Private Sector Provision as an "Escape Valve": The Mexico Diabetes Experiment

Ari Bronsoler[*]   Jonathan Gruber[†]   Enrique Seira[‡]

Google   *MIT and NBER*   *ITAM, MSU and JPAL*

September 12, 2023

Abstract

Public health systems are dominant in much of the world, but often face fiscal constraints that lead to rationing of care. As a result, private sector healthcare providers could in theory beneficially supplement public systems, but evaluating the benefits of private alternatives has been challenging. We evaluate a private supplement to the free public health system for one of the world's deadliest health problems, diabetes. We estimate enormous impacts of the private supplement, increasing the share of those treated who are under control by 69%. This effect arises through both improved treatment compliance and health behavior. We find diabetes complications fall in the short run, and that the net costs of this intervention are one-third of the gross costs. The returns to private care do not appear to reflect more productive delivery but rather more attachment to medical care, offering lessons for improving the public system.

---

[*]Behavioral Economics, Google, New York City 10011, United States. arib91@gmail.com

[†]Department of Economics, Massachusetts Institute of Technology. Cambridge 02142. United States. gruberj@mit.edu

[‡]Corresponding author. Department of Economics, Michigan State University. East Lansing 48824-1038. Telephone 517-355-7583. Fax 517-432-1068. Email: enrique.seira@gmail.com

1 Introduction

Many countries of the world feature universal or near-universal public provision of either health insurance or direct health care. Such public provision has become a major strain on the fiscal capacity of nations, with health care as a share of government spending rising from 4.8% in 2000 to 8.4% in 2020.[1] This has led to limits on public financing, from reductions in public reimbursement rates to health care providers, to tightly binding supply limits that create large waiting lines for some services, with controversies around these limits arising in the U.S. (Lopez (2015)), the UK (Campbell and Duncan, 2022), Mexico (INSP, 2011), and many other countries around the world.[2] In a review of several papers on waiting times to access health care McIntyre and Chow (2020) write that "existing data suggest waiting time presents a significant barrier to health care access for a range of health services". OECD (2020) writes that "Long waiting times for health services have been an important policy issue in most OECD countries for many years as they generate dissatisfaction for patients because the expected benefits of diagnoses and treatments are postponed, and the pain and discomfort remain while people wait."[3]

In many countries, private sector health care options provide an "escape valve" from such public sector limitations. Private sector care is highly demanded despite the fact that it is typically excluded from reimbursement and therefore must be paid out of pocket (or privately insured). If there are private options that are particularly cost effective, governments may want to consider either reimbursing those services or expanding public services to mimic them. In essence, private innovation in health care provides a learning laboratory for constrained public systems.

But accessing this learning laboratory is challenging. Use of private providers is endogenous, making it difficult to use cross-sectional variation in source of care to learn about the effectiveness

---

[1] Average health care spending as share of government spending for Australia, Canada, Chile, China, Colombia, France, Germany, Italy, Japan, Mexico, New Zealand, Netherlands, Norway, South Korea, and Sweden. Available at Our World in Data https://ourworldindata.org/grapher/public-health-expenditure-share-gdp-owid Accessed on October 11, 2022.

[2] There is growing discontent with waiting times at public health providers around the world. In the United States, a scandal regarding long waiting times in the Phoenix VA revealed that "patients had to wait an average of 115 days to be seen by a primary care provider... CNN reported that as many as 40 veterans died while on wait lists at the Phoenix VA hospital" (Lopez, 2015). England's National Health Service is also under strain, in 2022 "The total number of people in England waiting for hospital treatment rose again to a record high of 6.8 million at the end of July – almost one in eight of the population", up from 4.2 in 2007 (Campbell and Duncan, 2022). In Canada, the Fraiser Institute reports that in 2021 the median waiting time for a patient between referral from a general practitioner and treatment was 25.6 weeks, 175% higher than in 1993 when it started recording it (Moir and Barua, 2021). And in Mexico, patients have to wait 29.3 weeks on average from their first doctors appointment to their surgical procedure (INSP, 2011).

[3] See OECD (2020), and https://www.oecd.org/els/health-systems/waiting-times.htm. Saturation may not be a worldwide phenomenon, however. One referee noted for instance that the average Nigerian primary healthcare clinic sees 1 patient a day.

1

of private care. Randomized trials are also very difficult in this context due to the challenges of identifying and recruiting those interested in using private care, and the enormous samples needed to obtain sufficient first stage power. As a result, potential cost-effective, or event cost-reducing, changes to the public sector go untapped.

In this paper, we present an approach to tapping those private sector learning opportunities in a cost-effective way. The context for our analysis is one of the most important public health issues facing developing countries: diabetes. After decades of being primarily concerned with undernutrition around the world, policy makers are shifting their focus to this new problem that arises from both improper eating and overconsumption. There were 4.2 million deaths due to diabetes complications in 2019. Worldwide prevalence rates have risen from 4.7% in 1980 to 9.3% in 2019 and the increase has been most rapid for developing nations; the rate of diabetes is now higher in low-income nations than in high-income nations (Saeedi et al., 2019).

We focus on Mexico, one of the nations hardest hit by the growth in diabetes. Diabetes prevalence has risen from 6.7% in 1994 to 11% in 2018, and, signaling poor control, the country has almost twice the mean diabetic hospital admission rate among the OECD countries. Indeed, diabetes is the second most common cause of death in Mexico and is among the top five causes of disability. Alongside these poor outcomes, Mexico spends enormous amounts combating this illness: direct spending on diabetes represents 1.1% of GDP, and 10% of the entire budget of the Ministry of health (OECD, 2019; Barraza-Lloréns et al., 2015; INEGI, 2019b; INSP, 2018). Due to dissatisfaction with the diabetes care provided by the free public system, nearly 20% of diabetics report getting treatment at a private institution for their diabetes. This raises the key questions of whether this private care is improving health outcomes, whether it is cost effective in doing so, and how it delivers such savings.

In this paper, we implement a novel *deniers randomization* approach to cost-effectively provide a causal estimate of enrollment in private diabetes care. At its heart, deniers randomization shows that, under some weak assumptions, we can estimate the same LATE with a much smaller budget by filtering out some always and never takers from the sample. We run this experiment in partnership with a private provider of comprehensive diabetes care in Mexico, Clinicas de Azucar (CdA). CdA runs a chain of clinics that provide a wide range of services to diabetics, ranging from blood sugar measurement to medical interventions to nutritional counseling. This service has a flat rate and is fairly expensive, at a cost of $7000 pesos ($350 USD) per year, or 5% of median Mexican family yearly income, at the time of our experiment.[4]

---

[4]Estimation based on monthly earnings from the 5th and 6th decile in the distribution in INEGI (2018). The cost of CdA would represent 5.4% of income for the 5th decile and 4.5% for the sixth decile. The cost of CdA has risen recently to $8000 pesos.

Our sample construction starts with individuals who undergo a free initial comprehensive diabetes evaluation at CdA, which is part of the standard selling mechanism that CdA uses. We randomize all of those patients into treatment and control groups and survey them at baseline. After their evaluation, following business as usual, diabetic individuals are offered the opportunity to enroll in CdA at full-price (those who take it at full price we call always takers); among those who decline and are willing to hear additional offers (the ones who say no are never takers), an additional opportunity to enroll at 40% of the baseline price is presented to treatments but not controls.

Our experimental sample consists of those who were not screened out as always or never takers. Among the ones offered the discount, 49% end up using the CdA service, compared to 21% enrollment for the control group that does not get this offer (but may receive subsequent marketing from CdA). Due partly to the COVID outbreak, we have attrition in following up our experimental sample, but we were able to partially compensate by visiting each patient at their home for the follow-up survey and blood samples. As a result, a wide variety of tests show no evidence of either differential attrition or bias to our results from attrition.

We find a striking positive effect of the CdA intervention: the implied effect of participation is to lower blood sugar levels (measured by glycated hemoglobin or HbA1c) by a full point (relative to a control mean of 8.5%), and to increase the share of those treated who are under control by 69%. This is a huge impact, which according to the widely cited UKPDS study, could reduce microvascular diabetes complications by 35%. Moreover, this impact is at the upper end of estimated effect sizes from other diabetes interventions reviewed in a recent meta-analysis (Pimouguet et al., 2011a).

We show that this effect arises through a number of changes in behavior, including greater compliance with recommended medications and substitution for less invasive treatments; some change in behavior such as exercise and diet. Importantly we document that the treatment induced more frequent visits to medical providers, and as discussed below we find extra visits improve health.

We find that diabetic complications fall, even in the short run. We also find significant heterogeneity, with those who have worst baseline blood sugar control seeing the largest benefit. Since close to 80% of the patients in the experiment had previously been diagnosed with diabetes and 70% had access to public care, our results highlight the importance of additional care to complement the existing public system; we also find that the effects are not coming solely from enrolling those who had no care at baseline.

We show that the CdA intervention was highly cost effective —and perhaps even cost saving— through reductions in the cost of the public health system. Some of these savings arise from reduced

use of public primary care. But the larger source of savings is the direct positive fiscal externality from improved private care in terms of reduced (publicly paid) hospital spending. Adding these components, we estimate that the net benefit of this supplement in terms of money is between 65 and 105% of its gross costs. At the same time, our estimated health benefits are an order of magnitude greater than the gross costs of the program.

This raises the natural question of *how* the private sector program is providing such cost effective care. In the limit, this would require multiple analyses of different private providers or of variation in the elements of private provision. But we show that combining our data with that from the major public insurer (IMSS) is sufficient to decompose the findings into a more productive delivery of care per visit or simply more quantity of care that is delivered equally effectively. We use variation from the distance to public clinics, to quasi-experimentally estimate the marginal returns per visit to IMSS diabetes care. We find that, after controlling for differential observable selection into private sector care, the returns to public and private care per visit are in fact comparable. This suggests that the returns from CdA reflect the ability of the private vendor to encourage more care, which is consistent with the fact that both offer similar diabetes care services. We confirm this conclusion by showing that our treatment effects are in fact highest where public clinics are most crowded.[5]

Taken together, our findings suggest that private delivery of diabetes care had major benefits in Mexico. It led to improved health and significant offsets to public sector hospital expenditures, and at standard values of additional years of life, it was highly cost-effective. But the source of the improvements was not necessarily better technology for diabetes care, but rather encouraging more care in a case where cost effective care appears constrained. Moreover, the effects arise in the context of a flat payment contract that could, in principle, lead to under provision of care since complication costs are still carried by the public service. This suggests multiple paths forward for governments seeking to improve their diabetes care, ranging from outsourcing care, to improving the attractiveness of the public option.

Our findings contribute to the long-standing discussion on public vs. private healthcare provision. In a systematic review of the literature, Basu et al. (2012) document that the private sector is usually not more efficient, accountable, or medically effective than the public sector but offers better waiting times and hospitality towards patients. Das et al. (2012) and Das et al. (2016) document through standardized patient comparisons in India that both the public and private sector offer similar (low) quality services, with public sector physicians being more prepared but private sector practitioners compensating with more effort per appointment. Knutsson and Tyrefors (2022) find

---

[5]Recent research suggests that hospital crowding reduces care quality e.g. Hoe (2022).

that private ambulances in Sweden perform better than public ones on contracted measures such as response times, and reduce costs, but perform worse on non-contracted measures. In the U.S. context, two recent studies evaluate the role of private options relative to health care for the nation's military and veterans, with mixed outcomes; Frakes et al. (2021) find that children of military personnel born in the private sector have higher costs but better outcomes than those born on military bases, while Chan et al. (2020) find that those receiving emergency care at Veterans Administration hospitals see lower costs and better outcomes.

We contribute to this literature by showing that in our context where public healthcare is overstretched, there was a large health benefit of adding a (subsidized) private option, even for those with free public healthcare, and even if the private option is not necessarily of higher quality per unit of delivery. Second, we run one of the largest double-blind diabetes field experiments to date, and show that standard diabetes care with simple and cheap technology can have an enormous impact on reaching normal sugar levels.[6] Third, we introduce a new cost-effective approach to randomization in evaluating private alternatives for public services delivery, as well as an innovative approach to combining experimental with observational data to decompose the source of efficiency gains. Finally, we show that large fiscal externalities pay many-fold for the subsidy we implemented, contributing to the literature on fiscal externalities from health interventions (Chandra et al. (2010) for a review of this literature).

Our paper proceeds as follows. Section 2 provides background on diabetes in general and the Mexican context, on the Mexican public health care system, and briefly summarizes the conceptual framework for the study. Section 3 describes CdA and the design of the Mexico Diabetes Experiment. Section 4 presents our basic results on outcomes, mechanisms, and heterogeneity. Section 5 estimates the spillovers onto the public sector and the total social value of the improved care. Section 6 then explores the relative efficacy of public and private sector care. Section 7 concludes.

2 Diabetes and the Mexican Health Care Context

2.1 Diabetes Consequences and Measurement

Over the past 25 years, one of the fastest growing public health problems around the world has been diabetes. Diabetes is a progressive and often-fatal disease with no known cure. It can attack every

---

[6]Meta reviews of different kinds of interventions can be found in Ismail et al. (2004); Umpierre et al. (2011); Ajala et al. (2013); Pimouguet et al. (2011b). Notable exceptions to small sample size studies are those focused on trying to understand the correct level of HbA1c for diabetics, like the UKPDS 35 study (King et al., 2001)

organ in the body, resulting in higher risk of heart failure, stroke and poor circulation, which can lead to amputation of extremities, kidney failure, retinopathy and death. Those with Type I diabetes do not produce insulin, which turns glucose (sugar) into energy; those with Type II diabetes do not respond to insulin appropriately and do not make sufficient amounts of insulin. Worldwide, more than 450 million people are estimated to have diabetes.

While diabetes cannot currently be cured, it can be brought under control by following diet and exercise recommendations, closely monitoring blood sugar levels, and adjusting prescriptions accordingly.[7] Unlike several other chronic illnesses such as AIDS or Hepatitis C, diabetes can be easily and cheaply managed with relatively inexpensive medicine; Metformin, which is the most common pill used to control early-stage diabetes, costs under 2 dollars/month (Clinic, 2020).

The gold standard for measuring diabetes status is glycated hemoglobin. A hemoglobin A1c (HbA1c) test measures the amount of blood sugar (glucose) attached to hemoglobin, the part of red blood cells that carries oxygen from lungs to the rest of the body. An HbA1c test shows what the average amount of glucose attached to hemoglobin has been over the past three months; a three month average is used because that is typically how long a red blood cell lives. Individuals are diagnosed as diabetic with an HbA1c level of over 6.5% and diabetic patients are recommended to keep their levels below 7% (NIDDK, 2018). A more accessible and easier to use instrument to monitor blood sugar is the glucometer, which captures the blood sugar levels at any one point in time. While this measure has significantly more variance than HbA1c, it does not require lab processing and patients can use it in the privacy of their own home. Normal levels for this measurement are under 100 mm/hg and a patient will be diagnosed if she gets two fasting measurements over 125 (WHO, 2021).

2.2   The Mexican Health Care System

Health care in Mexico is provided primarily by several public sector institutions. The largest is Instituto Mexicano del Seguro Social (IMSS), the single payer insurance plan for formal sector workers in the country. This program covers formal workers and their families as well as students but also offers a voluntary enrollment option which makes up under 1% of beneficiaries. Every private employer that hires a new employee is required to enroll him/her to IMSS. This service is paid for in 3 parts: on average, the government contributes 5.3% of employees base wages, employers contribute 16.5% and employees another 2.5%.[8] IMSS also covers various benefits beyond health like childcare and workers' compensation for work-related injuries. IMSS runs its own 1522

---

[7]Saeedi et al. (2019) and IDF (2019).
[8]Social Security Law (1995).

primary care clinics, 248 acute care hospitals, and 61 specialty hospitals (IMSS, 2018). Smaller but similar public options for particular sectors such as government workers (ISSSTE), the navy (SEMAR), the army (SEDENA), and for workers of the state-owned oil company (PEMEX). In general, those having a registered salaried job and their families —roughly half of the population— receive their health care through one of these schemes.

Informal workers recieve health care through Seguro Popular. This program recently changed its name to The Wellness Institute (INSABI) and currently covers close to 60 million people, although everyone is eligible to enroll (SSP, 2018). While this service expansion benefits workers in the informal sector, survey evidence shows that Mexicans prefer other options.[9] Today, 83% of the population reports being affiliated to one of these systems, with 45% at formal sector systems, 38% at INSABI, and 0.6% holding private insurance. Total health care spending in Mexico is 5.6% of GDP (INEGI, 2017, 2019a), and public sector spending is close to 3%[10].

Private health care is accessible for whoever is willing to pay and is unrelated to employment status. However in 2020 less than 10% of the Mexican population had private health insurance, and about 3% used private sector hospitals.[11]

Despite the availability of free public health care, diabetes remains a major problem in Mexico. The public health care system has responded with a number of policy efforts, including a program to encourage annual checkups, large scale programs to encourage active lifestyles, and a new tax on sugary drinks and high-caloric foods in 2014.[12] Despite these efforts, obesity rates have not receded in Mexico, and diabetes diagnosis rates have remained at 11% of the population since 2012 (INSP, 2018, 2012; Ángel Rivera Dommarco et al., 2018). Diabetes patients who do not have the disease under control face much higher risks of hospitalization and disability. Mexico has twice the rate of hospitalizations per diabetic rate than the OECD average, and diabetes was the second highest cause of death in the country in 2019 (OECD, 2019; Ojeda, 2019).

While public health care is free, widespread dissatisfaction with the quality and waiting times of the public sector has caused the rapid growth of private health care systems in Mexico. This private care is primarily focused at the primary level; the share of private medical offices went from 5% in 1990 to over 30% in 2020 (SSP, 2021; INSP, 2018). In 2018, 18.2% of diabetics reported getting treatment on private institutions. Patients affiliated to either the formal sector or INSABI

---

[9]Based on data from the health and nutrition survey in 2018 that asks about satisfaction with a service. Data shows that the following percentage of patients believe the service is very good: 40% from private, 26% for IMSS and ISSTE and 20% for SP.

[10]https://tinyurl.com/mv7esafc)

[11]https://www.inegi.org.mx/temas/derechohabiencia/ and https://www.forbes.com.mx/solo-1-de-cada-10-mexicanos-tiene-seguro-de-gastos-medicos/.

[12]Aguilar et al. (2021) and Colchero et al. (2017).

programs can get their medications for free in the pharmacies of their clinics, but often choose to pay a small amount to reduce wait times by going to private pharmacies instead.[13]

The high and growing popularity of private options does not necessarily suggest more effective treatment of diabetes. The available studies suggest that IMSS and Seguro Popular have beneficial health effects on hypertension, diabetes and infant mortality Rivera-Hernandez et al. (2016); Conti and Ginja (2023); Knaul et al. (2012), and on glycemic control (Doubova et al., 2019). But if the private sector has developed a more cost-effective treatment strategy, this provides a low-cost learning opportunity for the public sector. The government could, for example, try to improve outcomes and/or save money by outsourcing care to the private sector. Indeed, the genesis of this project was an effort by the head of IMSS, a U.S. trained economist, to undertake such outsourcing. His project was ultimately terminated due to political opposition by the union of public sector physicians, largely on ideological grounds.[14] Alternatively, the government could go "inside the box" to understand what elements of the private option make it particularly effective and try to incorporate them into public care. Taking advantage of these lessons requires a careful evaluation of the private sector alternative —which is why we designed the Mexico Diabetes Experiment.

2.3   Conceptual Framework

The key conceptual question facing our paper is how a model like CdA could possibly improve health outcomes.[15] Indeed, it is not obvious that supplemental private care will improve the outcomes of diabetics in Mexico. The costs of downstream care (hospitalizations) are borne by the state, not the upstream private provider, reducing incentives for quality and quantity in private provision. There are forms of outcome-based contracting in which the government could engage to improve the private sector incentives, but in our case the financing of private sector care is a pure capitation payment from the patient. In theory, this aspect could lead the private provider to under-provide care on the margin and negatively affect patients, but we find strikingly large positive effects on the number of doctors visits for diabetics, as well as positive impacts on their health.

Two different models could deliver the result that a private supplement improves outcomes. The first is a model with underinvestment in public health care for political economy reasons (Lizzeri and Persico, 2001). Indeed, it is clear that Mexico's public health care system "lacks infrastructure, and has important supply shortages and long waiting times" (Rubli, 2023). Referring to IMSS, the New York Times writes that "preventive medicine is something unknown for a system whose

---

[13]Health Federal Law (1983); Rubli (2023)

[14]See El Economista. Accessed on October 24, 2022.

[15]We are grateful to the editor and referees for suggesting the inclusion of this section

main characteristic is long lines."[16] The Director of Administration of IMSS said that IMSS has an insufficient number of beds and of medical personnel with one-third of the OECD average, and that this generates waiting times of twice the national average.[17] As a result, even though Mexico has "free" public health service, patients often rely on out-of-pocket expenses to cover medical services.[18] Entry by private clinics or pharmacies has occurred not only in Diabetes care but also to deal mostly with respiratory diseases (Rubli, 2023).

With critical underinvestment, the marginal benefit of extra care could exceed the marginal cost, creating incentives for private clinics like CdA to enter and serve the underserved demand at a profit, while simultaneously causing improvements in health for patients. Thus our findings that (a) enrollment in the sugar clinics helps those with diabetes achieve better outcomes, (b) this is because of greater use of the clinic, (c) results are stronger for patients attending more overcrowded IMSS clinics.

Alternatively, relying on a different framework that does not appeal to supply constraints, Hart et al. (1997) explain that the private sector may have an advantage in producing more welfare at lower costs given that it faces higher-powered incentives.[19] In their framework the private sector is more effective than the public sector in producing a good or service when consumers can assess quality and buy directly from the supplier. Our results are consistent with this framework because diabetes and progress in its control are verifiable by a simple HbA1c blood test, and although there is no explicit contract around HbA1c, clients check progress often and can leave CdA if there is little progress and communicate to other potential clients their experience with CdA. Even if quality is unobservable, in Hart et al. (1997) the private sector still dominates when opportunities for cost reduction are small (as is the case with a standard and cheap technology like diabetes control) and the government employees have weak incentives (as is likely the case in IMSS).[20] We don't distinguish between the supply constraints and incomplete contracts frameworks, but it is likely that both forces are at play.

---

[16] https://tinyurl.com/56y7j5mx. This is not a problem in Mexico alone. In a review of several papers on waiting times to access health care across countries McIntyre and Chow (2020) write that "existing data suggest waiting time presents a significant barrier to health care access for a range of health services". OECD (2020) writes that "Long waiting times for health services have been an important policy issue in most OECD countries for many years as they generate dissatisfaction for patients because the expected benefits of diagnoses and treatments are postponed, and the pain and discomfort remain while people wait."

[17] https://tinyurl.com/mskxf5fk.

[18] Out-of-pocket expenditure as a percentage of current health expenditure is 42% in Mexico while it is only 12% for high-income countries (see https://tinyurl.com/mvundpf2).

[19] Knutsson and Tyrefors (2022) find some support for this theory. They find that for hard-to-verify health states, private ambulances in Sweeden tend to have less quality in the diagnosis of patients, arising from lower-quality staffing, presumably in order to cut costs.

[20] In their model residual control rights of profits from cost reductions reside on the entrepreneur but not on the bureaucrat, giving the former greater incentives to cut cost, but this results in lower quality, a cost to the consumer which is not internalized completely by the entrepreneur.

3 The Mexico Diabetes Experiment

3.1 Clinicas de Azucar

One of the private providers of disease care management is Clinicas del Azucar ("Sugar Clinics"). This chain of clinics was founded in the state of Monterrey by U.S.-educated health care entrepreneur Javier Lozano. The first clinic was established in 2011, and the chain has grown to 24 clinics in 5 states.

The Sugar Clinics are a chain of specialized diabetes clinics that provide affordable and comprehensive care. Each patient pays fixed-cost membership fees allowing him/her to have unlimited access to diagnostics, labs and consultations for one year. One of the main selling points of the clinics is that a patient can receive a full diabetes checkup with nutritionist assessments, and recommendations for diet, exercise and medication in under 90 minutes, avoiding several visits and long wait-times in the public sector. The government does not subsidize, hire, or have any relationship with CdA, it is 100% private. Appendix A gives more detail on the Sugar Clinics.

The sugar clinics do not appear to offer a revolutionary type of care nor add many benefits to what is already available for free in the public sector. The main advantage is that a patient can go to any of the branches whenever it suits them to get the care they need without having to wait. This one-stop feature of CdA saves cost and hassle, which is particularly important for diabetics given their physical limitations and the centrality of timely treatment. While public sector hospitals usually have several months of waiting time to get an appointment, in CdA the service is walk-in and immediate. Table OA-1 compares the services provided by the largest public healthcare provider IMSS and CdA and examines how much it would cost IMSS to provide the same services that CdA offers, according to their reported per unit cost. We can see that both suppliers offer similar services and that IMSS would spend at least 20% more to provide the same services.

The approach used by CdA parallels a disease management program generally applied internationally for chronic-obstructive pulmonary disease, certain types of cancer, and diabetes; a similar approach is used by Joslin Diabetes Center in Boston and Apollo Sugar Clinics in India.[21] We therefore view our project as evaluating more generally the provision of privatized diabetes care. Although we cannot say with certainty whether the results from CdA extrapolate to other private providers, our findings on mechanisms in Section 4 suggest that the effects may be quite general.

Non-causal estimates of the impact of CdA are very promising. Before/after estimates from

---

[21]IQEHC (2007).

CdA indicate that enrollees see their HbA1c levels fall by 2 points relative to baseline, and such an effect appears to be lasting. Based on these promising findings, we partnered with CdA to design an experimental intervention to assess the causal impact of their program, with funding from Eli Lilly and Company. We preregistered in the AEA registry.[22]

3.2 Experimental protocol

The past literature suggests three natural ways to set up such an experimental intervention. The first, which we call "overall randomization", involves finding a representative sample of Mexico's diabetics and incentivizing a random sub-sample of them to get CdA service. This would have been extremely hard and expensive as it would have required us to test a large sample of individuals to assess whether they are diabetic. A second approach to inducing differential use of CdA services would be what we call "visitors randomization", which would involve randomizing all those who arrived at CdA for a free screening into a treatment arm that receives a subsidy to enroll in the program. Properly designed, this would allow us to estimate the LATE among those who are interested in CdA services, and who were convinced to enroll by our incentive. The disadvantage of this approach, however, is that it is very underpowered when the private option has a high baseline enrolment rate without the enrollment incentive. At CdA, for example, among those who make an initial visit, 41% voluntarily pay for the program immediately with no discount. Moreover, another 26% show no interest in enrolling in the program at any discount.[23] This 67% combined rate of always and never-takers severely limits the power of the first stage; moreover, it also implies that we would be giving an incentive to the 31% of patients that we know would use the program anyway.

We therefore introduce a third new approach we call deniers randomization. It consists of screening out some always-takers and/or never-takers from the sample. Under some (weak) assumptions which we discuss below, deniers randomization estimates the same LATE that would be estimated with visitors randomization, but with a fraction of observations needed to achieve a given statistical power. Figure OA-4 in the Appendix shows that the extra power gained by deniers randomization allowed us to operate at a budget that was one-tenth of what would have been required by visitors randomization.

---

[22]https://www.socialscienceregistry.org/trials/3589. Table OA-23 reports other pre-specified results. Neither Lilly nor CdA had any say on the hypothesis, methods, or testing. Moreover, our findings are not in the commercial interest of Lilly since we find that treated patients need less insulin.

[23]Since CdA uses free-screenings as a selling mechanism, some people take advantage and seem to come only for the free checkup with no interest of buying or even listening to prices and offers. This behavior is not uncommon and CdA is aware that it happens.

3.2.1 Steps

Our experiment proceeds in several steps, as illustrated in Figure 1. When patients enter CdA clinics, they fill out the baseline survey while waiting for the free checkup process to begin. After they undergo a checkup, individuals meet with a physician who discusses their diagnosis and suggests a potential care package at CdA. We remove from our sample people who were not diagnosed with diabetes. Then, individuals meet with a sales force associate who offers the CdA service at the standard full price. Some of them did enroll at full price and some refused to enroll. We call those that enroll at full price the always-takers; they do not form part of our experimental sample.

After signaling no clear interest in buying at full price, the sales associate asked the client if she was willing to wait 10 seconds to search in the system for discounts. Some clients said they were not interested at all in any discount, we call these never-takers and remove them from the experiment as well.[24] Those that said they could wait were consulted for treatment status in the computer by the salesforce. We observe in the data whether a person was consulted or not and define our experimental sample based on that variable, since always-takers had already bought and did not need to be consulted and never-takers were not interested in any offers. If the individual showed up as being in the treatment group —which happened 50% of the time— the sales associate would then offer her a 60% discount.[25]

We were particularly concerned that the sales force might not wait until after the enrollee declined the full-price membership to offer them the discount lottery, thereby including most always-takers and reducing the power of our deniers randomization. We pursued three strategies to address this concern. First, we offered higher bonuses to the sales force for full price than for discounted sales. Second, we carefully instructed the agents on the importance of first ensuring a lack of interest in the full-price membership before offering the discount. Third, we had bi-weekly meetings with the entire sales force where we reinforced this and presented hypothetical cases where they actively participated. Treatment status was blind to the nurses and doctors operating the clinics, and to the CdA team, making it difficult to differentiate the treatment itself as a function of the experimental arm.[26] Spillovers are unlikely. On average we had only 3 clients per branch per day making it unlikely that heard offers we made to other clients. CdA personnel told us that they did

---

[24]CdA offers a free checkup as a marketing device. It is well known to CdA management that many of the people who go do not intend to buy in the first place, and only go for the checkup, that is why they are unwilling to listen to any kind of discount. Removing always-takers and never-takers saves on sample size and increases our statistical power. Table OA-18 compares always-takers, never-takers and our experimental sample. Our sample has patients with lower income and lower education than the other two groups, 0.8 points higher HbA1C levels than the never-takers but 0.6 points lower than always-takers. They are similar in public health coverage, age, and trust in CdA.

[25]We chose the 60% discount based on a pilot run with CdA which showed that such a discount yielded a 48% take up rate.

[26]As one referee pointed out, the fact that LATE is similar to OLS is also consistent with this.

not get abnormal requests for discounts and in the data we don't see any trend of more denials at full price.

The benefits of deniers randomization are clear, but what assumptions are needed for the LATE estimated by deniers randomization to be the same as that estimated by visitors randomization? Angrist et al. (1996) show that under the assumptions of stable unit treatment value, the exclusion restriction, relevance, and monotonicity, a randomized encouragement design identifies the average treatment effect for compliers (LATE). Because LATE uses only compliers for identification, excluding always-takers and never-takers from the sample would not change the LATE estimate. In Appendix B we show that the LATE estimated from the deniers randomization approach is equivalent to the LATE estimated from visitors randomization under three additional assumptions: (i) No exclusion of would-be compliers. That is, we screen out only sets of always-takers and/or never-takers from the sample; (ii) No direct causal effect of screening on enrollment, (iii) No direct causal effect of screening procedure on the outcome of interest. Appendix B discusses why these assumptions are likely satisfied in our experiment.

We believe deniers randomization could be applied in private clinics serving patients with chronic diseases such as hypertension, Asthma, AIDS, Coronary Artery Disease, and Chronic Kidney Disease, among others.These chronic diseases are sufficiently well understood that there are standardized guidelines and medication to follow. The procedure requires having a seamless and quick screening procedure to enroll clients, as well as making subsidized offers to those that deny service at a higher price. There are several private clinics and hospitals in the developing world where it could potentially be implemented (e.g. Narayaha Health in India, Bumrunghrad in Thailand, etc).

3.3   Sample

We recruited individuals over the period from June 2019 to February 2020, Figure 2 summarizes our recruitment. We approached 7,882 individuals who showed up at CdA offices for a free checkup. All of these individuals were asked to fill out a baseline survey while waiting to receive the checkup. The survey contained basic demographic questions, questions on the use of health systems, on personal health, habits, trust in service providers, among others. 94% of both the treatment and control arms completed at least part of the baseline survey.[27] Moreover, we can see that among patients who filled out the survey, the sales force screened out 67% of the individuals: the 41% who were always-takers of the initial offer, and the 26% who were never-takers, uninterested in our

---

[27]Overall, we have 87% of surveys with complete demographic information which we use on our main specifications, and our results do not change when restricting the sample to full surveys.

discount offer. Therefore, 33% were consulted for treatment/control status in the second step.

The last row of Figure 2 shows that the balance of our sample along a number of dimensions is excellent and that the rate of patients screened-out is similar across both treatment and control.[28] We ended up with 1226 individuals in the treatment arm, and 1184 individuals in the control arm. They are very well balanced in terms of baseline HbA1c, weight and age. Importantly, there is a significant difference in the odds of using the services provided by CdA. Among the controls, 21% eventually enrolled in CdA, probably because of additional efforts made by CdA's marketing.[29] Use was 28 points higher, among the treatment group.

After an eight-month enrollment period that ended on February 2020, we conducted a follow-up starting in June 2020 with professional surveyors and nurses visiting each patient's home to implement the questionnaire and take a blood-sample that allowed us to measure HbA1c, the key diabetes control metric. Our survey team was composed of part of the personnel who do Mexico's main health survey (ENSANUT), they are highly trained and complied with stringent health protocols. Our follow up was delayed by a month due to COVID, after which we were able to undertake the in-home survey done with proper COVID precautions.

We did face sizeable attrition, much of which is due to incorrect addresses (close to 30 percent of the sample) and individuals who moved (12 percent of the sample). These should not be impacted by treatment, as addresses were obtained before treatment assignment. Removing these individuals, we have a response rate of 75%, which reflects the enormous effort our survey team put into visiting each patient in their home to get survey results and blood samples. Nevertheless, given that this is still far from 100%, we engage in a battery of tests to ensure that differential attrition is not driving our results, which we summarize here; Appendix C has more details on the tests.

To begin with, Table 1 regresses a dummy for follow-up data collection on our treatment indicator, we find no evidence of differential attrition between treatment and control groups. Table 2 then shows also that there are no differences on a large battery of baseline characteristics between treatment and control, which is also comforting. The difference in HbA1c is an insignificant 0.15 points, which is only 1.6% of the mean and roughly one-eighth of our estimated treatment effect; the difference in the share out of control is an insignificant 0.02, which is below 10% of the control mean and roughly one-tenth of our estimated treatment effect. The sample is also very well balanced on demographics and type of insurance coverage.

---

[28] 40.3% and 41.5% were always takers for the treatment group and control group respectively.

[29] Following typical CdA practice, people who did not purchase would be called several times during the next two weeks to encourage enrollment. This follow up is orthogonal to treatment status, as treatment status was blind to the marketing department. These subsequent contacts would also offer additional small (10-20%) discounts over the full price.

In the Appendix, Table OA-4 shows that treatment by itself and interacted with covariates cannot predict attrition as we cannot reject the null that all their coefficients are zero (p-values > 0.50). We also show that our results are robust to the inverse probability weighting strategy suggested by Molina-Millan and Macours (2021). Lee bounds for the ITT and IV estimates are wide exclude zero, so that we can reject no effect of our intervention. We also implement the method of Behaghel et al. (2015), which allows for non-parameteric identification of average treatment effects even with differential unobservable attrition. It yields very similar results. Finally, matching with outside administrative data on work and on voting (where we can find those that did not respond the follow up survey) shows no significant differences between treatment and control groups. This set of tests leaves us confident that attrition is not biasing our findings.

More broadly, Table 2 also shows that our treatment and control samples are fairly unhealthy. Among controls, the mean level of HbA1c is 9.35, well above the control level of 7, and 76% of our sample is out of control. Mean BMI is 31, which is outside the recommended range of [18.5-25] and means that the average person who shows up to CdA is obese. The mean age of the control group is 52.4 years, and 32% are male; the lower male share may reflect the willingness of men to buy at full price, while wives may want to consult with spouses as well as the fact that women are more keen to answer our follow-up. Roughly 75% have access to public health systems. The self-reported level of trust in CdA is higher than that of alternative providers.[30]

To assess external validity, we can compare our sample to the full sample of diabetics surveyed in the Mexican health and nutrition survey (ENSANUT) on Table OA-17. We find that our sample has a higher HbA1c level and is more likely to be getting treated for their diabetes, but is less likely to have hypertension and has a similar overall level of complications.

4   The Causal Impacts of CdA

4.1   Effect on Blood Sugar

We implement our evaluation of the experiment through a straightforward regression framework. We initially estimate OLS models of the effect of HbA1c of the form:

---

[30]The sample that did answer our follow-up is composed of more women, a bit poorer and less educated individuals in general, but both samples appear to have the same health on average and the same access to public institutions. Importantly, there is no indication of differential attrition across any dimension. This comparison is shown in Table OA-4 in Appendix F.

$$Y_{ijt} = \beta_0 + \beta_1 \text{Use}_i + \Gamma_j + \psi_t + \chi X_i + \varepsilon_{ijt} \qquad (1)$$

where Y_{ijt} is HbA1c values for individual i who enrolled in clinic j on month t, U_i is an indicator for using CdA services (an endogenous variable), Γ_j and ψ_t capture clinic and month of enrollment fixed effects and χ_i includes controls for baseline HbA1c, BMI, gender, age, schooling, and income. These controls are missing for about 100 patients, lowering our regression sample size to 939; results are almost identical if we exclude the 5 variables and use the somewhat larger sample. Standard errors are robust. To estimate casual effects, we use the standard two stage least squares approach of instrumenting CdA usage by being in the treatment group to estimate the LATE.

Our main results are shown in Table 3. Column 1 shows the OLS estimate of using the Cda services on blood sugar, which indicates that CdA lowers blood sugar by -0.98 points, off a base of 8.54 points. This is much lower than the 2-point estimate what CdA told us they were finding using before/after comparison. An important part of the difference is improvement among those who do not use CdA but may go elsewhere for care —in fact, we see that our control population improves 0.9 points relative to baseline. Another part of the difference could arise through selection of those who stay in the program, and therefore continue to be measured by CdA's internal metrics, and those who leave (these later ones are also included in our evaluation since we go to their homes, but not in the internal CdA analysis).

Our results do not incorporate the value of the initial evaluation done by CdA. Our estimate compares treatments and controls conditional on the initial evaluation. But if the initial evaluation itself has value in terms of helping potential enrollees understand how to manage their diabetes, it could lead to some decline relative to their level of blood sugar at entry to the clinic.[31] As a result, our experimental effect is likely a lower bound on the total impact of interacting with CdA.

The rest of the table shows the coefficients on the control variables. By far the most important control variable is baseline HBA1c, with each point in baseline HbA1c associated with a 0.5 percentage point level of HbA1c at follow up. The other significant relationship is with income, where being in the second lowest income group is associated with HbA1c that is 0.57 points higher. Interestingly, we find little impact of BMI conditional on baseline blood sugar. The other coefficients are also insignificant.

Column 2 shows the corresponding IV estimates, where we instrument use of CdA with the treatment indicator. We find an effect that is comparable, but about 10% bigger than the naive

---

[31]Note that this is not a particular limitation of our deniers randomization approach —any approach which conditions on individuals arriving at CdA and being evaluated would suffer from this problem.

OLS, at -1.1 points. Another way to illustrate the effects of the intervention is to look at the share of individuals who have their blood sugar under control, defined as a level of HbA1c below 7. Columns 3 and 4 present these results. For this outcome, our IV estimate shows that 22% of individuals are brought under control by the intervention, which is more than two thirds of the baseline rate of control in our sample.[32]

These are very large effects. For example, a set of recent meta-analysis shows that, on average psychological interventions reduce HbA1c by 0.32, physical activity interventions by 0.67-0.89, self-monitoring of blood-sugar by 0.39, dietary approaches between 0.12 and 0.47 and disease management programs similar to the one we are working with show reductions in HbA1c between 0.38 and 0.45 points, less than half of what we find. This effect is also due to something more than increased adherence as the papers estimating the effect of adherence on outcomes report an effect that is about 40% of the one we find.[33] A widely cited study notes that a reduction of the magnitude we find in Table 3 is sufficient to reduce complications by 35% and reduce mortality by 4% (King et al., 2001; Arnold and Wang, 2014). Moreover, increasing the fraction of patients with HbA1c under control by 69% makes an enormous difference for life expectancy; each year a patient spends with diabetes out of control is estimated to reduce life expectancy by 100 days (Heald et al., 2020).

4.2 Mechanisms

Our survey results allow us to explore a variety of mechanisms through which CdA may have had its effects. The results of this analysis are shown in Table 4, using the IV version of equation 1. All dependent variables come from the responses to the follow up survey; the number of observations varies.[34]

We find that CdA enrollment leads to a very large increase in medical care. Total physician visits increase by 2.6, which is 40% of the control mean.[35] Column 2 shows the number of visits

---

[32]As explained above, the treatment condition was blind to the clinic, making differential service for treatment and control extremely unlikely. The concordance between OLS and LATE helps with this concern as well (we thank a referee for pointing this out).

[33]Ismail et al. (2004), Umpierre et al. (2011) Ajala et al. (2013), Pimouguet et al. (2011b) Krapek et al. (2004) Krass et al. (2015)

[34]The appendix contains two strategies to deal with multiple testing. The first creates a standardized index by families of outcomes as suggested Anderson (2008) and used by Banerjee et al. (2015); Kling et al. (2007), Table OA-20 shows these results. In the second, we use the False Discovery Rate for multiple testing adjustments (Anderson, 2008). Table OA-21 shows the results. Multiple testing corrections do not change our claims.

[35]Notably, this increase happens across the intensive and not the extensive margin; only 17% of patients in the control group report getting no care while 12.5% of patients in the treatment group report the same. It is also not due to the treatment individuals being told that they are diabetics –80% already knew this, and the control group also received the diagnostic.

that involve specific checkups on potential complications from diabetes, and we find that these almost double relative to baseline. The number of visits to diabetes specialists rises by almost 50%. Clearly, an important mechanism for our HbA1c results is that patients are getting a higher *quantity* of care. We hypothesize that part of the reason is that CdA has less hassle cost and waiting time, both to get an appointment and while at the clinic. First, CdA does not require a previous appointment whereas for our sample the average waiting time for IMSS is 28 days. Second, once at the clinic, the wait time while at CdA is 24 minutes for our sample, and 49 minutes per appointment at IMSS. Moreover, only one visit to CdA is needed to see a doctor, a psychologist and a nutritionist, while separate appointments for each are needed at IMSS. Below, we show that the effect of CdA is larger where the public sector clinics are more crowded, suggesting that wait times play an important role in this context.

Since CdA does not bear the cost of downstream hospital care, and the clinics receive a lump sum payment, one might be concerned that there is an incentive to skimp on quality of care – yet we find large improvements in health. Ex-post, there are two possible reasons why supplemental care does have such positive impacts. First, with low marginal costs of additional visits, and quantity-sensitive patient demand, there is a strong incentive for these clinics to not constrain visits. In that case CdA may have only weak incentives to engage in quantity rationing. At the same time, the marginal costs of additional visits are low –particularly relative to the high health benefits that we document.[36] If patient demand for health services is sensitive to unlimited access to doctors (one reason CdA attributes to their success), there are dynamic incentives to provide a high-quality product – at least in terms of patient experience and engagement, which we argue is the main driver of health gains from this model.

Second, it is possible that consumers (and not CdA) drive the utilization decision, and a one-time payment (per month or per year) allows patients to visit the more convenient CdA clinics, substantially increasing use. Once again, given that (as we show later) our results imply that quantity of care, not quality of care, is the centerpiece of health improvement, reducing marginal barriers to additional care (both in time and money terms) may be a cost-effective way to improve health. This does not imply that alternative contracting arrangements cannot be better, but it does suggest that the capitation payment is enough to drive health improvements in our context. Finally, our context may be an inefficient situation where public healthcare supply is outstripped by demand. In such a context, even a potentially sub-optimal contract like capitation may deliver health improvements.

---

[36]The salary of a doctor employed in CdA is about $20,000 pesos per month ($1000 USD), and of a nurse is $9000 pesos per month approximately ($450 USD), the cost of a blood test is less than $100 pesos, and a typical doctor a nurse serves around 20 patients per day. Thus, per visit, just on variable costs, it costs about less than $70 pesos, and with the HbA1c test it may cost say $170 pesos to serve a patient. Medication (including insulin) is sold separately.

We also find evidence for the key mechanism of drug compliance. In the follow up survey we ask individuals if they are likely to stop medication if they "feel good". The right answer to this question is clearly no: diabetes cannot be cured, and ongoing medication remains very inexpensive compared to the underlying health risks. Yet 25% of enrollees in the control group answer yes, indicating inappropriate use of medication. We find that the estimated effect of CdA is a reduction of this response by 22%.

Corresponding to this finding, we find a dramatic shift in how individuals intervene to manage their blood sugar. We see a significant rise in the use of cost effective (typically generic) blood sugar-reducing medication, with use rising by 25% from a baseline of 73%. Correspondingly, we see a 15% decline in the use of insulin, or 100% of baseline. This is an important finding for the efficiency of diabetes management, since insulin is more costly than sugar-reducing medication and is used at later stages of disease progression.[37]

As noted earlier, CdA provides a full suite of interventions, including nutritional and exercise advice. Behaviors such as eating and exercise are notoriously difficult to influence (Jager, 2003). Yet we find suggestive evidence that, in the short term at least, CdA influenced these behaviors positively. Our follow up survey includes a dichotomous variable for whether individuals report dieting or engaging in exercise. We find that the odds of both exercise and diet rise by 14%, and that there is a marginally significant increase in the odds of either diet or exercise by 26%, from a mean of 73%. We also find insignificant declines in the number of sodas and cigarettes consumed per week.

In the Appendix we explore a wide variety of dimensions along which we might find heterogeneous impacts. While our estimates are somewhat imprecise, we find no significant heterogeneity across age, sex, BMI, schooling or income[38]. We do find that there is significant heterogeneity by baseline level of HbA1c. Sicker individuals at baseline improve more than their healthier counterparts.

To be clear, our result is for a particular private provider, and may not generalize to the entire private sector. It is possible that patients are selecting CdA because it is the best private provider, and as a result any positive outcomes reflect selection and don't represent the average treatment effect in the entire private sector. While we cannot completely rule out this selection explanation, two pieces of evidence point to the effects we find being general. The first is that our findings

---

[37]According to IMSS guidelines, a patient that is under control or close to it should take Metformin (HbA1c<8%), Glibenclamida if the HbA1c between 8-8.5% and insulin if HbA1c is above 8.5%. Based on IMSS costs, a month of Metformin or Glbenclamidia are similar at around 1USD per month, but a month of insulin treatment costs $25 USD.

[38]The lack of results on income heterogeneity (and the fact that the subsidy is a small percentage of their income) suggests that our treatment does not just operate through an income effect.

appear driven by improved access to care, not a superior technology for treating patients once in care. Second, a main mechanism for the improvement in outcomes is drug maintenance. Both of these points should be readily addressable for other firms in the private sector.

For a longer exploration of external validity and the comparison of our sample with other populations see the online Appendix.

5   Fiscal Externalities on the Public Sector

In many public systems such as Mexico's, privatized care is provided at lower levels of care, but the costs of higher levels of care are primarily borne by the public sector. In our sample, for example, 74% of individuals who utilized a hospital report a public option as their main health provider. As a result, improved primary care through private diabetes clinics may impart a significant fiscal externality on the public sector.

Indeed, the hospital costs of diabetes to the Mexican public insurance system are enormous. Total costs amount to 2.25% of GDP, with 1.1% being direct medical costs (the remainder being indirect costs such as lost wages and value of years of life lost), and 87% of the direct costs are due to complications. These costs primary arise from a series of complications associated with diabetes such as diabetic retinopathy, diabetic foot and diabetes kidney disease.[39] Bringing blood sugar levels under control significantly reduces the risk of such complications —leading to reduced hospital costs (Barraza-Lloréns et al., 2015). The fact that Mexico has twice the OECD average diabetic hospitalization rate highlights that there is much to gain from improving early-stage care.

We investigate this in two ways. First, we directly estimate the effect of CdA on self-reported diabetic complications. This is challenging since we only have a short follow up time for our CdA enrollees, so that we will meaningfully understate the long-term impacts on complications. Second, we use a simulation model based on our estimated reduction in blood sugar, combined with the best estimates of the marginal impacts of reduced blood sugar on future complications.

Column 1 of Table 5 shows our estimated treatment effects for complications, which are de-

---

[39]*Diabetic retinopathy*: is a diabetes complication that affects eyes. It's caused by damage to the blood vessels of the light-sensitive tissue at the back of the eye (retina). At first, diabetic retinopathy may cause no symptoms or only mild vision problems. Eventually, it can cause blindness. *Diabetic foot*: Diabetes can damage your nerves or blood vessels. Nerve damage from diabetes can cause you to lose feeling in your feet. You may not feel a cut, a blister or a sore. Foot injuries such as these can cause ulcers and infections. Serious cases may even lead to amputation. *Kidney disease*: Diabetes can damage the blood vessels in your kidneys. When the blood vessels are damaged, they do not work as well. When your kidneys are damaged, they can't filter blood like they should, which can cause wastes to build up in your body.

fined as the sum of the early symptoms a person might experience: worsening sight, hands and feet tingling. Despite the short follow-up, we find a significant reduction in complications, with enrollment associated with a decline of complications of -0.25 off a base of 1.4. To put this result in context, we compute the estimated decline in complications that we would expect over time based on our blood sugar reductions. According to the widely cited UKPDS study, a one point reduction in HbA1c could lower the likelihood of complications by 35% –significantly higher than the 18% we find experimentally, which is unsurprising given our short term follow-up.

We next turn to estimating the fiscal externality. we summarize our calculations here and provide details in the Appendix F. Because the effects of CdA, and therefore the downstream savings, differ by baseline level of HbA1c, we divide our sample by starting HbA1c. For each category, we calculate the conditional local average treatment effect (CLATE) in terms of HbA1c for each of these bins using Athey et al. (2019)'s methodology. To go from HbA1c effects to health complications and hospitalizations, we use the complications incidence tables by level of HbA1c from the UKPDS 35 study (King et al., 2001). Finally, to estimate the peso savings to the system from reduced hospitalizations we use the complication-specific hospital cost data from Barraza-Lloréns et al. (2015) for 2013 in Mexico, updated for inflation.

As illustrated in Appendix Figure OA-8, we find significant medical savings from CdA, particularly for those with the highest blood sugar levels. Savings typically rise with HbA1c levels both because (a) the rate of complications is higher for those with higher blood sugar and (b) our CLATE estimates show larger effects of our intervention for those with higher blood sugar at baseline. We find that at high levels of blood sugar, the downstream savings from CdA are larger than the $7000 peso subscription price for the program; overall, we estimate that reduced hospital expenditures amount to 55% of the cost of CdA.

Of course, these calculations are not precise. On the one hand, we assume that the marginal reductions in blood sugar from our intervention have the same impact as the average reduction in blood sugar used by UKPDS 35. On the other hand, we only consider the direct fiscal externality from hospitalizations averted and do not consider the savings from care substitution (with the CdA visits displacing public visits). To address the latter, we directly measure care substitution in the last two columns of Table 5. The second column measures visits to a public insurance provider. This falls by 0.11 visits, or about one fifth of the control group mean at baseline, but the estimate is insignificant. One problem with this measure, however, is that COVID-19 may have had the impact of reducing all medical care use, for both treatments and controls, mitigating any estimated displacement effect. The next column uses a measure which may be more reliable, whether the respondent considers the public health system to be their main health care provider. While this variable has essentially the same mean, our estimated effect is nearly twice as large, and amounts

to 40% of baseline mean.

Adding the savings from reduced IMSS visits to our offset of hospital spending, we find that medical savings offset between 65% and 105% of the costs of CdA (depending on which measure of reduced visits we use). In either case, the net cost of this incremental care through CdA is much less than the gross costs.

So far this calculus does not consider the value of improved care. In Table 3, we showed that patients enrolled in CdA are 22% more likely to be in control than those who do not enroll. The medical literature estimates that every year that a diabetic patient spends out of control reduces life expectancy by 100 days (Heald et al., 2020). Typical estimates of the value of a life-year in international contexts is in the $50,000-$100,000 USD range, which would suggest that this intervention is worth $3000-$6000 USD per person, many multiples of gross or net program costs, which are around $350 USD a year.[40]

While estimating the value of life in this particular population is beyond our scope, it is worth noting that even this estimate excludes the valuation of the large reduction in morbidity —which does not only save the public sector costs but improves quality of life. This exercise does not take into consideration positive health externalities to family members which have been recently found in the literature (Fadlon and Nielsen, 2019; Chen et al., 2022).

6 Why Does the Private Sector Improve Outcomes?

The striking finding that enrolling in CdA dramatically improves health outcomes for diabetics that already have access to free public care raises the key question of *why* the private sector is doing a better job than the public sector addressing the medical needs of diabetics. One potential explanation is that our finding is driven by heterogeneity in the type of public insurance. While there is technically universal public coverage in Mexico, the care delivered by IMSS (the formal sector social insurance program) is typically perceived to be much higher quality than that delivered by the residual public welfare program, Seguro Popular. As a result, if our findings are driven by those individuals in our sample who are not formally employed and have to rely on Seguro Popular, this may reflect the lower quality of care in that public program. However, we find no evidence that the impact of CdA is driven by informal workers —if anything, the opposite appears to be true.[41]

---

[40]Lee et al. (2009) argues that the value per year of quality life is $129,00 but $50,000-$100,00 has been de facto international standard. For the calculation we multiply the effect on likelihood of control, times 100/365 days, times $50,000-$100,000.

[41]A regression of HbA1c on treatment interacted with having informal insurance shows a positive but insignificant interaction —suggesting that our results are not driven by larger impacts among those informally insured (see last

Even if the higher quality public programs were not performing as well as private care, the difference must be explained by either quality or quantity differences between the two platforms. On the one hand, it could be the case that CdA is providing a higher quality of service per interaction. On the other hand, perhaps CdA is doing more to attract diabetics to interact with care, improving outcomes through increased quantity of medical interactions.

To separate these hypotheses, we extend our analysis by incorporating data from the IMSS program, the largest formal sector health care system in Mexico, between 2010 and 2015. We worked with IMSS to collect administrative data for every primary care visit. This is a novel data set that has not been previously exploited for economics research. The data combines several large administrative datasets. The first is annual testing data; all IMSS enrollees are supposed to have a checkup that includes blood sugar once per year as part of the PrevenIMSS program and we have administrative data for all checkups recorded in that period (IMSS, 2013).[42] Second, we have administrative data from every primary care visit, which includes ICD-10 diagnosis codes. Third, we have exact location data for families' homes. Fourth, we have an infrastructure dataset that includes geocoded locations for the primary care clinics providing care through IMSS. In order to estimate the marginal return to care, we create a sample of all patients who have been treated for diabetes at IMSS and who had at least one PrevenIMSS checkup, leaving us with a sizeable sample of 440,000 diabetics. Moreover, 160 thousand have a second PrevenIMSS appointment one year later, which allows us to track their blood sugar dynamics.

We find a variety of evidence that there is a significant role for a higher quantity of care with the private sector option. Table 4 showed directly that the treatment group has more visits to doctors. Table OA-22 finds a much higher level of trust in CdA than their alternative care provider, suggesting a higher willingness to engage with the provider. Moreover, we find that randomized enrollment in CdA strengthens trust in the program.[43]

Use of private sector clinics may be more frequent due to greater convenience of this source of care. Our follow up survey asked treatments and controls about their waiting time for care. Column 3 of Table OA-22 shows our IV estimates for waiting time, and the impacts are striking:

---

column Table OA-19).

[42]See also https://tinyurl.com/yutc5xjv and https://tinyurl.com/54r7d9ew. Not all patients may go however. Even though this policy may be costly our results suggest that visits could be a good investment.

[43]In Table OA-22 we explore whether the CdA trust advantage grows as a result of treatment. The first column regresses the difference in the self-reported trust between CdA versus the alternative at follow up against our instrumented "Using CdA" indicator, controlling for the baseline *difference*. The second column carries out the same exercise for a different variable, which measures whether enrollees trust *the diagnoses* that comes from CdA as opposed to their current health provider. Both variables have a mean difference above 2 points on a ten point scale, and IV regressions of this gap in beliefs on being a CdA user shows that the gap increases significantly with use of CdA. These results highlight that patients are more engaged with the service once they use it —which may lead them to get care more often.

a reduction of 30 minutes in waiting time, or more than half of the baseline mean. Note also that this is an estimate per visit; the total time difference will be even larger in the public sector since while CdA offers a single visit one-stop-shop model of care, the public sector usually requires the patient to visit several times: to get tested, to speak with a nutritionist, to get medicines, etc. As we mentioned above in Mexico's public healthcare patients have to wait many weeks or months for each of appointment. This waiting is compounded by having multiple visits in public care compared to CdA.[44]

To further explore whether making care more accessible is a key mechanism driving our results, we focus on the subsample of our treatment and control groups that is enrolled in IMSS, and we consider the heterogeneity of our treatment effect by the level of IMSS clinic "saturation", a direct measure of how hard it is to access care at IMSS. If indeed more access to care is driving our findings, we would expect our effect to be bigger among the IMSS users who have to utilize a more saturated clinic.[45] We define our saturation measure as follows. According to IMSS the highest capacity per medical office is 4 patient visits per hour and that at 85% of this quantity the medical office is saturated but still functional. We thus classified a medical office as saturated if it is operating above that rate.[46] It is important to note that IMSS patients cannot choose which clinic to go to, these are assigned bureaucratically by IMSS based on the shortest distance from the home address, so patient demand plays no role in the definition of saturation.

Using data on visits from 2015 for 31 clinics in Monterrey and nearby regions, we match each patient to the closest IMSS clinic to their home and exploit administrative data to define clinic saturation. We begin by dividing the number of patients served at the clinic by the number of medical offices in that clinic. We take as a benchmark the 15 minutes per visit recommended by IMSS guidelines. We then label a clinic as "saturated" if it receives on average more than 85% of the maximum 4 visits per hour they can handle at maximum capacity per year. We then rerun our outcomes regressions, interacting our treatment dummy with a dummy for the closest clinic being saturated; we redo the exercise for non-IMSS users as a placebo, matching non-IMSS users to the nearest IMSS clinic.

---

[44]OECD (2020) and https://www.oecd.org/els/health-systems/waiting-times.htm shows not only that Mexico is one of the worst countries in terms of waiting times, but also that patients themselves often declare that long wait times are the main reason for having unmet care needs. See also (INSP, 2011).

[45]This comparison assumes that the more saturated clinics do not themselves deliver very different quality of care that less saturated clinics. If the care delivered at more saturated clinics is lower than at less saturated clinics, then part of the response we see here may be through differences in quality and not quantity. Of course, if saturation arises because there is more use of the highest quality clinics, the bias would go in the opposite direction to the effect we find.

[46]We obtain the same result in terms of sign using a linear definition for saturation instead of the threshold model, but is noisier and not statistically significant at conventional levels. A threshold model may be a better approximation to accelerating decreases in care quality the more saturated the clinic is.

In the first column of Table 6, we show that our treatment effect is larger for IMSS users when the clinics are more saturated, consistent with the notion that it is those who face the largest barriers to IMSS care who benefit most from CdA. Moreover, in the second column we show that such a relationship does not exist for the patients who are not enrolled at IMSS; in fact the point estimate goes in the opposite direction. We see this evidence as supporting the hypothesis that improvements due to CdA arise through more care.

Finally, we attempt to quantitatively disentangle these two channels with a quasi-experimental estimate of the marginal return to public care. We use our combined CdA and IMSS data to estimate the marginal returns to additional IMSS care versus care from CdA. To assess the marginal returns to IMSS, we use variation in the distance of individuals from their IMSS clinic. Individuals in IMSS are assigned to a local clinic based on fixed geographic designations, and as a result, the distance from homes to an IMSS clinics varies substantially. Figure OA-9 shows the distribution of distances from individual homes to IMSS clinics.

We restrict our analysis sample further to patients with two measures of blood sugar from PrevenIMSS, one year apart. We can use these measurements to assess whether more care during the intervening year induces improved outcomes —instrumenting the amount of care received with distance from an IMSS clinic.

In particular, we will estimate models of the following form:

$$Y_{i,j,t} = \beta_0 + \beta_1 N_{t,i} + \beta_2 Y_{t-1,i} + \Gamma_j + \psi_t + \chi X_i + \varepsilon_{t,i,j} \tag{2}$$

Where the dependent variable $Y_{i,j,t}$ is the level of blood sugar for individual i who got his checkups at clinic j at time t, $N_{t,i}$ is the number of visits to IMSS clinics in the 12 months after the first blood sugar measurements, Γ_j captures clinic fixed effects, ψ_t captures month fixed effects and χ_i are demographic control variables (gender, age and age-squared). We instrument the number of visits with the distance from residence to the assigned IMSS clinic.[47]

This quasi-experimental approach faces two key identification concerns. The first is that distance is correlated with underlying health. We address this by controlling for baseline blood sugar at time $t-1$, so that we are assessing the impact of visits on the improvement in blood sugar. Of course, this does not solve the underlying identification problem if those who live near IMSS clinics are on differential underlying health trajectories than are those who live far away. But the inclusion of clinic fixed effects control for any neighborhood factors that might drive such trends.[48]

---

[47]Several other papers use distance as an instrument in the health care domain, e.g. Gupta et al. (2021).

[48]One may be concerned with reverse causality, i.e. saturation being a function of patients liking high-quality care.

25

The second concern is that the measurement itself may be correlated with distance —e.g. those who live farther away may be differentially likely to get their blood sugar measured. This is a particular concern given that only 160,000 out of 440,000 patients have a second yearly checkup. We address this directly by assessing whether the odds of blood sugar measurement is itself correlated with distance in Table 7. Appendix F.4 contains a battery of extra tests. It regresses all the covariates we observe (sex, age, capillary glucose at baseline) with distance, and we find no worrisome correlation. To control for observables at a very local level we added extensive controls for locality-level characteristics from Mexico's Population Census[49]. Results were virtually unchanged. Finally, to control for unobserved time-invariant characteristics we included more than one thousand municipality dummies and the results are robust.

The results of our analysis are shown in Table 7. The first column shows the first stage estimate of the impact of distance on the number of IMSS visits. The coefficient is highly significant, indicating that each 30 kilometers of distance results in 0.1 fewer visits. The second column tests for selection in having a blood measurement. In fact, we see no evidence of a correlation between likelihood of second checkup and distance. While some people may get less diabetes care because it is too far, whenever they show up at the clinic, for any reason, they are asked to get a PREVENIMSS screening if they have not done so in the last year. Since PREVENIMSS is an independent module of primary care, there is no differential attrition.

We then turn to causal estimates of the impact of visits on blood sugar. Since PREVENIMSS captures capillary blood sugar measurements rather than HbA1c, we utilize that metric instead for our analysis. To compare to our earlier findings, our experimental results from Table 3 are equivalent to a reduction from 226 to 197 in capillary blood sugar.[50]

We begin by estimating equation 2 for our CdA intervention. That is, we regress capillary blood sugar levels —which we also measured in our baseline and follow-up surveys— on the number of CdA visits, controlling for baseline capillary blood sugar. We instrument number of visits with our treatment indicator, so that we are essentially measuring the total treatment effect as a function of number of visits. In this specification we are assuming linear impacts of each additional CdA visit. The third column of Table 6 shows that each CdA visit reduces capillary blood sugar by 8 points.

In this case, all else constant, we should expect *better* health performance among clients in saturated clinics, however we find the opposite. Second, recall that patients *cannot* choose which clinic to go to. The clinic is assigned by IMSS bureaucratically to the patient as a function of his/her address, assigning them to that with the shortest distance. So, patient demand plays no role.

[49]There are close to 200,000 localities in Mexico, so controls are fine-grained.

[50]We utilize a conversion from Nathan et al. (2008) that estimates that each point reduced in HbA1c is equivalent to a 28.7 reduction in capillary blood sugar. This conversion fares well when applied to our experimental sample since we find a reduction of 29 units in capillary blood sugar and a 1.1 point reduction in HbA1c.

The fourth column estimates equation 2, instrumenting by distance to an IMSS clinic, to estimate the return on each marginal visit at IMSS. We find that each additional visit provides a benefit of 5 points. While significant, this is less than two-thirds as large as the estimate for CdA (although the differences between CdA and IMSS are not statistically significant). This suggests that part of the reason for a larger effect for CdA is more effectiveness per visit (although the difference is small); moreover, as noted earlier, we potentially understate the impact of CdA because this treatment-control comparison excludes any impacts of the initial evaluation.

But this result does not account for potential selection on treatment effectiveness. In fact, those who sign up for CdA have considerably higher blood sugar than the typical person in IMSS, while the average baseline blood sugar in Cda is 225, at IMSS it is 135. And we showed earlier that the effect of CdA is larger for those with higher blood sugar —the same may be true for IMSS. To assess this, we re-estimate the regression for IMSS from column 4, but reweighting the sample by baseline blood sugar to make them more comparable. The final column of Table 7 shows that doing so dramatically increases the estimated IMSS treatment effect, which more than doubles. Indeed, this estimate is higher than the comparable CdA estimate, although not significantly so. Thus, these results do not indicate that CdA's impacts arise through a better "technology," at least in terms of the returns per visit.

Thus, a suite of evidence supports the notion that the higher quantity of care received at CdA, and not higher quality per unit, is driving the results. We see more use of care at CdA, driven at least partly by shorter wait times. We see that the effects of CdA are largest where access to IMSS is most restricted. And we find that the estimated return to a visit to IMSS is comparable to the estimated return to a CdA visit.[51]

Of course, this analysis is not as compelling as our purely experimental findings that CdA improves outcomes. Ideally, governments could conduct follow up analyses of private sector options that experimentally vary the basket of services to understand exactly which aspects of the alternative are delivering better outcomes. That said, this section demonstrates that a novel combination of experimental and observational data can offer strongly suggestive evidence that can guide further policy development in the public sector.

---

[51] Visits are important for at least three reasons. First medication optimization. Several types of medication could be used and different patients respond differently to each of them (Smith et al., 2010). We are told by CdA doctors that adjustment is common and that without the adjustment progress is much more limited. Second, visits are also used to adjust diets. Third, monitoring and motivation. Keeping track of HbA1c progress during frequent visit is critical to achieving good results (Martens et al., 2021). Studies show that left to their own devices, patients check their sugar levels much less often than optimal (Rossi et al., 2018) and as a result make little progress in decreasing HbA1c (Leelarathna et al., 2022). There are correlational studies showing that for diabetics more visits to the doctor is correlated with decreases in HbA1c (Al-Nozha, 2014; Morrison et al., 2011), but more causal evidence is needed.

7 Conclusion

The tight fiscal constraints on public health care systems around the world has led to a growth in private sector alternatives as an opt-out. These private sector alternatives are often viewed with suspicion by public health systems, when in fact they could be providing an additional benefit not only directly to patients, but in the form of fiscal externalities as well. They also provide a learning opportunity for the public sector to better understand what works in treating critical diseases. Accessing this learning opportunity has been difficult, as it is challenging to design studies that evaluate private options in an empirically compelling framework. This paper introduced such a framework, relying on a novel deniers randomization strategy to run a trial of the private provision of diabetes care to a publicly insured population in Mexico.

Our findings are striking: supplementary private care causes a highly significant and large reduction in the blood sugar levels of diabetics, increasing blood sugar control by more than two-thirds. We estimate that this occurred through improved use of medication, more frequent medical treatment, and more diet and exercise. This sizeable reduction in blood sugar was associated even within the first year with reduced diabetes complications. These large health effects suggest that this supplemental private service was highly cost-effective. We estimate that two-thirds or more of the cost of the private program are offset by reduced public primary and (especially) hospital care, and that the estimated health benefits are many multiples of either gross or net costs.

Interestingly, our results also suggest that the strong performance of this private sector alternative was not because of dramatic improvement in care modality, but rather through stronger attachment of patients to the private alternative due to shorter wait times and other advantages over public clinics. We see more use of care for the treatment group and we see that the effects of CdA are largest where access to IMSS is most restricted. We also find that the estimated return to a visit to IMSS is comparable to the estimated return to a CdA visit. This is consistent with diabetes being a condition that is well understood and where best practice guidelines and standard cheap medication/technology exists.

This suggests that much of the gains from privatization in this context could actually be captured by the public sector itself by improving access to care for its enrollees. For example, dedicating some space in public clinics to provide more timely and easily accessed care for diabetics could end up being both health-improving and cost-saving for the Mexican public health system. Indeed, we see this paper as an example of how the private sector can provide a learning model for public sector systems as they try to improve their care delivery.

In practice, however, there are two reasons why we think that public sector expansion is unlikely

to be as successful as CdA has been. First, recent developments in public health care in Mexico have been moving in the wrong direction. Health spending under IMSS has been cut and Seguro Popular has now recently been eliminated, leaving about 30 million people without health care Argen (2020).

Second, Appendix Table OA-1 we show that the *cost* IMSS pays for its inputs is larger than the *price* CdA charges for its service, and the latter includes CdA markup, so it is likely that IMSS is more inefficient at providing this service. Consistent with Hart et al. (1997) and given that progress toward diabetes control is verifiable through HbA1c tests, this is one case where the private sector may do better than the public sector. In fact CdA and pharmacy doctors for other diseases are expanding rapidly to satisfy demand Rubli (2023). IMSS may actually save money by subcontracting: both, directly, since the private sector seems cheaper according to our information, but also through the fiscal externality we discussed in the paper.

At the same time, a concern with the private approach, particularly given its high costs, is the furthering of inequality in health outcomes in Mexico. However we must not lose sight that the system is de facto already highly unequal, with many low-income and middle-income families having limited access; private suppliers help ease some of these constraints.[52] CdA does not exclusively serve the high-income population; about 36% of its patients are low-income (Figure OA-6).

Our approach does have some limitations. We examine a particular program, CdA, and outcomes may not extend to other alternatives that are designed differently. Moreover, our conclusion that the gains from privatization are simply in increased care, and not in substantial changes to the technology of care, may reflect the particular case of diabetes, where there is a standard and cost-effective course of treatment. For other diseases with less standardized and/or more expensive treatment modalities, private delivery may or may not offer gains in the quality as well as the quantity of care. Additionally, we are not considering the benefits of the information provided by CdA: among respondents to our baseline survey, 24% had diabetes and did not know, the equivalent number for those with IMSS is 22%. Finally, we cannot estimate general equilibrium effects. We conjecture that a first-order effect of an expansion of CdA or a similar provider would likely be that those diabetics not served in the status quo would receive medical care. A second-order effect would involve freeing-up capacity in public care, improving wait times. Because HbA1c is easily verifiable the market will likely experience an increase in efficiency as well, if some patients switch from IMSS to CdA.

Despite these caveats, however, our study provides a framework for estimating the effects of

---

[52]According to the New York Times (https://tinyurl.com/56y7j5mx, our translation from Spanish): "The Mexican middle classes gave up trying to receive care in public hospitals and the poorest use them think that they are only going to die there".

private alternatives to public health care. We focus here on diabetes, one of the world's deadliest chronic conditions, in Mexico, one of the countries where this problem is the largest. In contexts where the public sector is underfunded and faces strong capacity constraints a private health care provider could ease constraints and increase health without inefficiently incurring in excess fixed costs[53] The estimates suggest that CdA's intervention is worth $3000-$6000 USD per person, and each branch serves hundreds of them. The fixed cost (of about $20,000-$40,000 USD per branch) seems small compared to the potential benefits.

A natural question is the extent to which our findings can generalize to other diseases. Besides diabetes, there are several other critical medical conditions for which standardized medication and guidelines exist and where patient compliance with treatment is a major barrier for health improvement. These include hypertension, Asthma, AIDS, Coronary Artery Disease, and Chronic Kidney Disease. It seems likely that this type of model could apply in many of those cases as well.

---

[53]In our context it takes months and often even a year to get appointments at IMSS (https://tinyurl.com/3sa68r3w, https://tinyurl.com/djxrbhmv and https://sanluispotosi.quadratin.com.mx/principal/tarda-el-imss-en-slp-hasta-un-ano-para-agendar-citas/).

Acknowledgements

Funding and Conflict of Interest

Tables

Table 1: Attrition

	Answered follow up		
	(1)	(2)	(3)
Treatment	0.03	0.03	0.01
	(0.02)	(0.02)	(0.02)
Branch FE	No	Yes	Yes
Enrollment Month Fe	No	Yes	Yes
Basic Controls	No	No	Yes
Observations	2,410	2,410	2,042
R-squared	0.001	0.01	0.11
Mean dep. var.	0.43	0.43	0.43

Notes: This table presents the results from running a regression where the dependent variable is a dummy of answering follow-up and the independent variable is the treatment group dummy. The first column reports the regression without any controls, the second column controls for branch and enrollment month fixed effects, and the third column also includes basic controls: age, gender, HbA1c, BMI, schooling and income. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Balance Table for Those Measured at Follow Up

Variable	Control N	Control Mean/SE	Treatment N	Treatment Mean/SE	Difference C-T
Panel A: Demographics					
Age	509	52.39	558	53.06	-0.67
		(0.50)		(0.52)	(0.73)
Male	509	0.32	558	0.35	-0.03
		(0.02)		(0.02)	(0.03)
% High income	509	0.56	558	0.56	-0.01
		(0.02)		(0.02)	(0.03)
% High School or more	457	0.37	485	0.39	-0.02
		(0.02)		(0.02)	(0.03)
Panel B: Health and Health services					
HbA1c	508	9.35	556	9.51	-0.15
		(0.11)		(0.11)	(0.16)
BMI	509	31.07	557	30.83	0.24
		(0.27)		(0.25)	(0.37)
Has IMSS, ISSSTE or Seguro Popular	509	0.73	558	0.74	-0.01
		(0.02)		(0.02)	(0.03)
Percentage that use IMSS, ISSSTE or Seguro Popular	509	0.74	558	0.74	0.00
		(0.02)		(0.02)	(0.03)
HbA1c out of control (HbA1c>7)	508	0.76	556	0.78	-0.02
		(0.02)		(0.02)	(0.03)
Panel C: Beliefs					
Trust to improve following CdA recomendations (0-10)	432	9.43	467	9.54	-0.12
		(0.07)		(0.06)	(0.09)
Trust to improve following current health provider recomendations (0-10)	414	7.89	460	7.93	-0.03
		(0.13)		(0.13)	(0.18)
Trust in CdA diagnosis (0-10)	444	9.33	469	9.37	-0.04
		(0.07)		(0.07)	(0.09)
Trust in current health provider diagnosis (0-10)	411	7.93	450	7.88	0.05
		(0.13)		(0.13)	(0.18)

Notes: This table presents balance for patients who answered our follow-up. *(1) % High income*: Is an indicator variable equal to 1 if the person lived in the zip codes that the AMAI classifies as above middle income. This variable is an administrative variable and is based on the place of residency of each potential client at CdA. *(2) Has IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to be affiliated to a public service. *(3) Percentage that use IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to go for medical attention to the public services in our baseline survey. *(4) Trust to improve following CdA recommendations (0-10)*: is a variable that measures the trust in improving with CdA treatment. *(5) Trust to improve following current health provider recommendations (0-10)*: is a variable that measures the trust in improving with their current health provider. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Effect on HbA1c

	HbA1c OLS	HbA1C IV	I(HbA1C<7) OLS	I(HbA1C<7) IV
	(1)	(2)	(3)	(4)
Use Cda	-0.98***	-1.12***	0.15***	0.22**
	(0.13)	(0.41)	(0.03)	(0.09)
HbA1c Bl	0.50***	0.50***		
	(0.03)	(0.03)		
I(HbA1c<7) Bl			0.51***	0.52***
			(0.04)	(0.04)
BMI	-0.02	-0.02	0.00	0.00
	(0.01)	(0.01)	(0.00)	(0.00)
Age	-0.00	-0.00	0.00	-0.00
	(0.01)	(0.01)	(0.00)	(0.00)
Gender	-0.17	-0.17	0.04	0.03
	(0.14)	(0.14)	(0.03)	(0.03)
Elementary School	0.28	0.30	0.06	0.05
	(0.42)	(0.43)	(0.09)	(0.09)
Secondary School	0.43	0.44	0.04	0.03
	(0.42)	(0.43)	(0.09)	(0.09)
High School	0.05	0.06	0.05	0.04
	(0.44)	(0.45)	(0.09)	(0.09)
Tecnica o Normal	0.22	0.23	0.10	0.09
	(0.44)	(0.44)	(0.09)	(0.09)
University	-0.08	-0.06	0.14	0.12
	(0.47)	(0.48)	(0.10)	(0.10)
Income C+	0.26	0.25	-0.04	-0.04
	(0.31)	(0.31)	(0.12)	(0.12)
Income C	0.24	0.23	-0.06	-0.05
	(0.29)	(0.29)	(0.11)	(0.12)
Income D+	0.57*	0.56*	-0.15	-0.14
	(0.31)	(0.31)	(0.12)	(0.12)
Income D or Lower	0.41	0.40	-0.09	-0.08
	(0.32)	(0.32)	(0.12)	(0.12)
Observations	939	939	939	939
R-squared	0.36	0.35	0.26	0.22
F		93.26		93.33
First coeff		0.301		0.300
Mean dep. var	8.538	8.538	0.322	0.322

Notes: This table shows the results of estimating equation 1: $Y_{ijt} = \beta_0 + \beta_1 \text{Use}_i + \Gamma_j + \psi_t + \chi X_i + \varepsilon_{ijt}$. The first column presents the OLS estimates where $HbA1c_i$ is the dependent variable. The second column instruments Use_i of CdA with the treatment dummy, and so measures a LATE. The third and fourth columns replicate the two previous ones using as a dependent variable an indicator of sugar being under control, that is $I(HbA1c_i < 7)$. In all regressions we control for branch and month fixed effects, as well as basic demographics controls: age, sex, BMI, baseline HbA1c or controlled diabetes at baseline, schooling, and income class. Schooling controls are self-reported from our baseline survey and income class controls come from CdA administrative data. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$.

Table 4: Mechanism

	Visits to Doctor	# Special Check ups	# Specialists	Stop med if feels good	Takes Pills	Takes Insulin	I(Exercise)	I(Diet)	Diet+Exercise	# Cigaretts	# Sodas
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Use Cda (instrumented)	2.59**	1.32***	0.64**	-0.22**	0.25***	-0.15**	0.14	0.14	0.26*	-0.72	-0.06
	(1.06)	(0.38)	(0.27)	(0.09)	(0.09)	(0.06)	(0.10)	(0.10)	(0.15)	(0.80)	(0.14)
Branch FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Months since enrollment FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls Basic	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	878	817	940	815	913	913	940	940	940	774	857
R-squared	0.04	0.03	0.10	0.05	0.09	0.14	0.05	0.06	0.07	0.45	0.15
F	98.82	83.46	96.13	91.35	99.83	99.84	96.10	97.31	97.12	78.83	104.7
First coeff	0.321	0.307	0.306	0.323	0.315	0.316	0.306	0.308	0.308	0.307	0.332
Mean dep. var	6.502	1.408	1.383	0.249	0.727	0.134	0.289	0.436	0.725	0.911	0.841

This table presents estimates for specification 1 on different self-reported behaviors. For all of these specifications we report the local average treatment effect (LATE) that we causally estimated from our experiment instrumenting Use of CdA with the treatment arm. We control for branch and month fixed effects as well as basic demographics (age, sex, BMI, baseline HbA1c, schooling and income). *(1) Visits to Doctor*: captures how many diabetes-related visits to the doctor the person did the last year. *(2) # Special Check Ups*: diabetes-related checkups the person got during the last year, i.e eyes, kidney, foot, and blood checkups. *(3) # Specialists*: captures the number of diabetes-related specialist doctors the patient met with in a regular visit to their health provider. *(4) Stop Medication if feels good*: is an indicator variable that takes the value of 1 if the patient declared that he suspended his medication whenever he starts felling well. *(5) Takes Pills and (6) Takes Insulin*: both are indicator variables that equal 1 if the person declared that he took pills and insulin, respectively, as part of his treatment, we also control for their baseline measure. *(7) I(Exercise) and (8) I(Diet)*: both are indicator variables that equal 1 if the person declared that he did exercise and/or was involved in a diet as an effort to take care of his health, respectively. (9) Diet+Exercise: is the sum of I(Diet) and I(Exercise). (10) # Cigarrets and (11) #Sodas: capture the number of cigarettes and sodas consumed in a regular day by the person, we also control for their baseline measure. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Averted complications and Health Provider Substitution

	Total Diabetes Complications	Use of Public Service	Public Service as Principal Provider
	(1)	(2)	(3)
Use Cda	-0.25*	-0.11	-0.25**
	(0.13)	(0.10)	(0.10)
Branch FE	Yes	Yes	Yes
Enrollment Month Fe	Yes	Yes	Yes
Basic Controls	Yes	Yes	Yes
Observations	939	940	909
R-squared	0.05	0.09	0.14
F	93.31	96.06	95.34
First coeff	0.301	0.306	0.310
Mean dep. var	1.383	0.611	0.640

Notes: This table presents the results from the effect of getting CdA treatment on complications, utilization of the public sector health provider, and on listing the public sector as the main provider. We control for branch and month fixed effects as well as basic demographics (age, sex, BMI, baseline HbA1c, schooling and income). *(1) Total diabetes complications*: is a variable that sums the short run complications related to diabetes experienced by the person, i.e eyes, feet, and hand tingling. We also control by the total complications at baseline. *(2) Use of Public Service*: is an indicator variable equal to one if the person declared to have been to IMSS, Seguro Popular or ISSSTE in the previous year for any medical reason, we also control for the baseline value of this variable. *(3) Public Service as Main Provider*: Is an indicator variable equal to one if the person declared a public service (IMSS, Seguro Popular, ISSSTE) as their principal health provider, we control for their health affiliation at baseline. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Effect on HbA1c by saturation for IMSS enrollees

	HbA1c IMSS Members	HbA1c non-IMSS-members
	(1)	(2)
Treatment	0.22	-0.17
	(0.37)	(0.54)
High saturated clinic	0.88***	-0.15
	(0.31)	(0.47)
Treat x High sat. Clinic	-0.85**	0.05
	(0.41)	(0.60)
Branch FE	Yes	Yes
IMSS only	Yes	No
Observations	515	292
R-squared	0.28	0.30
Mean dep. var.	8.60	8.47

Notes: This table presents the heterogeneity results of the effect of CdA on HbA1c by saturation of IMSS clinics in Nuevo León. We define a clinic as saturated if they have at least an 85% flow of maximum capacity on average. That is, if they have at least 3.4 patients per hour per office open on average over a full year. The first column reports the heterogeneity estimates for IMSS population while the second column reports the same estimates for patients that do not report getting access to IMSS, which serves as a placebo. We control by HbA1c and branch fixed effects. We focus only in Nuevo León because Coahuila has very few IMSS clinics in Torreón and Saltillo, so we could not exploit the clinics heterogeneity there. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Comparison IMSS vs Cda Effect

	IMSS		CdA	IMSS	
	Number of medical visits	I(12 months follow up)	Capillary Glucose	Capillary Glucose	Capillary Glucose
	(1)	(2)	(3)	(4)	(5)
Distance (km)	-0.0030***	0.0001			
	(0.0005)	(0.0000)			
Number of medical visits			-8.39**	-5.07*	-12.69**
			(4.07)	(2.69)	(6.43)
Observations	160,035	439,287	1,067	160,035	137,308
R-squared	0.11	0.05	0.17	-0.06	-0.28
F			60.31	35.29	12.96
First coeff			1.284	-0.003	-0.002
Instrument			Discount	Distance	Distance (W)
Mean dep. var.	5.25	0.36	214.90	129.16	129.16

Notes: All columns except (3) estimated in IMSS data. The first column shows regression of number of visits at IMSS on distance from the clinic. The second column shows regression of dummy for having a follow-up blood sugar measurement on distance. The third column shows IV regression in CdA data where we regress capillary glucose on number of visits, instrumented by treatment indicator. The fourth column shows an IV regression of capillary glucose on the number of IMSS visits, instrumented by distance. The fifth column repeats this exercise but reweighting the sample so that the baseline distribution for capillary blood sugar matches that of CdA. All specifications have branch and month fixed effects and basic controls (age, age squared, initial capillary glucose, and gender). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figures

Figure 1: Recruitment Process
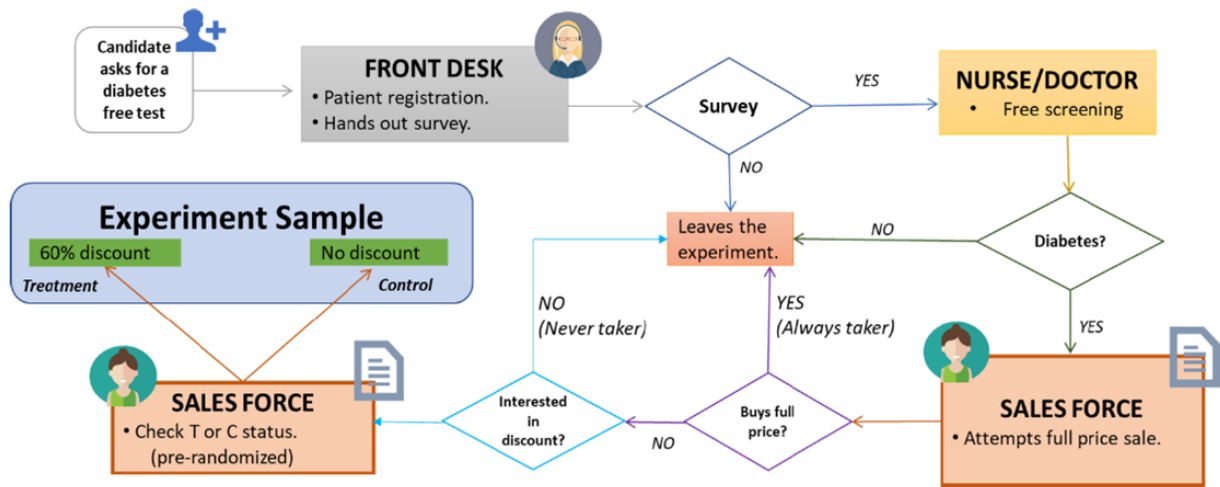
Notes: This figure represents the process through which a patient was included in our experiment. The patient would first go through the regular free-screening that CdA ususally offers and continue as a potential candidate unless she refused to fill out a survey. Then, if the patient was diagnosed as diabetic, the salesforce would try to sell a membership at full-price to that person. If the patient bought at full price, the person would leave the experiment, since we would know that person is an always-taker. If the person was not interested in buying a full-price membership, then the salesforce would offer the chance to win a 60% discount from our study. This is the first point in which we would modify the regular flow of patients within CdA. If the person said they were not interested at that price either, then we would know that such a patient was a never-taker. However, if the person said she was interested, then a button on the computer would reveal the treatment status to the salesforce and they would be able to offer the 60% discount if the patient was in the treatment group.

Figure 2: Recruitment Summary and Timeline

(a) Recruitment Summary
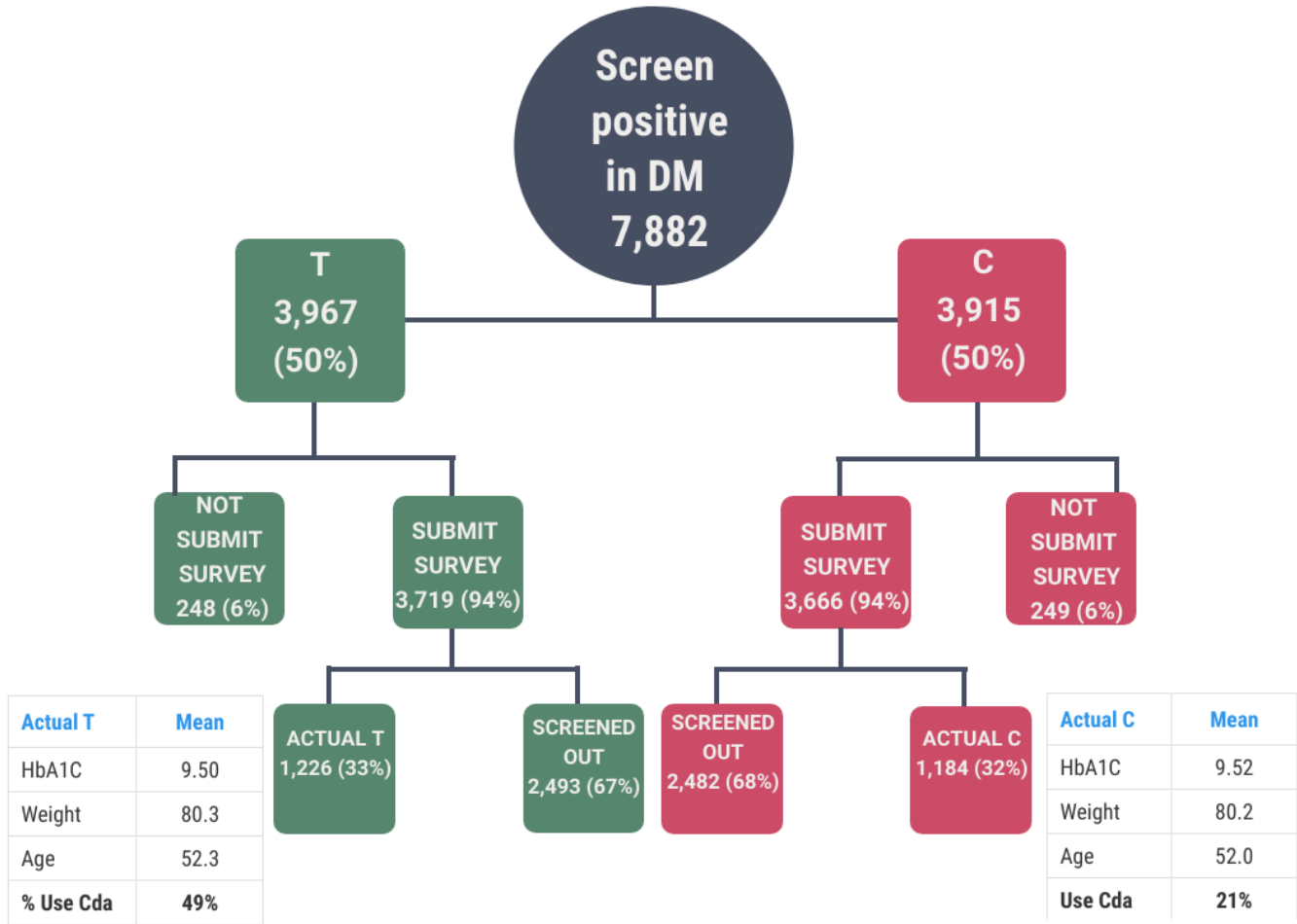
(b) Experiment Timeline
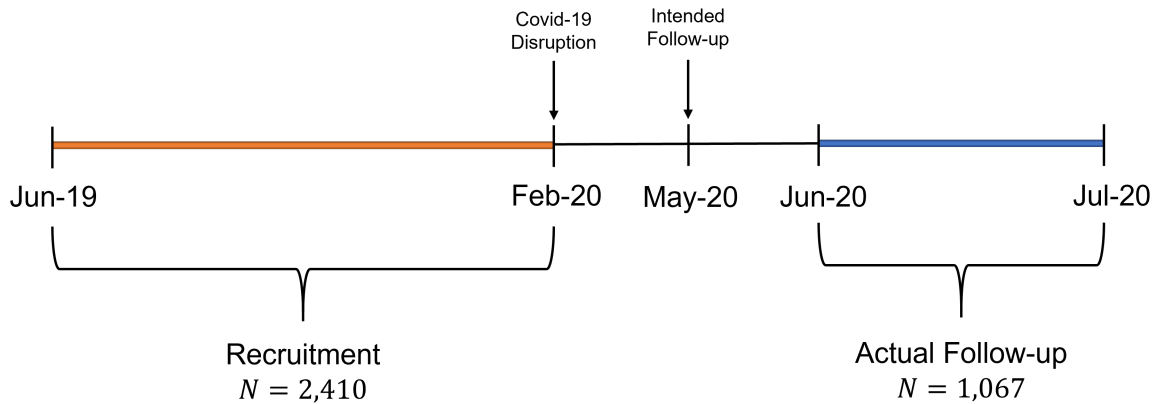
Notes: This figure summarizes our 8 month recruitment results as well as the timeline. We can see that there were nearly 8,000 diabetes patients that inquired about CdA, that our randomization was done evenly among treatment and control groups, and that 94% of the patients were willing to answer our baseline survey. Moreover, we can see that through our deniers randomization design we were able to screen out 67% of the sample as always-takers or never-takers, which significantly increased our power for this experiment.

References

AGUILAR, A., E. GUTIERREZ, AND E. SEIRA (2021): "The Effectiveness of Sin Food Taxes: Evidence from Mexico," *Forthcoming Journal of Health Economics.* [7]

AJALA, O., P. ENGLISH, AND J. PINKNEY (2013): "Systematic review and meta-analysis of different dietary approaches to the management of type 2 diabetes," *The American Journal of Clinical Nutrition*, 97. [5, 17]

AL-NOZHA, O. (2014): "Diabetes care and control: the effect of frequent visits to diabetes care center," *Ann Saudi Med*, 34. [27]

ANDERSON, M. L. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103, 1481–1495. [17, OA - 32, OA - 33, OA - 34]

ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN. (1996): "Identification of causal effects using instrumental variables." *Journal of the American statistical Association*, 434. [13, OA - 6]

ARGEN, D. (2020): "Farewell Seguro Popular," *The Lancet*, 395. [29]

ARNOLD, L. W. AND Z. WANG (2014): "The HbA1c and all-cause mortality relationship in patients with type 2 diabetes is J-shaped: a meta-analysis of observational studies," *Society for Biomedical Diabetes Research*, 11. [17]

ATHEY, S., J. TIBSHIRANI, AND S. WAGER (2019): "Generalized random forests," *The Annals of Statistics*, 47, 1148 – 1178. [21, OA - 31, OA - 37]

BANERJEE, A., E. DUFLO, N. GOLDBERG, D. KARLAN, R. OSEI, W. PARIENTÉ, J. SHAPIRO, B. THUYSBAERT, AND C. UDRY (2015): "A multifaceted program causes lasting progress for the very poor: Evidence from six countries," *Science*, 348, 1260799. [17, OA - 33]

BARRAZA-LLORÉNS, M., V. J. G. BARRÓN, C. H. VIVEROS, F. J. P. GUZMÁN, AND E. CRABLE (2015): *Carga Económica de la Diabetes Mellitus en México, 2013*, Fundación Mexicana para la Salud, A.C. [2, 20, 21, OA - 37, OA - 38]

BASU, S., J. ANDREWS, S. KISHORE, PANJABI, AND D. STUCKLER (2012): "Comparative Performance of Private and Public Healthcare Systems in Low- and Middle-Income Countries: A Systematic Review," *PLOS MEDICINE*, 6. [4]

BEHAGHEL, L., B. CRÉPON, M. GURGAND, AND T. LE BARBANCHON (2015): "Please Call Again: Correcting Non-response Bias in Treatment Effect Models," *The Review of Economics and Statistics*, 97, 1070–1080. [15, OA - 11, OA - 20, OA - 26]

CAMPBELL, D. AND P. DUNCAN (2022): "Record 6.8m people waiting for hospital treatment in England." *The Guardian.* [1]

CARD, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127–1160. [OA - 40]

CHAN, D., D. CARD, AND L. TAYLOR (2020): "Is There a VA Advantage? Evidence from Dually Eligible Veterans," *Stanford University mimeo.* [5]

CHANDRA, A., J. GRUBER, AND R. MCKNIGHT (2010): "Patient cost-sharing and hospitalization offsets in the elderly," *American Economic Review*, 100. [5]

CHEN, Y., P. PERSSON, AND M. POLYAKOVA (2022): "The Roots of Health Inequality and the Value of Intrafamily Expertise," *American Economic Journal: Applied Economics*, 14, 185–223. [22]

CLINIC, M. (2020): "Type 2 diabetes," Last accessed 9 March 2021, https://www.mayoclinic.org/diseases-conditions/type-2-diabetes/diagnosis-treatment/drc-20351199#:~:text=Metformin%20(Fortamet%2C%20Glumetza%2C%20others,body%20uses%20insulin%20more%20effectively. [6]

COLCHERO, M. A., J. A. RIVERA, B. M. POPKIN, AND S. WEN (2017): "Sustained consumer response: evidence from two-years after implementing the sugar sweetened beverage tax in Mexico," *Health Affairs*, 36. [7]

CONTI, G. AND R. GINJA (2023): "Who Benefits from Free Health Insurance?" *Journal of Human Resources*, 58, 146–182. [8]

DAS, J., A. HOLLA, V. DAS, M. MOHANAN, D. TABAK, AND B. CHAN (2012): "In Urban And Rural India, A Standardized Patient Study Showed Low Levels Of Provider Training And Huge Quality Gaps," *Health Affairs*, 31. [4]

DAS, J., A. HOLLA, A. MOHPAL, AND K. MURALIDHARAN (2016): "Quality and Accountability in Health Care Delivery: Audit-Study Evidence from Primary Care in India," *American Economic Review*, 106. [4]

DOUBOVA, S., D. PINTO-MASIS, R. PÉREZ-CUEVAS, V. BORJA-ABURTO, AND A. HUBBARD (2019): "Application of machine learning methodology to assess the performance of DIABETIMSS program for patients with type 2 diabetes in family medicine clinics in Mexico," *BMC Medical Informatics and Decision Making*, 19. [8]

FADLON, I. AND T. H. NIELSEN (2019): "Family Health Behaviors," *American Economic Review*, 109, 3162–91. [22]

FRAKES, M., J. GRUBER, AND T. JUSTICZ (2021): "Public and Private Options in Practice: The Military Health System," *National Bureau of Economic Research*, No. w28256. [5]

GUPTA, A., S. T. HOWELL, C. YANNELIS, AND A. GUPTA (2021): "Does Private Equity Investment in Healthcare Benefit Patients? Evidence from Nursing Homes," Working Paper 28474, National Bureau of Economic Research. [25]

HART, O., A. SHLEIFER, AND R. W. VISHNY (1997): "The Proper Scope of Government: Theory and an Application to Prisons*," *The Quarterly Journal of Economics*, 112, 1127–1161. [9, 29]

HEALD, A. H., M. STEDMAN, M. DAVIES, M. LIVINGSTON, R. ALSHAMES, M. LUNT, G. RAYMAN, AND R. GADSBY (2020): "Estimating life years lost to diabetes: outcomes from analysis of National Diabetes Audit and Office of National Statistics data," *Cardiovascular Endocrinology Metabolism*, 9. [17, 22]

HOE, T. P. (2022): "Does Hospital Crowding Matter? Evidence from Trauma and Orthopedics in England," *American Economic Journal: Economic Policy*, 14, 231–62. [4]

IDF (2019): *IDF Diabetes Atlas 9th Edition*, International Diabetes Federation. [6]

IMSS (2013): "Modelo de prevención en diabetes mellitus (PREVENIMSS)," 5ta Conferencia Científica Anual sobre Síndrome Metabólico, https://www.slideshare.net/sindromemetabolico/modelo-de-prevencin-en-diabetes-mellitus-prevenimss. [23]

——— (2018): "Informe al Ejecutivo Federal y al Congreso de la Unión 2017-2018," . [7]

INEGI (2017): "Encuesta Nacional de Empleo y Seguridad Social: Principales Resultados," Last accessed 10 March 2021, https://www.inegi.org.mx/contenidos/programas/eness/2017/doc/presentacion_eness_2017.pdf. [7]

——— (2018): "Encuesta Nacional de Ingresos y Gastos de los Hogares," . [2]

——— (2019a): "Cuenta satélite del Sector Salud de México, 2019," Last accessed 10 March 2021, https://www.inegi.org.mx/temas/saludsat/. [7]

——— (2019b): "¿De qué mueren los mexicanos?" Last accessed 9 March 2021, http://cuentame.inegi.org.mx/poblacion/defunciones.aspx?tema=P. [2]

INSP (2011): "Tiempos de Espera y Productividad en Rastreadores Clínicos Seleccionados en Hospitales del Sector Público," . [1, 24]

——— (2012): "Encuesta Nacional de Salud y Nutrición 2016," . [7]

——— (2018): "Encuesta Nacional de Salud y Nutrición 2018," . [2, 7]

IQEHC (2007): "What are disease management programs (DMP's)?" Last accessed 10 March 2021, https://www.ncbi.nlm.nih.gov/books/NBK279412/#:~:text=Disease%20management%20programs%20(DMPs)%20are,called%20%E2%80%9Cstructured%20treatment%20programs.%E2%80%9D. [10]

ISMAIL, K., K. WINKLEY, AND S. RABE-HESKETH (2004): "Systematic review and meta-analysis of randomised controlled trials of psychological interventions to improve glycaemic control in patients with type 2 diabetes," *The Lancet*, 363. [5, 17]

JAGER, W. (2003): *Breaking 'bad habits': a dynamical perspective on habit formation and change*, University of Groningen. [19]

KING, P., I. PEACOCK, AND R. DONNELLY (2001): "The UK prospective diabetes study (UKPDS): clinical and therapeutic implications for type 2 diabetes," *British Pharmacological Society Journals*. [5, 17, 21, OA - 37, OA - 38]

KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ (2007): "Experimental analysis of neighborhood effects," *Econometrica*, 75, 83–119. [17, OA - 33]

KNAUL, F. M., E. GONZÁLEZ-PIER, O. GÓMEZ-DANTÉS, D. GARCÍA-JUNCO, H. ARREOLA-ORNELAS, M. BARRAZA-LLORÉNS, R. SANDOVAL, F. CABALLERO, M. HERNÁNDEZ-AVILA, M. JUAN, D. KERSHENO-BICH, G. NIGENDA, E. RUELAS, J. SEPÚLVEDA, R. TAPIA, G. SOBERÓN, S. CHERTORIVSKI, AND J. FRENK (2012): "The quest for universal health coverage: achieving social protection for all in Mexico," *The Lancet*, 380, 1259–1279. [8]

KNUTSSON, D. AND B. TYREFORS (2022): "The Quality and Efficiency of Public and Private Firms: Evidence from Ambulance Services*," *The Quarterly Journal of Economics*, 137, 2213–2262. [4, 9]

KRAPEK, K., K. KING, S. S. WARREN, K. G. GEORGE, D. A. CAPUTO, K. MIHELICH, AND E. M. HOLST (2004): "Medication adherence and associated hemoglobin A1c in type 2 diabetes," *Annals of Pharmacotherapy*, 38. [17]

KRASS, I., P. SCHIEBACK, AND T. DHIPPAYOM (2015): "Adherence to diabetes medication: a systematic review," *Diabetic Medicine*, 32. [17]

LAW (1983): "Ley General de Salud," http://dof.gob.mx/nota_detalle.php?codigo=4652777&fecha=07/02/1984. [8]

——— (1995): "Ley del Seguro Social," http://www.imss.gob.mx/sites/all/statics/pdf/leyes/LSS.pdf. [6]

LEE, C., G. CHERTOW, AND S. ZENIOS (2009): "n Empiric Estimate of the Value of Life: Updating the Renal Dialysis Cost-Effectiveness Standard," *American Economic Review*, 12. [22]

LEE, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 76, 1071–1102. [OA - 11]

LEELARATHNA, L., M. L. EVANS, S. NEUPANE, G. RAYMAN, S. LUMLEY, I. CRANSTON, P. NARENDRAN, K. BARNARD-KELLY, C. J. SUTTON, R. A. ELLIOTT, V. P. TAXIARCHI, G. GKOUNTOURAS, M. BURNS, W. MU-BITA, N. KANUMILLI, M. CAMM, H. THABIT, AND E. G. WILMOT (2022): "Intermittently Scanned Continuous Glucose Monitoring for Type 1 Diabetes," *New England Journal of Medicine*, 387, 1477–1487. [27]

LIZZERI, A. AND N. PERSICO (2001): "The Provision of Public Goods under Alternative Electoral Incentives," *The American Economic Review*, 91, 225–239. [8]

LOPEZ, G. (2015): "The VA scandal of 2014, explained." *Vox*. [1]

MARTENS, T., R. BECK, R. BAILEY, K. RUEDY, P. CALHOUN, A. PETERS, R. POP-BUSUI, A. PHILIS-TSIMIKAS, S. BAO, G. UMPIERREZ, G. DAVIS, D. KRUGER, A. BHARGAVA, L. YOUNG, J. MCGILL, G. ALEPPO, Q. NGUYEN, I. OROZCO, W. BIGGS, K. LUCAS, W. POLONSKY, J. BUSE, D. PRICE, AND BERGENSTAL (2021): "Effect of Continuous Glucose Monitoring on Glycemic Control in Patients With Type 2 Diabetes Treated With Basal Insulin: A Randomized Clinical Trial," *JAMA*, 325. [27]

MCINTYRE, D. AND C. K. CHOW (2020): "Waiting Time as an Indicator for Health Services Under Strain: A Narrative Review," *Inquiry: The Journal of Health Care Organization, Provision, and Financing*, 57. [1, 9]

MOIR, M. AND B. BARUA (2021): "Waiting Your Turn Wait Times for Health Care in Canada, 2021 Report," . [1]

MOLINA-MILLAN, T. AND K. MACOURS (2021): "Attrition in Randomized Control Trials: Using tracking information to correct bias," *Paris School of Economics working paper*, 395. [15, OA - 10]

MORRISON, F., M. SHUBINA, AND A. TURCHIN (2011): "Encounter Frequency and Serum Glucose Level, Blood Pressure, and Cholesterol Level Control in Patients With Diabetes Mellitus," *Archives of Internal Medicine*, 171. [27]

NATHAN, D., J. KUENEN, R. BORG, H. ZHENG, D. SCHOENFELD, R. J. HEINE, AND A.-D. A. G. S. GROUP (2008): "Translating the A1C assay into estimated average glucose values," *American Diabetes Association*. [26]

NIDDK (2018): "The A1C Test & Diabetes," Last accessed 9 March 202, https://www.niddk.nih.gov/health-information/diagnostic-tests/a1c-test#:~:text=The%20A1C%20test%20measures%20the,level%20is%20below%205.7%20percent.https://medlineplus.gov/lab-tests/hemoglobin-a1c-hba1c-test/#:~:text=A%20hemoglobin%20A1c%20(HbA1c)%20test%20measures%20the%20amount%20of%20blood,the%20rest%20of%20your%20body. [6]

OECD (2019): "Diabetes Care," Last accessed 10 March 2021, https://www.oecd-ilibrary.org/sites/402e2e10-en/index.html?itemId=/content/component/402e2e10-en. [2, 7]

——— (2020): "Waiting times for health services 2019-2020," *OECD report*. [1, 9, 24]

OJEDA, R. (2019): "Diabetes en México: así se comporta la segunda causa de muerte en el país," Last accessed 10 March 2021, https://www.milenio.com/ciencia-y-salud/diabetes-en-mexico-2020-estadisticas-y-porcentaje. [7]

PIMOUGUET, C., M. L. GOFF, R. THIÉBAUT, J. F. DARTIGUES, AND C. HELMER (2011a): "Effectiveness of disease-management programs for improving diabetes care: a meta-analysis," *CMAJ-JAMC*, 183. [3]

——— (2011b): "Systematic review and meta-analysis of different dietary approaches to the management of type 2 diabetes," *CMAJ*, 183. [5, 17]

RIVERA-HERNANDEZ, M., M. RAHMAN, V. MOR, AND O. GALARRAGA (2016): "The Impact of Social Health Insurance on Diabetes and Hypertension Process Indicators among Older Adults in Mexico," *Health Services Research*, 51. [8]

ROSSI, M., G. LUCISANO, AND A. CERIELLO (2018): "Real-world use of self-monitoring of blood glucose in people with type 2 diabetes: an urgent need for improvement," *Acta Diabetol*, 55. [27]

RUBLI, A. (2023): "Trade-offs between access and quality in healthcare: Evidence from retail clinics in Mexico," *Journal of Public Economics*, 224, 104938. [8, 9, 29]

SAEEDI, P., I. PETERSOHN, P. SALPEA, B. MALANDA, AND S. KARURANGA (2019): "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition," *Diabetes Research and Clinical Practice*, 157. [2, 6]

SMITH, RJ AND, N. D., S. ARSLANIAN, L. GROOP, AND R. J. RIZZA, RA (2010): "Individualizing therapies in type 2 diabetes mellitus based on patient characteristics: what we know and what we need to know." *Clin Endocrinol Metab*, 95. [27]

SSP (2018): "Afiliación histórica del Sistema de Protección Social en Salud," Last accessed 10 March 2021, https://www.gob.mx/salud/seguropopular/documentos/afiliacion-historica-del-sistema-de-proteccion-social-en-salud#:~:text=A%20diciembre%20de%202018%2C%2053,la%20poblaci%C3%B3n%20total%20del%20pa%C3%ADs. [7]

——— (2021): "Catálogo de Clave Única de Establecimientos de Salud," Last accessed 10 March 2021, http://www.dgis.salud.gob.mx/contenidos/intercambio/clues_gobmx.html. [7]

UMPIERRE, D., P. A. B. RIBEIRO, C. K. KRAMER, C. B. LEITÃO, A. T. N. ZUCATTI, M. J. AZEVEDO, J. L. GROSS, J. P. RIBEIRO, AND B. D. SCHAAN (2011): "Physical activity advice only or structured exercise training and association with HbA1c levels in type 2 diabetes: a systematic review and meta-analysis," *JAMA*, 305. [5, 17]

WALTERS, C. R. (2018): "The Demand for Effective Charter Schools," *Journal of Political Economy*, 126, 2179–2223. [OA - 40]

WHO (2021): "Mean fasting blood glucose," Last accessed 9 March 2021, https://www.who.int/data/gho/indicator-metadata-registry/imr-details/2380#:~:text=The%20expected%20values%20for%20normal,and%20monitoring%20glycemia%20are%20recommended. [6]

WOOLDRIDGE, J. (2002): "Inverse probability weighted M-estimators for sample selection, attrition, and stratification," *Portuguese Economic Journal*, 117. [OA - 10]

ÁNGEL RIVERA DOMMARCO, J., A. COLCHERO, M. L. FUENTES, T. G. DE COSÍO MARTÍNEZ, C. A. A. SALINAS, G. H. LICONA, AND S. BARQUERA (2018): *La Obesidad en Mexico Estado de la política públicay recomendaciones-para su prevención y control*, INSP. [7]

Private Sector Provision as an "Escape Valve": The Mexico Diabetes Experiment

Ari Bronsoler, Jonathan Gruber, Enrique Seira

Appendix − For Online Publication

Contents

A Description of CdA

This appendix presents additional context on the service that CdA provides. CdA is based on a one-stop-shop model for diabetes care that provides every service that a patient needs to monitor their health in one visit. The patient usually comes to the clinic and checks in with a receptionist at the front desk. The patient is then directed to visit 4 different stations in which she is reviewed for eye complications, then takes a blood test, gets vitals measured and is reviewed for diabetic foot, then visits the dietician and finally meets with the doctor to overview their condition. Finally the patient checks out at the front desk where she can acquire any medicine that she needs as well as a diverse set of diabetes related merchandise such as sugar-free chocolate and diabetes specialized shoes. Figure OA-1 summarizes this process.

Figure OA-1: CdA Service
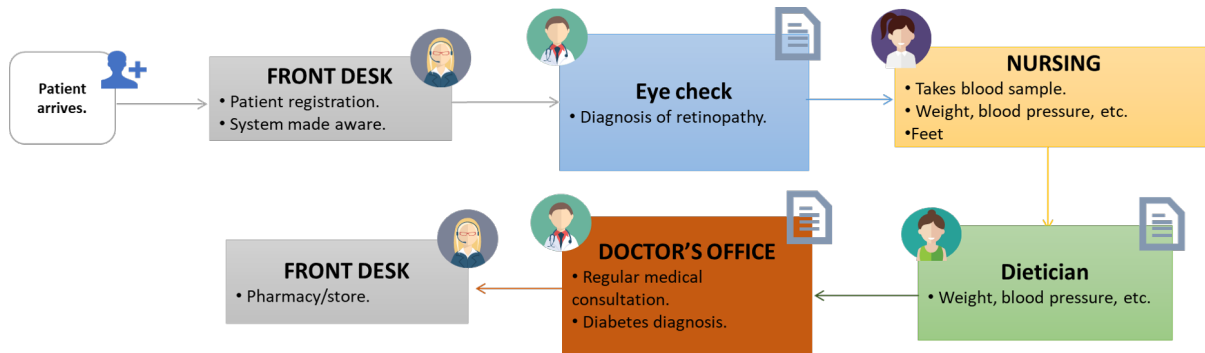
Notes: This figure summarizes the stations that a patient goes to while visiting CdA.

CdA has been expanding quickly over the last few years and aims to be accessible across Mexico. Their intention is to allow every patient to show up at the clinic that best suits them, whenever it best suits them. So for example, a patient could choose to get their first follow-up next to their workplace at clinic A during their lunch break and the next follow-up on a weekend at a clinic that is closer to their home. Figure OA-2 presents a map of CdA branches along with IMSS clinics as of December 2019 in the Monterrey area. Moreover, Table OA-1 compares the services provided at CdA to those provided by IMSS

Figure OA-2: CdA Branches in Monterrey

Notes: This figure summarizes branch locations for CdA. We include IMSS lcoations to highlight widespread access

Two of the authors and several the RAs actually walked incognito to CdA and experienced it. The first thing we noted was that you can just walk in, with no appointment, which makes Cda very convenient. This first visit is free. This is not possible in IMSS. Second, the service is all-encompassing and sequential, as in linear production lines as described above. The frequency of the meetings is a function of the state of the patients, but for patients who are not grave, appointments are scheduled every 3 months.

CdA's CEO explains their model as follows "Clínicas del Azúcar model eliminates many of the barriers to diabetes care by offering a patient-centric approach that emphasizes convenience. Nearly all of the clinics are located in retail locations with large anchors, such as supermarkets... this one-stop-shop model incorporates a high-quality, multidisciplinary health team under one roof." "Lozano says he wants to correct the common misperception that diabetics don't want to change or adhere to treatment plans. Often, it's simply difficult to get care or medication, or to access healthy food."[54] Finally, we would like to say that within the same one-stop-shop CdA has a store with low-sugar foods, shoes for diabetics, and a pharmacy that sells medication (medication is not included in the subscription price).

The CdA treatment effect we estimate is, therefore, a package of these services and we have no way to separately identify the contributions of the nurse versus the doctor versus the nutritionist, or having the pharmacy in the same place. Because the medications, the HbA1c diagnosis, and the recommendations are very standard[55], the medical technology used by CdA is not different. What is different is the higher convenience of using their service. This is particularly important in the Mexican context which, as we explained above, is underfunded and capacity constrained. IMSS requires multiple meetings one with a nutriologist, another for a blood sample, another for recommendations, and then going to a pharmacy to buy the medicines. Each of these appointments is hard to schedule.

---

[54]https://alum.mit.edu/slice/chain-mexican-clinics-aims-make-diabetes-care-easy.

[55]CdA follows The American Diabetes Association's Standards of Care in Diabetes for instance.

Below we present some pictures of how a clinic looks on the outside and inside below.

Figure OA-3: Extra CdA Branch pictures

Notes: This figure exemplifies how a typical CdA branch looks. The nurse, the pharmacy, the sales person and doctor mini offices, and the outside within a supermarket.

Table OA-1: IMSS vs CdA

Service	CdA	IMSS	IMSS costs
Doctor	5 visits if control 12 if not	5 visits if control 12 if not	3990
Nutrition	Personalized plan	General advice	NA
Psychology	2 visits	0 visits	2582
General blood	1/year	1/year	103
HbA1c	Quarterly	Quarterly	515
Lipids	2/year	2/year	206
Microalbumina	1/year	1/year	103
Foot	Every 3 months	Every 3 months	0
Eye	Once a year	Once a year	1416
Total Cost (pesos)	7500		8915

Notes: This table compares services provided by CdA and IMSS as well as the spending that IMSS would incur in case it would provide the same services as CdA does, according to their public per-unit cost report (Link). Numbers are based on a controlled patient that goes 5 times per year, which is the average usage we observe on IMSS administrative data. We can see that they generally offer similar care bundles, with CdA offering personalized nutrition plans and psychology care. We do not have data on desegregated costs for Cda. However, $7,500 pesos is the final price for the client, which already includes CdA's costs and profits.

B   Deniers Randomization

B.1   Power gains and cost savings from deniers randomization

The advantage of our deniers randomization approach over entry randomization is an enormous gain in efficiency in the experiment, both because our first stage is more powerful, and because we need not survey all always-takers and never-takers which do not contribute to our "compliers" estimates. Failing to remove the 67% of patients we identify as always or never-takers would have made our first stage substantially weaker and we would have required a much larger sample, which would in turn imply more spending on discounts and many more follow-up surveys. Since our follow-up was performed at patients' homes to prevent differential attrition, running extra surveys would be prohibitively expensive.

We illustrate the sample size savings with a simulation displayed in Figure OA-4. It plots the sample size needed to achieve 80% power and 95% confidence for different effect sizes. To fit our context, we assume that 38% of the individuals are always-takers, 53% are never-takers and only 9% are compliers. Moreover, we assume that a filter could exclude 82% of always-takers and 68% of never-takers, which is what ours does in the field. So for visitors randomization, out of every 100 individuals, there would be 38 always-takers, 53 never-takers, and only 9 compliers. If instead we use our deniers randomization approach, our sample consists of only 7 (38*0.18) always-takers, 17 (53*0.32) never-takers, and 9 compliers. Thus, our first stage coefficient would be 0.27, as opposed to the 0.09 first stage we would obtain from full visitors randomization.

Figure OA-4 shows that we need around 2500 total observations to detect an effect of 0.4 standard deviations, while visitors randomization would require around 25,000. The differences are even more striking

for smaller effects sizes more commonly found in the literature. This allowed us to operate at a budget that was one-tenth of what would have been required by visitors randomization —alternatively at the same budget, deniers randomization delivered three times the power of visitors randomization.

Figure OA-4: Power Calculations
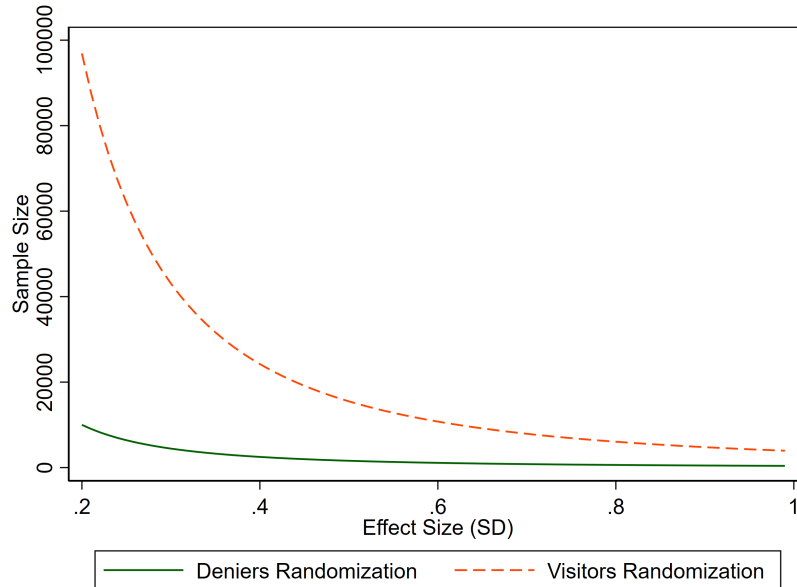
Notes: This figure highlights the sample needed to capture an effect with 80% power and an $\alpha = 0.05$ under visitors randomization and through a deniers randomization. To fit our context, we assume that 38% of the individuals are always-takers, 53% are never-takers and only 9% are compliers. Moreover, we assume that a filter could exclude 82% of always-takers and 68% of never-takers, which is what ours does on the field. So out of every 100 individuals, the researchers would get 38 always-takers, 53 never-takers and 9 compliers. After applying the deniers randomization procedure, only 7 (38*0.18) always-takers, 17 (53*.32) never-takers and 9 compliers would enter the experiment. Thus, the researchers would have to focus on 33/100 visitors only and have a first stage power of 27% within their experimental sample. A big improvement over applying visitors randomization and working with a 9% first stage effect.

B.2   Assumptions for visitor's LATE to equal deniers LATE

The benefits of deniers randomization are clear, but the LATE estimated from it may not be the same as the LATE estimated from visitor's randomization. In this appendix, we first informally describe the assumptions needed the equivalence between the two LATEs and then go through a more formal argument.

First some notation. Let Y_i capture the outcome for individual i, D_i represent the enrollment decision by individual i, and let there be a binary instrument Z where Z_i, in our case random allocation into treatment or control. Moreover, assume that there is a screening process S that marks $S_i = 1$ if the individual is screened in and 0 otherwise. The outcome of individual i is a function of decision to enroll, instrument and screening: $Y_i(Z, D, S)$. Similarly, the decision to enroll is a function of the instrument and the screening: $D_i(Z, S)$. For simplicity of exposition, we will denote $y_i(D_i)$ as the outcome when an individual i gets enrolled or not and $D_i(Z_i)$ as the enrollment decision as a function of the instrument.

Using the setup of Angrist et al. (1996) we know that the LATE result is the following:

$$\beta^{\text{LATE}} = \frac{\mathbb{E}[y_i|Z_i = 1] - \mathbb{E}[y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \mathbb{E}[y_i(1) - y_i(0)|D_i(1) - D_i(0) = 1]$$

Note that the LATE uses only compliers for identification. Excluding always-takers and never-takers from the sample would not change β^{LATE}.

We show that the LATE estimated from the deniers randomization approach is equivalent to the LATE estimated from visitor's randomization under three assumptions additional from the ones required for the LATE result:

1. *No exclusion of would-be compliers.* That is, we screen out only sets of always-takers and/or never-takers from the sample, and in doing this we do not screen out compliers.

2. *No direct causal effect of screening on enrollment.* The screening procedure does not itself directly affect enrollment decisions, say by changing the preferences that the visitor has for CdA.

3. *No direct causal effect of screening procedure on the outcome of interest.* For instance, the screening procedure itself does not have a direct effect on long term blood sugar levels.

The intuition is if the procedure does not change the set of compliers, and given that LATE uses only this sub-population, both estimators should give the same result. In general, whenever the screening procedure is quick and painless, as in our context, these assumptions are likely to hold. But if these assumptions do not hold then deniers randomization may estimate a different parameter than visitors randomization that may or may not be interesting.

We designed the procedure to make the three additional assumptions as likely to hold as possible. We ensured that all clients who rejected the offer at full prices were consulted for treatment status within 10 seconds (unless the client was not interested), and immediately offered the discount if they were in the treatment group. We verified that the sales force was not excluding would-be compliers by using mystery shoppers to test our process. This makes assumption one plausible.

Given that neither treatment nor controls experience denial of service, and that the screening is very standard and natural —i.e., a quote of the price— it is unlikely that there is a direct causal effect of screening on enrollment in CdA (Assumption 2). The visitor only needs to say she is not interested to buy in order to trigger the offer for the treatment group immediately. One way our screening procedure could affect the decision to enroll is if, by rejecting the offer at full price, the visitor herself affects her own preferences for CdA. This would certainly not be true in the neoclassical framework where demand has a negative slope and elicitation of demand does not change preferences.

The third assumption is that the screening procedure does not directly affect outcomes of interest, in particular HbA1c. This is satisfied since the deniers randomization certainly does not affect the quality of health services received at CdA; the sales force clicks on a screen to consult status, but they have no further interaction with the patients and nurses and doctors are blind to the treatment arm allocation.

Of course, both visitor randomization and deniers randomization occur once individuals express interest in CdA, and as such they do not necessarily capture the effect for the average diabetic in Mexico. We discuss issues of external validity below.

Result: Deniers LATE Equivalence

Let the following assumptions hold:

1. Stable Unit Treatment Value Assumption (SUTVA): for Z, Z', D, D', if $Z_i = Z_i'$, then $D_i(Z, S) = D_i(Z', S)$ and if $Z_i = Z_i'$, $D_i = D_i'$ then $Y_i(Z, D, S) = Y_i(Z', D', S)$ for all S.

2. Exclusion restriction: $Y(Z, D, S) = Y(Z', D, S)$ for all Z, Z' and for all D, S.

3. Relevance (nonzero average causal effect of Z on D): $E[D_i(1, S) - D_i(0, S)] \neq 0$.

4. Monotonocity (no defiers): $D_i(1, S) \geq D_i(0, S)$ for all S.

On top of these assumptions, that are sufficient for estimating a causal LATE under a randomized instrument in visitors , we assume that:

5. $\Pr(D_i(1) - D_i(0)|S_i = 0) = 0$. That is, no compliers are screened out by the screening process.

6. Screening does not affect enrollment decisions: $D(Z, S) = D(Z, S')$ for all S, S'.

7. Screening does not affect outcomes: $Y(Z, D, S) = Y(Z, D, S')$ for all S, S' and for all D.

Under the previous assumptions,

$$\beta_{IV}^{\text{Deniers}} = \frac{\mathbb{E}[y_i|Z_i = 1, S_i = 1] - \mathbb{E}[y_i|Z_i = 0, S_i = 1]}{\mathbb{E}[D_i|Z_i = 1, S_1 = 1] - \mathbb{E}[D_i|Z_i = 0, S_i = 1]} = \mathbb{E}[y_i(1) - y_i(0)|D_i(1) - D_i(0) = 1] = \beta_{IV}^{\text{Visitors}}.$$

Proof. Writing the potential outcome under being assigned to the instrument or not in terms of being screened in or screened out we get

$$\mathbb{E}[y_i|Z_i = 1] - \mathbb{E}[y_i|Z_i = 0] = \Pr(S_i = 1)(\mathbb{E}[y_i|Z_i = 1, S_i = 1] - \mathbb{E}[y_i|Z_i = 0, S_i = 1])$$
$$+ \Pr(S_i = 0)(\mathbb{E}[y_i|Z_i = 1, S_i = 0] - \mathbb{E}[y_i|Z_i = 0, S_i = 0]).$$

Expressing each one in terms of enrolling (D), we get

$$\Pr(S_i = 1)(\mathbb{E}[D_i(1)y_i(1) + (1 - D_i(1))y_i(0)|Z_i = 1, S_i = 1] - \mathbb{E}[D_i(0)y_i(1) + (1 - D_i(0))y_i(0)|Z_i = 0, S_i = 1])$$

$+\Pr(S_i = 0)(\mathbb{E}[D_i(1)y_i(1) + (1 - D_i(1))y_i(0)|Z_i = 1, S_i = 0] - \mathbb{E}[D_i(0)y_i(1) + (1 - D_i(0))y_i(0)|Z_i = 0, S_i = 0]).$

By independence of outcomes with respect to Z_i, we obtain

$$\Pr(S_i = 1)\mathbb{E}[(D_i(1) - D_i(0))(y_i(1) - y_i(0))|S_i = 1] + \Pr(S_i = 0)\mathbb{E}[(D_i(1) - D_i(0))(y_i(1) - y_i(0))|S_i = 0].$$

Let us define the set of always takers AT as the subset of individuals i such that $D_i(1) = D_i(0) = 1$, the set of never takers NT as the subset of individuals i such that $D_i(1) = D_i(0) = 0$, the set of compliers C as the subset of individuals i such that $D_i(1) - D_i(0) = 1$ and the set of defiers δ as the subset of individuals i such that $D_i(1) - D_i(0) = -1$. Noting that the set of defiers $\delta = \emptyset$ because of monotonicity, and that $D(1) - D(0) = 0$ for $i \in AT \cup NT$, we get the following expressions:

$$\Pr(S_i = 1)\mathbb{E}[(D_i(1) - D_i(0))(y_i(1) - y_i(0))|S_i = 1] = \Pr(S_i = 1)\mathbb{E}[y_i(1) - y_i(0)|i \in C, S_i = 1]\Pr(i \in C|S_i = 1) \text{ and}$$
$$\Pr(S_i = 0)\mathbb{E}[(D_i(1) - D_i(0))(y_i(1) - y_i(0))|S_i = 0] = \Pr(S_i = 0)\mathbb{E}[y_i(1) - y_i(0)|i \in C, S_i = 0]\Pr(i \in C|S_i = 0)$$

Since $\Pr(i \in C|S_i = 0) = 0$, the second expression is 0. Thus,

$$\Pr(S_i = 1)(\mathbb{E}[y_i|Z_i = 1, S_i = 1] - \mathbb{E}[y_i|Z_i = 0, S_i = 1]) = \Pr(S_i = 1)\mathbb{E}[y_i(1) - y_i(0)|i \in C, S_i = 1]\Pr(i \in C|S_i = 1)$$

Therefore, as $\Pr(i \in C|S_i = 1) = \mathbb{E}[D_i|Z_i = 1, S_1 = 1] - \mathbb{E}[D_i|Z_i = 0, S_i = 1]$

$$\beta_{IV}^{Deniers} = \frac{\mathbb{E}[y_i|Z_i = 1, S_i = 1] - \mathbb{E}[y_i|Z_i = 0, S_i = 1]}{\mathbb{E}[D_i|Z_i = 1, S_1 = 1] - \mathbb{E}[D_i|Z_i = 0, S_i = 1]} = \mathbb{E}[y_i(1) - y_i(0)|(D_i(1) - D_i(0) = 1, S_i = 1]$$

Now, given that screening does not affect outcomes nor enrollment decisions and that if an individual is a complier she is screened in, the conditionality on the right hand side in terms of S is redundant. So,

$$\beta_{IV}^{Deniers} = \frac{\mathbb{E}[y_i|Z_i = 1, S_i = 1] - \mathbb{E}[y_i|Z_i = 0, S_i = 1]}{\mathbb{E}[D_i|Z_i = 1, S_1 = 1] - \mathbb{E}[D_i|Z_i = 0, S_i = 1]} = \mathbb{E}[y_i(1) - y_i(0)|(D_i(1) - D_i(0) = 1] = \beta_{IV}^{Visitors}$$

$\square$

C Attrition

In the paper we have presented 2 tables studying attrition, presented here also for completeness. One that regresses a dummy for attrition on the treatment indicator and shows no differential attrition (Table OA-2, reproduced here for ease). A second one shows that among the attrited sample, treatment and control groups are balanced on observables (Table OA-3 reproduced here for ease). A third examines if the determinants of attrition are different in the control vs the treatment sample (Table OA-4). It predicts attrition by interacting the treatment with baseline covariates. We test and reject that the treatment and control groups have different determinants of attrition. This strongly suggests that we have no compositional differences resulting from treatment.

We present 5 more exercises below. Before these, it is important to understand that we defined attrition very conservatively. Table OA-5 disaggregates non-response by reason.

Exercise 1: Table OA-5 shows that close to one-quarter of the attrition arises from incorrect addresses[56], 12 percent from people who moved, and about half of a percent from people who died. Only 12 percent actually refused the survey. Refusal is not different between the control and the treatment group. This makes a story of endogenous refusal driving the results much less likely. Most of the reasons for not finding people (e.g. incorrect address) are unlikely to be driven by treatment, as the addresses were obtained before treatment assignment. Removing wrong addresses leaves us with response rates of 61%. Further removing the ones who moved results in response rates of 74%. Thus removing non-response that is likely unrelated to treatment results in much smaller attrition rates.

Another logistical fact that makes us confident that differential attrition is not driving the results is that instead of asking people to come to the clinic to take a blood sample, we had a survey team of nurses go home to home. We believe this practice could mitigate self-selection issues. Part of the reason for the low denial rates of our survey during covid was that we provided them with a free blood sample analysis at their doorstep.

Exercise 2: Inverse propensity weighting (IPW). Several papers, including the one you cite, proposed to use inverse propensity-score re-weighting to deal with treatment effect biases from attrition Molina-Millan and Macours (2021); Wooldridge (2002). We implement the strategy of Molina-Millan and Macours (2021) to weight regressions using the propensity score. Concretely, we feed Lasso all covariates and interactions with treatment from Table OA-4 to find a subset of important predictors of attrition. Then, in a second step, use the selected covariates in a Logit model to predict attrition. Finally, we use the Logit-predicted propensity to attrit as weights in our main OLS estimation. Table OA-7 displays the results. The estimates are very similar and lie

---

[56]In Mexico addresses are notoriously deficient with street names and numbers often missing or undefined. Recruiting people at the clinic rather than at home had this unexpected problem of having the wrong addresses.

within the confidence intervals of the estimates we had presented in the paper.[57]

The IPW method relies on selection-on-observables assumptions. The following two methods don't.

Exercise 3: Lee bounds. Our first method that does not rely on selection on observables is Lee bounds (Lee, 2009). Table OA-9 computes Lee bounds for the ITT estimate and for the IV. In both cases the bounds exclude zero, meaning that we can reject the null of no improvement in HbA1c, although they are somewhat wide as is common with lee bounds. Our result is qualitatively robust in the sense that CdA improved HbA1c levels.[58]

Exercise 4: Resistance to being surveyed. In this exercise, we follow the methodology of Behaghel et al. (2015) to survey attrition. Their model allows both for heterogeneous treatment effects and for an *unobserved* variable to affect response rates differentially for the treatment and control groups. With the aid of an additional variable —survey effort, proxied by the number of survey visits— they obtain non-parametric identification of average treatment effects in spite of differential attrition in unobservables, provided that the attrition mechanism can be represented by the latent variable threshold-crossing selection model, a common model in the literature. In the spirit of the marginal treatment effects literature, the availability of the number of visits enables the researcher to eliminate those excess marginal respondents in the treatment group.[59]

Figure OA-5 replicates their Figure 1 for our data. Our survey protocol established that nurses/enumerators would visit houses up to 4 times. If by the fourth time the survey was not completed, the survey was considered not doable. Most of the surveys were obtained on the first try. However, a non-negligible fraction required subsequent visits.

Table OA-10 applies their method to our sample. Because the treatment group had higher response rates, the procedure requires eliminating observations from the treatment group to make the response rate as close as possible to that of the control group. This involves eliminating those respondents we had to call 4 times in the treatment group. Eliminating them makes the response rates almost identical: 43.3 vs 43.7. Applying their method we again find very similar results, with point estimates that are within the confidence intervals of our previous estimates.

Exercise 5: Matching with administrative datasets. Referee 1 proposed matching subjects (attrited and non-attrited) to administrative datasets. Making progress in this regard involved substantial work, including access to confidential information through legal agreements. We were able to merge our survey data with two different data sets. The first is the Social Security data in Mexico, which enables us to observe if the person is employed in the formal sector in April 2023 and their average wage in that month. The second administrative

---

[57]Because some of the variables Lasso selected have missing values, the sample is smaller. Table OA-8 replicates the main table of the paper without IPW in this smaller sample.

[58]Columns 3 and 4 are not exactly the same as in the paper as we are not controlling for branch fixed effects or other regressors in the lee bounds excersise.

[59]Behaghel et al. (2015) uses more information than Lee bounds (namely the number of visits) and achieves tighter bounds.

data set is from the electoral institute, and it enables us to observe if the patient voted in the Federal June 2021 election, if they filed any procedure with the electoral institute (INE) up to April 2023, if they were discharged from the electoral voting list (up to April 2023).[60] INE also implements a small survey when a procedure is done, so we observe self-reported variables on being employed and self-employed.

One challenge is that we only have individual identifiers for 1225 out of 2410 baseline survey patients (Table OA-11), although the treatment/control arms are balanced in characteristics of persons with non-missing identifiers (see Table OA-12).[61] For this exercise we are forced to work with the sample with identifiers, which should be orthogonal to treatment as they were asked before treatment. Tables OA-14 and OA-15 use INE and Social Security administrative data to report means, number of observations[62], standard deviations, and p-values for the test of equality of means across treatment and control groups. Table OA-14 does not condition on response on the follow-up survey, while Table OA-15 focuses only on the subset of individuals that did not respond the follow up (and therefore on those we could not observe in the survey data). We find no statistically significant differences the labor market, in voting, or in electoral procedures. This is consistent with no differential attrition, providing additional evidence that our results are not biased.

Table OA-2: Attrition

| | $\mathbb{I}$(Answered follow-up) | | |
	(1)	(2)	(3)
Treatment	0.03	0.03	0.01
	(0.02)	(0.02)	(0.02)
Branch FE	No	Yes	Yes
Enrollment Month Fe	No	Yes	Yes
Basic Controls	No	No	Yes
Observations	2,410	2,410	2,042
R-squared	0.00	0.01	0.11
Mean dep. var	0.430	0.430	0.430

Notes: This table presents the results from running a regression where the dependent variable is a dummy of answering follow-up and the independent variable is the treatment group dummy. The first column reports the regression without any controls, the second column controls for branch and enrollment month fixed effects, and the third column also includes basic controls: age, gender, HbA1c, BMI, schooling and income. Columns 4 to 6 are analogous but remove patients that whose address was wrong. Finally, columns 7 to 9 remove further remove patients that moved. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

---

[60]The electoral institute issues the main ID for Mexicans. More than 90 million Mexicans (the overwhelming majority of adults) have INE identifications. INE data capture changes in addresses, renewals of IDs, and any such procedure. A discharge can occur for violating the law or by dying.

[61]Comparing those with vs without identifiers in Table OA-13, we find that those with identifiers are younger, richer, and more educated. Although we do not find differences in HbA1c.

[62]The number of observations differs based on whether the IDs are found in the different datasets.

Table OA-3: Balance Table for Those Measured at Follow Up

Variable	N	Control Mean/SE	N	Treatment Mean/SE	Difference C-T
Panel A: Demographics					
Age	509	52.39	558	53.06	-0.67
		(0.50)		(0.52)	(0.73)
Male	509	0.32	558	0.35	-0.03
		(0.02)		(0.02)	(0.03)
% High income	509	0.56	558	0.56	-0.01
		(0.02)		(0.02)	(0.03)
% High School or more	457	0.37	485	0.39	-0.02
		(0.02)		(0.02)	(0.03)
Panel B: Health and Health services					
HbA1c	508	9.35	556	9.51	-0.15
		(0.11)		(0.11)	(0.16)
BMI	509	31.07	557	30.83	0.24
		(0.27)		(0.25)	(0.37)
Has IMSS, ISSSTE or Seguro Popular	509	0.73	558	0.74	-0.01
		(0.02)		(0.02)	(0.03)
Percentage that use IMSS, ISSSTE or Seguro Popular	509	0.74	558	0.74	0.00
		(0.02)		(0.02)	(0.03)
HbA1c out of control (HbA1c>7)	508	0.76	556	0.78	-0.02
		(0.02)		(0.02)	(0.03)
Panel C: Beliefs					
Trust to improve following CdA recomendations (0-10)	432	9.43	467	9.54	-0.12
		(0.07)		(0.06)	(0.09)
Trust to improve following current health provider recomendations (0-10)	414	7.89	460	7.93	-0.03
		(0.13)		(0.13)	(0.18)
Trust in CdA diagnosis (0-10)	444	9.33	469	9.37	-0.04
		(0.07)		(0.07)	(0.09)
Trust in current health provider diagnosis (0-10)	411	7.93	450	7.88	0.05
		(0.13)		(0.13)	(0.18)

Notes: This table presents balance for patients who answered our follow-up. *(1) % High income*: Is an indicator variable equal to 1 if the person lived in the zip codes that the AMAI classifies as above middle income. This variable is an administrative variable and is based on the place of residency of each potential client at CdA. *(2) Has IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to be affiliated to a public service. *(3) Percentage that use IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to go for medical attention to the public services in our baseline survey. *(4) Trust to improve following CdA recommendations (0-10)*: is a variable that measures the trust in improving with CdA treatment. *(5) Trust to improve following current health provider recommendations (0-10)*: is a variable that measures the trust in improving with their current health provider. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA-4: Testing for differential attrition

	Attrition			
	(1)	(2)	(3)	(4)
Treatment	0.11	0.00	0.02	-0.13
	(0.20)	(0.22)	(0.22)	(0.28)
HbA1c	-0.00	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
(HbA1c)*Treatment	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
BMI	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
(BMI)*Treatment	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Age	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
(Age)*Treatment	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Male	-0.16***	-0.14***	-0.14***	-0.14***
	(0.03)	(0.03)	(0.03)	(0.03)
(Male)*Treatment	0.03	0.02	0.02	0.03
	(0.04)	(0.04)	(0.04)	(0.05)
High income		-0.09***	-0.10***	-0.11***
		(0.03)	(0.03)	(0.04)
(High income)*Treatment		0.01	0.01	0.02
		(0.05)	(0.05)	(0.05)
High school or more		-0.08**	-0.08**	-0.07**
		(0.03)	(0.03)	(0.04)
(High school or more)*Treatment		0.07	0.07	0.07
		(0.04)	(0.05)	(0.05)
Has IMSS, ISSSTE, or Seguro Populat			0.03	0.02
			(0.04)	(0.04)
(Has IMSS, ISSSTE, or Seguro Popular)*Treatment			-0.04	-0.02
			(0.05)	(0.06)
% That uses IMSS, ISSSTE, or Seguro Popular			0.01	0.02
			(0.04)	(0.04)
(% That uses IMSS, ISSSTE, or Seguro Popular)*Treatment			0.02	-0.00
			(0.05)	(0.06)
Trust to improve with CdA				0.00
				(0.02)
(Trust to improve with CdA)*Treatment				0.02
				(0.03)
Trust to improve with other				0.01
				(0.01)
(Trust to improve with current)*Treatment				-0.02
				(0.01)
Trust in CdA diagnosis				0.00
				(0.02)
(Trust in CdA diagnosis)*Treatment				0.00
				(0.03)
Trust in current diagnosis				-0.01
				(0.01)
(Trust in CdA diagnosis)*Treatment				0.00
				(0.01)
Observations	2,274	2,042	2,042	1,755
R-squared	0.097	0.103	0.104	0.111
Branch FE	Yes	Yes	Yes	Yes
Enrollment Month Fe	Yes	Yes	Yes	Yes
Mean dep. var	0.430	0.430	0.430	0.430
SD dep. var	0.495	0.495	0.495	0.495
F stat (H_0: all treatment interactions=0)	0.811	0.750	0.645	0.780
F test P value	0.518	0.609	0.740	0.671

Notes: This table presents a comparison between the patients who did answer our follow-up and the ones who did not. Attrition is a dummy variable taking the value of 1 if the patient answered the follow up survey. *% High income*: is an indicator variable equal to 1 if the person lived in zip codes that the AMAI classifies as above middle income. This variable is an administrative variable and is based on the place of residency of each potential client at CdA. *Has IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to be affiliated to a public health provider. *Percentage that use IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to go for medical attention to the public health provider at baseline. *Trust to improve following CdA recommendations (0-10)*: is variable that measures the trust in improving with CdA treatment. *Trust to improve following current health provider recommendations (0-10)*: Is a variable that measures the trust in improving with their current health provider. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA-5: Non-response decomposition

	Treatment	Control	t-test p-value
Baseline N	1226	1184	
Final N	558	509	
Non-response decomposition (%)			
(1) Wrong address	26.59	30.74	0.02
(2) Refused	12.56	12.75	0.89
(3) Moved	12.81	10.39	0.06
(4) Died	0.57	0.59	0.95
(5) Other	1.88	2.36	0.40
Response rates (%)			
All in denominator	45.51	42.99	0.21
Subtracting (1)	61.89	62.07	0.94
Subtracting (1) and (3)	74.97	73.03	0.40

Notes: This table decomposes the reasons for attrition.

Table OA-6: Differential attrition, different deffinitions of attrition

	𝕀(Answered follow-up)					
	Removing wrong address			Removing wrong address and moved		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.00	-0.00	-0.01	0.02	0.02	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Branch FE	No	Yes	Yes	No	Yes	Yes
Enrollment Month Fe	No	Yes	Yes	No	Yes	Yes
Basic Controls	No	No	Yes	No	No	Yes
Observations	1,720	1,720	1,465	1,440	1,440	1,241
R-squared	0.00	0.01	0.09	0.00	0.02	0.09
Mean dep. var	0.621	0.621	0.621	0.730	0.730	0.730

Notes: This table presents the results from running a regression where the dependent variable is a dummy of answering follow-up and the independent variable is the treatment group dummy. The first column reports the regression without any controls, the second column controls for branch and enrollment month fixed effects, and the third column also includes basic controls: age, gender, HbA1c, BMI, schooling and income. Columns 1 to 3 remove patients that whose address was wrong. Finally, columns 4 to 6 remove further remove patients that moved. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA-7: Main results using IPW

	HbA1c OLS	HbA1c IV	$\mathbb{I}$(HbA1c<7) OLS	$\mathbb{I}$(HbA1c<7) IV
	(1)	(2)	(3)	(4)
Use CdA	-1.00***	-1.20***	0.15***	0.21**
	(0.14)	(0.44)	(0.03)	(0.09)
HbA1c Bl	0.48***	0.49***		
	(0.03)	(0.04)		
$\mathbb{I}$(HbA1c<7) Bl			0.51***	0.52***
			(0.04)	(0.04)
BMI	-0.02*	-0.02*	0.00	0.00*
	(0.01)	(0.01)	(0.00)	(0.00)
Age	-0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.00)	(0.00)
Gender	-0.26*	-0.26*	0.06*	0.05*
	(0.16)	(0.16)	(0.03)	(0.03)
Elementary School	0.27	0.32	0.02	0.01
	(0.54)	(0.57)	(0.11)	(0.12)
Secondary School	0.46	0.49	-0.00	-0.01
	(0.55)	(0.56)	(0.11)	(0.12)
High School	0.07	0.10	-0.01	-0.01
	(0.57)	(0.59)	(0.12)	(0.12)
Tecnica o Normal	0.17	0.20	0.06	0.05
	(0.56)	(0.57)	(0.12)	(0.12)
University	-0.07	-0.03	0.09	0.08
	(0.61)	(0.62)	(0.13)	(0.13)
Income C+	0.16	0.15	-0.02	-0.02
	(0.38)	(0.37)	(0.11)	(0.12)
Income C	0.14	0.13	-0.04	-0.04
	(0.36)	(0.35)	(0.11)	(0.11)
Income D+	0.51	0.50	-0.13	-0.13
	(0.38)	(0.37)	(0.11)	(0.11)
Income D or Lower	0.26	0.25	-0.06	-0.06
	(0.40)	(0.39)	(0.11)	(0.12)
Observations	864	864	864	864
R-squared	0.35	0.32	0.25	0.22
F		97.47		97.10
First coeff		0.324		0.322
Mean dep. var	8.538	9.522	0.322	0.322

Notes: This table shows the results of estimating equation $Y_{ijt} = \beta_0 + \beta_1 \text{Use}_i + \Gamma_j + \psi_t + \chi X_i + \varepsilon_{ijt}$ using inverse probability weights. The first column presents the OLS estimates where $HbA1c_i$ is the dependent variable. The second column instruments Use_i of CdA with the treatment dummy, and so measures a LATE. The third and fourth columns replicate the two previous ones using as a dependent variable an indicator of sugar being under control, that is $I(HbA1c_i < 7)$. In all regressions we control for branch and month fixed effects, as well as basic demographics controls: age, sex, BMI, baseline HbA1c or controlled diabetes at baseline, schooling, and income class. Schooling controls are self-reported from our baseline survey and income class controls come from CdA administrative data. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$.

Table OA-8: Main results using the same sample as with IPW

	HbA1c OLS	HbA1c IV	$\mathbb{I}$(HbA1c<7) OLS	$\mathbb{I}$(HbA1c<7) IV
	(1)	(2)	(3)	(4)
Use CdA	-0.95***	-1.15***	0.14***	0.21**
	(0.14)	(0.42)	(0.03)	(0.09)
HbA1c Bl	0.50***	0.51***		
	(0.03)	(0.03)		
$\mathbb{I}$(HbA1c<7) Bl			0.52***	0.53***
			(0.04)	(0.04)
BMI	-0.01	-0.02	0.00	0.00
	(0.01)	(0.01)	(0.00)	(0.00)
Age	0.00	0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.00)	(0.00)
Gender	-0.23	-0.23	0.05*	0.05
	(0.15)	(0.15)	(0.03)	(0.03)
Elementary School	0.26	0.30	0.06	0.05
	(0.43)	(0.43)	(0.09)	(0.09)
Secondary School	0.45	0.46	0.04	0.03
	(0.43)	(0.43)	(0.09)	(0.09)
High School	0.07	0.09	0.03	0.03
	(0.45)	(0.45)	(0.09)	(0.09)
Tecnica o Normal	0.20	0.22	0.10	0.09
	(0.44)	(0.45)	(0.09)	(0.09)
University	-0.01	0.02	0.13	0.12
	(0.48)	(0.49)	(0.10)	(0.10)
Income C+	0.23	0.22	-0.03	-0.03
	(0.33)	(0.33)	(0.12)	(0.12)
Income C	0.25	0.23	-0.04	-0.04
	(0.31)	(0.31)	(0.11)	(0.12)
Income D+	0.56*	0.54*	-0.13	-0.12
	(0.33)	(0.33)	(0.12)	(0.12)
Income D or Lower	0.38	0.36	-0.07	-0.06
	(0.34)	(0.34)	(0.12)	(0.12)
Observations	864	864	864	864
R-squared	0.36	0.34	0.26	0.23
F		95.27		94.83
First coeff		0.315		0.314
Mean dep. var	8.538	9.301	0.329	0.329

Notes: This table shows the results of estimating equation $Y_{ijt} = \beta_0 + \beta_1 \text{Use}_i + \Gamma_j + \psi_t + \chi X_i + \varepsilon_{ijt}$ using the same sample used in inverse probability weights, Table OA-7. The first column presents the OLS estimates where $HbA1c_i$ is the dependent variable. The second column instruments Use_i of CdA with the treatment dummy, and so measures a LATE. The third and fourth columns replicate the two previous ones using as a dependent variable an indicator of sugar being under control, that is $I(HbA1c_i < 7)$. In all regressions we control for branch and month fixed effects, as well as basic demographics controls: age, sex, BMI, baseline HbA1c or controlled diabetes at baseline, schooling, and income class. Schooling controls are self-reported from our baseline survey and income class controls come from CdA administrative data. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA-9: Lee Bounds

	HbA1c ITT (1)	𝕀(HbA1C<7) ITT (2)	HbA1c IV (3)	𝕀(HbA1C<7) IV (4)
Treatment	-0.29*	0.05*		
	(0.15)	(0.03)		
Use CdA			-0.98**	0.18*
			(0.50)	(0.10)
Observations	939	939	939	939
R-squared	0.004	0.003	0.018	0.001
Lower Lee bound	-0.615	0.014	-1.922	0.043
Upper Lee bound	-0.143	0.072	-0.461	0.224
Control mean	8.538	0.322	8.538	0.322
Control SD	2.386	0.468	2.386	0.468

Notes: This table presets the ITT and IV estimates with their respective Lee bounds. Column 1 presents the ITT of our treatment on HbA1c at follow-up. Column 2 presents the ITT of our treatment on 𝕀(HbA1c<7) a dummy that takes the value of one if the patients' HbA1c level at follow-up is under 7. Column 3 instruments the use of CdA services with our treatment and presents the results for HbA1c at follow-up. Finally, column 4 instruments the use of CdA services with our treatment and presents the results for 𝕀(HbA1c<7) a dummy that takes the value of one if the patients' HbA1c level at follow-up is under 7. and 4

Table OA-10: Main results using Behaghel et al. (2015)

	HbA1c OLS	HbA1C IV	$\mathbb{I}$(HbA1C<7) OLS	$\mathbb{I}$(HbA1C<7) IV
	(1)	(2)	(3)	(4)
Use Cda	-1.01***	-1.06**	0.15***	0.19**
	(0.13)	(0.41)	(0.03)	(0.09)
HbA1c Bl	0.51***	0.51***		
	(0.03)	(0.03)		
$\mathbb{I}$(HbA1c<7) Bl	-0.01	-0.01	0.00	0.00
	(0.01)	(0.01)	(0.00)	(0.00)
BMI	0.00	0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.00)	(0.00)
Age	-0.21	-0.21	0.04	0.04
	(0.14)	(0.14)	(0.03)	(0.03)
Gender	0.32	0.33	0.06	0.05
	(0.41)	(0.42)	(0.09)	(0.09)
Elementary School	0.48	0.48	0.03	0.02
	(0.41)	(0.41)	(0.08)	(0.09)
Secondary School	0.22	0.23	0.02	0.02
	(0.43)	(0.44)	(0.09)	(0.09)
High School	0.28	0.29	0.09	0.09
	(0.43)	(0.44)	(0.09)	(0.09)
Tecnica o Normal	-0.01	-0.01	0.12	0.11
	(0.46)	(0.47)	(0.10)	(0.10)
University	0.21	0.21	-0.04	-0.04
	(0.31)	(0.31)	(0.12)	(0.12)
Income C+	0.25	0.25	-0.06	-0.06
	(0.28)	(0.28)	(0.11)	(0.12)
Income C	0.59*	0.58*	-0.16	-0.15
	(0.30)	(0.30)	(0.12)	(0.12)
Income D+	0.27	0.26	-0.07	-0.07
	(0.32)	(0.32)	(0.12)	(0.12)
Income D or Lower			0.52***	0.52***
			(0.04)	(0.04)
Observations	917	917	917	917
R-squared	0.38	0.36	0.27	0.23
F		92.34		92.28
First coeff		0.304		0.302
Mean dep. var	8.538	9.353	0.322	0.322

Notes: This table shows the results of estimating equation $Y_{ijt} = \beta_0 + \beta_1 \text{Use}_i + \Gamma_j + \psi_t + \chi X_i + \varepsilon_{ijt}$ using only the patients according to Behaghel et al. (2015). The first column presents the OLS estimates where $HbA1c_i$ is the dependent variable. The second column instruments Use_i of CdA with the treatment dummy, and so measures a LATE. The third and fourth columns replicate the two previous ones using as a dependent variable an indicator of sugar being under control, that is $I(HbA1c_i < 7)$. In all regressions we control for branch and month fixed effects, as well as basic demographics controls: age, sex, BMI, baseline HbA1c or controlled diabetes at baseline, schooling, and income class. Schooling controls are self-reported from our baseline survey and income class controls come from CdA administrative data. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$.

Table OA-11: Identifier existence and Follow-up survey completion

		Follow-up survey		
		0	1	
ID exists	0	674	461	1135
	1	673	607	1280
		1345	1068	2415

Notes: This table presents a cross tabulation of whether the ID used in the match existed, and whether follow-up survey was completed. For subjects with baseline survey.

Table OA-12: Balance Table for treatment and control groups, conditional on having ID used in Match with Admin data

Variables	N	Control Mean/(SE)	N	Treatment Mean/(SE)	Pairwise t-test P-value C-T
Panel A: Demographics					
Age	577	50.19 (0.47)	587	50.43 (0.49)	0.73
Male	612	0.44 (0.02)	613	0.41 (0.02)	0.37
% Incomer higher than C	612	0.65 (0.02)	613	0.67 (0.02)	0.30
% High School or more	562	0.51 (0.02)	574	0.47 (0.02)	0.19
Panel B: Health and Health services					
HbA1c	610	9.47 (0.10)	611	9.45 (0.10)	0.88
HbA1c out of control (HbA1c>10)	610	0.78 (0.02)	611	0.79 (0.02)	0.82
BMI	611	31.21 (0.26)	613	31.57 (0.25)	0.32
Has IMSS, ISSSTE or Seguro Popular	612	0.77 (0.02)	613	0.80 (0.02)	0.16
Percentage that use IMSS, ISSSTE or Seguro Popular	612	0.78 (0.02)	613	0.81 (0.02)	0.10
Panel C: Beliefs					
Trust in CDA diagnosis (0-10)	542	9.27 (0.07)	560	9.31 (0.07)	0.69
Trust in current health provider diagnosis (0-10)	502	7.86 (0.12)	540	7.94 (0.12)	0.64
Trust to improve following CDA recomendations (0-10)	534	9.38 (0.07)	558	9.49 (0.06)	0.24
Trust to improve following current health provider recomendations (0-10)	509	7.80 (0.13)	546	7.93 (0.12)	0.46

Notes: This table presents a comparison between the control and treatment group for patients for whom we have CURPs, the Mexican national identifyer. *% High income*: is an indicator variable equal to 1 if the person lived in zip codes that the AMAI classifies as above middle income. This variable is an administrative variable and is based on the place of residency of each potential client at CdA. *Has IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to be affiliated to a public health provider. *Percentage that use IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to go for medical attention to the public health provider at baseline. *Trust to improve following CdA recommendations (0-10)*: is variable that measures the trust in improving with CdA treatment. *Trust to improve following current health provider recommendations (0-10)*: Is a variable that measures the trust in improving with their current health provider. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$.

Table OA-13: Balance table comparing characteristics of those with versus without ID for matching with Admin data

Variable		Dont have ID (1)		Have ID (2)	(1)-(2) Pairwise t-test
	N	Mean/(SE)	N	Mean/(SE)	P-value
Panel A: Demographics					
Age	1118	53.97 (0.35)	1164	50.31 (0.34)	0.00***
Male	1185	0.41 (0.01)	1225	0.42 (0.01)	0.40
% Incomer higher than C	1185	0.58 (0.01)	1225	0.66 (0.01)	0.00***
% High School or more	1012	0.36 (0.02)	1136	0.49 (0.01)	0.00***
Panel B: Health and Health services					
HbA1c	1182	9.59 (0.08)	1221	9.46 (0.07)	0.21
HbA1c out of control (HbA1c>10)	1112	0.79 (0.01)	1221	0.78 (0.01)	0.85
BMI	1184	30.27 (0.17)	1224	31.39 (0.18)	0.00***
Has IMSS, ISSSTE or Seguro Popular	1185	0.77 (0.01)	1225	0.79 (0.01)	0.31
Percentage that use IMSS, ISSSTE or Seguro Popular	1185	0.69 (0.01)	1225	0.80 (0.01)	0.00***
Panel C: Beliefs					
Trust in CDA diagnosis (0-10)	979	9.30 (0.05)	1102	9.29 (0.05)	0.88
Trust in current health provider diagnosis (0-10)	908	7.95 (0.09)	1042	7.90 (0.08)	0.66
Trust to improve following CDA recomendations (0-10)	964	9.40 (0.05)	1092	9.43 (0.05)	0.61
Trust to improve following current health provider recomendations (0-10)	938	7.93 (0.09)	1055	7.87 (0.09)	0.64

Notes: This table presents a comparison between the patients for whom we have CURPS and those that we do not. *% High income*: is an indicator variable equal to 1 if the person lived in zip codes that the AMAI classifies as above middle income. This variable is an administrative variable and is based on the place of residency of each potential client at CdA. *Has IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to be affiliated to a public health provider. *Percentage that use IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to go for medical attention to the public health provider at baseline. *Trust to improve following CdA recommendations (0-10)*: is variable that measures the trust in improving with CdA treatment. *Trust to improve following current health provider recommendations (0-10)*: Is a variable that measures the trust in improving with their current health provider. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$.

Table OA-14: Balance using IMSS and INE administrative data

	N	Control (1) Mean/(SE)	N	Treatment (2) Mean/(SE)	(1)-(2) Pairwise t-test P-value
Panel A: Demographics					
Sex	615	0.44 (0.02)	619	0.42 (0.02)	0.40
Panel B: Labor					
Wage*	145	535.50 (36.76)	144	499.11 (36.42)	0.48
$\mathbb{I}$(Formal worker)*	640	0.23 (0.02)	640	0.23 (0.02)	0.95
$\mathbb{I}$(Self-employed)	615	0.09 (0.01)	619	0.09 (0.01)	0.59
Panel C: Civic behavior					
$\mathbb{I}$(Voted in 2021)	625	0.57 (0.02)	625	0.60 (0.02)	0.28
$\mathbb{I}$(ID procedure)	624	0.81 (0.02)	624	0.82 (0.02)	0.56

Notes: This table presents the balance between treatment and control using Social Security data and Electoral Authority data from Mexico. Variables marked with an * come form the Social Security data, other variables come from the Mexican Electoral Authority.

Table OA-15: Balance on Admin Data: for those that did not answer followup

	N	Control (1) Mean/(SE)	N	Treatment (2) Mean/(SE)	(1)-(2) Pairwise t-test P-value
Panel A: Demographics					
Sex	330	0.55 (0.03)	323	0.47 (0.03)	0.04**
Panel B: Labor					
Wage*	86	577.26 (52.28)	80	564.55 (58.32)	0.87
𝕀(Formal worker)*	342	0.25 (0.02)	331	0.24 (0.02)	0.77
𝕀(Self-employed)	330	0.12 (0.02)	323	0.10 (0.02)	0.43
Panel C: Civic behavior					
𝕀(Voted in 2021)	337	0.54 (0.03)	325	0.54 (0.03)	0.85
𝕀(ID procedure)	337	0.81 (0.02)	325	0.82 (0.02)	0.86

Notes: This table presents the balance between treatment and control using Social Security data and Electoral Authority data from Mexico. Variables marked with an * come form the Social Security data, other variables come from the Mexican Electoral Authority.
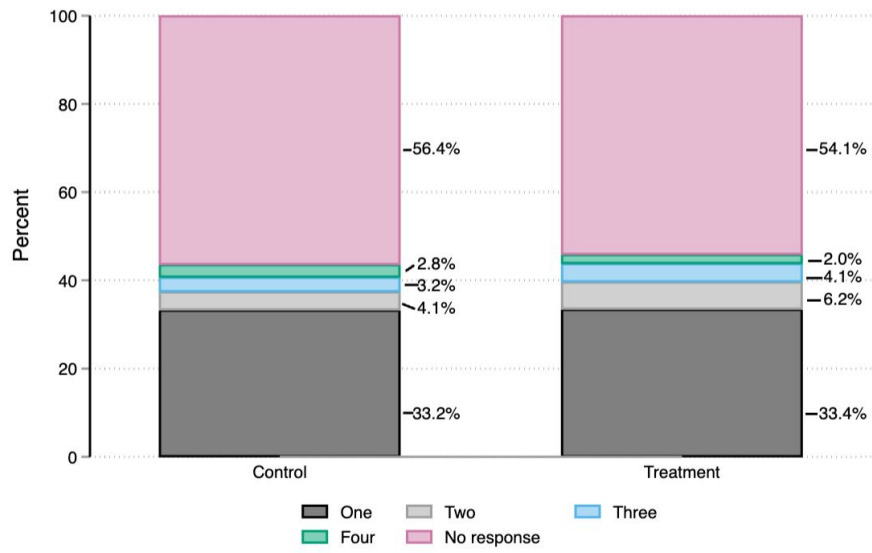
Figure OA-5: Visits

Figure 1: Visits

Notes: This figure presents the distribution of visits to try to reach the patients in the treatment and control groups, and is analogous to Figure 1 in Behaghel et al. (2015).

D External Validity

In this appendix we explore external validity.

D.1 Comparison with other populations

We first present a comparison within CdA patients, we then turn to compare how our sample looks compared to that of diabetics surveyed in the Mexican health and nutrition survey (ENSANUT).

Table OA-16 presents descriptive statistics of CdA patients that are in our experiment and the other CdA patients who were interested in a membership and had a screening. We can see that patients in our experimental sample are more likely to be female, are poorer and less educated. In terms of health they are in a similar condition and other patients are 5% more likely to utilize public services. Lastly, they look similar in terms of beliefs, both in the servicce of CdA and the other providers.

Table OA-17 compares CdA patients that are in our experiment and the average diabetic surveyed in the 2018 ENSANUT. We can see that patients in our experimental sample are younger and more likely to be female. In terms of health, patients in our sample are in a worse condition in terms of HbA1c, but less likely to have hypertension. Both samples have similar BMI. Moreover, we can see that our sample is more likely to take pills and insuline, and also more likely to diet and exercise. They look similar in terms of complications, even though our sample is twice as likely to utilize IMSS services.

Table OA-16: CdA Patients Comparison

Variable	N	Others Mean/SE	N	Experiment Mean/SE	Difference O-E
Panel A: Demographics					
Age	4503	51.6604 [0.1830]	2282	52.1017 [0.2478]	-0.4412
Male	5118	0.4416 [0.0069]	2410	0.4154 [0.0100]	0.0262**
% Incomer higher than C	4792	0.6511 [0.0069]	2410	0.6191 [0.0099]	0.0320***
% High School or more	4738	0.5317 [0.0073]	2148	0.4288 [0.0107]	0.1029***
Panel B: Health and Health Services					
HbA1c	4828	9.4441 [0.0380]	2403	9.5215 [0.0513]	-0.0773
BMI	4840	30.8465 [0.0862] [344.7927]	2408	30.8403 [0.1250] [581.7519]	0.0062
Has IMSS, ISSSTE or Seguro Popular	4844	0.7642 [0.0061]	2339	0.7790 [0.0086]	-0.0147
Percentage that use IMSS, ISSSTE or Seguro Popular	5118	0.7900 [0.0057]	2410	0.7411 [0.0089]	0.0489***
Panel C: Beliefs					
Trust in CdA diagnosis (0-10)	4595	9.3280 [0.0207]	2081	9.2989 [0.0338]	0.0291
Trust in current health provider diagnosis (0-10)	4329	7.9857 [0.0389]	1950	7.9241 [0.0604]	0.0616
Trust to improve following CdA recomendations (0-10)	4551	9.4959 [0.0186]	2056	9.4193 [0.0332]	0.0767**
Trust to improve following current health provider recomendations (0-10)	4378	7.9589 [0.0402]	1993	7.8946 [0.0619]	0.0643

Notes: This table presents descriptive statistics of CdA patients that are in our experiment and the CdA patients who were interested in a membership and had a screening. We can see that patients in our experimental sample are more likely to be female, poorer and less educated. In terms of health they are in a similar condition. Patients not in our experimental sample are 5% more likely to utilize public services. Lastly, they look similar in terms of beliefs, both in the servicce of CdA and the other providers. * $p < 0.10$, ** $p < 0.05$, $***p < 0.01$.

Table OA-17: CdA vs Average in ENSANUT

| Variable	N	CdA Mean/SE	N	ENSANUT Mean/SE	Difference C-T
Panel A: Demographics					
Age	1067	52.74 [0.36]	2120	56.79 [0.30]	-4.05***
Male	1067	0.33 [0.01]	2120	0.38 [0.01]	-0.05***
Panel B: Health					
HbA1c	1023	9.43 [0.08]	2120	7.51 [0.05]	1.92***
BMI	1024	31.03 [0.19]	1199	31.03 [0.18]	0.00
Hypertension	1067	0.38 [0.01]	2120	0.43 [0.01]	-0.05**
Panel C: Health care and expenses					
Takes pills	1025	0.61 [0.02]	2120	0.58 [0.01]	0.03*
Takes insuline	1025	0.23 [0.01]	2120	0.14 [0.01]	0.09***
I(Diet)	1025	0.33 [0.01]	2120	0.27 [0.01]	0.06***
I(Exercise)	1025	0.28 [0.01]	2120	0.06 [0.01]	0.22***
Panel D: Diabetic complications					
Ulcers	1025	0.04 [0.01]	2120	0.06 [0.01]	-0.01
Heart attack	1025	0.01 [0.00]	2120	0.01 [0.00]	0.00
Panel E: Treatment at IMSS					
IMSS	1025	0.57 [0.02]	2120	0.23 [0.01]	0.34***

Notes: This table presents Descriptive statistics of CdA patients that are in our experiment and the average diabetic surveyed in the 2018 health and nutrition survey (ENSANUT). We can see that patients in our experimental sample are younger and more likely to be female. In terms of health, patients in our sample are in a worse condition in terms of HbA1c, but less likely to have hypertension. Both samples have similar BMI. Moreover, we can see that our sample is more likely to take pills and insulin, and also more likely to diet and exercise. They look similar in terms of complications, even though our sample is twice as likely to utilize IMSS services. * $p < 0.10$, ** $p < 0.05$.

	Always takers 1		Never takers 2		Our sample 3		t-test Difference (1)-(2)	t-test Difference (1)-(3)	t-test Difference (2)-(3)	F-test for joint orthogonality
	N	Mean/SE	N	Mean/SE	N	Mean/SE				
Age	2998	51.58 [0.22]	2488	51.72 [0.25]	2282	52.10 [0.25]	-0.15	-0.52	-0.38	0.29
Male	3311	0.48 [0.01]	2792	0.38 [0.01]	2410	0.42 [0.01]	0.11***	0.07***	-0.04***	0.00***
% High income	3208	0.63 [0.01]	2445	0.68 [0.01]	2410	0.62 [0.01]	-0.06***	0.01	0.06***	0.00***
% High shcool or more	3044	0.51 [0.01]	2576	0.56 [0.01]	2148	0.43 [0.01]	-0.05***	0.08***	0.13***	0.00***
HbA1c	3236	10.09 [0.04]	2461	8.69 [0.05]	2403	9.52 [0.05]	1.40***	0.57***	-0.83***	0.00***
BMI	3239	30.74 [0.11]	2472	30.97 [0.12]	2408	30.84 [0.12]	-0.23	-0.10	0.13	0.37
Weight	3241	81.71 [0.33]	2473	81.67 [0.36]	2408	80.23 [0.37]	0.03	1.48***	1.44***	0.00***
Height	3240	1.63 [0.00]	2473	1.62 [0.00]	2339	1.61 [0.00]	0.01**	0.02***	0.01***	0.00***
I(Takes pills daily)	3311	0.54 [0.01]	2792	0.42 [0.01]	2338	0.52 [0.01]	0.13***	0.03**	-0.10***	0.00***
I(Takes insulin)	3311	0.18 [0.01]	2792	0.13 [0.01]	2339	0.20 [0.01]	0.05***	-0.02	-0.07***	0.00***
Has IMSS, ISSSTE or Seguro Popular	3242	0.76 [0.01]	2474	0.77 [0.01]	2339	0.78 [0.01]	-0.00	-0.02	-0.01	0.32
Percentage that use IMSS, ISSSTE or Seguro Popular	3311	0.79 [0.01]	2792	0.79 [0.01]	2410	0.74 [0.01]	0.00	0.05***	0.05***	0.00***
Trust to improve following CdA recomendations	2946	9.53 [0.02]	2461	9.45 [0.03]	2056	9.42 [0.03]	0.08**	0.11***	0.03	0.01***
Trust to improve following current health provider recomendations	2826	7.76 [0.05]	2375	8.14 [0.05]	1993	7.89 [0.06]	-0.38***	-0.14*	0.25***	0.00***
Trust in CdA diagnosis	2971	9.35 [0.03]	2494	9.31 [0.03]	2081	9.30 [0.03]	0.04	0.05	0.01	0.48
Trust in current health provider diagnosis	2797	7.81 [0.05]	2341	8.20 [0.05]	1950	7.92 [0.06]	-0.39***	-0.11	0.28***	0.00***

Notes: This table presents a comparison between always takers, never takers, and our sample. *% High income*: is an indicator variable equal to 1 if the person lived in zip codes that the AMAI classifies as above middle income. This variable is an administrative variable and is based on the place of residency of each potential client at CdA. *Has IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to be affiliated to a public health provider. *Percentage that use IMSS, ISSSTE or Seguro Popular*: is an indicator variable equal to 1 if the person declared to go for medical attention to the public health provider at baseline. *Trust to improve following CdA recommendations (0-10)*: is variable that measures the trust in improving with CdA treatment. *Trust to improve following current health provider recommendations (0-10)*: Is a variable that measures the trust in improving with their current health provider. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure OA-6: Comparing socio-economic levels: Cda patients vs Mexican population
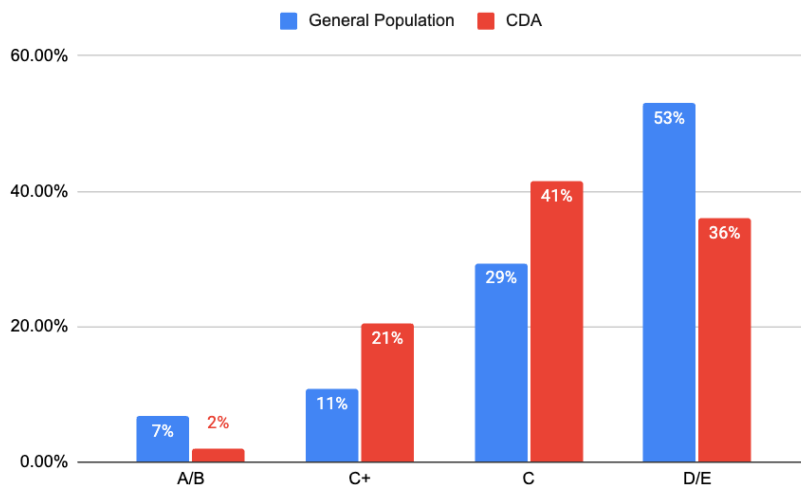
Notes: This figure compares socio-economic levels Cda patients vs Mexican population. CdA does not collect income information but does collect locations where their patients come from and uses the methodology of the Mexican Association of Public Opinion Agencies (AMAI, https://www.amai.org/NSE/index.php?queVeo=preguntas) to classify them (red, all of CdA patients, not just the experimental. sample). AMAI also provides estimates for Mexico as a country (blue).

D.2 Conditional Average Treatment Effects

To assess a form of external validity, we use Athey et al. (2019)'s method of honest causal forests to estimate the conditional average treatment effects (CATE) more flexibly.[63] This method enables the estimation of the average treatment effect for each patient in the experiment, according to the partition of the covariate space they belong to (i.e. the trees leafs). We then use these estimates to predict treatment effects for the subjects in our experiment versus the subjects we called "always-takers" which are not in our experimental sample, based on their observed covariates at baseline.[64] Figure OA-7 compares the CATE of both distributions, and shows that both distributions share a common support, although the distribution of effects is less dispersed for always-takers than that of our complier sample. On average the predicted effect is -0.334 for the always-takers and -0.327 for the compliers.

Figure OA-7: CATE: Compliers vs. always-takers
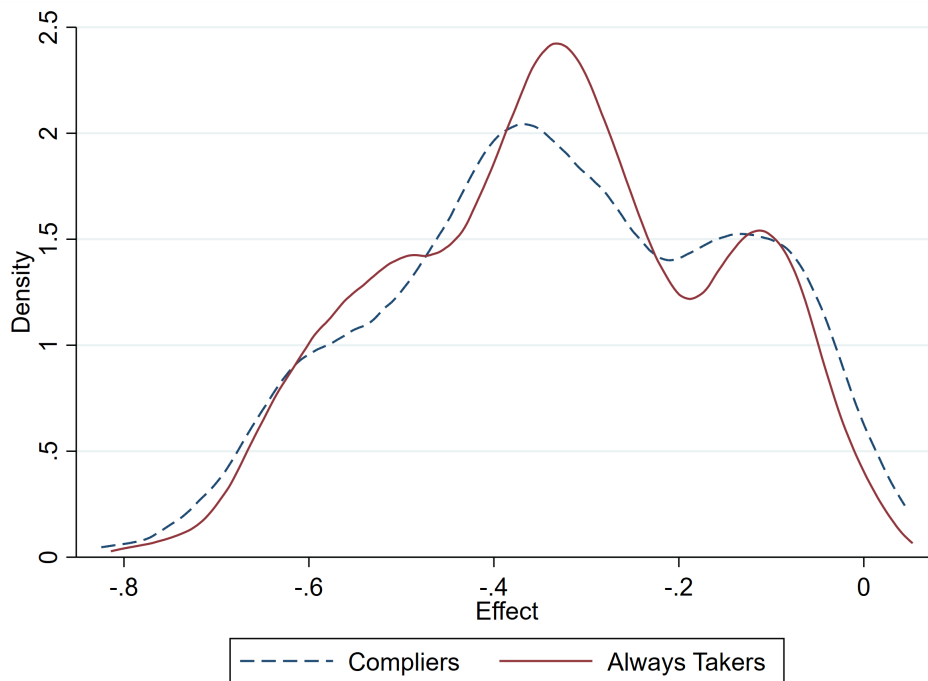
Notes: This figure compares the distribution of the conditional average treatment effect we estimate based on Athey et al. (2019) for our sample and how such an effect looks if we extrapolate to the rest of patients that showed interest in CdA. We can see both distributions look quite similar. Note that here we only utilize the reduced form estimates from being assigned to control or treatment and not the IV since we cannot know the endogenous choice that a regular patient would have made on whether or not to take up the offer by CdA. All the persons out of the experiment where assigned a zero in the treatment variable. We include as covariates: gender, HbA1c, BMI, age, social insurance, and the clinic where the appointment took place. We omit the education variable because of missing values.

---

[63]The conditional average treatment effect is just the average treatment effect for a subsample of the population that share certain values of the covariates, $\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]$. For more detail we refer the reader to Athey et al. (2019).

[64]These covariates include body mass index, HbA1c, age, sex, income, office branch where baseline was conducted, and whether the individual has IMSS, ISSSTE, Seguro Popular, private, none, or other.

E   Heterogeneity

We explored a wide variety of dimensions along which we might find heterogeneous impacts. To model heterogeneity, we estimate the following model:

$$Y_{ijt} = \beta_0 + \beta_1 T_i + \beta_2 H_i + \beta_3 T_i * H_i + \Gamma_j + \psi_t + \chi X_i + \varepsilon_{ijt} \tag{3}$$

where Y_{ijt} is HbA1c values for individual i who enrolled in clinic j on month t, T_i is an indicator for being part of the treatment group, H_i is a dummy for whether the patient has a value higher than the median of the heterogeneity variable. Γ_j and ψ_t capture clinic and month of enrollment fixed effects. The coefficient of interest for this analysis is β_3. The results are shown in Table OA-19.

While our estimates are somewhat imprecise, we find no significant heterogeneity across age, sex, BMI, schooling or income[65]. We do find that there is significant heterogeneity by baseline level of HbA1c. Sicker individuals at baseline improve more than their healthier counterparts.[66]

Table OA-19: Heterogeneity Effect

Dep Var HbA1c follow up

	Age	Male	Income	Schooling	HbA1c	BMI	Informal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.51**	-0.40**	-0.29*	-0.36*	-0.16	-0.21	-0.40**
	(0.21)	(0.17)	(0.15)	(0.19)	(0.12)	(0.20)	(0.16)
Regressor	-0.66***	-0.00	-0.50*	-0.57**	2.26***	-0.34	0.08
	(0.20)	(0.22)	(0.29)	(0.22)	(0.19)	(0.21)	(0.23)
Treatment X regressor	0.45	0.27	-0.10	0.20	-0.42*	-0.20	0.31
	(0.27)	(0.30)	(0.36)	(0.30)	(0.25)	(0.28)	(0.32)
Branch FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	No	No	No
Observations	1,067	1,067	1,067	942	1,064	1,066	1,067
R-squared	0.03	0.02	0.02	0.03	0.21	0.03	0.02
Mean dep. var.	8.54	8.54	8.54	8.54	8.54	8.54	8.54

Notes: This table presents the heterogeneity results of the effect of getting CdA treatment on health, estimated using specification 3 and an interaction with the covariate listed at the top of each column. For each of these, we report the treatment, covariate, and interaction coefficients. In all regressions we control for branch fixed effects. For age, HbA1c and BMI we split by above the median. For income and schooling we split for those considered middle or upper class and those with high school or higher education respectively. Informal: is an indicator variable equal to 1 if the person declared to be affiliated to the Seguro Popular or to not have social insurance to the medical staff at CdA. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

---

[65]The lack of results on income heterogeneity (and the fact that the subsidy is a small percentage of their income) suggests that our treatment does not just operate through an income effect.

[66]When we use the False Discovery Rate for multiple testing adjustments (Anderson, 2008) we reach the same conclusion: we still find that doctor visits increased (q-value = 0.03) and find no effect on health behaviors (diet, exercise, etc).

F   Other Tables and Figures

F.1   Multiple testing

Table 4 tests multiple hypothesis. Here we present two strategies to deal with multiple testing. The first creates a standardized index by families of outcomes as suggested Anderson (2008) and used by Banerjee et al. (2015); Kling et al. (2007), Table OA-20 shows these results.

In the second, we use the False Discovery Rate for multiple testing adjustments (Anderson, 2008). Table OA-21 shows the results using q-values associated with the False-Discovery Rate for multiple testing, along with the typical p-values for comparison. Using both methods we reach the same conclusion as before. We find that doctor visits increased and find suggestive effect on health behaviors (diet, exercise, etc).

Table OA-20: Mechanisms Index

	Doctor visits (1)	Index of Medicines usage (2)	Self-care (3)
Use CdA (Instrumented)	0.82***	0.70***	0.27
	(0.21)	(0.20)	(0.19)
Branch FE	Yes	Yes	Yes
Months since enrollment FE	Yes	Yes	Yes
Basic Controls	Yes	Yes	Yes
Observations	940	914	940
R-squared	0.13	0.07	0.17
F	96.13	100.7	95.97
First coeff	0.306	0.317	0.306
Mean dep. var	0.00	0.00	0.00

Notes: This table presents indexes calculated following Anderson (2008). That is, we first standardized variables by subtracting the mean and dividing by is standard deviation. We then sum these standardized variables, weighting them by the inverse of their variances. We do this by using the *swindex* command in Stata which implements the generalized least-squares method of index construction. For all of these specifications, we report the local average treatment effect (LATE) that we causally estimated from our experiment instrumenting Use of CdA with the treatment arm. We control for branch and month fixed effects as well as basic demographics (age, sex, BMI, baseline HbA1c, schooling and income). The Index of doctor visits is made from the following three variables: *(1) Visits to Doctor*: captures how many diabetes-related visits to the doctor the person did the last year. *(2) # Special Check Ups*: diabetes-related checkups the person got during the last year, i.e eyes, kidney, foot, and blood checkups. *(3) # Specialists*: captures the number of diabetes-related specialist doctors the patient met with in a regular visit to their health provider. The Index of medicines usage if made from the following variables: *(1) Stop Medication if feels good*: is an indicator variable that takes the value of 1 if the patient declared that he suspended his medication whenever he starts felling well. *(2) Takes Pills and (3) Takes Insulin*: both are indicator variables that equal 1 if the person declared that he took pills and insulin, respectively, as part of his treatment, we also control for their baseline measure. The Index of Self-care is made from the following variables: *(1) I(Exercise) and (2) I(Diet)*: both are indicator variables that equal 1 if the person declared that he did exercise and/or was involved in a diet as an effort to take care of his health, respectively. (3) Diet+Exercise: is the sum of I(Diet) and I(Exercise). *(4) # Cigarettes and (5) #Sodas*: capture the number of cigarettes and sodas consumed in a regular day by the person, we also control for their baseline measure. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA-21: Mechanisms q-values Anderson (2008)

	Visits to Doctor (1)	# Special Check ups (2)	# Specialist (3)	Stop med If feels good (4)	Takes pilss (5)	Takes Insuline (6)	$\mathbb{I}$(Excercise) (7)	$\mathbb{I}$(Diet) (8)	Diet + Excercise (9)	# Cigarettes (10)	# Sodas (11)
Use CdA (instrumented)	2.59**	1.32***	0.64**	-0.22**	0.25***	-0.15**	0.14	0.26*	0.14	-0.72	-0.06
	(1.06)	(0.38)	(0.27)	(0.09)	(0.09)	(0.06)	(0.10)	(0.15)	(0.10)	(0.80)	(0.14)
Observations	878	817	940	815	913	913	940	940	940	774	857
R-squared	0.04	0.03	0.10	0.05	0.04	-0.00	0.05	0.07	0.06	0.45	0.15
Branch FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Months since enrollment FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Original p-values	0.02	0.00	0.02	0.02	0.01	0.02	0.17	0.08	0.18	0.37	0.66
q-values	0.03	0.01	0.03	0.03	0.03	0.03	0.12	0.06	0.12	0.17	0.29
F	98.82	83.46	96.13	91.35	99.78	99.78	96.10	97.12	97.31	78.83	104.7
First coeff	0.321	0.307	0.306	0.323	0.316	0.316	0.306	0.308	0.308	0.307	0.332
Mean dep. var	6.502	1.408	1.383	0.249	0.727	0.134	0.289	0.725	0.436	0.911	0.841

Notes: For all of these specifications we report the local average treatment effect (LATE) that we causally estimated from our experiment instrumenting Use of CdA with the treatment arm. We control for branch and month fixed effects as well as basic demographics (age, sex, BMI, baseline HbA1c, schooling and income). *(1) Visits to Doctor*: captures how many diabetes-related visits to the doctor the person did the last year. *(2) # Special Check Ups*: diabetes-related checkups the person got during the last year, i.e eyes, kidney, foot, and blood checkups. *(3) # Specialists*: captures the number of diabetes-related specialist doctors the patient met with in a regular visit to their health provider. *(4) Stop Medication if feels good*: is an indicator variable that takes the value of 1 if the patient declared that he suspended his medication whenever he starts felling well. *(5) Takes Pills and (6) Takes Insulin*: both are indicator variables that equal 1 if the person declared that he took pills and insulin, respectively, as part of his treatment, we also control for their baseline measure. *(7) I(Exercise) and (8) I(Diet)*: both are indicator variables that equal 1 if the person declared that he did exercise and/or was involved in a diet as an effort to take care of his health, respectively. *(9) Diet+Exercise*: is the sum of I(Diet) and I(Exercise). *(10) # Cigarrets and (11) #Sodas*: capture the number of cigarettes and sodas consumed in a regular day by the person, we also control for their baseline measure. Robust standard errors in parentheses. The q-values for multiple testing are calculated following Anderson (2008). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F.2 Extra tables

Table OA-22: Effect on Trust and Waiting Time

	Dif Trust in Improving (1)	Dif Trust in Diagnosis (2)	Waiting time (2)
Use Cda	1.92***	1.48**	-30.33***
	(0.59)	(0.59)	(9.02)
Branch FE	Yes	Yes	Yes
Enrollment Month Fe	Yes	Yes	Yes
Basic Controls	Yes	Yes	Yes
Observations	682	676	739
R-squared	0.02	0.02	-0.05
F	82.97	92.88	82.49
First coeff	0.330	0.348	0.325
Mean dep. var	2.107	2.292	43.46

Notes: This table presents the results from the effect of getting CdA treatment on trust and waiting time. The first column captures the local average treatment effect (LATE) on the difference in the self-reported trust that the patient will improve their health through CdA, as opposed to through their health provider at baseline (the difference in the rows shown in Table 2), we also control by the baseline measure of the variable which is define as the difference from the baseline survey on the same questions. The second column carries out the same exercise for patients' trust in diagnoses that come from CdA as opposed to their health provider at baseline, we also control by the baseline measure of the variable which is define as the difference from the baseline survey on the same questions. The third column reports our results on waiting time. On our follow-up survey we asked how long do patients usually wait at the doctor they usually go to get diabetes care and estimated the difference based on our experiment, instrumenting going to CdA rather than to other providers by being in the treatment group. We winzorized at 5% the upper tail to omit implausible waiting times. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA-23: Other results in the pre-analysis plan

	Weight (1)	BMI (2)	Visits to hospital (3)	Expenditure on medicine (4)	Expenditure on doctors (5)	Expenditure on hospital (6)	Annual expenditure (7)
Use CdA (Instrumented)	0.70	0.06	-5.50	-186.97	362.02	-2,754.42	975.41
	(1.03)	(0.52)	(6.34)	(176.01)	(1,215.37)	(3,035.87)	(5,444.20)
Observations	924	924	638	444	240	318	131
R-squared	0.918	0.840	-0.018	0.022	0.128	-0.034	0.198
Branch FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Enrollment Month Fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var	78.74	31.64	1.868	291.4	739.1	352.7	5173
SD dep. var	16.33	6.130	23.37	539.2	2079	6203	8316

Notes: This table presents the results for regressions of the form of equation 1 on different self-reported behaviors. For all of these specifications we are reporting the local average treatment effect (LATE) that we causally estimated from our experiment using instrumental variables. In all regressions we control for branch and month fixed effects as well as basic demographics (age, sex, BMI, baseline HbA1c, schooling and income). *(1) Weight*: captures the weight of the patient in kilograms. *(2) BMI*: the body mass index of the patient. *(3) Visits to hospital*: is a variable that captures the times a patient had to visit a hospital in the last 12 months. *(4) Expenditure on Medication*: captures the out-of-pocket expenditure on medication. *(5) Expenditure on doctors*: captures the out-of-pocket expenditure on doctors. *(6)Expenditure on hospital*: captures the out-of-pocket expenditure on hospital visits. (7) Annual expenditure: captures the out-of-pocket annual expenditure. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F.3 A bit more detail on fiscal externality calculation

In this subsection of the appendix we provide a bit mode detail on Figure OA-8. Because the effects and therefore the savings differ by baseline level of HbA1c, we first classify our sample into 10 baseline HbA1c value bins $j = 5$ to 14, and allocate a person to the bin according to her starting HbA1c. We then calculate a conditional local average treatment effect (CLATE) in terms of HbA1c for each of these bins using Athey et al. (2019)'s methodology, and call it $\widehat{CLATE}_j$. Second, we calculate the average HbA1c observed at follow up for the control group in each bin j, $HbA1C_{C_j}^{followup}$. We sum $\widehat{CLATE}_j$ to $HbA1C_{C_j}^{followup}$ to estimate level of HbA1c for the treatment group in each bin: $HbA1C_{T_j}^{followup} = HbA1C_{C_j}^{followup} + \widehat{CLATE}_j$.

To go from HbA1c effects to health complications and hospitalizations, we use the complications incidence tables by level of HbA1c from the UKPDS 35 study (King et al., 2001).[67] We define averted hospitalizations as the difference of hospitalizations between treatment and control group for each bin given their level of $HbA1C^{followup}$. Finally, to estimate the peso savings to the system from reduced hospitalizations we use the complication-specific hospital cost data from Barraza-Lloréns et al. (2015) for 2013 in Mexico updated for inflation.

---

[67]For each group, we interpolate linearly between the two HbA1c integer numbers that the incidence tables in the paper report. Since dynamics of the effect of higher blood sugar for each complication vary, we estimate incidence for each main complication separately. We focus on neuropathy, ulcers, amputations, ophthalmic complications, diabetic coma, nephropathy, stroke, and heart attack.

Figure OA-8: Savings from Hospitalizations Averted per Patient/Year (comparing T vs C)
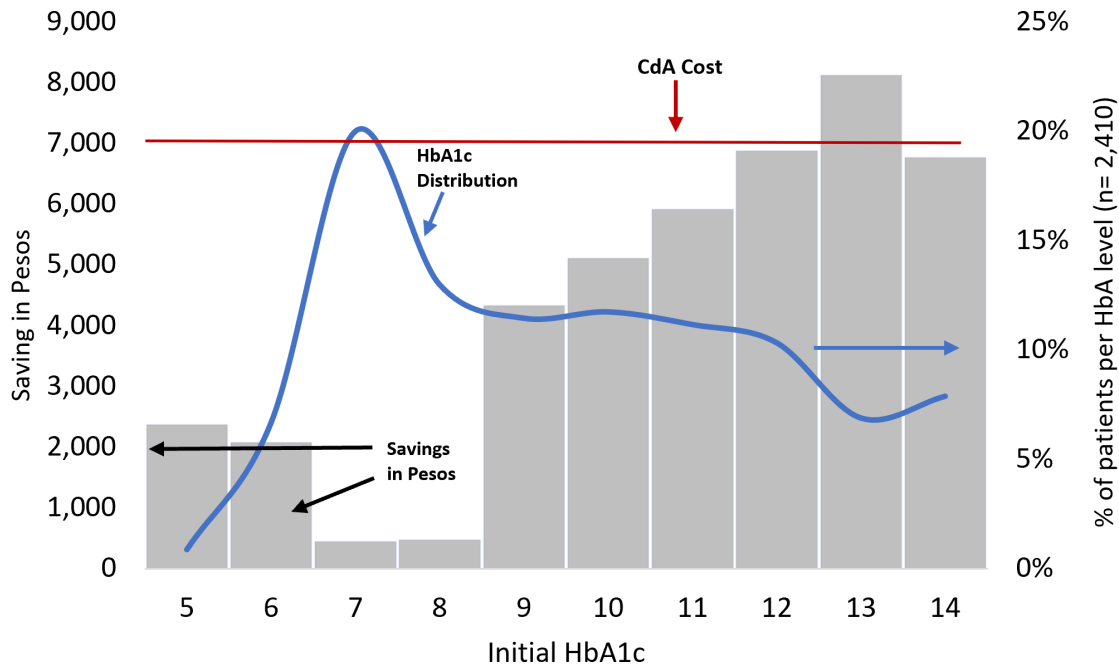
Notes: This figure shows the yearly savings we would observe from averted hospitalizations based on the reductions in HbA1c that we causally estimate along with estimates from the medical literature and public spending data from the government. The gray bars are the yearly savings in pesos estimated for each initial HbA1c level (left axis). The blue curve corresponds to the distribution of baseline HbA1c of the patients (right axis). The red line marks the cost of CdA. Specifically, we follow six steps recognizing the important heterogeneity in impacts by baseline blood sugar levels. First, we classify our sample into HbA1c baseline value bins, using the smaller nearest integer. Second, for those in the control group, in each bin we averaged their HbA1c observed at follow up ($HbA1C_{C_j}^{followup}$). Third, to go from HbA1c levels to yearly health complications, we use the complications incidence tables by level of HbA1c from the widely cited UKPDS 35 study (King et al., 2001). This gives us the estimated yearly complications for the control group by bin. Fourth, we apply an analogous method for the treatment group. We start from their baseline level and add in the conditional local average treatment effect (CLATE) of the respective bin, where the estimate is done as above but separately for each bin. That is for each bin j we calculate the number $HbA1C_{T_j}^{followup} = HbA1C_{T_j}^{baseline} + CLATE_j$, and map these to health complications. Fifth, we define averted hospitalizations as the difference in expected hospitalizations between treatment and control group for each bin. Finally, to estimate the yearly savings to the system from reduced hospitalizations, we multiply the cost of each complication by the yearly averted hospitalizations using the cost data from Barraza-Lloréns et al. (2015) for 2013 in Mexico updated for inflation. Overall, we see that on average 55% of the yearly costs would be recuperated by averted hospitalizations and that CdA is a potential savings mechanism for the government for the sickest patients. It is important to note that this figure is not considering substitution away from public sector services nor externalities from emptier clinics.

Figure OA-8 shows the results. The bins are represented in the horizontal axis, and the thin blue curve displays the distribution of initial HbA1c. As can be seen, a large fraction of clients are not under control at baseline (Initial HbA1c>7), this fraction is displayed on the vertical axis of the right. The vertical bars represent the savings in pesos (left Y-axis) estimated per patient *year* for each bin. For instance, those in the 13-HbA1c bin save more than $8000 pesos per year just in the cost of hospitalizations from complications, while those in the 12-HbA1c bin save almost $7000 pesos. Savings typically rise with HbA1c levels both because (a) the rate of complications is higher for those with higher blood sugar and (b) our CLATE estimates show larger effects of our intervention for those with higher blood sugar at baseline. For comparison, the horizontal red line plots the annual subscription price of CdA of $7000 pesos. Our results show that at high levels of blood sugar, the program essentially pays for itself in reduced hospital spending as the savings of those between 11

and 14 initial HbA1c averages close to $7000 pesos; overall, we estimate that reduced hospital expenditures amount to 55% of the cost of CdA. Note that both the benefits and the costs are measured in pesos per year, so that we do not require assumptions of the durability of effects over time. The bottom of the Figure displays for each bin the average savings as a fraction of the CdA annual subscription.

The above calculation does not consider visits. At baseline, our control group has 5 visits on average to their public provider in the year before joining CdA. The direct cost to IMSS of each visit is $800 pesos. In addition, IMSS data shows that the total cost of maintaining care for an under-control diabetic is $9000 pesos per year. So a 20% reduction in visits would save $800-$1800 pesos per year; if we use the larger estimate from the third column of Table 5, we would have savings of $1600-$3600 pesos per year just from lowering visits.

Recall that for our sample the cost of CdA treatment for one year was $7000 pesos. Using our estimate of offsetting hospital spending on complications, as well as our lowest estimate of offsetting primary care expenditures, roughly 65% of the costs of CdA are offset by reduced public sector costs; at the upper bound, the offset is 105%. In either case, the net cost of this incremental care through CdA is much less than the gross costs.

F.4 Distance as an instrument

Figure OA-9: Distance to Clinic Distribution
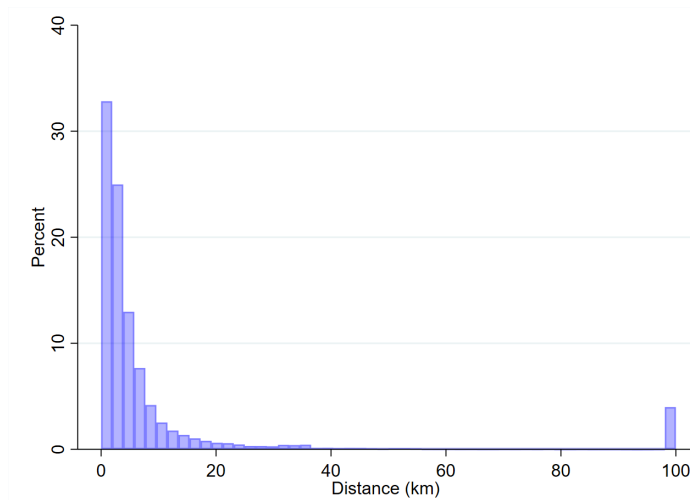
Notes: This figure shows the distribution in distance form home to the clinic for the IMSS patients we utilize in our regressions. We winzorized the distances higher than 100 km.

Distance may be correlated with unobservables in the regression, and although it is an instrument often used in education economics (Walters, 2018; Card, 2001), as in this literature we had to make the identification assumption that distance to clinics is not correlated with potential outcomes. We wrote in the paper that distance may be correlated with underlying health, and that "We address this by controlling for baseline blood sugar at time $t-1$, so that we are assessing the impact of visits on the *improvement* in blood sugar. Of course, this does not solve the underlying identification problem if those who live near IMSS clinics are on differential underlying health trajectories than are those who live far away. But the inclusion of clinic fixed effects control for any neighborhood factors that might drive such trends." We now perform 3 more exercises.

Exercise 1. Correlation between instrument and demographics. We correlate distance with observable variables at the client level. In the IMSS dataset we have only a limited number of predetermined covariates unfortunately. Table OA-24 regresses these covariates (sex, age, capillary glucose at baseline) with distance. We find no correlation with sex and capillary glucose. We do find a negative correlation between distance and age, but it is tiny: an increase of 10 km from the clinic is associated with a decrease of 0.07 years of age. This lends more credibility to the identification assumption.

Exercise 2. Adding extensive controls for locality-level characteristics from Mexico's Population Census. One may worry that people who live farther away liven in places that are very different from those that are

Table OA-24: Correlation between covariates and instrument

	Sex	Age	Capillary glucose
	(1)	(2)	(3)
Distance	-0.000	-0.007***	0.002
	(0.000)	(0.002)	(0.009)
Observations	160,035	160,035	160,035
R-squared	0.027	0.067	0.108
Branch FE	Yes	Yes	Yes
Month Fe	Yes	Yes	Yes
Mean dep. var	0.394	60.59	134.5

living closer. We were able to use GPS coordinates to merge our data with the 2020 Mexican population census polygons. This census reports variables at the level of the locality; Mexico has 185,000+ localities. The census reports several covariates at the locality level. Share of people with access to social security, mean schooling, share people with a physical disability, share of households with internet, share of households with automobile, and population at the locality. We include all these 6 variables in the distance regressions in an effort to control for omitted variables. Table OA-25 shows that including them barely changes the estimates, again lending more credibility to our identification strategy.

Table OA-25: Comparison IMSS vs Cda Effect: Extensive Controls

	IMSS		CdA	IMSS	
	Number of medical visits	$\mathbb{I}$(12 months follow up)	Capillary Glucose	Capillary Glucose	Capillary Glucose
	(1)	(2)	(3)	(4)	(5)
Distance (km)	-0.0027***	0.0000			
	(0.0005)	(0.0000)			
Number of medical visits			-8.53**	-4.76*	-12.94**
			(4.06)	(2.87)	(6.50)
Observations	156,083	429,222	1,053	156,083	133,872
F			60.80	30.56	12.81
First coeff			1.301	-0.003	-0.002
Instrument			Discount	Distance	Distance (W)
Extensive controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var	5.273	0.364	214.9	129.2	129.2

Notes: All columns except (3) estimated in IMSS data. The first column shows regression of number of visits at IMSS on distance from the clinic. The second column shows regression of dummy for having a follow-up blood sugar measurement on distance. The third column shows IV regression in CdA data where we regress capillary glucose on number of visits, instrumented by treatment indicator. The fourth column shows an IV regression of capillary glucose on the number of IMSS visits, instrumented by distance. The fifth column repeats this exercise but reweighting the sample so that the baseline distribution for capillary blood sugar matches that of CdA. All specifications have branch and month fixed effects and basic controls (age, age squared, initial capillary glucose, and gender). Extensive controls include share of people with access to social security, mean schooling, share people with a physical disability, share of households with internet, share of households with automobile, and population at the locality size. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Exercise 3. Controlling for municipality fixed effects. By controlling for municipality fixed effects we can control for unobservable time-invariant characteristics that may affect HbA1c at follow-up. This limits the distance variation we use —as it only uses distance variation within geographic units— but substantially increases the parametrization of the regression and its robustness to unobserved factors at the municipality level. In Columns 1, 4, and 5 we use a data set that contains 1,509 municipalities. Column 2 data set contains 1,704 municipalities. Finally, Column 3 data has 23 municipalities in it. Results are reported in Table OA-26. In spite of including this large set of fixed effects results are qualitatively analogous and quantitatively not far.[68]

Table OA-26: Comparison IMSS vs Cda Effect: Municipality FE

	IMSS		CdA	IMSS	
	Number of medical visits	$\mathbb{I}$(12 months follow up)	Capillary Glucose	Capillary Glucose	Capillary Glucose
	(1)	(2)	(3)	(4)	(5)
Distance (km)	-0.0028***	0.0001			
	(0.0006)	(0.0001)			
Number of medical visits			-8.29**	-7.75**	-9.97*
			(3.98)	(3.40)	(5.57)
Observations	159,749	439,050	1,052	159,749	137,024
F			62.59	25.23	15.28
First coeff			1.342	-0.003	-0.003
Instrument			Discount	Distance	Distance (W)
Municipality FE	Yes	Yes	Yes	Yes	Yes
Mean dep. var	5.27	0.36	214.9	129.2	129.2

Notes: All columns except (3) estimated in IMSS data. The first column shows regression of number of visits at IMSS on distance from the clinic. The second column shows regression of dummy for having a follow-up blood sugar measurement on distance. The third column shows IV regression in CdA data where we regress capillary glucose on number of visits, instrumented by treatment indicator. The fourth column shows an IV regression of capillary glucose on the number of IMSS visits, instrumented by distance. The fifth column repeats this exercise but reweighting the sample so that the baseline distribution for capillary blood sugar matches that of CdA. All specifications have branch and month fixed effects and basic controls (age, age squared, initial capillary glucose, and gender). All specifications include municipality fixed effects. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

---

[68]Unfortunately, we do not observe in our data how catchment areas are defined and cannot carry out an analysis of comparing people in neighboring catchment areas that go to different clinics. Note however that there are more municipalities in our data than clinics, so that controlling for municipality fixed effects could be already doing a lot of the work that catchment area fixed effect could do.