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

Claire Boone*

November 4, 2022

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 (BP) $\geq$140/90 mmHg, in Chile's public health care system. Hypertension is the single most important risk factor for cardiovascular disease, but diagnosis can be challenging because BP 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 BP 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.

*Job market paper. University of California Berkeley, School of Public Health. **cboone@berkeley.edu**.
I am grateful to Paul Gertler, Ziad Obermeyer, Jon Kolstad, and Ulrike Malmendier for their guidance and support. I thank the Primary Care Division at Chile's Ministry of Health, and in particular Irma Vargas and Ximena Barros-Rubio. Priyanka Taneja, Madhav Nekkar, and Francesco Crivelli provided excellent research assistance.

1 Introduction

Clinical practice guidelines aim to standardize medicine using evidence-based recommendations, yet large inefficiencies in care delivery remain (Abaluck et al., 2020, 2016; Chandra and Staiger, 2014; Mullainathan and 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 and 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 and 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 presssure (BP) is at least 140/90 mmHg - both missed- and over-diagnosis are common because accurately measuring BP is challenging (Baron, 2018). Instantaneous BP fluctuates around the patient's long-term or true BP, which is a strong predictor of risk for cardiovascular events (Muntner Paul et al., 2019). Simple actions can change BP 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 BP 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 BP under the 140 threshold, which would otherwise be smooth.[1] I estimate this behavior with a bunching estimator and interpret it as a behavioral

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

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'.

Specifically, I use estimate the magnitude of bunching in BP - 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.

I find approximately 10% of clinics exhibit discretion in the diagnosis of hypertension, and at these clinics up to 62% of patients expected to have tested just above the diagnostic threshold are instead recorded just below it. 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, there is no difference in the probability of new hypertension diagnosis or prescription at high vs. low discretion clinics, indicates that the value of BP recorded in the electronic health record does guide decision-making, whether it was manipulated by the provider or not. Untreated high BP raises a patient's risk of adverse cardiac events. 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 I find 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. These results suggest not that discretion itself reduces cardiovascular hospitalization, but that at high discretion there is a more accurate classification of patients into hypertension negative and positive based on their 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 and Kahneman, 1974).

Indeed, I find that patients with characteristics representative of high cardiovascular risk - older males - are less likely to be sorted under the threshold. Compared to the clinical practice guidelines, where BP is the only determinant of hypertension status, by using the representativeness heuristic providers add information to their 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 find emergency room providers are subject to end digit bias - an age-related heuristic that leads to sub-optimal test decisions. These recent papers all find that heuristic thinking can lead to biased decisions and poor outcomes for some patient groups. In the setting of hypertension diagnosis and treatment I instead find that heuristic thinking contributes to more accurate decisions. 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. 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 and Zhang, 2021).

It is well studied that when incentives or policies change discontinuously at a threshold, behavior can be distorted (Kleven, 2016). This distortion can occur even when the decision-maker dies not benefit directly, but instead distorts their decisions to change outcomes for a beneficiary such as a patient or a student. This paper is closest to Diamond and Persson (2016), who use bunching methods to identify and understand teacher discretion in national test grading. The authors find that teachers selectively increase some students' scores from

below to above the pass threshold. Teachers do this for students who scored lower than their own average on previous test (i.e. students who 'had a bad test day') (Diamond and Persson, 2016). This behavior is analogous to my finding that providers round down the test result of a patient who 'had a bad BP test day'.

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.

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.

2 Setting and Data

2.1 Health and Healthcare in Chile

Chile is an OECD country with a similar burden of disease to 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 States, and 24.2% in Canada (Garies et al., 2019; Lanas et al., 2020; Ostchega and 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 preventative and acute medical services(Bastías et al., 2008; Bitrán et al., 2010; Bossert and Leisewitz, 2016). Primary care is obtained either at a clinic, or a low-complexity hospital. Low-complexity provide a combination of primary care and some in-patient services in rural areas that have a single medical center. 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

(Aguilera et al., 2015). FONASA, the public insurer, assigns patients a copayment level based on their income. 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,b). 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 goals and metrics are not clear and change frequently (JLN, 2017a,b). 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.

2.2 Medical Context

An estimated 1.4 billion people have high BP (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' BP 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 BP $\geq$140/90 mmHg. Hypertension can be prevented and treated with medication and behavior change including regular exercise and improving diet.

Accurately measuring BP is challenging as instantaneous or short-term BP fluctuates around the patient's long-term or true BP. BP 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 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 BP is often measured with error. Here, 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 BP, including rounding down and the over-representation of numbers that end in zero,[2] which have long been documented in the medical literature (Alcocer et al., 1997; Wilcox, 1961). An example of end digit bias in BP measurement is shown in figure 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).

2.3 Data

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., BP, 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, 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 (de Salud Gobierno de Chile, 2017). 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.

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

To compare the distribution of BP in the EHR to a nationally representative distribution I use Chile's National Health Survey (ENS, 2016-2017 wave), which is also representative of the population at the municipality level.

2.3.2 Sample

To measure discretion and its impact, I include visits where BP could be diagnosed, from patients aged 18-79 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 1). I use each patient's first observed visit to avoid over-representing BP measures that are under 140/90 mmHg, which could impact bunching estimates.

I also exclude visits where BP was not recorded (2.4% of visits), and patients under 18 or over 80 (9.8%). Patients 80 and older are subject to a different diagnostic threshold and goal for controlled BP. 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 A1 for full details on sample selection.

Tables 2 and 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 BP 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 Empirical Strategy

I first describe the intuition behind bunching estimation. 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.

3.1 Measuring Provider Discretion With Bunching

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 and Waseem, 2013; Saez, 2010).

I study bunching of BP 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 BP 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 BP as a running variable in which to study bunching offers several advantages. First, because hypertension lacks symptoms and can be deadly BP 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.

3.1.2 Bunching Estimation

Examining raw distributions of BP 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 2). 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 and Waseem, 2013). In BP 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 BP 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 BP for high and low rounding clinics, respectively. For each clinic I separately estimate:

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

Where $n_{c,i}$ is the number of observations in bin i at clinic c, k is the order of the polynomial,[3] p_i is the BP 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 $\mathbb{1}\{y_i = t\}$ are indicators for if bin i is in the excluded region around the threshold. $\mathbb{N}$ is the set of natural numbers, R is a vector of numbers that end in zero, omitting 140 (the threshold) $R = \{8, 9, 10, 11, 12, 13, 15, 16, ..., 20\}$.

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 and Waseem, 2013). To quantify the magnitude of discretion, I compare the predicted density, (1) to the observed empirical density. Figure 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

[3]For 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.

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

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

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.

3.1.3 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 BP is a biological measurement this assumption is easily satisfied (see figure A1 for the expected distribution of BP).

The second assumption is that missing mass is not caused by selection. It is recommended that BP is measured at every primary care visit making selection into testing unlikely. BP 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 BP measurement (table A1).

Third, manipulation is one-sided and bounded, that is, providers are only rounding BP from above to below the threshold and only for individuals in some range of BP. Because hypertension status is only determined by one threshold, there is no reason for providers to systematically round BP 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 BP 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 BP, if rounded or measured imprecisely, is systematically rounded down (Greiver et al., 2019; Myers et al., 2011). In addition, it is well established the BP 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 BP measurement (de Salud Gobierno de Chile, 2017). Of the 8 common errors listed, 7 lead to an overstatement of BP which can be as large as 20 mmHg, and only one leads to an understatement of BP, of 7 mmHg (see appendix table A2 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 and 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 BP 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.

3.2 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. I construct three other variables that measure discretion, and all lead to very similar result (see appendix 8.1 for details).

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

Clinical decisions include 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 BP. 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 A3 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 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:

$$Y_{i,c} = \alpha + \beta_1 Discretion_c + \beta_2 BP < 140_i + \beta_3 (Discretion_c * BP < 140_i) + \lambda_i + \gamma_c + \epsilon_{ic} \tag{2}$$

Where $Discretion_c$ is the normalized excess mass in clinic c, $BP < 140_i$ is an indicator for if patient i's recorded systolic BP 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 enrolees 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 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.

4 Results

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 2 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 some BP 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 BP readings at or above 140 indicate the same clinical action, that is, to diagnose BP, bunching at 140 is not as clinically meaningful and likely reflects rounding instead of discretion in clinical decision-making.

Figure 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. This means up to 62.2% of patients who are predicted to have BP at or just above 140 mmHg - and who therefore should be diagnosed with hypertension - are instead recorded as having BP 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.

The observed pattern of bunching is a combination of several possible behaviors. In the simplest case, the provider measures BP $\geq$140 but writes down a number under 140 in the EHR. Or, the provider may have taken multiple BP 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 (de Salud Gobierno de Chile, 2017). 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 BP reading below 140 for the patient, indicating they have used their discretion to change the test result.

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 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 BP recorded below the diagnostic threshold (Discretion x 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 BP 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* BP (whether manipulated or not) with respect to the threshold is meaningful, and encodes the provider's belief about their risk.

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 BP raises a patient's risk of adverse cardiac events. Thus, if providers are *incorrectly* reclassifying true positives as false positives we would see a higher rate of hospitalization for cardiac events in the months following the primary care visit in the group affected by discretion. Instead, we see the opposite. Table 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 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 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), shown as positive coefficients in row 2 of table 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.[4] 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 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 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

[4]The group of hospitalizations included here are mutually exclusive to stroke, heart attack, and heart failure.

either side of the diagnostic threshold in terms of health I would ideally measure effects on patient mortality, however mortality is unobserved. In appendix 8.2 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. Table A7 shows similar impacts of discretion on cardiovascular hospitalizations in this sample.

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? 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, i.e. use the representativeness heuristic (Tversky and 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 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.

The age thresholds of 55 and 65 are likely to be salient to providers. While age is not explicitly in the clinical practice guidelines for diagnosing hypertension, ages 55 and 65 appear frequently in Chile's guidelines for managing other cardiovascular diseases (de Salud Gobierno de Chile, 2017). Similar effects were not seen for ages 50, 60, and 70 (table A9). 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.

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 0 to 62% of patients' BP measurements near the diagnostic threshold were manipulated. Providers selectively round some patients' BP down, turning a positive test into a negative. I argue that the uncertainty in BP 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 BP.

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. Specifically, providers 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 BP 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 chose 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 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, within a clinic providers 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 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 and Ullrich, 2021; Mullainathan and 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.

6 References

Abaluck, J., L. Agha, D. C. C. Jr, D. Singer, and D. Zhu (2020). Mastering the Art of Cookbook Medicine: Machine Learning, Randomized Trials, and Misallocation. pp. 71.

Abaluck, J., L. Agha, C. Kabrhel, A. Raja, and A. Venkatesh (2016, December). The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care. *American Economic Review 106*(12), 3730–3764.

Adiyaman, A., N. Tosun, L. D. Elving, J. Deinum, J. W. M. Lenders, and T. Thien (2007, June). The effect of crossing legs on blood pressure. *Blood Pressure Monitoring 12*(3), 189–193.

Aguilera, I., A. Infante, H. Ormeño, and C. Urriola (2015, September). Implementation of the Universal Access with Explicit Guarantees (AUGE) reform Chile case Study. Technical report, World Health Organization.

Alcocer, L., G. Novoa, and D. Sotres (1997, October). 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.

Arnold, C., J. Koetsenruijter, J. Forstner, F. Peters-Klimm, and M. Wensing (2021, August). Influence of physician networks on prescribing a new ingredient combination in heart failure: a longitudinal claim data-based study. *Implementation Science 16*(1), 84.

Baron, R. B. (2018, June). Treating Blood Pressure Correctly by Measuring It Correctly. *JAMA Internal Medicine 178*(6), 860–861.

Bastías, G., T. Pantoja, T. Leisewitz, and V. Zárate (2008, December). Health care reform in Chile. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne 179*(12), 1289–1292.

Bitrán, R., L. Escobar, and P. Gassibe (2010, December). After Chile's Health Reform: Increase In Coverage And Access, Decline In Hospitalization And Death Rates. *Health Affairs 29*(12), 2161–2170.

Blomquist, S., W. Newey, A. Kumar, and C.-Y. Liang (2017, December). On Bunching and Identification of the Taxable Income Elasticity. Technical Report w24136, National Bureau of Economic Research.

Bossert, T. J. and T. Leisewitz (2016, January). Innovation and Change in the Chilean Health System. *New England Journal of Medicine 374*(1), 1–5.

Budnitz, D. S., N. Shehab, M. C. Lovegrove, A. I. Geller, J. N. Lind, and D. A. Pollock (2021, October). US Emergency Department Visits Attributed to Medication Harms, 2017-2019. *JAMA 326*(13), 1299.

Böhm, A.-K., U. Schneider, J. Aberle, and T. Stargardt (2021, May). Regimen simplification and medication adherence: Fixed-dose versus loose-dose combination therapy for type 2 diabetes. *PLOS ONE 16*(5), e0250993. Publisher: Public Library of Science.

Chan, D. C., M. Gentzkow, and C. Yu (2022, January). Selection with Variation in Diagnostic Skill: Evidence from Radiologists*. *The Quarterly Journal of Economics*, qjab048.

Chan, D. C. and J. Gruber (2020, May). Provider Discretion and Variation in Resource Allocation: The Case of Triage Decisions. *AEA Papers and Proceedings 110*, 279–283.

Chandra, A., E. Flack, and Z. Obermeyer (2021). The Health Costs of Cost-Sharing. pp. 66.

Chandra, A. and D. O. Staiger (2014). EXPERTISE, UNDERUSE, AND OVERUSE IN HEALTHCARE. pp. 25.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011, May). 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.

Coussens, S. (2018). Heuristic Thinking in the Emergency Department. pp. 59.

Currie, J. and J. Zhang (2021, June). Doing More with Less: Predicting Primary Care Provider Effectiveness. Technical Report w28929, National Bureau of Economic Research, Cambridge, MA.

Das, J., A. Holla, A. Mohpal, and K. Muralidharan (2016, December). Quality and Accountability in Health Care Delivery: Audit-Study Evidence from Primary Care in India. *American Economic Review 106*(12), 3765–3799.

de Salud Gobierno de Chile, M. (2017). Orientación Técnica Programa de Salud Cardiovascular.

Diamond, R. and P. Persson (2016, May). The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests. Technical Report w22207, National Bureau of Economic Research.

Donohue, J. M., H. Guclu, W. F. Gellad, C.-C. H. Chang, H. A. Huskamp, N. K. Choudhry, R. Zhang, W.-H. Lo-Ciganic, S. P. Junker, T. Anderson, and S. Richards-Shubik (2018, October). Influence of peer networks on physician adoption of new drugs. *PLOS ONE 13*(10), e0204826. Publisher: Public Library of Science.

Egan, B. M., S. E. Kjeldsen, G. Grassi, M. Esler, and G. Mancia (2019, June). 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.

Einav, L., A. Finkelstein, and P. Schrimpf (2015, May). The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D *. *The Quarterly Journal of Economics 130*(2), 841–899.

Einav, L., A. Finkelstein, and P. Schrimpf (2017, February). Bunching at the kink: Implications for spending responses to health insurance contracts. *Journal of Public Economics 146*, 27–40.

Elnaem, M. H., N. A. Irwan, U. Abubakar, S. A. Syed Sulaiman, M. E. Elrggal, and E. Cheema (2020, November). 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.

Garies, S., S. Hao, K. McBrien, T. Williamson, M. Peng, N. A. Khan, R. S. Padwal, H. Quan, A. A. Leung, and for Hypertension Canada's Research and Evaluation Committee (2019, March). 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.

Gertler, P. and C. Vermeersch (2013, May). Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes. Technical Report w19046, National Bureau of Economic Research, Cambridge, MA.

Greiver, M., S. Kalia, T. Voruganti, B. Aliarzadeh, R. Moineddin, W. Hinton, M. Dawes, F. Sullivan, S. Syed, J. Williams, and S. d. Lusignan (2019, January). Trends in end digit preference for blood pressure and associations with cardiovascular outcomes in Canadian and UK primary care: a retrospective observational study. *BMJ Open 9*(1), e024970. Publisher: British Medical Journal Publishing Group Section: Cardiovascular medicine.

Gruber, J., T. P. Hoe, and G. Stoye (2018, March). Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers. Technical Report w24445, National Bureau of Economic Research.

Hernández-Pizarro, H. M., C. Nicodemo, and G. L. Casasnovas (2020, October). Discontinuous system of allowances: The response of prosocial health-care professionals. *Journal of Public Economics 190*, 104248.

Huang, S. and H. Ullrich (2021). Physician effects in antibiotic prescribing: Evidence from physician exits. pp. 62.

JLN (2017a). Case Studies on Payment Innovation for Primary Health Care. Technical report.

JLN (2017b). Financing and Payment Models for Primary Health Care: Six Lessons from JLN Country Implementation Experience. Technical report.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*(1), 435–464. _eprint: https://doi.org/10.1146/annurev-economics-080315-015234.

Kleven, H. J. and M. Waseem (2013, May). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan*. *The Quarterly Journal of Economics 128*(2), 669–723.

Knaapen, L. (2014). Evidence-Based Medicine or Cookbook Medicine? Addressing Concerns over the Standardization of Care. *Sociology Compass 8*(6), 823–836. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/soc4.12184.

Lanas, F., M. S. Garcia, M. Paccot, J. Peña, P. Torres, T. Beaney, A. C. Ster, N. R. Poulter, and X. Xia (2020, August). May Measurement Month 2018: an analysis of blood pressure screening results from Chile. *European Heart Journal Supplements 22*(Supplement_H), H37–H39.

Mavrokonstantis, P. (2019). Introduction to the bunching Package. pp. 19.

Miao, F., B. Wen, Z. Hu, G. Fortino, X.-P. Wang, Z.-D. Liu, M. Tang, and Y. Li (2020, August). Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques. *Artificial Intelligence in Medicine 108*, 101919.

Mullainathan, S. and Z. Obermeyer (2021, December). Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care*. *The Quarterly Journal of Economics*, 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., and null null (2019, May). Measurement of Blood Pressure in Humans: A Scientific Statement From the American Heart Association. *Hypertension 73*(5), e35–e66. Publisher: American Heart Association.

Myers, M. G., M. Godwin, M. Dawes, A. Kiss, S. W. Tobe, F. C. Grant, and J. Kaczorowski (2011, February). Conventional versus automated measurement of blood pressure in primary care patients with systolic hypertension: randomised parallel design controlled trial. *The BMJ 342*.

Myers, M. G., M. Godwin, M. Dawes, A. Kiss, S. W. Tobe, and J. Kaczorowski (2012, August). Conventional versus automated measurement of blood pressure in the office (CAMBO) trial. *Family Practice 29*(4), 376–382.

Olenski, A. R., A. Zimerman, S. Coussens, and A. B. Jena (2020, February). Behavioral Heuristics in Coronary-Artery Bypass Graft Surgery. *New England Journal of Medicine 382*(8), 778–779. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMc1911289.

Ostchega, Y. and D. T. Nguyen (2020). Hypertension Prevalence Among Adults Aged 18 and Over: United States, 2017–2018. (364), 8.

PAHO (2020). Chile.

Piper, M. A., C. V. Evans, B. U. Burda, K. L. Margolis, E. O'Connor, and E. P. Whitlock (2015, February). 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. *Annals of Internal Medicine 162*(3), 192–204. Publisher: American College of Physicians.

Runnacles, J., A. Roueché, and P. Lachman (2018, February). The right care, every time: improving adherence to evidence-based guidelines. *Archives of Disease in Childhood - Education and Practice 103*(1), 27–33. Publisher: Royal College of Paediatrics and Child Health Section: E&P: Quality improvement.

Saez, E. (2010, August). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Singh, M. (2021, October). Heuristics in the delivery room. *Science 374*(6565), 324–329.

Socrates, T., P. Krisai, A. S. Vischer, A. Meienberg, M. Mayr, and T. Burkard (2021, January). Improved agreement and diagnostic accuracy of a cuffless 24-h blood pressure measurement device in clinical practice. *Scientific Reports 11*(1), 1143. 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.

Stergiou, G. S., E. O'Brien, M. Myers, P. Palatini, G. Parati, and t. S. B. S. A. Board (2020, March). STRIDE BP: an international initiative for accurate blood pressure measurement. *Journal of Hypertension 38*(3), 395–399.

Timmermans, S. and A. Mauck (2005, January). The Promises And Pitfalls Of Evidence-Based Medicine. *Health Affairs 24*(1), 18–28. Publisher: Health Affairs.

Tversky, A. and D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases. *185*, 10.

Vrijens, B., S. Antoniou, M. Burnier, A. de la Sierra, and M. Volpe (2017). Current Situation of Medication Adherence in Hypertension. *Frontiers in Pharmacology 8*, 100.

Wilcox, J. (1961, January). Observer Factors in the Measurement of Blood Pressure. *Nursing Research 10*(1), 4–17.

Williams, J. S., S. M. Brown, and P. R. Conlin (2009, January). Blood-Pressure Measurement. *New England Journal of Medicine 360*(5), e6. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMvcm0800157.

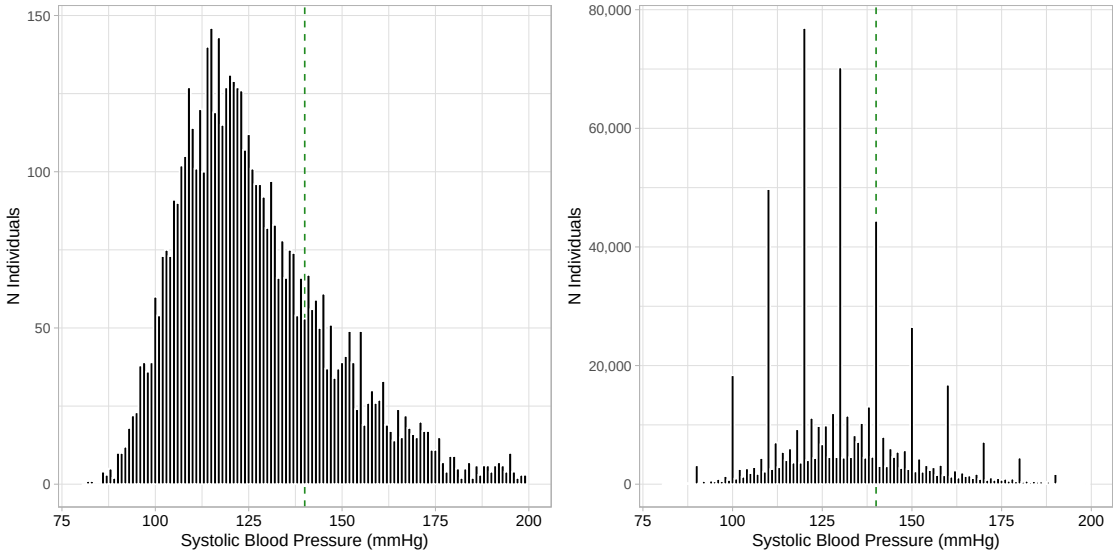
Wohlfahrt, P., R. Cífková, A. Krajcoviechová, P. Šulc, J. Bruthans, A. Linhart, J. Filipovský, O. Mayer, and J. J. Widimský (2020, April). Comparison of three office blood pressure measurement techniques and their effect on hypertension prevalence in the general population. *Journal of Hypertension 38*(4), 656–662.

Zheng, K., R. Padman, D. Krackhardt, M. P. Johnson, and H. S. Diamond (2010, May). 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.

7 Tables and Figures

Figure 1: Expected vs. Observed Distribution of Systolic BP

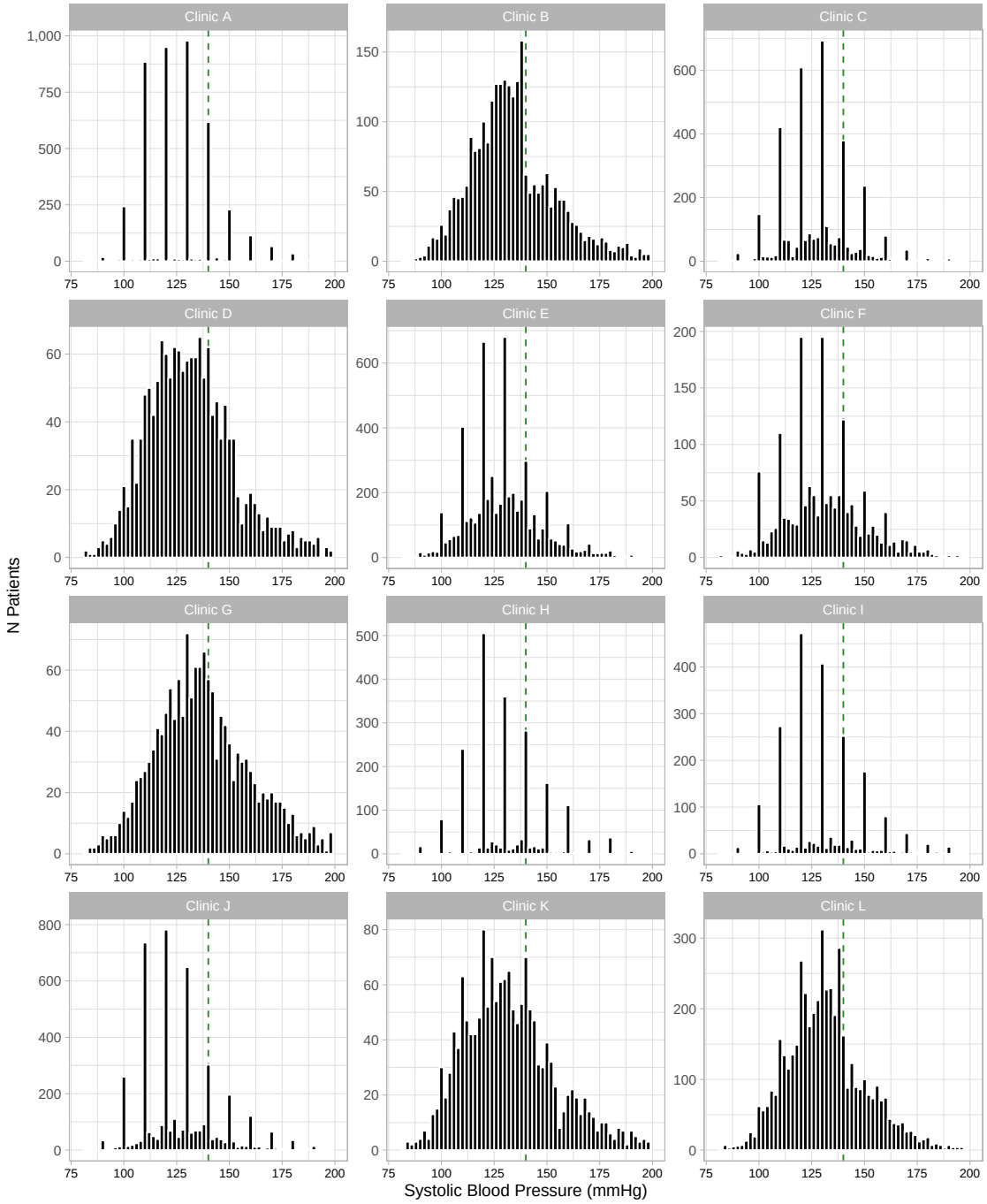
Panel A: National Health Survey. **Panel B**: Primary Care Records



Note: Comparison of expected (panel A) to observed (panel B) distribution of systolic BP in two large populations. Panel A shows the distribution of systolic BP fromres Chile's nationally representative health survey (ENS 2016-2017). Panel B shows the distribution of systolic BP 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 were diagnosed with hypertension before this instance of BP was measured.

back to section 2.2

Figure 2: Examples of the Distribution of Systolic BP at 12 Clinics

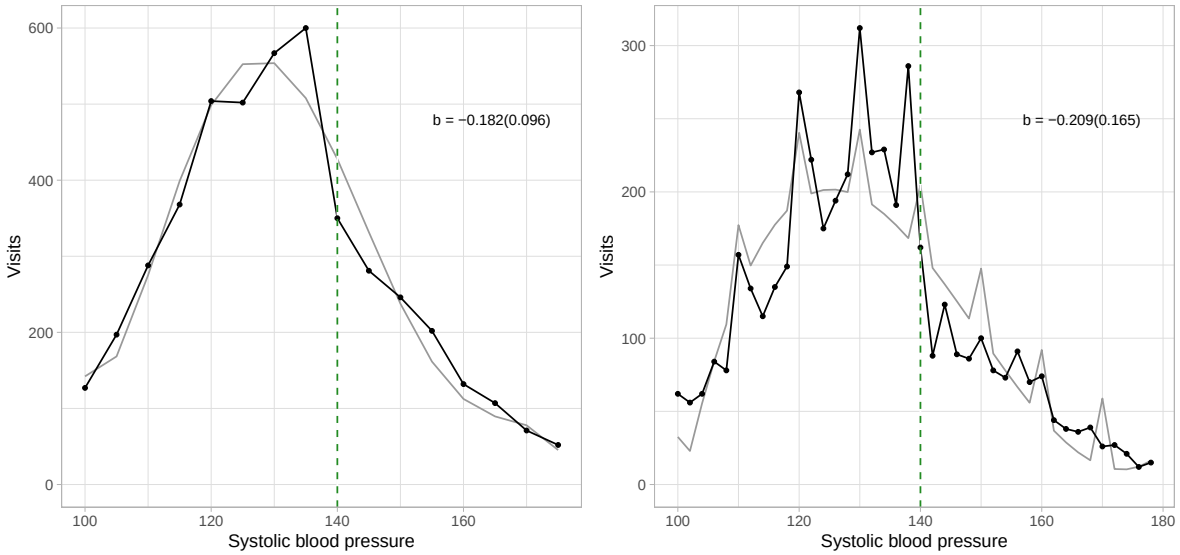


Note: Distributions of systolic BP from 12/257 primary care clinics from preventative care visits among patients who were not previously diagnosed with hypertension. Clinics A, C, E, F, H, I and J show substantial end digit bias: numbers that end in zero are over-represented. Clinics A, B, E and L show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

back to section 3.1.2
back to section 4.1

Figure 3: Example of counterfactual bunching estimation

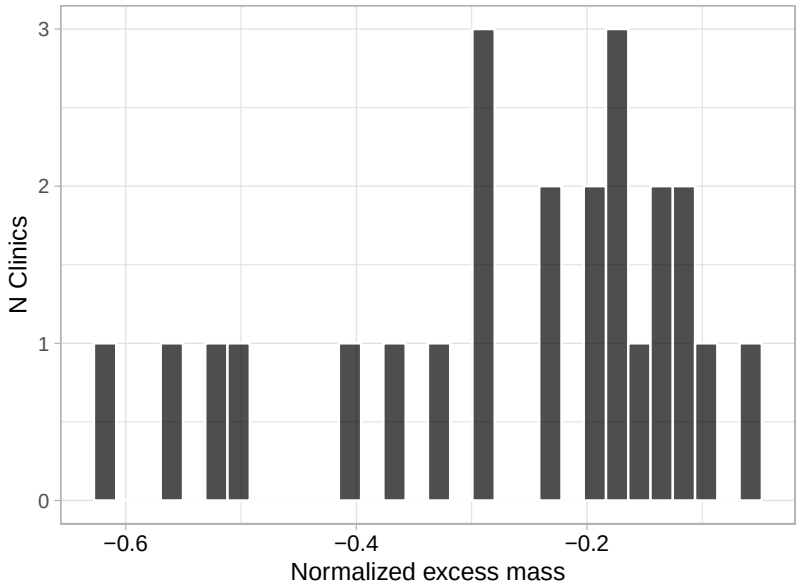
Panel A: Clinic with little rounding. **Panel B**: Clinic with high rounding



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).

back to section 4.1

Figure 4: Estimated Magnitude of Discretion of BP

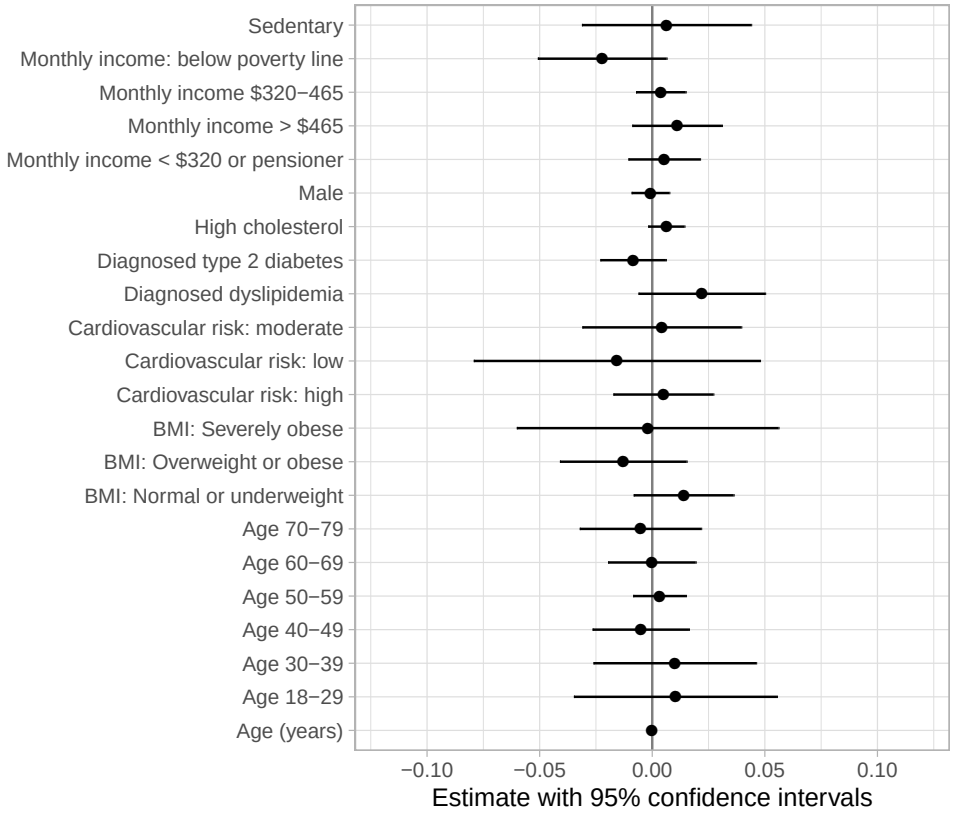


Note: Histogram of the magnitude of discretion (estimated using bunching) for 24 of 257 clinics. Only clinics with statistically significant negative bunching are included. A negative number indicates missing mass to the right of the threshold, and excess mass below it. Each b estimate is the number 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).

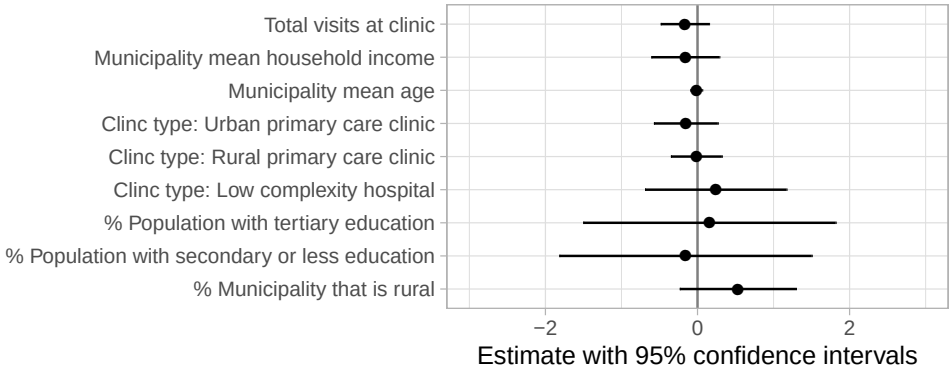
back to section 4.1

Figure 5: Balance of Patient and Clinic Characteristics by Magnitude of Discretion

Panel A: Patient Level Characteristics



Panel B: Clinic and Municipality Characteristics



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.

back to section 2.3.2

Table 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 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.

back to section 2.3.2

Table 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.

back to section 2.3.2

Table 4: Impact of Discretion on Hypertension Diagnosis and Prescription

	Hypertension Diagnosis	**Hypertension Prescription**
	(1)	(2)
Discretion x BP<140	−0.008	−0.019
	(0.005)	(0.014)
Discretion	−0.006	0.007
	(0.006)	(0.010)
BP<140	−0.251***	−0.169***
	(0.005)	(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 BP (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.

back to section 4.2

Table 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)
Discretion x BP<140	−0.018**	−0.039***	−0.061***	−0.028**	−0.040***	−0.036*	0.003	0.001	−0.001
	(0.009)	(0.012)	(0.019)	(0.012)	(0.013)	(0.019)	(0.006)	(0.010)	(0.013)
Discretion	0.018**	0.039***	0.060***	0.022*	0.033**	0.030	−0.005	−0.008	−0.011
	(0.009)	(0.012)	(0.017)	(0.012)	(0.014)	(0.020)	(0.005)	(0.008)	(0.010)
BP<140	−0.029***	−0.061***	−0.107***	−0.025***	−0.062***	−0.119***	0.010	0.013	0.003
	(0.005)	(0.009)	(0.014)	(0.009)	(0.013)	(0.019)	(0.007)	(0.010)	(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 BP (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.

back to section 4.3

Table 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	0.010	−0.027	−0.047	0.038	0.114
	(0.051)	(0.065)	(0.100)	(0.105)	(0.160)	(0.216)
Discretion	−0.004	−0.039	0.017	0.066	0.007	−0.152
	(0.047)	(0.063)	(0.092)	(0.102)	(0.160)	(0.230)
BP<140	0.008	−0.125*	−0.311***	0.003	−0.228*	−0.480***
	(0.048)	(0.065)	(0.088)	(0.091)	(0.124)	(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 BP (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.

back to section 4.3

Table 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***	−0.003	0.002	−0.007***	−0.005***
	(0.002)	(0.005)	(0.001)	(0.003)	(0.002)
Discretion	0.008***	−0.005	−0.001	−0.005	−0.008
	(0.002)	(0.006)	(0.003)	(0.005)	(0.005)
BP<140	−0.096***	0.007*	−0.019***	−0.141***	−0.127***
	(0.002)	(0.004)	(0.001)	(0.004)	(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 BP (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.

back to section 4.4

8 Appendix

Table A1: Inclusion Criteria for Primary Care Electronic Health Records Data

Inclusion criteria	**N patients**	**N clinics**	**% patients excluded**
-	1,448,800	552	-
BP 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%

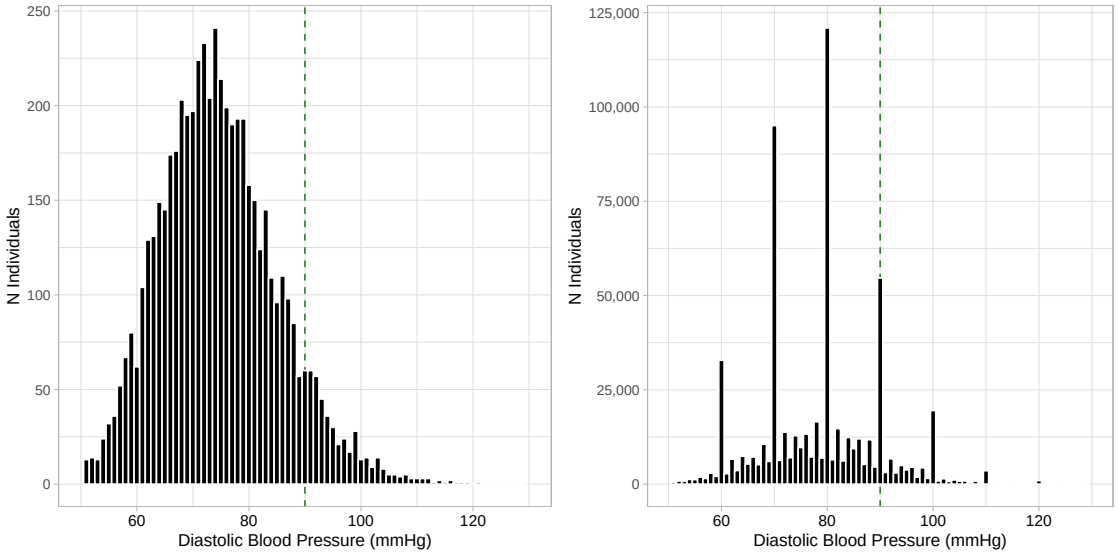
Note: Inclusion criteria and number and percent of patients excluded at each step. Large is defined a clinic above the median number of visits. The final sample is 619,907 patient-visits and 257 clinics.

Table A2: Chile's Clinical Practice Guidelines: Frequent errors that result in an inaccurate measurement of BP

Factor	**Discrepancy in BP systolic/diastolic (mmHg)**
Did not rest 5 minutes prior to measurement	10-20/
Talking or active listening during measurement	10-17/10-13
BP cuff over clothing	5-40/
BP cuff too tight	10-15/2-8
BP 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/

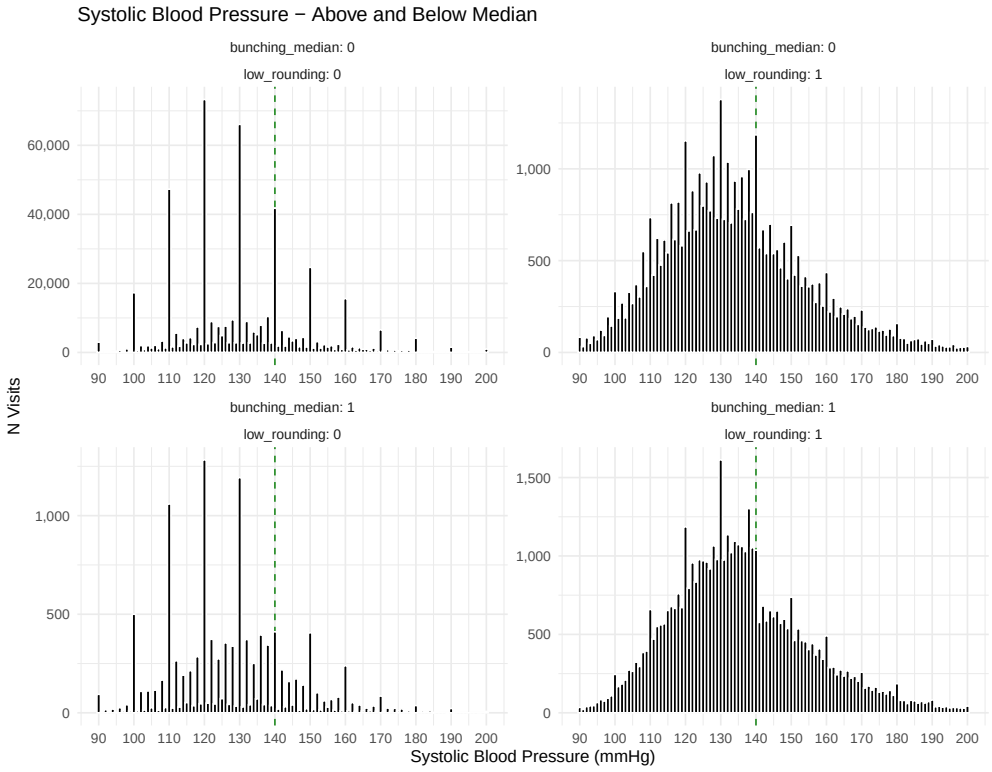
Note: This table is the author's own translation of appendix table 3 from Chile's 2017 clinical practice guidelines for cardiovascular disease management (Orientación Técnica Programa de Salud Cardiovascular 2017, page 37). If no number is included after the slash it means no change in diastolic BP is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

Figure A1: Expected vs. Observed Distribution of Diastolic BP
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 BP in two large populations. Panel A shows the distribution of diastolic BP from Chile's nationally representative health survey (ENS 2016-2017). Panel B shows the distribution of diastolic BP 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 were diagnosed with hypertension before this instance of BP was measured.

Figure A2: Histogram of Systolic BP by High vs. Low Discretion Clinics



Note: Histogram of systolic BP 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.

8.1 Construction of Variables

Parameterization of discretion:
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 A3, A4, and A5 display coefficients and 95% confidence intervals for all main outcomes estimated with each of the four independent variables.

Table A3: 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.

back to section 3.2

8.2 Impact of Discretion among Alive Patients

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 BP.

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 A4). 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 A5, and so in tables A6 and A7 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 A6). 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 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.

Table A4: 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.

back to section 4.3

Table A5: 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	−0.008
	(0.005)	(0.006)	(0.007)
Discretion	0.002	0.001	0.001
	(0.003)	(0.003)	(0.003)
BP<140	−0.018***	−0.019***	−0.019***
	(0.002)	(0.002)	(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 BP (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.

back to section 4.3

Table A6: 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	−0.107	−0.161
	(0.055)	(0.073)	(0.117)
Discretion	0.036	0.063	0.141
	(0.058)	(0.083)	(0.110)
BP<140	−0.020	−0.242***	−0.513***
	(0.051)	(0.074)	(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 BP (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.

back to section 4.3

Table A7: 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	**Heart attack**	**Heart failure**	**Stroke**	**Heart attack**	**Heart failure**	**Stroke**	**Heart attack**	**Heart failure**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Discretion x BP<140	−0.017	−0.020**	0.007	−0.036***	−0.025*	0.006	−0.057***	−0.026	−0.002
	(0.011)	(0.009)	(0.007)	(0.014)	(0.013)	(0.011)	(0.021)	(0.024)	(0.015)
Discretion	0.018*	0.014	−0.007	0.038***	0.022	−0.010	0.058***	0.022	−0.010
	(0.011)	(0.008)	(0.005)	(0.014)	(0.014)	(0.009)	(0.021)	(0.022)	(0.011)
BP<140	−0.027***	−0.019**	0.007	−0.064***	−0.055***	0.0001	−0.103***	−0.116***	−0.027*
	(0.006)	(0.010)	(0.007)	(0.010)	(0.014)	(0.011)	(0.015)	(0.021)	(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 BP (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.

back to section 4.3

Table A8: Selection by Patient Characteristics - Health History Questions

	Was asked about family history of:				**Patient has family history of:**			
	Any	**Heart disease**	**CVD**	**Dyslipidemia**	**Any**	**Heart disease**	**CVD**	**Dyslipidemia**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discretion x BP<140	−0.005	−0.004	−0.004	−0.004	−0.002	−0.0004	0.0001	−0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.002)	(0.002)	(0.0005)	(0.001)
Discretion	−0.001	−0.001	−0.004	−0.002	−0.008*	−0.008*	−0.001	−0.002
	(0.013)	(0.013)	(0.014)	(0.013)	(0.004)	(0.004)	(0.001)	(0.002)
BP<140	−0.022***	−0.022***	−0.021***	−0.021***	−0.012***	−0.008***	0.004***	−0.009***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.002)	(0.002)	(0.001)	(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

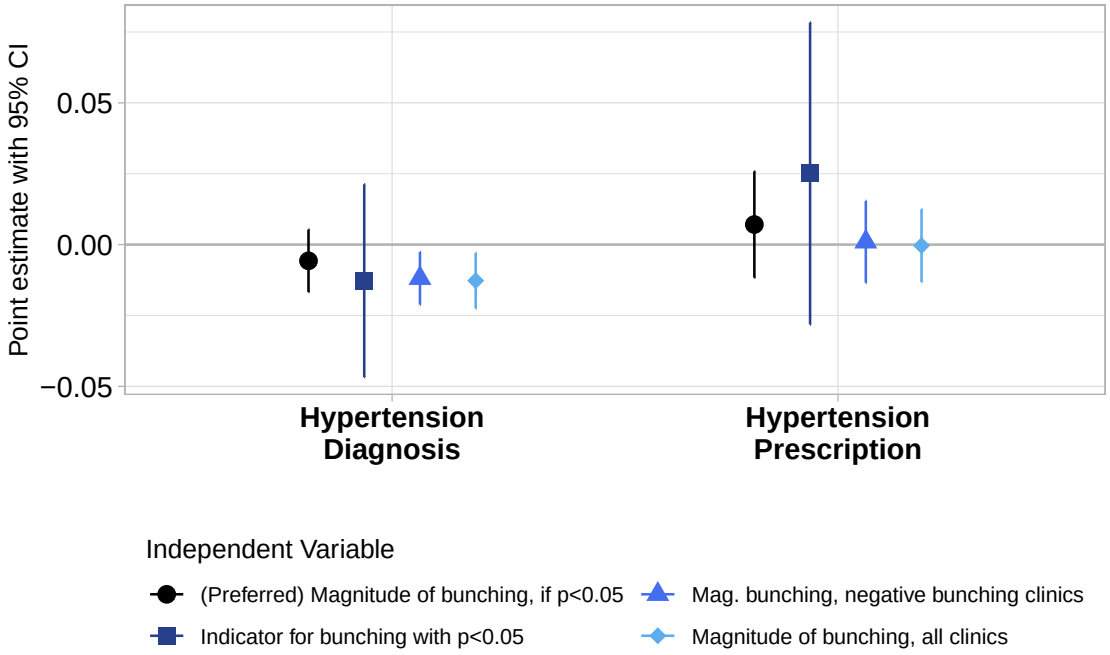
Note: *p<0.1; **p<0.05; ***p<0.01.This table presents difference-in-differences estimates of the impact of crossing the diagnostic threshold from above to below 140 mmHg at high vs. low discretion clinics on health history questions. BP< 140 is an indicator for if recorded systolic BP (BP) was below 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 A9: Selection by Patient Age

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

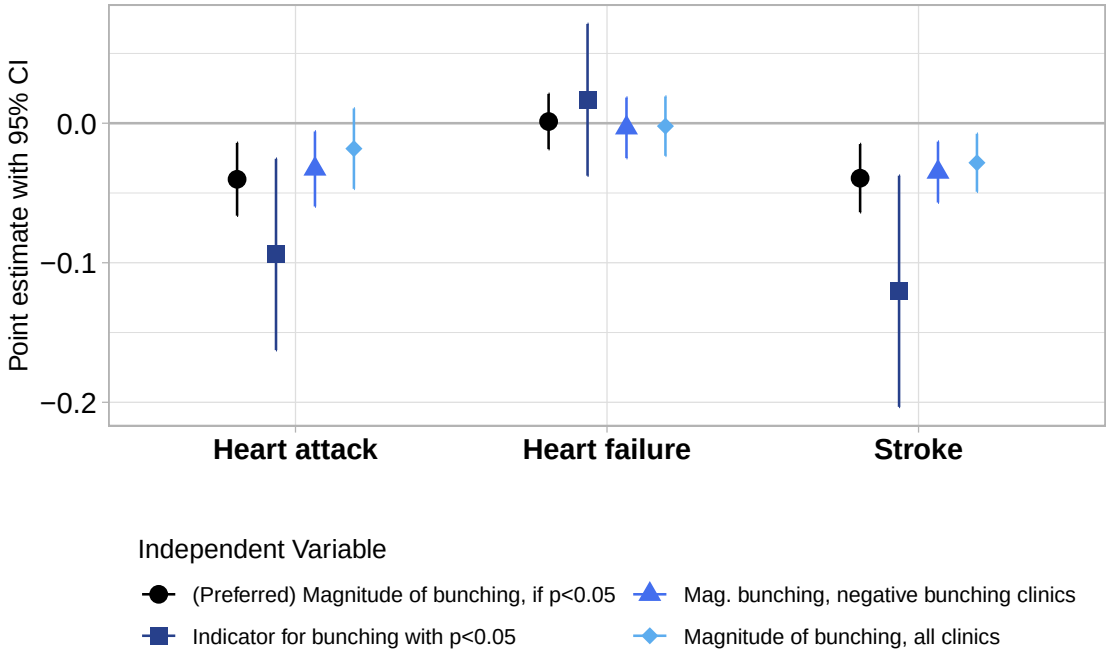
Note: *p<0.1; **p<0.05; ***p<0.01.This table presents difference-in-differences estimates of the impact of crossing the diagnostic threshold from above to below 140 mmHg at high vs. low discretion clinics on patient characteristics. BP< 140 is an indicator for if recorded systolic BP (BP) was below 140. 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 A3: Comparison of results under different parameterizations of the magnitude of discretion: adherence to clinical practice guidelines



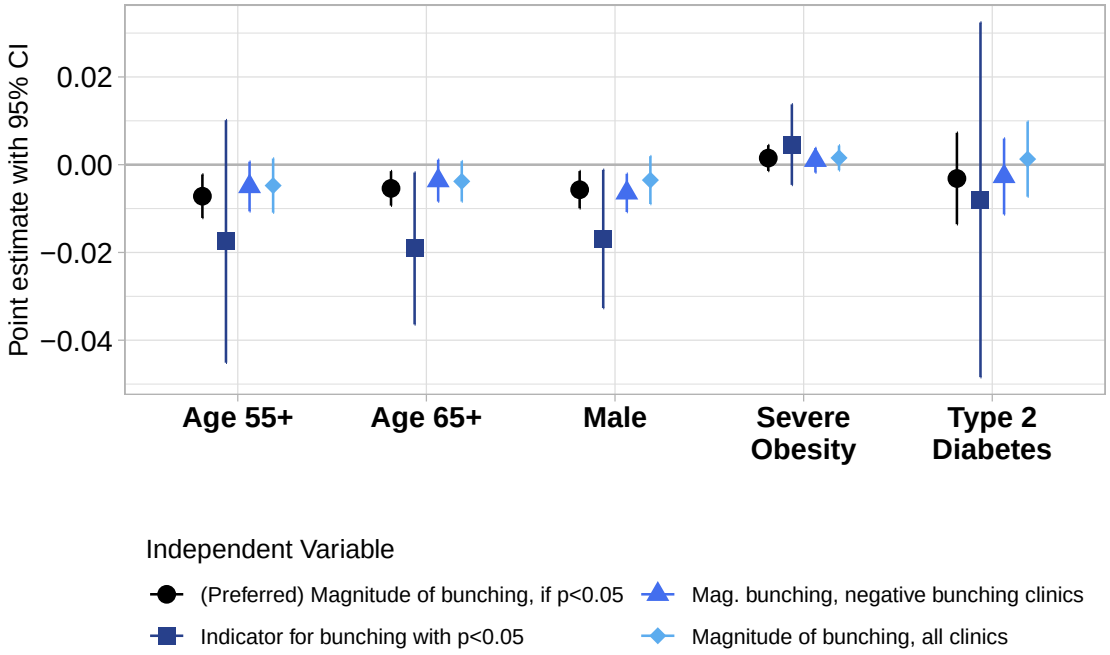
Note: This figure presents the impact of discretion on key outcomes, under different discretion specifications. Point estimates for (Discretion x BP<140) and 95% confidence intervals shown from equation (2) are shown. Each outcome is estimated four separate times with different treatment variables. The 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 A4: 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 (Discretion x 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 A5: 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 x 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.