2, I 21.3, I 21.4, I 21.9
Acute ischaemic heart disease	I 24.9
Heart failure	I 50.0, I 50.1, I 50.9
Hemorrhage	I 60.0, I 60.1, I 60.2, I 60.3, I 60.4, I 60.5, I 60.6, I 60.7, I 60.8, I 60.9, I 61.0, I 61.1, I 61.2, I 61.3, I 61.4, I 61.5, I 61.6, I 61.8, I 61.9, I 62.0, I 62.1, I 62.9
Cerebral infarction	I 63.0, I 63.1, I 63.2, I 63.3, I 63.4

Note: ICD-10 codes listed are included in outcomes cardiovascular-related hospitalization and cardiovascular-related mortality. All other ICD-10 codes are included in non cardiovascular-related hospitalization and non cardiovascular-related mortality outcomes.

Table A.2: Event study: ITT estimates of the effect of SMS reminders on visits and tests

Patients:	Diabetes		Hypertension	
	Visit (1)	Blood sugar test (2)	Visit (3)	Blood pressure test (4)
Pre 3	-0.057* (0.031)	-0.022 (0.028)	-0.015 (0.030)	-0.014 (0.030)
Pre 2	-0.033 (0.039)	-0.016 (0.038)	0.029 (0.024)	0.031 (0.024)
Pre 1	-0.035 (0.042)	-0.032 (0.043)	0.022 (0.026)	0.024 (0.027)
Post 0	0.020 (0.012)	0.008 (0.012)	0.020* (0.010)	0.022** (0.011)
Post 1	0.037** (0.015)	0.023 (0.015)	0.047*** (0.014)	0.048*** (0.014)
Post 2	0.051*** (0.017)	0.033** (0.015)	0.064*** (0.019)	0.069*** (0.019)
Post 3	0.040** (0.017)	0.018 (0.015)	0.053*** (0.020)	0.055*** (0.020)
Post 4	0.038** (0.016)	0.007 (0.015)	0.047** (0.019)	0.053*** (0.019)
Post 5	0.030* (0.017)	0.010 (0.015)	0.039** (0.020)	0.044** (0.019)
Post 6	0.028 (0.020)	0.009 (0.019)	0.027 (0.023)	0.032 (0.023)
Pre-trends p-val	0.257	0.512	0.021	0.023
Mean dep. var.	0.648	0.553	0.646	0.621
Observations	458,670	462,557	2,223,140	2,223,140
Clinics	313	313	310	310

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents event study estimates of the effect of SMS reminders estimated using the Borusyak, Jaravel, and Spiess (2021) estimator for staggered policy adoption. The treatment variable is an indicator equal to one in the semester of, and after a clinic implemented SMS reminders and zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. Visit is an indicator for if the patient had at least one visit to primary care in a given semester, zero otherwise. Blood sugar test is an indicator for if the patient had a hemoglobin A1c test or a blood glucose test in a given semester, zero otherwise.

Table A.3: Event study: ATE estimates of the effect of SMS reminders on visits and tests

Patients:	Diabetes		Hypertension	
	Visit (1)	Blood sugar test (2)	Visit (3)	Blood pressure test (4)
Pre 4	0.054* (0.032)	0.045 (0.029)	-0.025 (0.033)	-0.025 (0.032)
Pre 3	-0.003 (0.024)	0.021 (0.022)	-0.024 (0.029)	-0.021 (0.028)
Pre 2	0.006 (0.013)	0.021* (0.012)	-0.002 (0.017)	-0.003 (0.017)
Post 0	0.009 (0.025)	0.008 (0.024)	-0.000 (0.027)	-0.004 (0.027)
Post 1	0.059** (0.025)	0.043 (0.027)	0.070*** (0.023)	0.064*** (0.023)
Post 2	0.084*** (0.026)	0.068** (0.027)	0.100*** (0.029)	0.099*** (0.029)
Post 3	0.075*** (0.025)	0.051* (0.027)	0.094*** (0.028)	0.088*** (0.028)
Post 4	0.071*** (0.023)	0.040 (0.026)	0.083*** (0.027)	0.082*** (0.027)
Post 5	0.066*** (0.025)	0.042 (0.027)	0.072** (0.029)	0.070** (0.029)
Post 6	0.025 (0.029)	0.016 (0.030)	0.035 (0.038)	0.032 (0.037)
Mean dep. var.	0.644	0.553	0.643	0.618
Observations	497,949	497,949	2,411,117	2,411,117
Clinics	314	314	310	310

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents event study estimates of the effect of SMS reminders estimated using two-way fixed effect models because the treatment variable is a continuous measure of compliance. Here, the treatment variable is the share of visits sent a reminder, among visits eligible to receive a reminder, zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. Visit is an indicator for if the patient had at least one visit to primary care in a given semester, zero otherwise. Blood sugar test is an indicator for if the patient had a hemoglobin A1c test or a blood glucose test in a given semester, zero otherwise.

Table A.4: Event study: ITT estimates of the effect of SMS reminders on medication adherence among patients with type 2 diabetes

Patients:	Diabetes		
	Prescription at diagnosis (1)	Any prescription pick-up (2)	Medication adherence (3)
Pre 3	0.020 (0.035)	-0.080 (0.066)	-0.018* (0.010)
Pre 2	0.041 (0.045)	-0.063 (0.087)	-0.008 (0.012)
Pre 1	0.067 (0.054)	0.018 (0.093)	0.001 (0.013)
Post 0	-0.010 (0.021)	0.053** (0.026)	0.008 (0.006)
Post 1	0.030 (0.023)	0.044** (0.020)	0.007 (0.006)
Post 2	0.002 (0.023)	0.042** (0.021)	0.003 (0.006)
Post 3	0.025 (0.027)	0.027 (0.019)	0.007 (0.006)
Post 4	0.016 (0.024)	0.023 (0.019)	0.004 (0.006)
Post 5	0.032 (0.023)	0.002 (0.021)	0.003 (0.007)
Post 6	0.042 (0.027)	-0.009 (0.025)	-0.001 (0.008)
Pre-trends p-val	0.572	0.000	0.069
Mean dep. var.	0.447	0.281	0.033
Observations	73,766	223,318	221,923
Clinics	312	294	294

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents event study estimates of the effect of SMS reminders estimated using the Borusyak, Jaravel, and Spiess (2021) estimator for staggered policy adoption. The treatment variable is an indicator equal to one in the semester of, and after a clinic implemented SMS reminders and zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. The model in column 1 is estimated using each patient's first visit after diagnosis only. Models in columns 2 and 3 are estimated among patients with diabetes and hypertension, who had any prescription at their first visit for diabetes or hypertension, respectively. Medication adherence is an indicator for if the patient's proportion of days covered 80% or greater in a given semester, zero otherwise.

Table A.5: Event study: ATE estimates of the effect of SMS reminders on medication adherence among patients with type 2 diabetes

Patients:	Diabetes		
	Prescription at diagnosis (1)	Any prescription pick-up (2)	Medication adherence (3)
Pre 4	-0.068 (0.044)	-0.007 (0.076)	-0.001 (0.012)
Pre 3	-0.046* (0.027)	-0.084** (0.035)	-0.014 (0.010)
Pre 2	-0.020 (0.022)	-0.073*** (0.022)	-0.005 (0.007)
Post 0	-0.052 (0.041)	0.059 (0.046)	0.021* (0.013)
Post 1	0.036 (0.038)	0.084*** (0.027)	0.014 (0.010)
Post 2	-0.014 (0.038)	0.083*** (0.030)	0.011 (0.009)
Post 3	0.020 (0.040)	0.050 (0.031)	0.011 (0.010)
Post 4	0.012 (0.034)	0.034 (0.028)	0.003 (0.009)
Post 5	0.028 (0.037)	0.003 (0.034)	0.002 (0.011)
Post 6	0.036 (0.036)	-0.085** (0.039)	-0.011 (0.011)
Mean dep. var.	0.447	0.276	0.032
Observations	75,899	242,194	242,194
Clinics	314	313	313

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents event study estimates of the effect of SMS reminders estimated using two-way fixed effect models because the treatment variable is a continuous measure of compliance. Here, the treatment variable is the share of visits sent a reminder, among visits eligible to receive a reminder, zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. The model in column 1 is estimated using each patient's first visit after diagnosis only. Models in columns 2 and 3 are estimated among patients with diabetes and hypertension, who had any prescription at their first visit for diabetes or hypertension, respectively. Medication adherence is an indicator for if the patient's proportion of days covered 80% or greater in a given semester, zero otherwise.

Table A.6: Event study: ITT estimates of the effect of SMS reminders on medication adherence among patients with hypertension

Patients:	Hypertension		
	Prescription at diagnosis (1)	Any prescription pick-up (2)	Medication adherence (3)
Pre 3	0.060* (0.031)	-0.091* (0.048)	-0.020*** (0.006)
Pre 2	0.038 (0.041)	-0.059 (0.073)	-0.012 (0.010)
Pre 1	0.070 (0.052)	0.010 (0.076)	-0.003 (0.010)
Post 0	-0.008 (0.023)	0.054** (0.025)	0.008* (0.005)
Post 1	0.014 (0.022)	0.060*** (0.018)	0.012*** (0.004)
Post 2	0.041* (0.025)	0.052*** (0.019)	0.004 (0.004)
Post 3	0.012 (0.026)	0.039** (0.018)	0.008* (0.005)
Post 4	0.034 (0.027)	0.035* (0.018)	0.008* (0.004)
Post 5	0.026 (0.027)	0.026 (0.019)	0.007 (0.005)
Post 6	0.044 (0.036)	0.004 (0.021)	0.006 (0.005)
Pre-trends p-val	0.274	0.000	0.000
Mean dep. var.	0.473	0.283	0.036
Observations	327,169	1,114,915	1,109,167
Clinics	310	307	307

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents event study estimates of the effect of SMS reminders estimated using the Borusyak, Jaravel, and Spiess (2021) estimator for staggered policy adoption. The treatment variable is an indicator equal to one in the semester of, and after a clinic implemented SMS reminders and zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. The model in column 1 is estimated using each patient's first visit after diagnosis only. Models in columns 2 and 3 are estimated among patients with diabetes and hypertension, who had any prescription at their first visit for diabetes or hypertension, respectively. Medication adherence is an indicator for if the patient's proportion of days covered 80% or greater in a given semester, zero otherwise.

Table A.7: Event study: ATE estimates of the effect of SMS reminders on medication adherence among patients with type 2 diabetes

Patients:	Hypertension		
	Prescription at diagnosis (1)	Any prescription pick-up (2)	Medication adherence (3)
Pre 4	-0.074 (0.045)	0.012 (0.058)	0.002 (0.008)
Pre 3	-0.012 (0.029)	-0.071** (0.028)	-0.015** (0.006)
Pre 2	-0.033 (0.025)	-0.055** (0.022)	-0.007 (0.007)
Post 0	-0.079** (0.038)	0.053 (0.049)	0.015 (0.010)
Post 1	-0.028 (0.033)	0.084*** (0.027)	0.020** (0.008)
Post 2	0.016 (0.037)	0.089*** (0.027)	0.011 (0.007)
Post 3	-0.026 (0.038)	0.057** (0.025)	0.012* (0.007)
Post 4	0.034 (0.039)	0.042* (0.023)	0.007 (0.006)
Post 5	0.007 (0.043)	0.026 (0.031)	0.011 (0.008)
Post 6	0.016 (0.060)	-0.080** (0.035)	0.001 (0.008)
Mean dep. var.	0.473	0.278	0.036
Observations	328,037	1,213,287	1,213,287
Clinics	310	310	310

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents event study estimates of the effect of SMS reminders estimated using two-way fixed effect models because the treatment variable is a continuous measure of compliance. Here, the treatment variable is the share of visits sent a reminder, among visits eligible to receive a reminder, zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. The model in column 1 is estimated using each patient's first visit after diagnosis only. Models in columns 2 and 3 are estimated among patients with diabetes and hypertension, who had any prescription at their first visit for diabetes or hypertension, respectively. Medication adherence is an indicator for if the patient's proportion of days covered 80% or greater in a given semester, zero otherwise.

Table A.8: Event study: ITT estimates of the effect of SMS reminders on hospitalizations per 1000 population

Patients:	All			
	Cardiovascular		Non-cardiovascular	
	Hospital- ization (1)	In-hospital mortality (2)	Hospital- ization (3)	In-hospital mortality (4)
Pre 3	3.3731* (1.7673)	0.2572* (0.1508)	0.9623 (2.3300)	0.0659 (0.2029)
Pre 2	2.3101 (1.5443)	0.0696 (0.1719)	-1.1991 (2.4012)	0.0336 (0.2225)
Pre 1	3.2319* (1.6599)	0.0759 (0.1938)	-1.8521 (3.0194)	-0.2542 (0.2956)
Post 0	0.2332 (0.5914)	-0.1276 (0.0803)	0.3847 (1.0789)	-0.1365 (0.1182)
Post 1	0.9813 (0.6449)	-0.0170 (0.0921)	-0.2278 (1.2721)	0.0566 (0.1638)
Post 2	1.5051** (0.7604)	0.0199 (0.1039)	2.5294 (1.7181)	0.1637 (0.1672)
Post 3	1.9185*** (0.6745)	0.0897 (0.1035)	3.5019** (1.7011)	0.3010* (0.1540)
Post 4	1.7030** (0.6816)	-0.1637* (0.0932)	3.6366** (1.6028)	0.2699 (0.1860)
Post 5	1.4563** (0.6878)	0.0259 (0.1122)	1.9090 (1.7484)	0.1456 (0.1651)
Post 6	1.7348** (0.7744)	-0.0260 (0.1000)	2.3233 (1.7184)	0.1958 (0.1790)
Pre-trends p-val	0.177	0.355	0.567	0.350
Mean dep. var.	11.960	0.529	39.901	1.336
Observations	2,401,234	2,401,234	2,401,234	2,401,234
Clinics	315	315	315	315

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents event study estimates of the effect of SMS reminders on hospitalizations per 1000 population estimated using the Borusyak, Jaravel, and Spiess (2021) estimator for staggered policy adoption. The treatment variable is an indicator equal to one in the semester of, and after a clinic implemented SMS reminders and zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. All models also include fixed effects for male and 2-year age.

Table A.9: Event study: ATE estimates of the effect of SMS reminders on hospitalizations per 1000 population

Patients:	All			
	Cardiovascular		Non-cardiovascular	
	Hospital- ization (1)	In-hospital mortality (2)	Hospital- ization (3)	In-hospital mortality (4)
Pre 4	-2.501* (1.502)	-0.042 (0.131)	0.079 (0.182)	-2.705 (2.610)
Pre 3	0.950 (0.949)	0.206* (0.119)	0.151 (0.218)	-0.110 (1.467)
Pre 2	-0.713 (0.634)	-0.030 (0.098)	0.222 (0.153)	-0.882 (1.472)
Post 0	-0.505 (1.167)	-0.393** (0.174)	-0.059 (0.318)	-0.477 (2.232)
Post 1	1.104 (1.086)	-0.101 (0.207)	0.273 (0.345)	-0.813 (2.081)
Post 2	1.910* (1.132)	-0.004 (0.190)	0.525 (0.320)	2.050 (2.316)
Post 3	2.861*** (1.073)	0.043 (0.203)	0.621** (0.306)	2.072 (2.449)
Post 4	2.494** (0.970)	-0.270 (0.166)	0.523 (0.354)	3.218 (2.293)
Post 5	1.373 (1.058)	-0.125 (0.195)	0.500 (0.312)	1.224 (2.631)
Post 6	2.099* (1.108)	-0.119 (0.173)	0.501* (0.296)	2.210 (2.376)
Mean dep. var.	11.900	0.527	1.330	39.700
Observations	2,607,151	2,607,151	2,607,151	2,607,151
Clinics	315	315	315	315

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents event study estimates of the effect of SMS reminders estimated using two-way fixed effect models because the treatment variable is a continuous measure of compliance. Here, the treatment variable is the share of visits sent a reminder, among visits eligible to receive a reminder, zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. All models also include fixed effects for male and 2-year age.

Table A.10: Event study: ITT estimates of the effect of SMS reminders on mortality per 1000 population

Patients:	All	
	Cardiovascular mortality (1)	Non-cardiovascular mortality (2)
Pre 4	-5.304** (2.555)	-1.478*** (0.516)
Pre 3	-3.727*** (1.236)	-1.878*** (0.350)
Pre 2	-3.543*** (1.061)	-2.047*** (0.333)
Pre 1	-2.782*** (0.963)	-2.189*** (0.348)
Year 0	0.342 (0.379)	-0.905*** (0.328)
Year 1	0.700 (0.485)	-0.598 (0.401)
Year 2	0.650 (0.615)	0.263 (0.368)
Year 3	0.580 (0.713)	0.251 (0.406)
Year 4	0.452 (0.820)	0.245 (0.523)
Pre-trends p-val	0.017	0.000
Mean dep. var.	7.639	7.378
Observations	4,625,798	4,625,798
Counties	275	275

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents event study estimates of the effect of SMS reminders on hospitalizations per 1000 population estimated using the Borusyak, Jaravel, and Spiess (2021) estimator for staggered policy adoption. The treatment variable is an indicator equal to one in the year of, and after a county implemented SMS reminders and zero otherwise. All models include fixed effects for year, county, and years since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the county level. All models also include fixed effects for male and 2-year age.

Table A.11: Event study: ATE estimates of the effect of SMS reminders on mortality per 1000 population

Patients:	All	
	Cardiovascular mortality (1)	Non-cardiovascular mortality (2)
Pre 3	-1.770* (0.965)	-0.588* (0.304)
Pre 2	-1.590** (0.736)	-0.646** (0.285)
Year 0	-2.163*** (0.769)	-1.772** (0.794)
Year 1	-0.827 (1.193)	-0.553 (0.917)
Year 2	-1.439 (1.253)	0.662 (0.731)
Year 3	-1.915 (1.310)	0.915 (0.844)
Year 4	-2.454 (1.505)	0.898 (0.908)
Year 5	-2.804 (1.726)	1.534 (1.171)
Mean dep. var.	7.639	7.378
Observations	4,186,391	4,186,391
Counties	275	275

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents event study estimates of the effect of SMS reminders estimated using two-way fixed effect models because the treatment variable is a continuous measure of compliance. Here, the treatment variable is the share of visits eligible to receive a reminder, among visits in that year-county cell where a reminder should have been sent, zero otherwise. All models include fixed effects for year, county, and years since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the county level. All models also include fixed effects for male and 2-year age.

Table A.12: Event study: ITT estimates of the effect of SMS reminders on visits (TWFE model)

Patients:	Hypertension
	Visit (1)
Pre 4	-0.007 (0.028)
Pre 3	-0.014 (0.026)
Pre 2	0.015 (0.011)
Post 0	0.026** (0.010)
Post 1	0.045*** (0.013)
Post 2	0.067*** (0.017)
Post 3	0.065*** (0.018)
Post 4	0.061*** (0.018)
Post 5	0.053*** (0.018)
Post 6	0.027 (0.022)
Mean dep. var.	0.643
Observations	2,411,117
Clinics	310

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents event study estimates of the effect of SMS reminders estimated using a two-way fixed effects estimator for staggered policy adoption. The treatment variable is an indicator equal to one in the semester of, and after a clinic implemented SMS reminders and zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. Visit is an indicator for if the patient had at least one visit to primary care in a given semester, zero otherwise.

Table A.13: Difference-in-differences estimates of the effect of SMS reminders on visits (TWFE model)

Patients:	Hypertension
	Visit (1)
Post SMS	0.042*** (0.016)
Mean dep. var.	0.643
Observations	2,411,117
Clinics	310

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents event study estimates of the effect of SMS reminders estimated using a two-way fixed effects estimator for staggered policy adoption. The treatment variable is an indicator equal to one in the semester of, and after a clinic implemented SMS reminders and zero otherwise. All models include fixed effects for semester, clinic, and semesters since the patient was diagnosed with hypertension and/or diabetes. Robust standard errors are clustered at the clinic level. Visit is an indicator for if the patient had at least one visit to primary care in a given semester, zero otherwise.

Appendix B

Chapter 3 – Supplementary Tables and Figures

Chapter 3: Appendix

Table B.1: Inclusion criteria for primary care electronic health records data

Inclusion criteria	N patients	N clinics	% patients excluded
-	1,448,800	552	-
Blood pressure was measured	1,413,499	545	2.4%
No previous hypertension diagnosis or medication prescription	995,528	539	32.4%
Age 18-80	863,571	538	9.6%
Large clinics	755,653	257	12.5%
At least one year of follow-up time	619,907	257	18.0%

Note: Inclusion criteria and number and percent of patients excluded at each step. Large is defined a clinic above the median number of visits. The final sample is 619,907 patient-visits and 257 clinics.

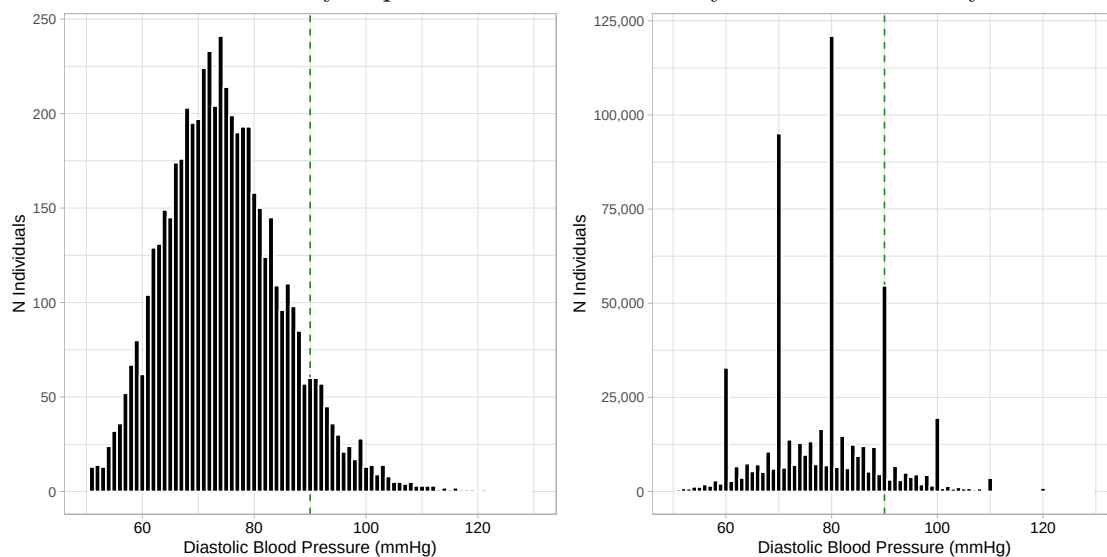
Table B.2: Chile’s clinical practice guidelines: Frequent errors that result in an inaccurate measurement of blood pressure

Factor	Discrepancy in blood pressure systolic/diastolic (mmHg)
Did not rest 5 minutes prior to measurement	10-20/
Talking or active listening during measurement	10-17/10-13
Blood pressure cuff over clothing	5-40/
Blood pressure cuff too tight	10-15/2-8
Blood pressure cuff too loose	-7/5
Ate, smoked, or drank coffee 30 minutes before measurement	6-20/
Back was not against the chair	6-10/
Crossed legs	5-8/

Note: This table is the author’s own translation of appendix table 3 from Chile’s 2017 clinical practice guidelines for cardiovascular disease management (Orientación Técnica Programa de Salud Cardiovascular 2017, page 37). If no number is included after the slash it means no change in diastolic blood pressure is expected. Discrepancy indicates the change in blood pressure relative to the patient’s ‘true’ blood pressure.

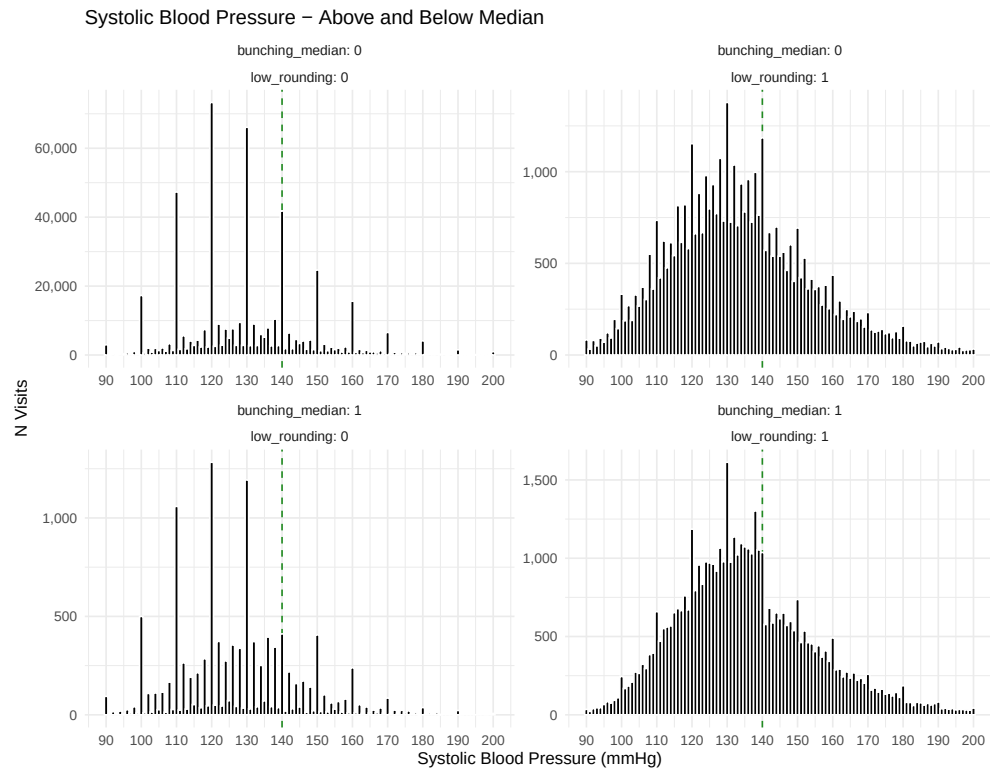
Figure B.1: Expected vs. observed distribution of diastolic blood pressure

Panel A: Chile's Nationally Representative Health Survey. **Panel B:** Primary Care Records



Note: Comparison of expected (panel A) to observed (panel B) distribution of diastolic blood pressure in two large populations. Panel A shows the distribution of diastolic blood pressure from Chile's nationally representative health survey (ENS 2016-2017). Panel B shows the distribution of diastolic blood pressure from primary care electronic health records (2016-2019). Substantial end digit bias, or rounding to zero, is seen, as well as some missing mass to the right of the diagnostic threshold. Both distributions exclude individuals who have been diagnosed with hypertension before this instance of blood pressure was measured.

Figure B.2: Histogram of systolic blood pressure by high vs. low discretion clinics



Note: Histogram of systolic blood pressure by whether the estimated magnitude of discretion is above or below the median. We can see that the magnitude of discretion indeed corresponds to the amount of discretion, visually, at both high and low rounding clinics.

Table B.3: Construction of hospitalization variables

Variable	Description	ICD-10 codes
Stroke	Any cerebral infarction	I63
Heart attack	Also called acute coronary syndrome. Includes myocardial infarction, and other range of conditions associated with sudden, reduced blood flow to the heart	I20, I21, I22, I24, I25
Congestive heart failure	Failure of the heart of pump or fill adequately.	I50
Non-cardiovascular	Hospitalization for any reason except stroke, heart attack, and heart failure	Any except I20, I21, I22, I24, I25, I50, I63
Type 2 Diabetes	Hospitalization for reasons directly related to type 2 diabetes	E10, E11, I12, N03, N04, N08, N18, Z79, Z99

Note: This table describes how hospitalization variables were constructed. Each variable is an indicator equal to one if the patient was hospitalized with a primary or secondary diagnosis listed in the ICD-10 column.

Table B.4: Sample of primary care 3, 6, and 12 months after initial visit

	Yes	No	% Yes	% No
Observed again after 3 months	529,497	90,410	85.4	14.6
Observed again after 6 months	512,475	107,432	82.7	17.3
Observed again after 12 months	477,079	142,828	77.0	23.0

Note: This table describes the number of patients who were observed a second time (or more) in the electronic health records or hospitalization data, 3, 6, or 12 months after their initial primary care visit. Because they were observed again these patients are considered living during this period.

Table B.5: Impact of discretion on patients' retention in the data

	Retained 3 mo.	Retained 6 mo.	Retained 12 mo.
	(1)	(2)	(3)
Discretion x BP<140	−0.004 (0.005)	−0.005 (0.006)	−0.008 (0.007)
Discretion	0.002 (0.003)	0.001 (0.003)	0.001 (0.003)
BP<140	−0.018*** (0.002)	−0.019*** (0.002)	−0.019*** (0.002)
Observations	619,907	619,907	619,907
Mean dep. var.	0.854	0.827	0.77
Clinics	257	257	257

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents difference-in-differences estimates of the impact of crossing the diagnostic threshold from above to below 140 mmHg at high vs. low discretion clinics on the probability that the patient was re-observed in the data within 3, 6, or 12 months. BP < 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of discretion at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit, and type of primary care clinic.

Table B.6: Discretion and all-cause hospitalization 3, 6, and 12 months after primary care visit among retained patients

	All-cause hospitalization		
	<3 months	<6 months	<12 months
	(1)	(2)	(3)
Discretion x BP<140	−0.081 (0.055)	−0.107 (0.073)	−0.161 (0.117)
Discretion	0.036 (0.058)	0.063 (0.083)	0.141 (0.110)
BP<140	−0.020 (0.051)	−0.242*** (0.074)	−0.513*** (0.097)
Observations	529,497	512,475	477,079
Mean dep. var.(%)	2.419	4.621	8.508
Clinics	257	257	257

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents difference-in-differences estimates of the impact of crossing the diagnostic threshold from above to below 140 mmHg at high vs. low discretion clinics on the probability that the patient was re-observed in the data within 3, 6, or 12 months. BP < 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of discretion at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit, and type of primary care clinic. Model 1 is estimated among patients who returned to primary care or were hospitalized 3 or more months after their initial primary care visit (85.4% of the full sample). Model 2; 6 or more months (82.7% of the full sample). Model 3; 12 or more months (77.0% of the full sample). Dependent variables are an indicator for hospitalization for any reason within 3, 6, or 12 months of the patient's initial primary care visit.

Table B.7: Discretion and cardiovascular hospitalization 3, 6, and 12 months after primary care visit among retained Patients

	Within 3 months			Within 6 months			Within 12 months		
	Stroke (1)	Heart attack (2)	Heart failure (3)	Stroke (4)	Heart attack (5)	Heart failure (6)	Stroke (7)	Heart attack (8)	Heart failure (9)
Discretion x BP<140	-0.017 (0.011)	-0.020** (0.009)	0.007 (0.007)	-0.036*** (0.014)	-0.025* (0.013)	0.006 (0.011)	-0.057*** (0.021)	-0.026 (0.024)	-0.002 (0.015)
Discretion	0.018* (0.011)	0.014 (0.008)	-0.007 (0.005)	0.038*** (0.014)	0.022 (0.014)	-0.010 (0.009)	0.058*** (0.021)	0.022 (0.022)	-0.010 (0.011)
BP<140	-0.027*** (0.006)	-0.019** (0.010)	0.007 (0.007)	-0.064*** (0.010)	-0.055*** (0.014)	0.0001 (0.011)	-0.103*** (0.015)	-0.116*** (0.021)	-0.027* (0.016)
Observations	529,497	529,497	529,497	512,475	512,475	512,475	477,079	477,079	477,079
Mean dep. var.(%)	0.027	0.106	0.055	0.063	0.198	0.109	0.124	0.368	0.195
Clinics	257	257	257	257	257	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences estimates of the impact of crossing the diagnostic threshold from above to below 140 mmHg at high vs. low discretion clinics on the probability that the patient was re-observed in the data within 3, 6, or 12 months. BP<140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of discretion at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit, and type of primary care clinic. Model 1 is estimated among patients who returned to primary care or were hospitalized 3 or more months after their initial primary care visit (85.4% of the full sample). Model 2; 6 or more months (82.7% of the full sample). Model 3; 12 or more months (77.0% of the full sample). Dependent variables are an indicator for hospitalization for any reason within 3, 6, or 12 months of the patient's initial primary care visit. Stroke includes any cerebral infarction (ICD-10 codes I63). Heart attack includes acute coronary syndrome, myocardial infarction, and any other condition associated with sudden, reduced blood flow to the heart (I20-22, I24-25). Congestive heart failure (CHF) is I50.

Table B.8: Selection by patient characteristics - health history questions

	Was asked about family history of:				Patient has family history of:			
	Any (1)	Heart disease (2)	CVD (3)	Dislip- idemia (4)	Any (5)	Heart disease (6)	CVD (7)	Dislip- idemia (8)
Discretion x BP<140								
	-0.005 (0.005)	-0.004 (0.005)	-0.004 (0.005)	-0.004 (0.005)	-0.002 (0.002)	-0.0004 (0.002)	0.0001 (0.0005)	-0.001 (0.001)
Discretion	-0.001 (0.013)	-0.001 (0.013)	-0.004 (0.014)	-0.002 (0.013)	-0.008* (0.004)	-0.008* (0.004)	-0.001 (0.001)	-0.002 (0.002)
BP<140	-0.022*** (0.005)	-0.022*** (0.005)	-0.021*** (0.005)	-0.021*** (0.005)	-0.012*** (0.002)	-0.008*** (0.002)	0.004*** (0.001)	-0.009*** (0.001)
Observations	619,907	619,907	619,907	619,907	360,053	378,879	359,547	372,205
Mean dep. var.	0.63	0.611	0.58	0.6	0.212	0.139	0.024	0.067
Clinics	257	257	257	257	257	257	257	257

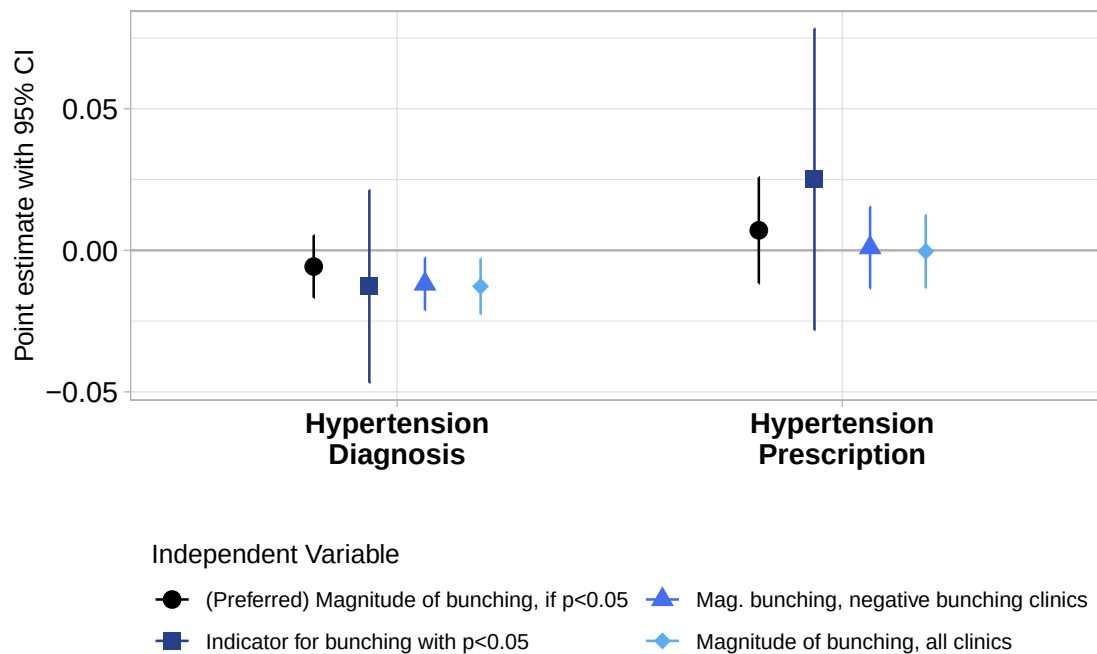
Note: * $p<0.1$; ** $p<0.05$; *** $p<0.01$. This table presents difference-in-differences estimates of the impact of crossing the diagnostic threshold from above to below 140 mmHg at high vs. low discretion clinics on health history questions. BP < 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of discretion at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Each dependent variable is an indicator for if the patient has the characteristic listed, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for year and quarter of primary care visit, and type of primary care clinic. Models 1-4 are indicators for if any answer to the health history question was recorded in the EHR. Models 5-8 are estimated among patients who did have the relevant health history question recorded in the EHR. CVD stands for cardiovascular disease.

Table B.9: Selection by patient age

	Age 50+	Age 60+	Age 70+
	(1)	(2)	(3)
Discretion x BP<140	−0.001 (0.002)	−0.003* (0.002)	−0.003 (0.002)
Discretion	0.006* (0.003)	−0.001 (0.002)	−0.007* (0.004)
BP<140	0.033*** (0.002)	−0.063*** (0.002)	−0.086*** (0.002)
Observations	619,907	619,907	619,907
Mean dep. var.	0.263	0.264	0.201
Clinics	257	257	257

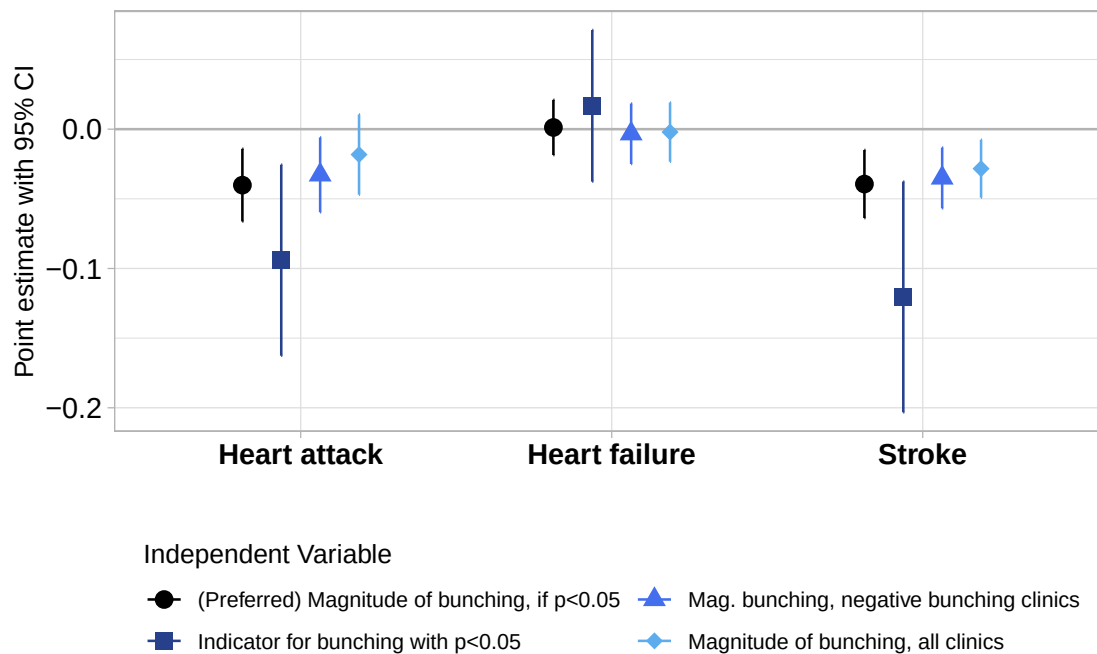
Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table presents difference-in-differences estimates of the impact of crossing the diagnostic threshold from above to below 140 mmHg at high vs. low discretion clinics on patient characteristics. BP < 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Bunch is a continuous variable for the estimated magnitude of discretion at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Each dependent variable is an indicator for if the patient is at least X years old, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for year and quarter of primary care visit, and type of primary care clinic.

Figure B.3: Comparison of results under different parameterizations of the magnitude of discretion: adherence to clinical practice guidelines



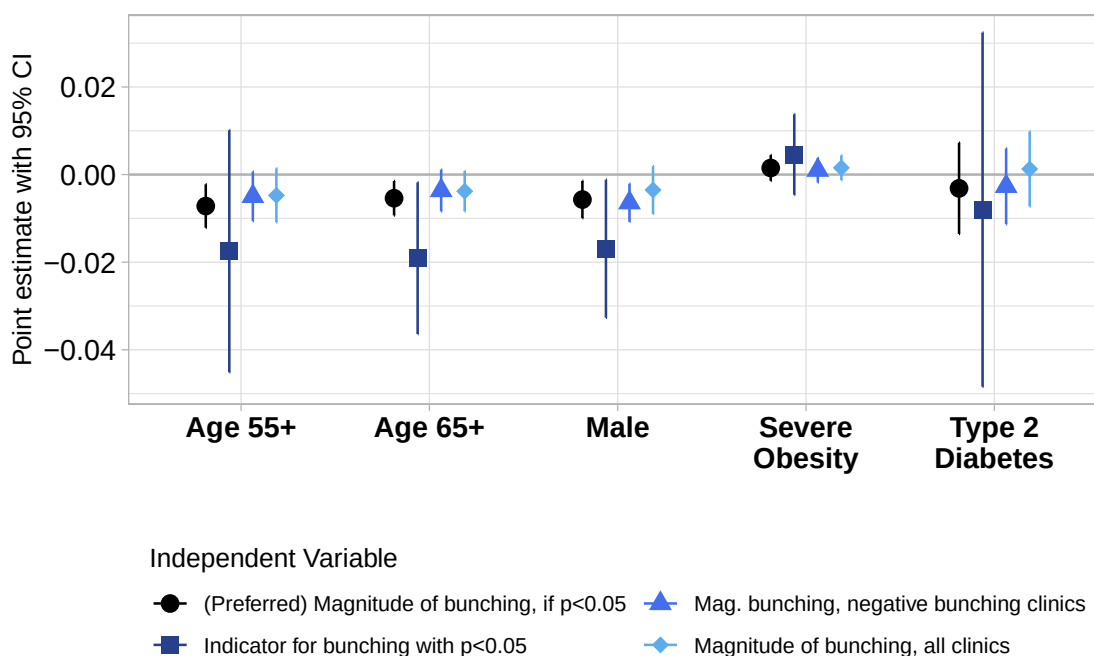
Note: This figure presents the impact of discretion on key outcomes, under different discretion specifications. Point estimates for $(\text{Discretion} \times \text{BP} < 140)$ and 95% confidence intervals shown from equation (2) are shown. Each outcome is estimated four separate times with different treatment variables. The preferred treatment variable, used everywhere else in this paper, is shown in black. This is the magnitude of discretion among clinics with negative discretion that was different from zero at $\alpha = 0.05$ significance level, zero otherwise. The treatment variable with the triangular marker is an indicator for if the clinic had a magnitude of negative discretion that was different from zero at $\alpha = 0.05$ significance level, zero otherwise. The treatment variable with the square marker is the magnitude of discretion for all clinics (regardless of statistical significance), normalized to mean zero, SD 1. The treatment variable with the cross marker is the magnitude of discretion for clinics with negative discretion (regardless of statistical significance), normalized to mean zero, SD 1. Diagnosis and prescription are indicators for if the patient received a new diagnosis or prescription at the visit.

Figure B.4: Comparison of results under different parameterizations of the magnitude of discretion: cardiovascular hospitalizations



Note: This figure presents the impact of discretion on key outcomes, under different discretion specifications. Point estimates for $(\text{Discretion} \times \text{BP} < 140)$ and 95% confidence intervals shown from equation (2) are shown. Each outcome is estimated four separate times with different treatment variables. The preferred treatment variable, used everywhere else in this paper, is shown in black. This is the magnitude of discretion among clinics with negative discretion that was different from zero at $\alpha = 0.05$ significance level, zero otherwise. The treatment variable with the triangular marker is an indicator for if the clinic had a magnitude of negative discretion that was different from zero at $\alpha = 0.05$ significance level, zero otherwise. The treatment variable with the square marker is the magnitude of discretion for all clinics (regardless of statistical significance), normalized to mean zero, SD 1. The treatment variable with the cross marker is the magnitude of discretion for clinics with negative discretion (regardless of statistical significance), normalized to mean zero, SD 1. Stroke, heart attack, and heart failure each are indicators for hospitalization with the event within six months of the primary care visit, and are each multiplied by 100 so they can be interpreted as percentages.

Figure B.5: Comparison of results under different parameterizations of the magnitude of discretion: patient characteristics



Note: This figure presents the impact of discretion on key outcomes, under different discretion specifications. Point estimates for (Discretion $\times$ BP < 140) and 95% confidence intervals shown from equation (2) are shown. Each outcome is estimated four separate times with different treatment variables. The preferred treatment variable, used everywhere else in this paper, is shown in black. This is the magnitude of discretion among clinics with negative discretion that was different from zero at $\alpha = 0.05$ significance level, zero otherwise. The treatment variable with the triangular marker is an indicator for if the clinic had a magnitude of negative discretion that was different from zero at $\alpha = 0.05$ significance level, zero otherwise. The treatment variable with the square marker is the magnitude of discretion for all clinics (regardless of statistical significance), normalized to mean zero, SD 1. The treatment variable with the cross marker is the magnitude of discretion for clinics with negative discretion (regardless of statistical significance), normalized to mean zero, SD 1. Each dependent variable is an indicator for if the patient has the characteristic. DM2 stands for previous diagnosis with type 2 diabetes. Sev. Obese is body mass index ≥ 35 .