I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14685

The Right to Health and the Health Effects of Denials

Sonia Bhalotra
Manuel Fernández

AUGUST 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14685

The Right to Health and the Health Effects of Denials

Sonia Bhalotra
University of Warwick and IZA

Manuel Fernández
Universidad de los Andes and IZA

AUGUST 2021

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Right to Health and the Health Effects of Denials*

We estimate the health costs of supply-side barriers to accessing medical care. The setting is Colombia, where citizens have a constitutional right to health care, but insurance companies that manage delivery impose restrictions on access. We use administrative data on judicial claims for health as a proxy for unmet demand. We validate this using the register recording all health service utilization, estimating that a one standard deviation increase in judicial claims is associated with pervasive decreases in utilization rates of between 0.25 and 0.71 standard deviations, including in medical consultations, procedures, hospitalizations and emergency care. These restrictions on access manifest in population health outcomes. We estimate that a one standard deviation increase in judicial claims increases the all-cause mortality rate by between 0.10 and 0.23 standard deviations. Increases in mortality are pervasive across causes, with the largest increase in deaths from certain cancers. They are also pervasive across the age and sex distribution but larger among individuals over the age of fifty and (weakly) among women and the low-income population.

JEL Classification: I11, I13, I18, K4

Keywords: health care, health insurance, mortality, right-to-health, litigation, universal-health-coverage, Colombia

Corresponding author:
Sonia Bhalotra
Department of Economics
University of Warwick
Coventry, CV4 7AL
United Kingdom
E-mail: sonia.bhalotra@warwick.ac.uk

* We acknowledge funding from ESRC grant ES/M010263/1 for the Human Rights Big Data and Technology Project (hrbdt.ac.uk) at the Human Rights Centre, University of Essex. We are grateful to Marty Gaynor, Paul Hunt, Clara Sandoval-Villalba, Bob Town, Simon Weidenholzer and Carmel Williams for many helpful discussions.

1 Introduction

Universal health coverage has recently garnered widespread support as a policy objective (Bloom et al., 2018). It is seen as marking the third great transition in public health, following the demographic and epidemiological transitions of the past two centuries (Rodin and de Ferranti, 2012). However it is increasingly recognized that there are weak links between coverage and population health outcomes. Some studies have highlighted low uptake (Banerjee et al., 2010), while others have highlighted poor service quality (Mohanan et al., 2014; Powell-Jackson et al., 2015; Kruk et al., 2018). We instead investigate supply-side denials or restrictions on access to care. The setting is Colombia where a state commitment to the right to health for all (Article 48) was written into the new constitution formulated in 1991, following which a new health care system that progressively reached nearly universal health coverage was put in place. The Constitution provides judicial protection of the right to health and a series of other human rights (Article 86), allowing citizens who fear that their rights are jeopardized because of restrictions to file claims for redress with the Constitutional Court. These judicial claims (tutelas) are costless, simple, can be filed with any judge within the local jurisdiction, and have preferential proceeding such that the judge is mandated to return a decision within ten days. This is a major 'experiment' in accountability.

The commitment to health coverage has been delivered, since 1993, through a managed competition model of health insurance which, at 90 percent, has one of the highest levels of comprehensive insurance coverage among developing economies. Health insurers are responsible for organizing provision of health services, but tight regulation constrains the extent to which insurers can vary the premium, the content of the benefits package, or risk-based enrollee selection, the strategic variables that are typically available to insurers to manipulate. Competition between insurers for enrollees is thus driven by the network of providers and service quality. In practice, the industry is highly concentrated at the municipality level.[1] This, together with the fact that insurers get a fixed fee per enrollee while the costs they incur depend on a variable demand for medical services, generates an incentive to impose barriers to restrict access to medical care as a cost-reduction strategy, especially when the punishments for doing so are small. A weak punishment regime reflects the weak administrative capacity that characterizes many low and middle income countries.

The evidence is accumulating to suggest that health insurers have been systematically restricting access of enrollees to required medical services, mostly through delays or denials of authorization for examinations, medicines, surgeries or treatments that have already been prescribed by a health professional (Maya, 2008; Corte Constitucional, 2008; Yamin and

[1] Municipalities are decentralized subdivisions of which there are 1,120 in the country.

Parra-Vera, 2009, 2010; Parra-Vera and Yamin, 2013; Arrieta-Gómez, 2018).[2] In response to these restrictions, individuals have pursued litigation. Between 2010 and 2016, more than 675,000 tutelas were filed against health insurers, a yearly average of 2.27 claims per 1,000 enrollees (henceforth the tutela rate). On average, 80% of health-related claims were decided in favor of claimants.

We obtained administrative data on the 675,000 tutelas (judicial claims) at the municipality-insurer-year level for 2010-2016.[3] In the analysis period there were an average of 80 health insurers per year operating across 1,120 municipalities. We merged these data with mortality rates by age, sex and cause of death, obtained from Vital Statistics registers. We also obtained, at the municipality-year level, administrative records that contain the universe of medical services provided, classified by type of diagnoses (e.g. neoplasms vs mental disorders) and type of service (consultations, procedures, hospitalization, emergencies).

A key feature of our study is that we use administrative data on judicial claims as a measure of unmet demand for health care.[4] We sketch a conceptual framework that captures the decision of insurers to deny health claims, and the decision of citizens to litgate when faced with a denial. This shows that the tutela rate (which is observed) is a scalar multiple of unmet demand (which is unobserved). We investigate impacts of unmet demand on mortality, a widely used marker of population health. We demonstrate the presumed mechanism by showing that utilization of medical services is decreasing in unmet demand. This validates our approach of proxying unmet demand with judicial claims. Importantly, it also makes it unlikely that our results for mortality are driven by demand side variation. This is because demand shocks will tend to move health service utilization and mortality in the same direction. In contrast, we expect supply side denials to reduce service utilization and increase mortality.

We nevertheless adopt an empirical strategy that increases our confidence that we isolate supply-side restrictions from demand-side factors. By virtue of using insurer-municipality-year data, we are able to control not only for cross-sectional heterogeneity at the insurer-municipality level but also for shocks at the insurer level and at the municipality level. To address the concern that there remain relevant omitted variables that vary at the insurer-municipality-time level, we additionally introduce an instrumental variables approach. Leveraging the fact that all insurers operate across multiple municipalities, we instrument the

[2]See also the periodical reports of the Colombian Ombudsman office (Defensoría del Pueblo, 2004, 2007, 2009, 2010, 2011, 2012, 2013, 2015); and a report by the Ministry of Health (Ministerio de Salud y Proteccion Social, 2018).

[3]These data are not publicly available. We obtained data on judicial claims- and also on complaints to the Health Superintendent by filing a petition under the Right to Information act.

[4]Unmet demand refers to health services requested (demanded) but not delivered because of supply-side barriers.

tutela rate at the municipality-insurer-year level with the tutela rate at the insurer-year level calculated using information from all municipalities in which the insurer operates, except the municipality of interest.[5] The identifying assumption is that insurer-specific (supply side) factors determine the denial rate in a way that is orthogonal to municipality-level demand for health care.[6] The exclusion restriction holds if the denial rate in municipalities outside the department where the index municipality is located, conditional on the structure of fixed effects and other time-varying controls, does not directly affect the mortality rate in the index municipality.

To provide an external check on the estimates obtained using tutela (judicial claims) data, we obtained an alternative measure of unmet demand, which is the number of complaints against health insurers made to the National Health Superintendent (Supersalud), the agency in charge of oversight of the system. Between 2013 and 2017, the years for which administrative data on complaints are available, there were 1.5 million individual complaints against health insurers, the majority of which were related to restrictions imposed by insurers on access to requested medical services. Importantly, the claims and the complaints are collected and managed by different agencies. An advantage of the complaints data is that we can identify the age and sex of the plaintiff in these data, which allows us to analyze heterogeneity in the effects of complaints on health outcomes along these dimensions.

As discussed, we investigate unmet demand directly by modelling health care utilization as a function of tutelas. As these data are only available at the municipality- year level, we modify the empirical strategy, using a leave-one-out shift-share IV. The premise remains that insurer-specific restriction rates are correlated across municipalities because of supply side determinants that are orthogonal to local demand for health care.

We find large effects of denials (restrictions) on population-level mortality rates. The marginal effects are similar irrespective of whether we use judicial claims (tutelas) or complaints as the measure of unmet demand. The results are also broadly similar whether or not we use an instrument, increasing our confidence that selection is not a major issue. Using the IV specification we estimate that a one standard deviation increase in the tutela rate is associated with an increase of between 0.10 and 0.23 standard deviations of the mortality rate (which corresponds to an increase of between 8.7 and 20.2 percent in the average mor-

[5] Allowing for correlated demand shocks between municipalities that are close to each other, we refine the instrument to include only municipalities outside the department in which the index municipality is located. We control for mean reversion and for the level and change in the market share of health insurers in the municipality-year.

[6] Natural examples of insurer-specific supply side factors are cost structures determined by different service provider contracts, and different managerial practices.

tality rate). These numbers translate to between 310 and 720 additional deaths per year.[7] Mortality rates are used as a measure of population health because they are consistently available over a long period of time from vital statistics data. However, mortality is an extreme measure that represents the tip of the iceberg of underlying morbidities. We might therefore regard our estimates for mortality as a lower bound of the impact of denials on population health.

The impact of denials on mortality is pervasive, being evident across cause of death, age, sex and broad income groups. The strongest impacts are on mortality rates due to certain cancers. Impacts are stronger among people over the age of 50 and (weakly) women. The standardized effects are larger in the subsidized regime (SR) that covers the lower income population, albeit the uncertainty associated with the parameters estimates is large enough that the difference relative to the contributory regime (CR) is not always statistically significant. We investigated whether unmet demand (tutelas) is greater in municipalities with weaker insurer level competition, and find no significant association. We similarly find no evidence that increased mortality emerges from municipalities where insurers face lower competition. This is consistent with a recently evolving literature on the industrial organization of health care markets which shows, for the American market, that the impact of insurer competition on welfare, negotiated provider prices, and premia is theoretically ambiguous (Ho and Lee, 2017). In our setting too, it is unclear that enforcing more competition will induce better provision, for instance because greater market power may enhance the bargaining power of insurers when negotiating contracts with providers, generating cost reductions that lower the incentive to impose restrictions on demand.

We proceeded to estimate the extent to which denials hamper health service utilization, a direct measure of denials, and the driver of increased mortality. We estimate that a one standard deviation increase of the tutela rate reduces the per capita number of *hospitalizations* by 0.54 standard deviations, of *emergencies* by 0.46 standard deviations, and of *consultations with health professionals* by 0.71 standard deviations. Estimates for *medical procedures* are not statistically significant when using the tutela rate, but when we use the complaints rate they are. Thus denials reduce utilization of most sorts of medical services. We also find that they reduce utilization across the range of common diagnoses. As indicated earlier, these results not only delineate a mechanism but also validate the use of tutelas (or complaints) as a measure of denials.

We discuss broadly related research in Section 7, but here we briefly delineate where the

[7] Although not directly comparable in concept or context, to benchmark these effect sizes, we note that studies analyzing Medicaid expansion in the US have found impacts of health insurance coverage on mortality in a wide range, from no effect (Finkelstein and McKnight, 2008; Weathers and Stegman, 2012) to a reduction in all-cause mortality of 6.1 percent (Sommers et al., 2012).

contributions of this paper lie. Our work is related to a literature studying the health effects of expanding health insurance coverage which highlights the importance of studying outcomes alongside access (see, among others, Baicker et al. (2013); Miller et al. (2013); Gruber et al. (2014); Goodman-Bacon (2018)) We contribute to this literature by identifying the extent to which slippage between coverage and access can hamper the realization of the potential benefits of coverage. Although the incentive structure that allow insurers to restrict access to health care is particular to the Colombian system, analogous problems arise whenever health insurance coverage does not guarantee effective access to medical care. For instance, studies of Medicaid have noted supply shortages and uncompetitive physician fees (Currie et al., 1995; Government Accountability Office, 2011; Rosenbaum, 2014). However, neither for the US, nor for any other country, are we aware of a study that identifies the size and spread of the unmet demand for medical services and the associated population-level health costs of denying medical care. We make a contribution on this front.

We further contribute to a literature studying how the use of litigation by patients can influence provision of health care. Economists have focused upon the effects of liabilities for malpractice on the quality and cost of medical care (Kessler and McClellan, 1996; Danzon, 2000; Currie and MacLeod, 2008; Frakes and Jena, 2016). We depart from this thread by focusing upon a context in which litigation is used not to deter medical malpractice, but to enforce access to medical care, supported by the constitutionalization of health as a human right. The right to health is explicitly recognized, to a greater or lesser extent, in the written constitutions of more than half of all UN member countries (Kinney and Clark, 2004; Backman et al., 2008; Hamel et al., 2015). However it is only in some cases that this notional commitment is accompanied by constitutional provisions to guarantee its enforcement, including the possibility of legal recourse. In Colombia, use of tutelas and complaints as an enforcement mechanism has become widespread, leading to what is referred as judicialization of health care (Yamin and Parra-Vera, 2009, 2010; Parra-Vera and Yamin, 2013). Although they seem less widespread in these countries, similar provisions exist in Argentina (Bergallo, 2011), Brazil (Biehl et al., 2009), Costa Rica and South Africa (Wilson, 2011). We provide what appears to be the first quantitative analysis of the content of the judicial claims instrument, using comprehensive administrative data.

Our study is particularly relevant in light of the current wave of expansion in health care provision in many countries, enhanced by a global agenda for universal health coverage. With public budgets often inadequate to the task, issues of effective delivery and accountability are increasingly important. Our results highlight the costs to population health and to the public purse associated with the incentives of providers being unaligned with constitutional commitments. There is too much regulation insofar as insurers are not allowed to set prices

or select enrollees and, at the same time, too little regulation insofar as insurers are not penalized when complaints against them are received and upheld.

The rest of this paper unfolds as follows. Section 2 outlines the relevant features of the Colombian health care system and the process of judicialization of access to health care. Section 3 describes the data sources. Section 4 discusses a theoretical framework. The econometric model and identification strategy are presented in Section 5. In Sections 6 and 7 we present and discuss the results and robustness checks. Section 8 concludes.

2 Institutional Background and Policy Context

Structure. The health care system in Colombia is organized as a tightly regulated managed competition model of health insurance.[8] Individuals enter the system by enrolling with a health insurer operating in their municipality of residence. The insurance company is responsible for guaranteeing enrollees access to required medical services through their independently contracted network of providers (e.g. hospitals, clinics, laboratories). Citizens can freely choose which insurer to enroll with, and which providers within the insurer network to use. There are two regimes that have a common structure but differ in their target populations: the contributory regime (CR) covers formal sector employees and those in self-employment that are able to pay, while the subsidized regime (SR) covers the low-income population. Affiliation to the CR is mandatory for all formal employees, while eligibility for the SR is means-tested.[9]

Insurers are constrained by regulation of three key strategic variables. First, insurers cannot set premia, but instead receive a standardized capitation payment per enrollee which is defined annually by the government. Second, they must offer a standardized compulsory benefit package (known as POS) that includes preventive care services and essential procedures and medications.[10] Finally, insurers cannot deny enrollment on the basis of

[8]For a characterization of the managed competition model of health insurance see Enthoven (1978); Londoño and Frenk (1997). We provide an overview of characteristics relevant for our analysis, an in-depth description is available in Gaviria et al. (2007); Maya (2008); Glassman et al. (2009). Analysis of the SR is provided in Miller et al. (2013).

[9]The contributory regime regime is financed by payroll contributions and corporate income taxes, while the subsidized regime is financed by 'solidarity' contributions from the CR and transfers from both subnational governments and the central government.

[10]If a required service or medication is not included in the benefits package, users have to pay for it unless it is deemed to be essential and they lack the resources to cover the cost. In such cases, the insurer has to guarantee the timely provision of the service and the government reimburses the insurer for the cost incurred.

demographics or pre-existing conditions, so they cannot directly engage in risk selection.[11]

The high coverage of the system is a marker of success. The average share of the population enrolled with an insurer in either the CR or the SR between 2010 and 2018 was 89 percent (Appendix Figure B.1), with half of the covered population in each regime. Between 2010 and 2018, there were 80 active insurers on average per year in the country, 48 in the CR and 32 in the SR. The average number of insurers per municipality was 6.5 in the CR, and 4.2 in the SR (Panel (a) of Appendix Figure B.2).

Restrictions. There is extensive evidence that insurers have been systematically imposing barriers to limit utilization of medical services (Maya, 2008; Abadia et al., 2009; Yamin and Parra-Vera, 2009; Rodríguez, 2012; Arrieta-Gómez, 2018). These barriers take the form of delays or denials of authorizations for exams, medicines, surgeries or treatments prescribed by an independent health professional.[12] This is a well known problem inside the country and press articles documenting these practices abound (see Appendix Figure B.3). That insurers have an incentive to impose these barriers is largely explained by the pay-structure they face: they are mostly financed by capitation payments, which are set at a constant rate for the year, but the costs they incur are variable and directly determined by the utilization of care services. It is then in the financial interest of insurers to limit utilization of services and medications so as to reduce variable costs. Although some limitations on health care use can be efficient (counterbalancing demand-side incentives leading to ex post moral hazard) others may be inefficient and harmful (Miller et al., 2013). The financial incentive is compounded by the inability of the agencies in charge of oversight of the health care system to enforce the law, which prohibits these practices (Maya, 2008; Corte Constitucional, 2008; Rodríguez, 2012).

Tutela Writs. In response to restrictions citizens have increasingly used judicial claims available to them under Colombian law, the most salient being the tutela writ. The Colombian Political Constitution, enacted in 1991, explicitly recognized a broad set of fundamental

[11]The capitation can vary by age, gender, municipality of residence and regime, but is common across insurers conditional on these variables. The composition of the POS is also common across insurers, being defined by the Ministry of Health and changed periodically. It is designed to be comprehensive enough to cover most health care needs of the population.

[12]For example, an insurer may impose an inappropriate administrative cost, pecuniary or other, that, if not met, leads to a delay or rejection of an authorization. They might exploit 'gray areas' in the definition of the benefits package, arguing that a part of a treatment, examination or medication was not explicitly included in the benefit package, so as to delay or deny the entire treatment.

rights for all its residents, among which is the right-to-health.[13] Crucially, the Constitution not only recognized the rights, it also created a legal enforcement mechanism, called the tutela writ. This is a legal claim designed to get express access to the judicial system when rights are violated. The claim is i) costless; ii) can be filed with any judge within the local jurisdiction; iii) need only contain the basic facts so that a judge can address the case; and iv) has preferential proceeding so that the judge has to return a decision in the first instance within 10 days.

The Constitutional Court of Colombia has stated explicitly that when insurers impose barriers to access health services, even those that are not included in the mandatory benefit package, it constitutes a violation of the right-to-health.[14] An individual can file a tutela against an insurer when a service is delayed or denied and, if the judge decides in favor of the claimant, the insurer is forced to provide it. No additional penalty is imposed upon the insurer beyond the mandate to authorize the services requested. Since some denials do not lead to tutela claims and some claims are not granted, this generates an incentive for the insurer to continue to restrict access. Due to its simplicity and the fast turnaround, the tutela has progressively become a very popular instrument in the country. The number of tutelas invoking *any* fundamental right increased from 0.3 to 12.7 per 1,000 inhabitants between 1992 and 2016 (See Panel (a) of Figure 1),[15] with tutela writs invoking the right-to-health accounting for 20 to 40 percent of all claims. The growth over time has been substantial: in 1999, the total number of tutelas presented to courts invoking the right-to-health was 21,301 (0.53 per 1,000 inhabitants), but by 2016 the number reached 163,977 (3.36 per 1,000

[13]The right-to-health is broadly defined as the right to access health services in a timely and efficient manner so as to allow for the preservation, improvement and promotion of health. See Articles 44, 48 and 49 of the Colombian Political Constitution.

[14]In section 3.2.1.3 of sentence T-760 of 2008, the Court states: "*The Constitutional jurisprudence has indicated that access to a required health service, contemplated in the mandatory plans, is an autonomous fundamental right. To this extent, the denial of health services contemplated in the POS is a violation of the fundamental right-to-health, therefore, it is a clearly enforceable and justiciable provision through a Tutela writ.*". Later on, discussing services not included in the mandatory benefit package, it states that "*when a person requires a health service that is not included in the Mandatory Benefit Plan, and lacks resources to cover its cost, the entities in charge of ensuring the provision of the service must adhere to their responsibility and, consequently, ensure access to it. However, it is the State that has to bear the cost of the service, since it has the obligation to guarantee the effective enjoyment of the right.*" Corte Constitucional (2008).

[15]Unless stated otherwise, the statistics reported in this section are taken from reports produced by the Colombian Ombudsman office (Defensoría del Pueblo, 2004, 2007, 2009, 2010, 2011, 2012, 2013, 2015) and the Ministry of Health and Social Protection (Ministerio de Salud y Proteccion Social, 2018).

inhabitants), a more than sevenfold increase in a span of 17 years.[16]

Four features of the data on tutelas invoking the right-to-health are noteworthy: first, between 65.9 and 82.4 percent are made against insurers (Panel (a) of Figure B.4).[17] Second, more than half of the services requested in tutela writs were part of the compulsory benefit package (Panel (b) of Figure B.4). This shows that a large share of litigation is to access services that people should be able to get without going to court. Third, between 72.2 and 86 percent of tutelas are decided in the first instance in favor of the claimant (Panel (c) of Figure B.4). This implies that, in the eyes of the Colombian jurisprudence, citizen claims are more often than not recognized as valid. Fourth, although the share of enrollees in each regime is approximately equal, only between 12.7 and 36.7 percent of tutelas against insurers are made by individuals in the subsidized regime (Panel (d) of Figure B.4). Thus the low-income population are less likely to use this judicial instrument to protect their rights.

Complaints to the Supersalud. Tutela writs are a powerful legal instrument in Colombia, but they are not the only mechanism that people have to enforce access to restricted medical services. Individuals can also make complaints (and requests) to the National Health Superintendent (Supersalud), the agency in charge of oversight of the health care system. The Supersalud has the capacity to apply sanctions against insurers, which can range from economic penalties to the removal of the license to operate, and can mediate in disputes between insurers and enrollees regarding access to requested services. The complaints process is simple: citizens can file complaints to the Supersalud in person, on a telephone, in writing, through an online platform or using the social network accounts (Facebook and Twitter) of the agency. The Supersalud notifies the defendant, who is asked to respond to the complaint. If mediation is necessary, the Supersalud makes a decision on the case within ten days.

Between 2013 and 2017, the years for which the data on complaints is available to us, people made 1,513,247 complaints against insurers, an average of 6.92 complaints per 1,000 enrollees in the system per year. This number is 2.7 times larger than the tutela rate during

[16]The extent to which individuals were using tutela writs invoking the right-to-health during the early 2000s led the Colombian Constitutional Court to make a major pronouncement: Ruling T-760 of 2008. In the ruling, the Court provided a thorough analysis of the problems that led to the judicialization of the health care system, and ordered a set of changes including i) full clarity about the procedures and medications that were included in the benefit packages; ii) periodical audits of the insurers in order to inform users of their performance; iii) strengthening the regulatory agency overlooking the system; and iv) a revision of the capitation payments that had been unchanged for several years. The number of tutelas invoking the right-to-health declined temporarily between 2008 and 2009, but by 2015 they were back to the pre-ruling levels. We can only study the period after the 2008 pronouncement because the tutela data are only available after then.

[17]Less than 4 percent are made against providers and between 15 and 30 percent are made against local authorities, governmental agencies or other actors in the system.

the same period[18] (Panel (b) of Figure 1). Nevertheless, Tutelas and complaints are different instruments that can both be used to enforce access to care services that were restricted. Panel (a) of Figure 2 provides a scatter plot of tutela against complaint rates for every insurer × municipality × year cell.[19] The correlation between the two instruments is 0.38.

3 Data

We use five sources of administrative data for the analysis. These data are available either at the individual level for the entire population (vital statistics), as insurer × municipality × year averages (enrollees, tutelas, complaints), or as municipal × year averages (health services register). There are 1,120 municipalities, 80 insurers on average per year,[20] and, in the most comprehensive dataset, we have information for 2010–2017.

Vital Statistics. As a marker of the population-level health costs of denials, we use overall, cause-specific, age-specific, and sex-specific mortality rates at the level of insurer × municipality × year between 2010 and 2017, which we calculate from individual data on mortality. We use mortality because it is one of the few objective, well-measured health outcomes available over time (Finkelstein and McKnight, 2008), and because the vital statistics registers include information on the municipality of residence of the deceased and the insurer with which he or she was enrolled, if enrolled.[21]

Enrollment. We use data on the total number of enrollees in each insurer × municipality × year during 2010–2017. This information is publicly available on the website of the Ministry of Health and Social Protection.

Tutelas. We use the number of tutelas filed against health insurers and invoking the right-to-health. We obtained these data from the ombudsman's office after we filed a formal petition supported by right-to-information legislation. The data covers the universe of health-related tutelas put forward to Colombian courts between the years 2010 and 2016. The information is available at the insurer × municipality × year level. At this disaggregated level, the data are restricted to the number of claims made, so we cannot differentiate by the type of medical care requested or whether the claims were successful.

[18] The higher rate of complaints relative to tutelas might reflect two differences between them: first, a complaint is easier to make. The Supersalud provides many different channels for users to make complaints, while to file a tutela you need to write and deliver a document to a judge. Second, for a judge to rule a tutela in favor of the claimant the individual not only has to show that access to a health service was restricted, but that this restriction violates their right to health.

[19] Between 2013 and 2016, the years for which the data on tutelas and complaints overlap, we have 44,619 observations at insurer × municipality × year level.

[20] If an insurer is operating in both the subsidized or contributory regime we treat it as two different entities.

[21] Information about the insurer of the deceased is only available starting in 2008.

Complaints. We use the universe of complaints put forward to the National Health Superintendent during 2013–2017. We can identify which insurance company is involved in the complaint, the main (broad) reason for the complaint, and the age and sex of the claimant. We only use complaints against health insurers. We further exclude complaints about administrative procedures unrelated to the provision of health services and any general information requests. These data were also obtained by us following a right to information request.

Utilization of Health Services. We use data that contain all medical services effectively supplied in each municipality × year during 2010–2017, obtained from the Ministry of Health and Social Protection of Colombia and publicly available through an on-line platform (SISPRO). This information can be disaggregated by type of service: consultation, hospitalization, procedure, or emergency; and by the medical diagnosis associated with it, classified in ICD-10 codes. This information cannot be disaggregated by insurer.

4 Theoretical Framework

In this section we discuss the theoretical framework that underlies our empirical strategy. A formalization is presented in Appendix Section A, here we provide a simplified schematic description focusing on the salient features of the problem. The aim is to make explicit the main factors that determine the decision of insurers to restrict access to a service, the decision of enrollees to file tutelas or complaints, and to characterize what is meant by unmet demand and how it relates to the prevalence of judicial claims.

The basic structure of the problem is summarized in the decision tree shown in Figure 3. To simplify, suppose there are only two periods, all agents are enrolled with a health insurer, and tutelas (not complaints) are the only instrument available. In the initial period individuals observe their health status, and, if required, get prescribed a treatment, procedure or medication. In the context of the Colombian health care system, these prescriptions can be authorized or denied/delayed (henceforth restricted) by the health insurer. Any medical service that is prescribed but not readily authorized is enforceable through litigation, but not all individuals choose to use legal instruments. This is captured by the separation in the branches leading to end node B on the one hand, and end nodes C and D on the other. Since only a fraction of tutelas are decided in favor of the plaintiff, i) some restricted services will not be delivered (i.e. end nodes B or C); and ii) some restricted services will be delivered as a result of litigation (i.e. end node D). In the final period, new health outcomes are realized.

Let α_i be the fraction of enrollees in insurer i that file tutelas conditional on having their requested services restricted. In terms of the decision tree, α_i corresponds to the ratio

of enrollees that end up in nodes C or D, relative to the total number of enrollees in nodes B, C, or D. It follows that

$$Tutela_i = \alpha_i Restriction_i, \tag{4.1}$$

where $Restriction_i$ is the fraction of enrollees that face restrictions and $Tutela_i$ is the fraction of enrollees that file tutelas. Equation 4.1 describes the direct link between the tutela rate, which we observe, and the restriction rate, which we do not observe.

Both α_i and $Restriction_i$ are equilibrium objects that jointly determine $Tutela_i$. We show in Appendix Section A that α_i is i) increasing in the (average) improvements in health agent's get from receiving the restricted care services; and ii) increasing in the (average) probability that litigation is decided in favor of the plaintiff. On the other hand, $Restriction_i$ is i) increasing in the (average) cost of providing the services; ii) decreasing in the (average) probability that litigation is decided in favor of the plaintiff; and iii) decreasing in α_i. These are all intuitive results.

As long as receiving medical care cannot worsen health, with the demand for medical care held constant, a larger (supply-side driven) restriction rate will translate into worse average health outcomes among the population. It follows from Equation 4.1 that an increase in the restriction rate that is not demand driven will be reflected in a higher tutela rate and associated with i) lower utilization of health care services and ii) a higher mortality rate. These are the two key empirical predictions of the model that we test.

The predicted associations will be attenuated by the fact that a fraction of services are delivered after favorable judicial rulings (node D). But the fact that some services are nevertheless restricted implies there is an unmet demand for health care, defined as the share of services that are requested but not delivered. More formally,

$$\begin{aligned} unmetDemand_i &= Restriction_i - ClaimsUpheld_i \\ &= \left(\alpha_i^{-1} - \beta_i\right) Tutela_i, \end{aligned} \tag{4.2}$$

where β_i is the fraction of enrollees that receive medical care as a result of litigation, and $\left(\alpha_i^{-1} - \beta_i\right) \geq 0$.

Panel (b) of Figure 2 shows the raw correlation between the (log) tutela and (log) mortality rates using municipality $\times$ insurer $\times$ year data. Panel (c) replicates the exercise using the complaint rate. Consistent with the model predictions, there is a clear positive association. We argue that this reflects the impact of restricting access to health services on

health outcomes, as described in Equation 4.1, which results in an unrealized demand for health care, as described in Equation 4.2. We now proceed to discuss the empirical strategy used to test this hypothesis.

5 Empirical Strategy

For expositional purposes we focus here on tutela rates, but each model is also estimated using complaint rates, an alternative and independently generated measure of unmet demand. We weight each observation by the square root of the number of enrollees in the insurer × municipality × year triplet. Since each rate is an average over the enrollees in the triplet, these weights are inversely proportional to the variance of each observation.[22]

5.1 Supply Restrictions and Mortality: Three-Way Fixed Effects and IV

We first estimate a model of the form:

$$Mort_{i,m,t} = \tau Tutela_{i,m,t} + \delta X_{i,m,t} + \phi_{m,t} + \theta_{i,t} + \varphi_{i,m} + \epsilon_{i,m,t}, \tag{5.1}$$

where $i \in \{1 \dots I\}$ indexes health insurers, $m \in \{1 \dots M\}$ indexes municipalities, and $t \in \{1 \dots T\}$ indexes years, $Mort_{i,m,t}$ is the mortality rate, $Tutela_{i,m,t}$ is the rate of tutela claims, $\phi_{m,t}$ are municipality specific flexible time trends; $\theta_{i,t}$ are health insurer specific flexible time trends; and $\varphi_{i,m}$ are municipality × health insurer fixed effects. $X_{i,m,t}$ are time-varying covariates. The main identification challenge is to isolate the variation in the tutela rate induced by supply-side restrictions from demand driven changes. Once that is done, the parameter τ can be interpreted as the effect of supply side restrictions to medical care access on mortality.

In the theoretical framework, we derived that $Tutela_{i,m,t} = \alpha_{i,m,t} Restriction_{i,m,t}$, where $\alpha_{i,m,t} \in [0, 1]$ is the fraction of enrollees that file tutelas conditional on having a service restricted. The equation makes clear that we need to consider factors influencing the restriction rate and the prevalence of litigation. The likely drivers of demand for both are absorbed by the fixed effects. In particular we are able to control not only for cross-sectional heterogeneity at the insurer-municipality level $\varphi_{i,m}$ but also for shocks at the insurer level $\theta_{i,t}$ and at the municipality level $\phi_{m,t}$.[23] To address the concern that there remain relevant omitted variables that vary at the insurer-municipality-time level, we additionally introduce

[22] All results are qualitatively unchanged if we estimate an unweighed regression.

[23] These controls capture relevant baseline differences and/or shocks to health including infections, demographics, information, efficiency of provision, system-level changes, and cultures of claiming.

an instrumental variables approach. We leverage the fact that all insurers operate across multiple municipalities. The instrument rests on the premise that insurer-specific (supply side) factors determine the restriction rate in a way that is orthogonal to local demand for health care. Natural examples of insurer-specific supply side factors are cost structures determined by different service provider contracts, and different managerial practices.[24] The instrument is the insurer-specific change in the tutela rate, calculated using the information from all municipalities in which the insurer operates, except the municipality of interest m:

$$Z_{i,m,t} = \sum_{k \neq m} \nu_{i,k,t} \Delta_t Tutela_{i,k,t}, \tag{5.2}$$

where $\nu_{i,k,t} \equiv \frac{N_{i,k,t}}{\sum_{k \neq m} N_{i,k,t}}$, and $N_{i,k,t}$ is the total number of enrollees in health insurer i, municipality k, and year t. In practice, we calculate $Z_{i,m,t}$ using only municipalities outside the department in which m is located to further isolate correlated demand shocks between municipalities that are close to each other.[25] The exclusion restriction holds if the restriction rate in municipalities outside the department where m is located, conditional on the structure of fixed effects described above, does not directly affect the mortality rate in municipality m.

The time-varying covariates $X_{i,m,t}$ include the one-period lagged mortality rate which captures average health outcomes in the previous year and accounts for the possibility of mean-reversion. Although this is not common practice because most municipalities have only a few insurers (Appendix Figure B.2), and bureaucratic barriers make switching costly, we want to account for the possibility that people switch to other insurers due to restrictions, changing the risk-profile of enrollees. To do this, we include the market share of the health insurer in the municipality-year[26], and the change in that market share from the previous year.

5.2 Supply Restrictions and Utilization of Health Services: Shift-Share - IV

The mechanism through which restrictions affect health outcomes is by limiting the utilization of health care services. We test this prediction using a similar strategy. As the health services data are only available at the municipal × year level, we modify the empirical strategy, using a leave-one-out shift-share IV. The premise remains that insurer-specific

[24] In terms of the model parameters, the cost structure would be captured by the distribution of c_i^x, and efficiency would be reflected in the value of $\tilde{c}_i$.

[25] Departments are broader political entities of which there are 32 in Colombia. They are analogous to states in the United States.

[26] Market share is defined as total share of enrollees the health insurer has within the municipality.

restriction rates are correlated across municipalities because of supply side determinants that are orthogonal to local demand for health care. The estimated model takes the form:

$$y_{m,t} = \pi Tutela_{m,t} + \gamma W_{i,m,t} + \phi_m + \epsilon_{m,t}, \tag{5.3}$$

where y_{mt} is a measure of health services used, including patient consultations, hospitalizations, procedures and emergencies, all defined on a per 1,000 enrollees basis. $W_{i,m,t}$ is a vector of controls, and ϕ_m is a municipality fixed effect. To construct the instrument, note that the municipal tutela rate is a weighted average of the tutela rates of each insurer operating in the municipality, where the weights correspond to market shares:

$$Tutela_{m,t} = \sum_i \nu_{i,m,t} Tutela_{i,m,t}, \tag{5.4}$$

where $\nu_{i,m,t}$ is the municipal share of enrollees of insurer i at time t. We can decompose $Tutela_{i,m,t}$ into an aggregate, insurer level rate, and an idiosyncratic, municipal level rate:

$$Tutela_{i,m,t} = Tutela_{i,t}^{\neg m} + \mu_{i,m,t}, \tag{5.5}$$

where $Tutela_{i,t}^{\neg m}$ is the tutela rate of insurer i at time t at the national level, calculated leaving out municipality m. As before, we calculate $Tutela_{i,t}^{-m}$ using only municipalities outside the department in which m is located. The instrument is then constructed to utilize variation from the aggregate rates ($Tutela_{i,t}^{\neg m}$), and not from the local idiosyncratic rates ($\mu_{i,m,t}$):

$$B_{mt} = \sum_i \nu_{i,m,t_0} Tutela_{i,t}^{-m}, \tag{5.6}$$

where ν_{i,m,t_0} is the municipal share of enrollees of insurer i in a baseline year $t_0 = 2010$. Defining the market share in the baseline year ameliorates concerns that people might switch insurers because of restrictions, which would endogenously change the risk-pool distribution.

Like all shift-share IV's, the instrument has two components: First, how exposed a municipality is to restrictions by an insurer, given by the respective market share ν_{i,m,t_0}. Second, how many restrictions were imposed by each insurer on average outside the municipality m, proxied by the tutela rate. The recent literature on shift-share IV shows that

validity of the instrument in this set-up can be argued in terms of the exposure variables (ν_{i,m,t_0}) (Goldsmith-Pinkham et al., 2020), or the aggregate shocks ($Tutela_{i,t}^{-m}$) (Borusyak et al., 2021). In our setting, both alternatives are plausible, but, following the discussion in the previous section, we argue that the restriction rate in municipalities outside the department where m is located is unrelated to local demand for health care, which implies exogeneity of the aggregate shocks. To further account for local demand changes and allow for mean reversion, we include in the vector of controls $W_{m,t}$ one-period lags of the mortality rate and of each of the four utilization rates.

6 Results

6.1 Restrictions and Mortality

The estimates are presented in Table 1. We show results with and without the instrument, and with and without the time-varying controls. We consistently provide estimates for both tutela and complaint rates as these provide alternative sources of variation in the same underlying variable (restrictions), see Section 2. The F-statistic on the excluded instrument in the first stage is shown and, in all cases, is above 30, which indicates the instrument has sufficient power. There is no evidence of mean reversion.

An increase in supply side restrictions on access to health services leads to an increase in the mortality rate that is robust to the alternative specifications and to both measures. It is both statistically significant and meaningful. The point estimates range between 0.117 (se 0.014) and 0.271 (se 0.053). To get a sense of magnitude, this implies that a one standard deviation (SD) increase in the tutela rates –an increase of 2.55 tutelas per 1,000 enrollees– is associated with an increase of between 0.10 and 0.23 SD of the mortality rate. The standardized estimates using complaints data are strikingly simialar- a one SD increase of the complaint rate –an increase of 4.01 complaints per 1,000 enrollees– is associated with an increase of between 0.12 and 0.21 SD of the mortality rate. These results corroborate the hypothesis that restrictions have important effects on average health outcomes, even on a health outcome as extreme as the mortality rate. The actual incidence of restrictions on population health will be larger as it will tend to include unmeasured impacts on morbidities that do not translate into mortality.

By Regime. In Table 2 we repeat the analysis, dividing the sample between the contributory (CR) and subsidized (SR) regimes, which have different target populations, the SR covering the most economically vulnerable segments of the population. Earlier we noted (Panel (b) of Figure 1) that tutela and complaint rates are significantly larger in the CR

than the SR, which suggests that the CR population either faces more restrictions or has a greater tendency to use judicial instruments. It is hard to identify which, but both are plausible. We display only the specifications with the full set of controls. The F-statistic on the excluded instrument of the first stage is above 10 in all cases.

In both regimes we find evidence that supply side restrictions lead to an increase in mortality rates. The standardized effects are larger in the SR, albeit the uncertainty associated with the parameters estimates implies that the differences are not always statistically significant. Going by the magnitudes, it appears that despite the extension of universal health cover to the lower income population, restrictions on access have more severe consequences in this group. This may be because they are less likely to make use of judicial instruments to gain access, or because their baseline health is worse making it more likely that any level of restriction translates into mortality in this group.

The IV results indicate that, in the CR, a one SD increase in tutelas and complaints respectively leads to increases of 0.09 and 0.16 SD of the mortality rate. In the SR, a one SD increase in each of the two measures leads to an increase of 0.38 and 0.41 SD deviations of the mortality rate respectively.

By Cause of Death. A notable feature of judicial claims is that people request all forms of medical care supported by a wide range of medical diagnoses. Panel (a) of Figure B.5) shows that of all tutelas between 2000 and 2015, 9.3% requested appointments with health specialists, 12.8% requested exams, 15.5% surgeries, 17.1% medications, and 18.0% treatments. Moreover, a recent report by the Ministry of Heath and Social Protection (Ministerio de Salud y Proteccion Social, 2018) shows that 11.6% of tutelas are supported by medical diagnoses of neoplasia, 9.9% refer to diseases of the nervous system and 8.7% to diseases of the circulatory system (Panel (b) of Figure B.5).

Although denials appear pervasive, some health conditions require more urgent or comprehensive care, making it plausible that the impact of restrictions varies by cause. We therefore investigate cause-specific mortality rates, see Figure 4. We do not have tutelas and complaints by cause, so we use the municipality-insurer-year level rate as before. To facilitate exposition, we focus on the 15 most prevalent causes of death according to the classification of diseases defined by the Colombian National Statistical Agency, which has 71 categories, and we only report results for the IV specification using tutela rates. The Figure displays both the coefficient estimates and the standardized effects.

There are two main takeaways from this exercise. First, all of the point estimates are positive, and most are statistically significant at standard levels, indicating that restrictions increase mortality rates pretty much across the spectrum. This is consistent with the wide

range of diagnoses supporting the tutela claims. Second, some conditions exhibit a stronger response to restrictions. The standardized effect on the mortality rate from certain forms of cancers and sequelae of accidents and assaults is between 0.12 and 0.13 standard deviations, and from cerebrovascular, respiratory and cancer of pancreas is close to 0.10 standard deviations. In contrast, the impacts on mortality from stomach cancer and pulmonary disease is relatively small, lying between 0.03 and 0.04 standard deviations.

Mortality by Age and Sex The information on complaints is available at the individual level, allowing us to identify both the age group and the sex of the person affected in the complaint (who is not necessarily the person who files the claim), so we can create group specific rates. The information on vital statistics is also available at the individual level, so we can construct group specific mortality rates. Combining the two sources we can estimate the three-way fixed effects model for different subsamples.[27]

As a starting point, Panels (a) and (b) of Figure B.6 show the average complaint and mortality rates for males and females and for four age groups: 0-5, 6-24, 25-49, and 50 or above. Two patterns are evident in the data: first, complaint rates have a U-shaped pattern across age groups. The lowest rate is at age 6-24, while the largest is at age 50-plus. Second, the complaint rate tends to be larger for females. In both cases, differences could be explained either by a higher propensity to make complaints, or by a higher prevalence of restrictions. It seems plausible that older people, who are more vulnerable to chronic disease which is relatively expensive to treat, face greater restrictions.

The estimated effects are presented in Figure 5. To facilitate exposition, we report results only for the instrumental variable specification with the full set of controls. The standardized effects show larger impacts of restrictions on older people. The impact on mortality rates for individuals above 50 years of age is close to 0.16 standard deviations. This is consistent with the conditions of older people being more sensitive to access to medical care. The standardized effect is slightly larger for females than for males (0.14 vs. 0.11 SD), although we cannot reject that they are equal at standard levels of significance.

Shift-Share IV As a robustness check on the parameter estimates and to improve comparisons with the results in the next section, Table 3 shows the results of estimating Equation 5.3 using the municipal-year (rather than the municipal-insurer-year) mortality rate as the dependent variable. For comparability, we report the same specifications as those in Table 1. The F-statistic on the excluded instrument of the first stage is above 10. Results using

[27] We cannot identify age and sex in the tutela data, but impacts of tutela and complaint rates on aggregated mortality rates were similar.

this alternative IV strategy are similar to those from the three-way fixed effects approach and the corresponding IV.[28]

For the instrumental variable specification that includes the full set of controls, the point estimates show that a one SD increase of the tutela rate is associated with an increase of 0.26 SD in the mortality rate. A similar movement of the complaint rate distribution is associated with an increase of 0.18 SD of the mortality rate. The corresponding numbers for the same specification in Table 1 were 0.23 and 0.12 standard deviations. The somewhat larger impacts observed once we aggregate across insurers to get municipality level mortality rates are consistent with the aggregated data incorporating externalities. Externalities may be generated by infection transmission that influences the demand for health, or information transmission that influences litigation conditional on demand.

Overall, the results described in this section provide compelling evidence that restrictions in access to health care are imposing significant health costs on the population. The identified increases in mortality rates are robust to varying the controls (two and three way fixed effects models, time-varying controls), the estimator (with and without IV, and varying the instrument) and the measure (tutela vs complaint rates). This stability of the estimates mitigates concerns that selection drives our results. In the next section, we investigate impacts of restrictions on the likely mechanism- the utilization of health care services. This allows us not only to verify the mechanism in principle but also to identify the size of these effects, and their distribution by service type.

Competition A mechanism that could explain the prevalence of supply-side restrictions is limited competition between insurers within municipalities. As shown in Panel (a) of Appendix Figure B.2, the average number of insurers per municipality was 6.5 in the CR, and 4.2 in the SR, but these numbers conceal important heterogeneities: approximately 10 percent of municipalities have a single insurer in the SR (Panel (b) Appendix Figure B.2), while between 10 and 20 percent of all municipalities have at most two insurers in the CR (see Panel (c) of Appendix Figure B.2). Moreover, the average cross-municipality Herfindahl-Hirschman Index (HHI) of market concentration, defined in terms of the share of enrollees, is 5,251 in the contributory regime and 5,463 in the subsidized regime. This is more than twice the lower threshold used by the Justice Department of the United States to classify an industry as highly concentrated.

To investigate the role of competition, we estimated Equation 5.3 interacting the tutela or complaint rates with the HHI index, or with alternative binary classifications of munici-

[28]The IV estimates are broadly similar. The OLS estimates are more sensitive (they are smaller), consistent with the fixed effects capturing considerably less heterogeneity in the more aggregate, municipal level, specifications.

palities based on the HHI index. We find no clear evidence that impacts of restrictions on mortality emerge predominantly from low-competition markets- the interaction term is, in general, statistically insignificant at standard levels across specifications.[29] This is not entirely surprising. For example, higher market concentration can generate economies of scale or increase the power that insurers have to negotiate lower prices with providers, reducing the incentive to use denials as a way to control costs. These sorts of trade-offs have been studied in the literature on the industrial organization of health care markets (Gaynor et al., 2015; Ho and Lee, 2017), and they suggest that enforcing more competition may not be enough to induce a better provision of the services.

6.2 Restrictions and Utilization of Health Services

Estimates of Equation 5.3 are presented in Tables 4 and 5. The first table considers hospitalizations and emergencies and the second considers consultations and procedures, all per 1,000 enrollees. We show results using both the fixed-effects OLS specification and the shift-share IV and, as before, for tutela and complaint rates. The F-statistic on the excluded instrument of the first stage is above 20. The mean and standard deviation of each variable (reported in the tables) shows that, on average (per 1,000 enrollees) in a year there are 42.02 (SD 22.55) hospitalizations, 130.23 (SD) emergencies, 2,362 (SD 1,095) consultations with health professionals, and 1,770 (SD 1,127) medical procedures.

The evidence is clear that supply side restrictions on access to medical care reduce service utilization and this result is not sensitive to specification. To get a sense of magnitude, a one standard deviation increase in the tutela rate reduces the rate of hospitalizations by 0.54 SD, of emergencies by 0.46 SD, and of consultations with health professional by 0.71 SD. A one SD increase in the complaint rate reduces the rate of hospitalizations by 0.63 SD, of emergencies by 0.25 SD, of consultations by 0.45 SD and of procedures by 0.32 SD. Although we do not detect a significant impact of the tutela rate on procedures in the IV specification, once we use the complaint rate we do.

The results are robust and the effect sizes, relative to the reported means and SD are large. These results indicate that restrictions on access are not temporary, or not only around delay, but that there is a lot of outright denial. Moreover, the evidence lines up with the results for mortality, delineating the mechanism and showing that it is active across the different margins analyzed.

[29] We also find that the descriptive association of tutelas with the HHI is not statistically significant. Results are available on request.

7 Discussion

Our work is broadly related to a literature studying the health effects of expanding health insurance coverage. Available research suggests that expanding health insurance increases medical service utilization (Feldstein, 1977; Manning et al., 1987; Currie and Gruber, 1996a; Gaviria et al., 2007; Finkelstein, 2007; Card et al., 2008; Finkelstein et al., 2012; Weathers and Stegman, 2012; Baicker et al., 2013; Miller et al., 2013; Gruber et al., 2014; Goodman-Bacon, 2018), but with mixed impacts on health outcomes. This highlights the importance of studying outcomes alongside access. Health outcomes that have been shown to improve with expansion of health insurance coverage in the United States include child mortality, birth weight (Currie and Gruber, 1996a,b; Camacho and Conover, 2013; Chou et al., 2014; Gruber et al., 2014; Goodman-Bacon, 2018), self-reported mental and physical health (Gaviria et al., 2007; Finkelstein et al., 2012; Weathers and Stegman, 2012; Baicker et al., 2013), and all-cause mortality, especially among older adults and minorities (Sommers et al., 2012). However, some studies find no effect on measured physical health outcomes (Baicker et al., 2013) including mortality (Finkelstein and McKnight, 2008; Weathers and Stegman, 2012). We contribute to this literature by identifying the extent to which slippage between coverage and access can hamper the realization of the potential benefits of coverage.

Although the incentive structure that allows insurers to restrict access to health care is particular to the Colombian system, analogous problems arise whenever health insurance coverage does not guarantee effective access to medical care. For example, there is evidence that Medicaid recipients in the United States face barriers to access ambulatory care because of low rates of physician participation and shortages of primary care providers (MAG, 1994; Currie et al., 1995; Government Accountability Office, 2011; Rosenbaum, 2014; Sonchak, 2015). Similarly, physicians treating Medicaid patients experience greater difficulty referring them to speciality care relative to the privately insured (Government Accountability Office, 2011; Felland et al., 2013). These are examples of supply-side constraints that can lead to an under-provision of medical services even when they are part of the benefit plan. However, neither for the US, nor for any other country, are we aware of estimates of the population-level health costs of denying medical care, and we make a contribution on this front.

We also contribute to a literature studying how the use of litigation by patients can affect the provision of health care. Economists have focused upon the effects of liabilities for malpractice on the quality and cost of medical care (Kessler and McClellan, 1996; Danzon, 2000; Kessler and McClellan, 2002; Currie and MacLeod, 2008; Frakes, 2013; Frakes and Jena, 2016). Results from this literature indicate that the fear of tort liabilities encourages physicians to engage in 'defensive medicine', administering treatments that potentially have

little medical impact but that vastly increase the cost of providing health care. Some studies show that reforms that reduce tort liabilities reduce both the number of lawsuits filed and medical expenditures (Kessler and McClellan, 1996; CBO and Congressional Budget Office of the United States, 2004), with no effect on mortality or the quality of care (Kessler and McClellan, 1996; Frakes and Jena, 2016), although there is also evidence that, in specific situations like childbirth, they can increase medical complications (Currie and MacLeod, 2008).

We depart from this literature by focusing upon a context in which litigation is used not to deter medical malpractice, but to enforce access to medical care, supported by constitutionalization of health as a human right. The idea of institutionalizing rights to health has legal foundations in the Universal Declaration of Human Rights of 1948,[30] and, more explicitly, in the International Covenant on Economic, Social and Cultural Rights (ICESCR) of 1966, a legally binding treaty signed and later ratified by 169 countries.[31] The right to health is explicitly recognized –to a greater or lesser extent– in the written constitutions of more than half of all UN member countries (Kinney and Clark, 2004; Backman et al., 2008; Hamel et al., 2015).

In some cases, this notional commitment is accompanied by constitutional provisions to guarantee its enforcement, including the possibility of legal recourse. In Colombia, as discussed, use of an enforcement mechanism has become widespread, leading to what is referred as 'judicialization of health care' (Yamin and Parra-Vera, 2009, 2010; Parra-Vera and Yamin, 2013). Although it seems less widespread in these countries, a similar provision has been made available in Argentina (Bergallo, 2011), Brazil (Gauri, 2004; Biehl et al., 2009, 2012), Costa Rica (Wilson, 2011; Norheim et al., 2014), and South Africa (Cooper, 2011).[32] Most of the literature studying judicial enforcement of the right to health is qualitative, often focusing on individual cases or a small group of cases. We provide what appears to be the first quantitative analysis, using unique data on judicial claims made to enforce access to medical

[30] Article 25 of the Declaration states that: "Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care..."

[31] The ICESCR declares in its Article 12 that "the States parties... recognize the right of everyone to the enjoyment of the highest attainable standard of physical and mental health." It is not, according to the standard interpretation, a right to be healthy, but a set of entitlements that include "the right to a system of health protection which provides equality of opportunity for people to enjoy the highest attainable level of health,... taking into account both the individual's biological and socio-economic preconditions and a State's available resources." (Alston, 2010, general comment 14, N. 8-9).

[32] The United States has not ratified the ICESCR, and its constitution makes no reference to health rights, but it has nevertheless experienced right-to-health litigation in specific segments of the population. In the 1976 landmark decision *Estelle vs. Gamble*, the Supreme Court of the United States established the right to medical care for prisoners after a Texas inmate, J. W. Gamble, sued the State Department of Corrections for lack of adequate medical treatment. The failure of correctional officials to honor these rights has resulted in protracted litigation involving hundreds of cases (Rold, 2008; Hamel et al., 2015).

care. Our study is particularly relevant in the context of sharp expansions in health care provision across countries, enhanced by a global push for universal health coverage. Where the incentives of providers are not necessarily aligned with constitutional commitments, questions of effective delivery and accountability become increasingly important.

8 Conclusion

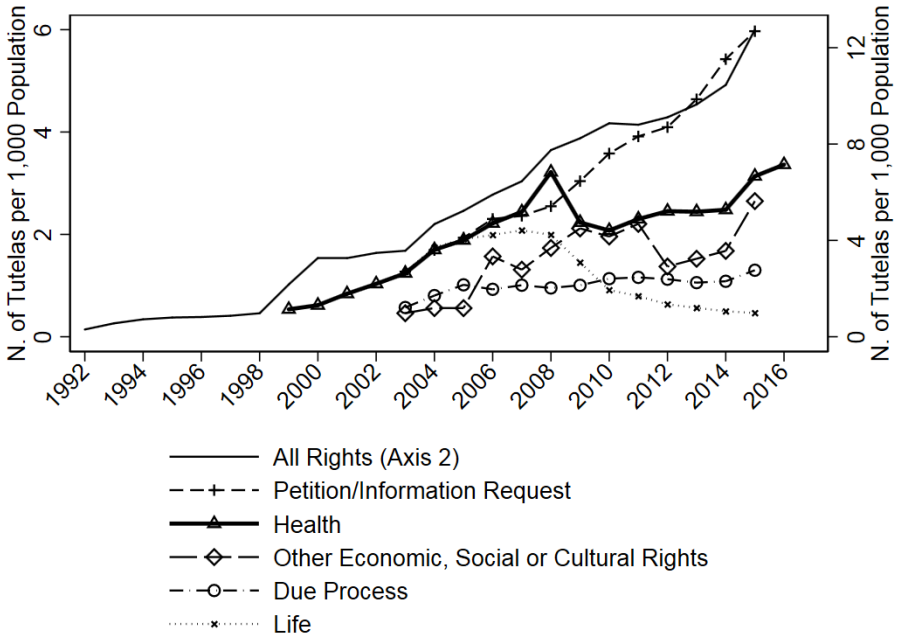
Our results provide evidence that supply-side restrictions on access to state-insured health care services manifest as very substantial reductions in medical consultations, hospitalizations, emergency care and (with more uncertainty) medical procedures. In line with this, we identify increases in mortality rates. The estimates are robust to varying the controls, the estimator and the measure of restrictions (tutela vs complaint rates). They are pervasive across social groups defined by economic and demographic characteristics, and across many different causes of death. The standardized effects are larger in the lower income population, among older people and (weakly) among women. Our results are relevant to newly emerging health care insurance systems in developing countries. Identifying an effective design for these new regimes is important both because public budgets are tight and because health and economic status are more intricately tied among the poor. Our results are also more widely relevant as richer countries are increasingly having to grapple with rising demand that current provisions are, in general, unable to satisfy.

Colombia has made a constitutional commitment to the right to health and acted to implement universal health care coverage ahead of many other countries. It has in place a well-functioning, widely accessed institution designed to protect citizen rights. Yet, the organization of delivery of health care services, which operates through insurance companies (that are mostly but not entirely privately owned) suffers design deficiencies. There is too much regulation insofar as insurers are not allowed to set prices or select enrollees and, at the same time, too little regulation insofar as insurers are not penalized when complaints against them are received and upheld. Our estimates provide a lower bound on the population level health gains that can be achieved through re-design of the system.

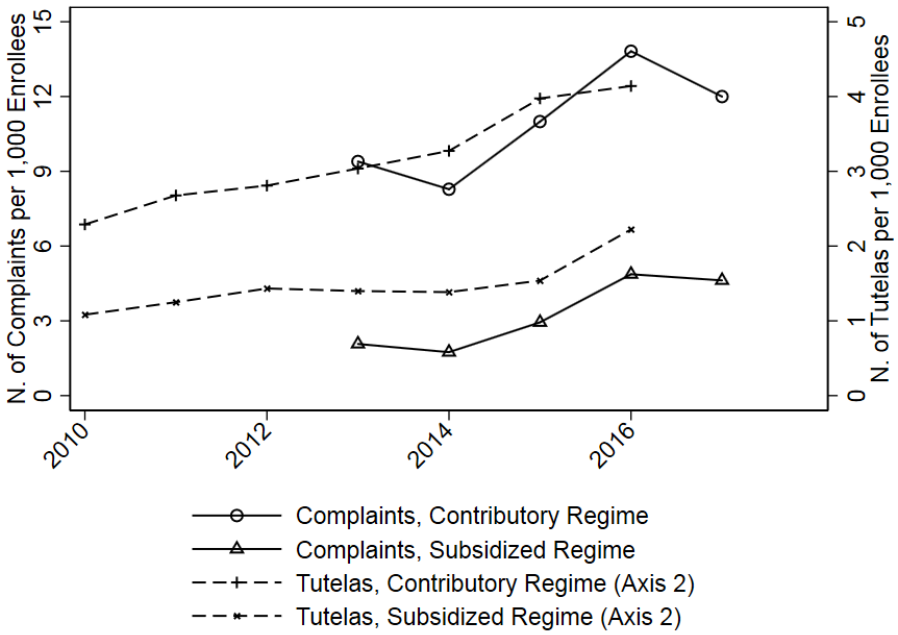
Tables and Figures

Figure 1: Prevalence of Tutelas and Complaints

(a) Tutelas per 1,000 Inhabitants by Fundamental Right Being Invoked



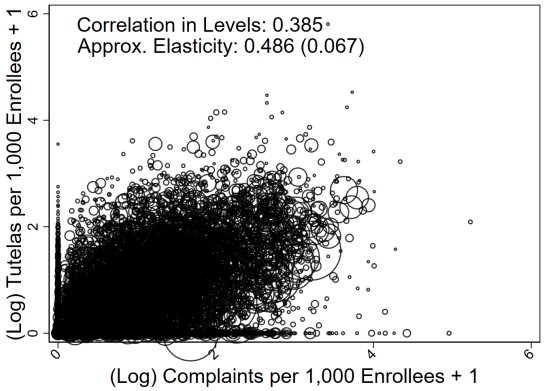
(b) Number of Complaints and Tutelas per 1,000 Enrollees by Regime



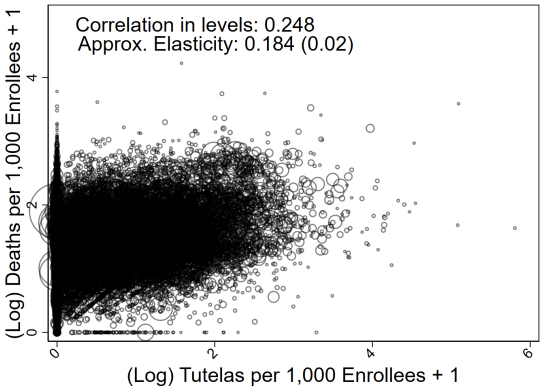
Notes: The numbers reported in in the right-to-health series of Panel (a) include complaints made against other actors in the system different from insurers (i.e. IPSs, Secretary of Health and Social Protection, local authorities). See footnote 16 for a discussion on the peak of health related tutelas in 2008. The numbers in Panel (b) are based on Defensoría del Pueblo (2004, 2007, 2009, 2010, 2011, 2012, 2013, 2015), and include only complaints and tutelas made against insurers.

Figure 2: Correlations between Mortality, Tutela and Complaint Rates within Municipality, Insurer and Year

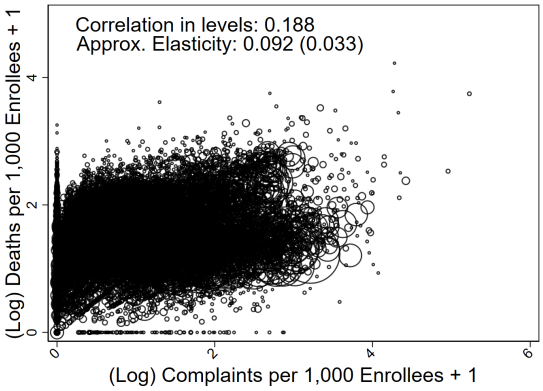
(a) Tutela and Complaint Rate



(b) Mortality and Tutela Rate



(c) Mortality and Complaint Rate



Notes: The scatter plot in Panel (a) shows the relation between the (log) of the tutela and (log) complaint rates (plus one) defined at the municipality × insurer × year level. The scatter plot in Panel (b) shows the relation between the (log) of the tutela and (log) mortality rates (plus one) defined at the same level. Panel (c) repeats the exercise but using the complaint rate. The size of each circle is proportional to the total share of enrollees. The approximate elasticity is measured by the slope a regression of the respective variables. The sources of data to construct the figure are described in Section 3.

Figure 3: Decision Tree

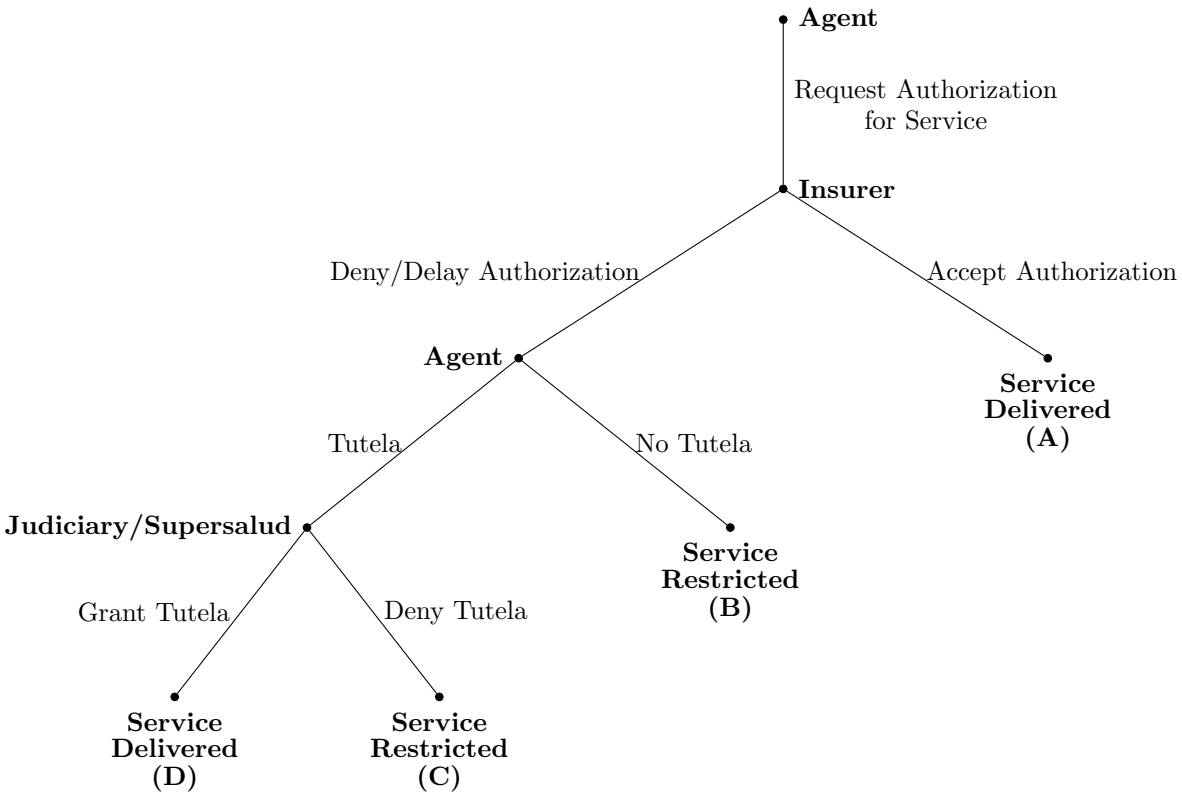
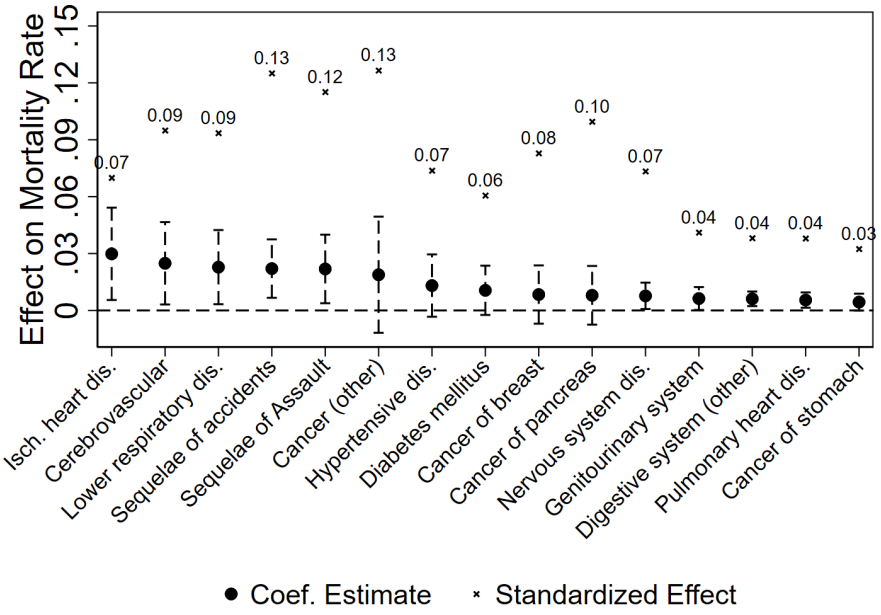


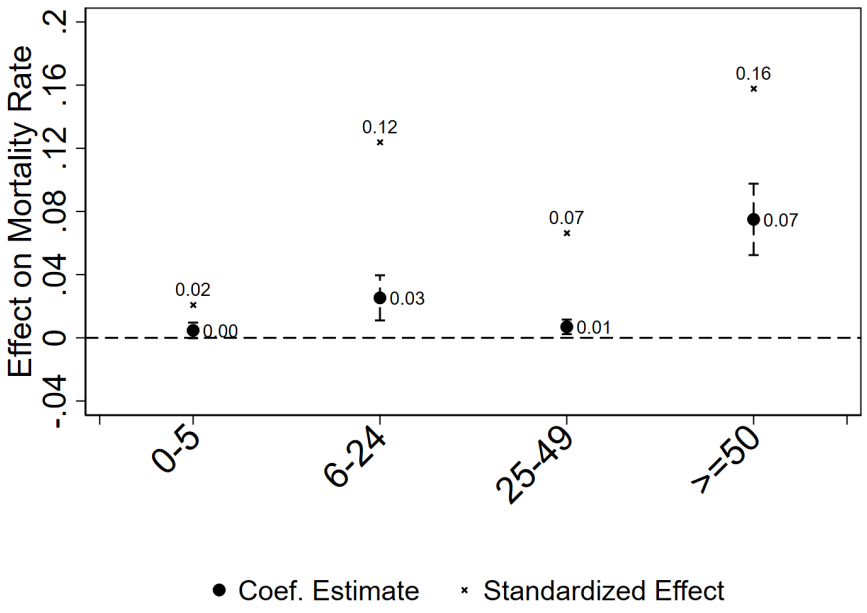
Figure 4: Effect of Restrictions on Access to Medical Care on Mortality Rates by Cause of Death



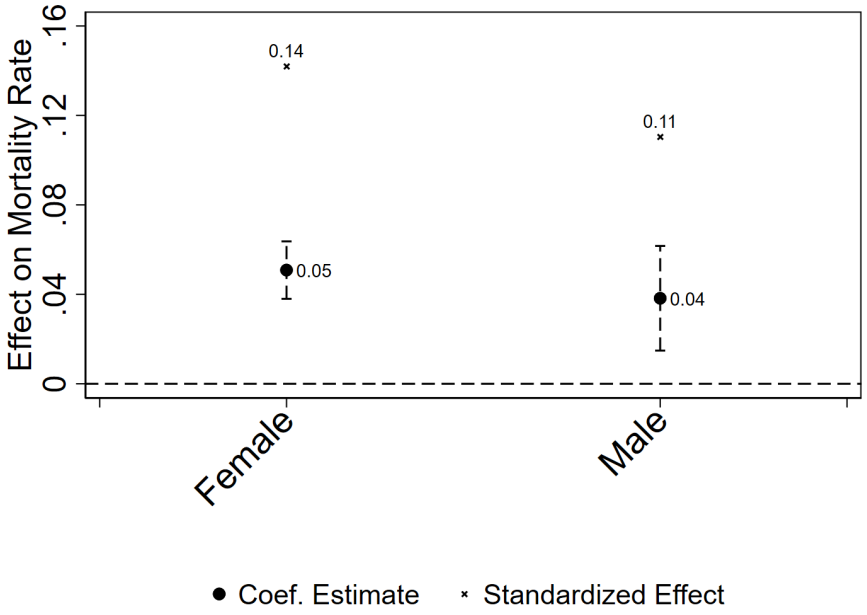
Notes: The Figure shows the estimated effect of supply side restrictions to access health care on cause-specific mortality rates. Restrictions are measured using the municipality × insurer × year tutela rate. The estimates correspond to the instrumental variable specification discussed in Section 5.1. The standardized effect is calculated as the effect of a one standard deviation increase of the tutela rate on the cause-specific mortality rate, divided by its standard deviation.

Figure 5: Effect of Restrictions on Access to Medical Care on Mortality Rates by Age and Sex

(a) By Age



(b) By Sex



Notes: The Figure shows the estimated effect of supply side restrictions to access to health care on age-specific (Panel (a)) and sex-specific (Panel (b)) mortality rates. The model specification uses complaint rates since the tutela data can not be disaggregated by age nor sex. The estimates correspond to the instrumental variable specification discussed in Section 5.1. The standardized effect is calculated as the effect of a one standard deviation increase of the complaint rate on the age or sex-specific mortality rate, divided by its standard deviation.

Table 1: Effect of Restrictions on Access to Medical Care on Mortality Rates: Three-Way Fixed Effects IV Model

	Dep. Var.: Mortality Rate (per 1,000 Enrollees) Subsidized and Contributory Regime							
	OLS (a)	IV (b)	OLS (c)	IV (d)	OLS (e)	IV (f)	OLS (g)	IV (h)
Tutela Rate	0.206***	0.206***	0.117***	0.271***				
	(0.041)	(0.046)	(0.014)	(0.053)				
Complaint Rate					0.179***	0.090**	0.100***	0.093**
					(0.044)	(0.029)	(0.013)	(0.037)
Observations	72,976	48,754	51,550	48,502	50,876	29,911	40,669	29,700
Summary Stats.								
Av. Mortality Rate	3.83	3.82	3.90	3.81	4.10	4.06	4.10	4.06
	(3.37)	(3.13)	(3.10)	(3.12)	(3.36)	(3.11)	(3.06)	(3.08)
Av. Tutela Rate	1.96	2.09	2.05	2.09				
	(2.55)	(2.63)	(2.63)	(2.63)				
Av. Complaint Rate					3.93	4.13	3.91	4.13
					(4.01)	(4.09)	(4.02)	(4.09)
Fixed Effecs								
Municipality × Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Insurer × Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality × Insurer	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls: lagged mortality, Insurer Market share, and Δ_t Insurer Market Share	No	No	Yes	Yes	No	No	Yes	Yes
First Stage								
F stat. First Stage		31.91		54.49		88.44		176.05
Standardized Effect	0.16	0.17	0.10	0.23	0.21	0.12	0.13	0.12

*** 1 percent ** 5 percent * 10 percent.

Notes: The Table shows the estimated effect of supply side restrictions to access health care on mortality rates. The IV estimates correspond to the three-way fixed effects IV model discussed in Section 5.1. Each observation is weighted by the square root of the the number of enrollees in the insurer × municipality × year triad. All rates are defined per 1,000 enrollees. We report cluster-robust standard errors, clustered by municipality. The standardized effect is calculated as the effect of a one standard deviation increase of the tutela/complaint rate on the mortality rate, divided by its standard deviation.

Table 2: Effect of Restrictions on Access to Medical Care on Mortality Rates by Regime: Three-Way Fixed Effects IV Model

	Contributory Regime				Subsidized Regime			
	OLS (a)	IV (b)	OLS (c)	IV (d)	OLS (e)	IV (f)	OLS (g)	IV (h)
Tutela Rate	0.102***	0.216**			0.231***	0.432***		
	(0.015)	(0.073)			(0.033)	(0.074)		
Complaint Rate			0.107***	0.075*			0.074**	0.326**
			(0.015)	(0.038)			(0.024)	(0.126)
Observations	31,853	31,853	22,144	16,173	18,375	14,928	17,518	12,343
Fixed Effecs								
Municipality × Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Insurer × Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality × Insurer	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls: lagged mortality, Insurer Market share, and Δ_t Insurer Market Share	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First Stage								
F stat. First Stage		46.66		199.87		11.89		10.57
Standardized Effect	0.08	0.16	0.12	0.09	0.22	0.41	0.08	0.38

*** 1 percent ** 5 percent * 10 percent.

Notes: The Table shows the estimated effect of supply side restrictions to access health care on mortality rates, conditional on regime. The IV estimates correspond to the three-way fixed effects IV model discussed in Section 5.1. Each observation is weighted by the square root of the the number of enrollees in the insurer × municipality × year triad. All rates are defined per 1,000 enrollees. We report cluster-robust standard errors, clustered by municipality. The standardized effect is calculated as the effect of a one standard deviation increase of the tutela/complaint rate on the mortality rate, divided by its standard deviation. Summary statistics on the mortality, tutela and complaint rates are shown at the bottom of Table 1.

Table 3: Effect of Restrictions on Access to Medical Care on Mortality Rates: Shift-Share Approach

	Dep. Var.: Mortality Rate (per 1,000 Enrollees) Subsidized and Contributory Regime							
	OLS (a)	IV (b)	OLS (c)	IV (d)	OLS (e)	IV (f)	OLS (g)	IV (h)
Tutela Rate	0.228***	0.297***	0.095***	0.181**				
	(0.023)	(0.105)	(0.015)	(0.059)				
Complaint Rate					0.053***	0.098**	0.053***	0.082**
					(0.011)	(0.032)	(0.013)	(0.026)
Observations	7,721	7,721	6,574	6,574	5,510	5,510	5,499	5,499
Fixed Effecs								
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls: lagged utilization and lagged mortality	No	No	Yes	Yes	No	No	Yes	Yes
First Stage								
F stat. First Stage		10.93		48.02		15.38		20.77
Standardized Effect	0.31	0.38	0.14	0.26	0.11	0.21	0.11	0.18

*** 1 percent ** 5 percent * 10 percent.

Notes: The Table shows the estimated effect of supply side restrictions to access health care on mortality rates. The estimates in this table use municipality-year data that aggregates over insurers. For this reason we replace the instrument in Tables 1 - 2 with the shift share IV discussed in Section 5.2. All rates are defined per 1,000 enrollees. Each observation is weighted by the square root of the the number of enrollees in the municipality × year. We report cluster-robust standard errors, clustered by municipality. The standardized effect is calculated as the effect of a one standard deviation increase of the tutela/complaint rate on the mortality rate, divided by its standard deviation. Summary statistics on the mortality, tutela and complaint rates are shown at the bottom of Table 1.

Table 4: Effect of Restrictions on Access to Medical Care on Utilization of Medical Services I: Shift-Share Approach

	Hospitalizations per 1,000 Enrollees				Emergencies per 1,000 Enrollees			
	OLS (a)	IV (b)	OLS (c)	IV (d)	OLS (e)	IV (f)	OLS (g)	IV (h)
Tutela Rate	-4.19***	-20.65***			-12.26***	-77.63**		
	(0.95)	(6.20)			(3.15)	(26.97)		
Complaint Rate			-4.32***	-10.92**			-16.07***	-17.64**
			(0.53)	(3.38)			(2.57)	(7.41)
Observations	6,570	6,570	5,460	5,460	6,574	6,574	5,493	5,493
Av. Dependent Variable	42.02	42.02	34.82	34.82	130.23	130.23	120.80	120.80
	(22.57)	(22.57)	(25.13)	(25.13)	(100.54)	(100.54)	(104.00)	(104.00)
Fixed Effecs								
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls: lagged utilization and lagged mortality	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First Stage								
F stat. First Stage		48.02		20.77		48.02		20.80
Standardized Effect	-0.11	-0.54	-0.25	-0.63	-0.07	-0.46	-0.23	-0.25

*** 1 percent ** 5 percent * 10 percent.

Notes: Table shows the estimated effect of supply side restrictions to access health care on utilization of care services. The estimates in this table use municipality-year data that aggregates over insurers. For this reason we replace the instrument in Tables 1 - 2 with the shift share IV discussed in Section 5.2. All rates are defined per 1,000 enrollees. Each observation is weighted by the square root of the the number of enrollees in the municipality × year pair. We report cluster-robust standard errors, clustered by municipality and year. The standardized effect is calculated as the effect of a one standard deviation increase of the tutela/complaint rate on the utilization rate, divided by its standard deviation. Summary statistics on the tutela and complaint rates are shown at the bottom of Table 1.

Table 5: Effect of Restrictions on Access to Medical Care on Utilization of Medical Services II: Shift-Share Approach

	Consultations per 1,000 Enrollees				Procedures per 1,000 Enrollees			
	OLS (a)	IV (b)	OLS (c)	IV (d)	OLS (e)	IV (f)	OLS (g)	IV (h)
Tutela Rate	-202.99***	-1318.45**			-138.26**	-27.66		
	(42.53)	(468.23)			(44.76)	(75.03)		
Complaint Rate			-215.81***	-364.09***			-57.26**	-255.11**
			(24.66)	(104.67)			(24.72)	(112.41)
Observations	6,574	6,574	5,499	5,499	6,573	6,573	5,493	5,493
Av. Dependent Variable	2326.44	2326.44	1984.81	1984.81	1770.12	1770.12	1572.65	1572.65
	(1095.44)	(1095.44)	(1173.62)	(1173.62)	(1127.84)	(1127.84)	(1161.11)	(1161.11)
Fixed Effecs								
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls: lagged utilization and lagged mortality	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First Stage								
F stat. First Stage		48.02		20.77		47.91		20.80
Standardized Effect	-0.11	-0.71	-0.27	-0.45	-0.07	-0.01	-0.07	-0.32

*** 1 percent ** 5 percent * 10 percent.

Notes: Table shows the estimated effect of supply side restrictions to access health care on utilization of care services. The estimates in this table use municipality-year data that aggregates over insurers. For this reason we replace the instrument in Tables 1 - 2 with the shift share IV discussed in Section 5.2. All rates are defined per 1,000 enrollees. Each observation is weighted by the square root of the the number of enrollees in the municipality × year pair. We report cluster-robust standard errors, clustered by municipality and year. The standardized effect is calculated as the effect of a one standard deviation increase of the tutela/complaint rate on the utilization rate, divided by its standard deviation.

References

Abadia, C. E., Oviedo, D. G., Abadía, E., and Oviedo, D. G. (2009). Bureaucratic Itineraries in Colombia. A theoretical and methodological tool to assess managed-care health care systems. *Social Science and Medicine*, 68(6):1153–1160.

Alston, P. (2010). The General Comments of the UN Committee on Economic, Social and Cultural Rights. Doc. E/C.12/2000/21.

Arrieta-Gómez, A. I. (2018). Realizing the fundamental right to health through litigation: The Colombian case. *Health and Human Rights*, 20(1):133–145.

Backman, G., Hunt, P., Khosla, R., Jaramillo-Strouss, C., Fikre, B. M., Rumble, C., Pevalin, D., Páez, D. A., Pineda, M. A., Frisancho, A., others, Tarco, D., Motlagh, M., Farcasanu, D., and Vladescu, C. (2008). Health systems and the right to health: an assessment of 194 countries. *The Lancet*, 372(9655):2047–2085.

Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., Schneider, E. C., Wright, B. J., Zaslavsky, A. M., and Finkelstein, A. N. (2013). The Oregon Experiment — Effects of Medicaid on Clinical Outcomes. *New England Journal of Medicine*, 368(18):1713–1722.

Banerjee, R., Ziegenfuss, J. Y., and Shah, N. D. (2010). Impact of discontinuity in health insurance on resource utilization. *BMC Health Services Research*, 10(1):195.

Bergallo, P. (2011). Argentina Courts and the Right to Health: Achieving Fairness Despite "Routinization" in Individual Coverage Cases? In Yamin, A. E. and Gloppen, S., editors, *Litigating Health Rights: Can Courts Bring More Justice to Health?*, chapter 3, pages 43–75. Harvard University Press.

Biehl, J., Amon, J. J., Socal, M. P., and Petryna, A. (2012). Between the court and the clinic: Lawsuits for medicines and the right to health in Brazil. *Health and Human Rights*, 14(1):E36–52.

Biehl, J., Petryna, A., Gertner, A., Amon, J. J., and Picon, P. D. (2009). Judicialisation of the right to health in Brazil. *The Lancet*, 373(9682):2182–2184.

Bloom, D. E., Khoury, A., and Subbaraman, R. (2018). The promise and peril of universal health care. *Science*, 361(6404):eaat9644.

Borusyak, K., Hull, P., and Jaravel, X. (2021). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies*, 88(24997).

Camacho, A. and Conover, E. (2013). Effects of subsidized health insurance on newborn health in a developing country. *Economic Development and Cultural Change*, 61(3):633–658.

Card, D., Dobkin, C., and Maestas, N. (2008). The impact of nearly universal insurance coverage on health care utilization: Evidence from medicare. *American Economic Review*, 98(5):2242–2258.

CBO and Congressional Budget Office of the United States (2004). The Effects of Tort Reform: Evidence from the States. Technical Report June, Congressional Budget Office.

Chou, S.-Y. Y., Grossman, M., and Liu, J.-T. T. (2014). The impact of National Health Insurance on birth outcomes: A natural experiment in Taiwan. *Journal of Development Economics*, 111:75–91.

Cooper, C. (2011). Health Rights Litigation: Cautious Constitutionalism. In Yamin, A. E. and Gloppen, S., editors, *Litigating Health Rights: Can Courts Bring more Justice to Health?*, chapter 8, pages 190–229. Harvard University Press.

Corte Constitucional (2008). SENTENCIA N° T-760 de 2008. July 31, 2008. Magistrado Ponente: Manuel José Cepeda.

Currie, J. and Gruber, J. (1996a). Health insurance eligibility, utilization of medical care, and child health. *Quarterly Journal of Economics*, 111(2):431–466.

Currie, J. and Gruber, J. (1996b). Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women. *Journal of Political Economy*, 104(6):1263–1296.

Currie, J., Gruber, J., and Fischer, M. (1995). Physician payments and infant mortality: evidence from Medicaid fee policy. *The American economic review*, 85(2):106–111.

Currie, J. and MacLeod, W. B. (2008). First do no harm? Tort reform and birth outcomes. *Quarterly Journal of Economics*, 123(2):795–830.

Danzon, P. M. (2000). Liability for medical malpractice. In *Handbook of Health Economics*, number PART B in Handbook of Health Economics, chapter 26, pages 1339–1404. Elsevier.

Defensoría del Pueblo (2004). La Tutela y el Derecho a la Salud: 1999-2003. Reporte Institucional.

Defensoría del Pueblo (2007). La tutela y el derecho a la salud periodo 2003–2005. Reporte Institucional.

Defensoría del Pueblo (2009). La Tutela y el Derecho a la Salud: 2006-2008. Reporte Institucional.

Defensoría del Pueblo (2010). La Tutela y el Derecho a la Salud 2009. Reporte Institucional.

Defensoría del Pueblo (2011). La Tutela y el Derecho a la Salud 2010. Reporte Institucional.

Defensoría del Pueblo (2012). La Tutela y el Derecho a la Salud 2011. Reporte Institucional.

Defensoría del Pueblo (2013). La Tutela y el Derecho a la Salud 2012. Reporte Institucional.

Defensoría del Pueblo (2015). La Tutela y los Derechos a la Salud y a la Seguridad Social 2014. Reporte Institucional.

Enthoven, A. C. (1978). Consumer-Choice Health Plan. *New England Journal of Medicine*, 298(12):650–658.

Feldstein, M. (1977). Quality Change and the Demand for Hospital Care. *Econometrica*, 45(7):1681–1702.

Felland, L. E., Lechner, A., and Sommers, A. (2013). Access to Specialty Care for Medicaid Patients: Policy Issues and Options. Technical report, The Commonwealth Fund.

Finkelstein, A. (2007). The aggregate effects of health insurance: Evidence from the introduction of medicare. *Quarterly Journal of Economics*, 122(1):1–37.

Finkelstein, A. and McKnight, R. (2008). What did Medicare do? The initial impact of Medicare on mortality and out of pocket medical spending. *Journal of Public Economics*, 92(7):1644–1668.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., and Group, O. H. S. (2012). The oregon health insurance experiment: Evidence from the first year. *Quarterly Journal of Economics*, 127(3):1057–1106.

Frakes, M. (2013). The impact of medical liability standards on regional variations in physician behavior: Evidence from the adoption of national-standard rules. *American Economic Review*, 103(1):257–276.

Frakes, M. and Jena, A. B. (2016). Does medical malpractice law improve health care quality? *Journal of Public Economics*, 143:142–158.

Gauri, V. (2004). Social rights and economics: Claims to health care and education in developing countries. *World Development*, 32(3):465–477.

Gaviria, A., Medina, C., Mejia, C., Mejía, C., McKenzie, D., and Soares, R. R. (2007). Assessing Health Reform in Colombia: From Theory to Practice. *Economía*, 7(1):29–72.

Gaynor, M., Ho, K., and Town, R. J. (2015). The Industrial Organization of Health-Care Markets. *Journal of Economic Literature*, 53(2):235–284.

Glassman, A. L., Escobar, M.-L., Giuffrida, A., Giedion, U., and Giedion, A. (2009). From few to many : ten years of health insurance expansion in Colombia. Technical report, Inter-American Development Bank and The Brookings Institution.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*.

Goodman-Bacon, A. (2018). Public insurance and mortality: Evidence from medicaid implementation. *Journal of Political Economy*, 126(1):216–262.

Government Accountability Office (2011). Medicaid and CHIP: Most Physicians Serve Covered Children but Have Difficult Referring Them for Specialty Care. Technical Report June, Government Accountability Office.

Gruber, J., Hendren, N., and Townsend, R. M. (2014). The great equalizer: Health care access and infant mortality in Thailand. *American Economic Journal: Applied Economics*, 6(1 A):91–107.

Hamel, M. B., Ruger, J. P., Ruger, T. W., and Annas, G. J. (2015). The Elusive Right to Health Care under U.S. Law. *New England Journal of Medicine*, 372(26):2558–2563.

Ho, K. and Lee, R. S. (2017). Insurer Competition in Health Care Markets. *Econometrica*, 85(2):379–417.

Kessler, D. and McClellan, M. (1996). Do doctors practice defensive medicine? *Quarterly Journal of Economics*, 111(2):352–390.

Kessler, D. and McClellan, M. (2002). Malpractice law and health care reform: Optimal liability policy in an era of managed care. *Journal of Public Economics*, 84(2):175–197.

Kinney, E. D. and Clark, B. A. (2004). Provisions for health and health care in the constitutions of the countries of the world. *Cornell International Law Journal*, 37(2):285–355.

Kruk, M. E., Gage, A. D., Joseph, N. T., Danaei, G., Garc\'\ia-Saisó, S., Salomon, J. A., García-Saisó, S., and Salomon, J. A. (2018). Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. *The Lancet*, 392(10160):2203–2212.

Londoño, J.-L. L. and Frenk, J. (1997). Structured pluralism: Towards an innovative model for health system reform in Latin America. *Health Policy*, 41(1):1–36.

MAG (1994). Access of Medicaid Recipients to Outpatient Care. *New England Journal of Medicine*, 330(20):1426–1430.

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., Leibowitz, A., and Marquis, M. S. (1987). Health insurance and the demand for medical care: evidence from a randomized experiment. *American Economic Review*, 77(3):251–277.

Maya, E. (2008). El derecho a la salud en la perspectiva de los derechos humanos y del sistema de inspección, vigilancia y control de quejas en materia de salud *. Technical Report 4, Procuraduría General de la Nación.

Miller, G., Pinto, D., and Vera-Hernández, M. (2013). Risk protection, service use, and health outcomes under Colombia's health insurance program for the poor. *American Economic Journal: Applied Economics*, 5(4):61–91.

Ministerio de Salud y Proteccion Social (2018). Informe de Cumplimiento de la Vigencia 2017 Orden 30 de la Sentencia T-760 DE 2008. Reporte Institucional Ministerio de Salud y Protección Social.

Mohanan, M., Bauhoff, S., La Forgia, G., Babiarz, K. S., Singh, K., and Miller, G. (2014). Effect of Chiranjeevi Yojana on institutional deliveries and neonatal and maternal outcomes in Gujarat, India: a difference-in-differences analysis. *Bulletin of the World Health Organization*, 92(3):187–194.

Norheim, O. F., Wilson, B. M., and Wilson, B. M. (2014). Health rights litigation and access to medicines: Priority classification of successful cases from Costa Rica's constitutional chamber of the supreme court. *Health and Human Rights*, 16(2):47–61.

Parra-Vera, O. and Yamin, A. (2013). La sentencia T-760 de 2008, su implementación e impacto: retos y oportunidades para la justicia dialógica. *Clérico, L., Aldao, M., Ronconi, L(coords.), Tratado de derecho a la salud.*, pages 2591–2649.

Powell-Jackson, T., Mazumdar, S., and Mills, A. (2015). Financial incentives in health: New evidence from India's Janani Suraksha Yojana. *Journal of Health Economics*, 43:154–169.

Rodin, J. and de Ferranti, D. (2012). Universal health coverage: The third global health transition? *The Lancet*, 380(9845):861–862.

Rodríguez, C. (2012). La Judicialización de la Salud: Síntomas, Diagnóstico y Prescripciones. In Bernal, O. and Gutierrez, C., editors, *La salud en Colombia: logros retos y recomendaciones*, chapter 12, pages 507–560. Ediciones Uniandes-Universidad de los Andes.

Rold, W. J. (2008). Thirty years after Estelle v. Gamble: A legal retrospective. *Journal of Correctional Health Care*, 14(1):11–20.

Rosenbaum, S. (2014). Medicaid Payments and Access to Care. *New England Journal of Medicine*, 371(25):2345–2347.

Sommers, B. D., Baicker, K., and Epstein, A. M. (2012). Mortality and Access to Care among Adults after State Medicaid Expansions. *New England Journal of Medicine*, 367(11):1025–1034.

Sonchak, L. (2015). Medicaid reimbursement, prenatal care and infant health. *Journal of Health Economics*, 44:10–24.

Weathers, R. R. and Stegman, M. (2012). The effect of expanding access to health insurance on the health and mortality of Social Security Disability Insurance beneficiaries. *Journal of Health Economics*, 31(6):863–875.

Wilson, B. (2011). Health Rights Litigation: Causes and Consequences. In Yamin, A. E. and Gloppen, S., editors, *Litigating Health Rights: Can Courts Bring more Justice to Health?*, chapter 6, pages 132–154. Harvard University Press.

Yamin, A. E. and Parra-Vera, O. (2009). How do courts set health policy? the case of the colombian constitutional court. *PLoS Medicine*, 6(2):0147–0150.

Yamin, A. E. and Parra-Vera, O. (2010). Judicial Protection of the Right to Health in Colombia: From Social Demands to Individual Claims to Public Debates. *Hastings International and Comparative Law Review*, 33(2):431.

A Restrictions on Access to Health Care Services and Litigation: A Simple Model

In this section we develop a simple decision model that captures the salient features of the health care system in Colombia. The aim of the model is to make explicit the main factors that determine the decision of insurers to restrict access to a service, and the decision of enrollees to file tutelas or complaints. There are two periods. In the initial period agents observe their health status, and, if required, get prescribed a treatment, procedure or medication. These prescriptions can be authorized or denied/delayed (henceforth restricted) by the health insurer. Any medical service that is prescribed but not readily authorized is enforceable through litigation, but not all individuals choose to use legal instruments.

Agents. There is a continuum of agents, each enrolled with a health insurer. Suppose the initial endowment of health of an agent is given by the value of the variable $h_0 \in H = \{h_0^1 \dots h_0^H\}$, where larger values indicate better health. For each health endowment there is a unique corresponding medical service, $x \in X = \{x^1 \dots x^H\}$ that a health professional prescribes, equal across agents.[33] We assume h_0, or alternatively x,[34] has a probability distribution f that is common across agents. If a service x is prescribed and delivered, the health level in the final period is $h^x \geq h^{\neg x}$, where $h^{\neg x}$ is the health level of an agent that was prescribed x but did not receive it.[35] The assumption is that receiving medical care is at least as good as not getting it.

If an insurer denies an authorization, an agent can use litigation to enforce its access. To simplify, we assume tutelas are the only legal instrument available. An agent j that files a tutela incurs an idiosyncratic cost $\eta_j \in \mathbb{R}_+$, expressed in health units, which is independent of the service requested and the value of which is unobserved to the insurer. Filing a tutela has no pecuniary cost, so we think of η_j as capturing effort (the effort required to learn about the process) and psychic costs (the stress of filing the claim). We assume η_j has a known cumulative distribution function G across the population. A tutela requesting a service x is decided in favor of the plaintiff with an exogenous and known probability $\beta^x \in [0, 1]$. This probability can vary depending on the characteristics of the service requested, given that some restrictions more clearly contravene an individual's right-to-health. Finally, suppose that agents only care about their health and wish to maximize a utility function U, where $U' > 0$ and $U'' < 0$.

If an agent j faces a restriction on a service x she will find it optimal to file a tutela if

$$\beta^x U\left(h^x - \eta_j\right) + (1 - \beta^x) U\left(h^{\neg x} - \eta_j\right) \geq U\left(h^{\neg x}\right), \tag{A.1}$$

or, alternatively, if

[33] Agents with the same health endowment get prescribed the same medical service.

[34] There is a one-to-one correspondence between H and X.

[35] We assume health evolves in a deterministic way, although adding a stochastic component delivers qualitatively similar results.

$$\beta^x \geq \frac{U\left(h^{\neg x}\right) - U\left(h^{\neg x} - \eta_j\right)}{U\left(h^x - \eta_j\right) - U\left(h^{\neg x} - \eta_j\right)}. \tag{A.2}$$

The inequality states that as long as $\Delta h^x \equiv h^x - h^{\neg x} > \eta_j$, that is, if the health gain from receiving the service is larger than the idiosyncratic health cost, there will be a value of β^x for which the agent finds it optimal to file a tutela.

The ratio on the right hand side of inequality A.2 is an increasing function of η_j, which takes a minimum value of zero at $\eta_j = 0$. This implies that for a given $\beta^x > 0$, there is a lower threshold of the idiosyncratic cost, $\underline{\eta^x}$, such that all agents that get x restricted and have $\eta_j \leq \underline{\eta^x}$ will file a tutela. Hence, from the point of view of the insurer, the probability that an agent will file a tutela if x is restricted is $\alpha^x \equiv G(\underline{\eta^x})$. Note $\underline{\eta^x}$ is determined by two exogenous factors in the model: i). how much improvement in health an agent gets from receiving the service (Δh^x); and ii). the probability that litigation for x is decided in favor of the plaintiff (β^x).

Health Insurers. Insurers, indexed by $i \in \{1, ..., I\}$, receive a capitation payment y for each enrollee, which is exogenously determined. If an insurer authorizes a requested service x, it pays the full cost c_i^x. The costs of service provision can vary across insurers reflecting the fact that they can contract with their own independent network of providers. If they restrict a service, and the affected agent files a tutela that is upheld, the insurer pays the full cost of the service plus an additional idiosyncratic cost $\tilde{c}_i \in \mathbb{R}_+$. We think of this idiosyncratic cost as capturing factors like the administrative costs of dealing with the claims and general managerial efficiency. There are no additional penalties reflecting the fact that each case is treated individually and generates no precedent.

An insurer receiving a request to authorize a service x finds it optimal to restrict access if

$$\alpha^x \left[\beta^x \left(y - c_i^x - \tilde{c}_i\right) + \left(1 - \beta^x\right) y\right] + \left(1 - \alpha^x\right) y \geq y - c_i^x \tag{A.3}$$

or, alternatively, if

$$c_i^x \geq \tilde{c}_i \frac{\alpha^x \beta^x}{\left(1 - \alpha^x \beta^x\right)}. \tag{A.4}$$

This implies that every authorization for services for which this inequality holds will be restricted by insurer i, regardless of who makes the request. That is, some treatments, procedures or medications will always face supply side barriers to access. Let's define the subset of restricted services by insurer i as $X_i^r \subseteq X$. Which services belong to X_i^r depends on three factors: i). the cost of providing the service (c_i^x); ii). the probability that, if x is restricted, an agent will file a claim that is upheld ($\alpha^x \beta^x$); and iii). the idiosyncratic cost $\tilde{c}_i$.

Restrictions and Tutela Rates. We can now characterize some of the aggregate quantities used in the empirical strategy. First, since every service $x \in X_i^r$ is restricted by insurer i, the fraction of its enrollees that face restrictions is equal to the fraction of enrollees that were prescribed those services:

$$Restriction_i = \sum_{x \in X_i^r} f(x). \tag{A.5}$$

Equation A.5 shows that, beyond the three factors that determine X_i^r, the restriction rate also depends on the demand for medical care, as defined by the probability distribution of services requested f.

Second, the fraction of enrollees of insurer i that will file a tutela is given by

$$Tutela_i = \sum_{x \in X_i^r} \alpha^x f(x). \tag{A.6}$$

Let $\alpha_i \equiv \frac{\sum_{x \in X_i^r} \alpha^x f(x)}{\sum_{x \in X_i^r} f(x)} \in [0,1]$ be the fraction of enrollees in insurer i that file tutelas conditional on having their requested services restricted. It follows that

$$Tutela_i = \alpha_i Restriction_i. \tag{A.7}$$

Equation A.7 allows us to use the tutela rate to proxy the restriction rate which is not observed. In terms of the decision tree of Figure 3, α_i corresponds to the ratio of enrollees that end up in nodes C or D, relative to the total number of enrollees in nodes B, C, or D.

Holding the demand for medical care constant, a larger restriction rate will translate into worse average health outcomes in the final period. We assumed receiving medical care was at least as good as not getting it ($\Delta h^x \geq 0$), so restrictions can only lead to a deterioration of population health. It follows from Equation A.7 that an increase in the restriction rate that is not demand driven will be reflected in a higher tutela rate and associated with i). lower utilization of health care services and ii). a higher mortality rate. These are the two key empirical predictions of the model. The predicted associations will be attenuated by the fact that a fraction

$$ClaimsUpheld_i = \sum_{x \in X_i^r} \beta^x \alpha^x f(x) \tag{A.8}$$

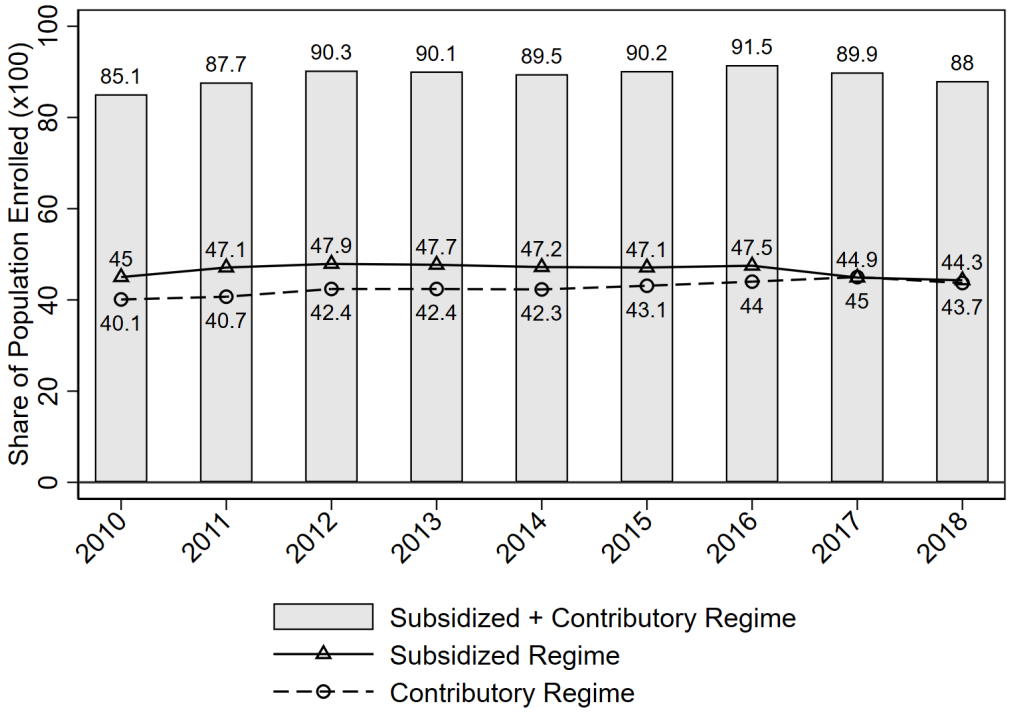
of services are delivered after favorable judicial rulings. But the fact that some services are nevertheless restricted implies there is an unmet demand for health care, defined as the share of services that are requested but not delivered. More formally,

$$\begin{aligned} unmetDemand_i &= Restriction_i - ClaimsUpheld_i \\ &= \left(\alpha_i^{-1} - \beta_i\right) Tutela_i, \end{aligned} \tag{A.9}$$

where $\beta_i \equiv \frac{\sum_{x \in X_i^r} \beta^x \alpha^x f(x)}{\sum_{x \in X_i^r} f(x)} \in [0,1]$ is the fraction of enrollees that receive medical care as a result of litigation, and $\left(\alpha_i^{-1} - \beta_i\right) \geq 0$.

B Online Appendix

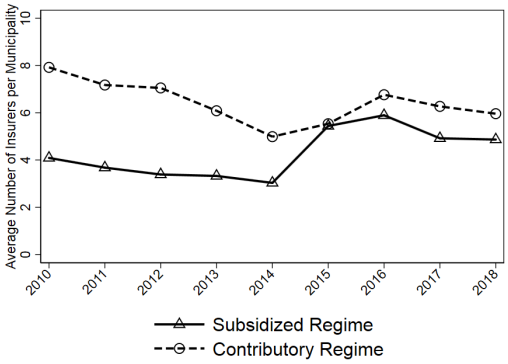
Figure B.1: Coverage of the Health System: Overall and by Regime



Notes: The bars show the share of the the population that is either in the contributory or subsidized regimes by year. The two lines report the share of population that is in the subsidized regime (solid) and contributory regime (dashed) by year.

Figure B.2: Competitive Structure of the Health Care System

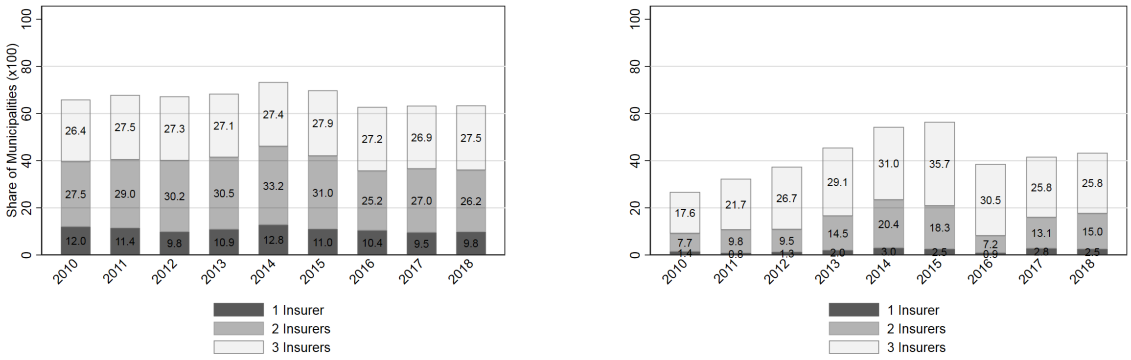
(a) Insurers per Municipality



Share of Municipalities with at Most One, Two or Three Insurers

(b) Subsidized Regime

(c) contributory Regime



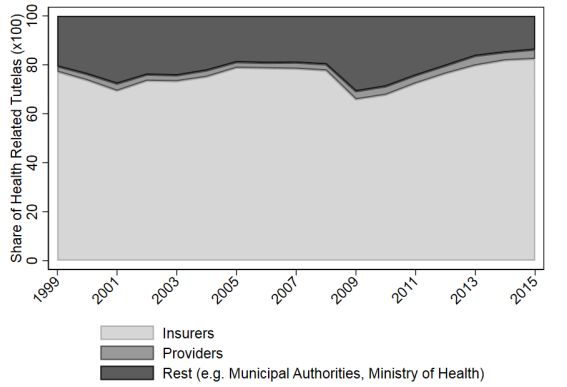
Notes: Panel (a) shows the average number of insurers per municipality and by regime. Panels (b) and (c) show the share ($\times$ 100) of municipalities that have at most 1, 2, or 3 insurers by regime.

Figure B.3: Examples of Press Coverage Documenting Insurers Limiting Access to Care Services

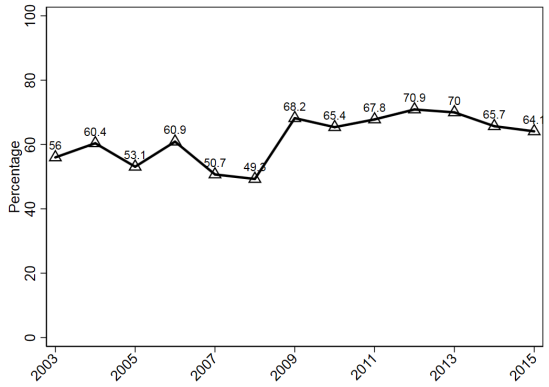
News Headline	Translation
	**El Universal (30/01/2015).** A patient dies waiting for the health insurer to authorize a medication in Cali.
	**El Tiempo (28/07/2020).** The ombudsman's office denounces heal insurers are not authorizing the delivery of medications. Cancer patients experience delays of 10 days for needed medications.
	**Semana (30/09/2014).** The `walk of death' takes the life a child. Even after a judicial intervention, the health insurer delayed attention to a child 22 months of old, causing his death.
	**Caracol (04/07/2016).** Young woman dies after asking in a video for an authorization to its health insurer. The women suffered from stomach cancer and needed special attention.
	**W Radio (26/02/2020).** ombudsman's office in Tunja worried because health insurers are systematically failing to comply. Patients are having to file tutelas.
	**Vanguardia (14/03/2017).** A child with cancer might lose his leg because of delays in surgery. The life of 14 years old minor, who suffers from bone cancer in his left knee, is at risk for delays in a surgery that the health insurer must approve. According to the kid's mother, the entity is denying authorizing the priority surgical intervention.
	**La FM (14/07/2019)**: Health insurers delivered medications for a patient with cancer after his death.

Figure B.4: Evolution of Tutelas Invoking the Right-to-Health by Defendant Body, Share of Services Requested Included in the Benefit Package, First Instance Decision by the Judge, and Regime

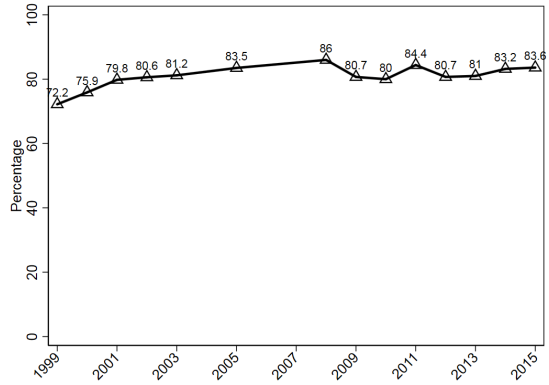
(a) Share of Health Related Tutelas by Defendant Body



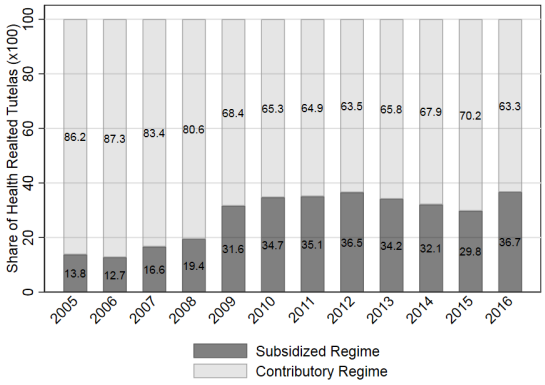
(b) Share of Services Demanded in Tutelas that are Included in the Benefit Package



(c) Share of Health Related Tutelas Decided in Favor of the Claimant



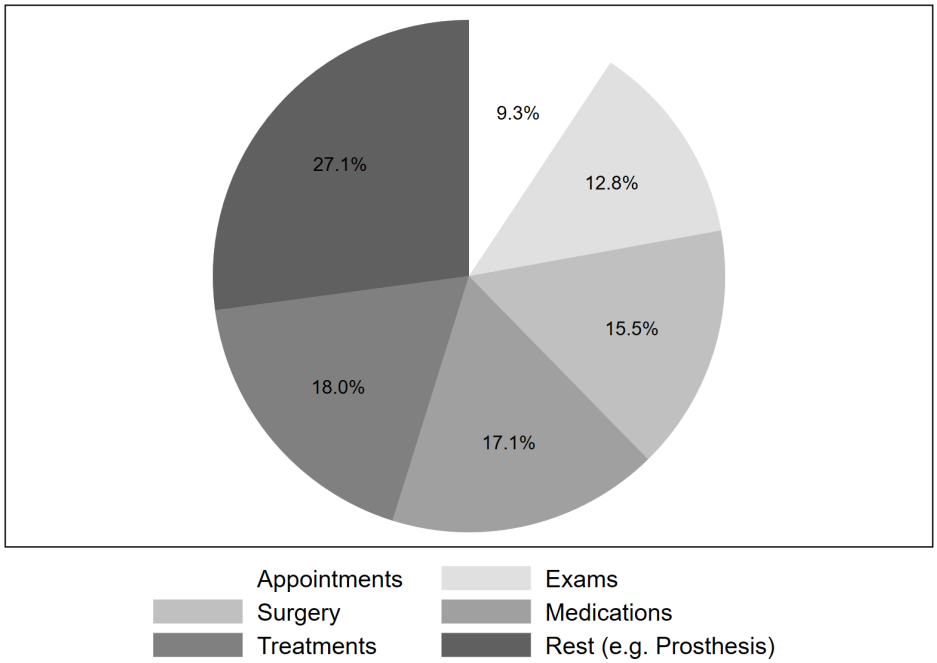
(d) Share of Health Related Tutelas by Regime



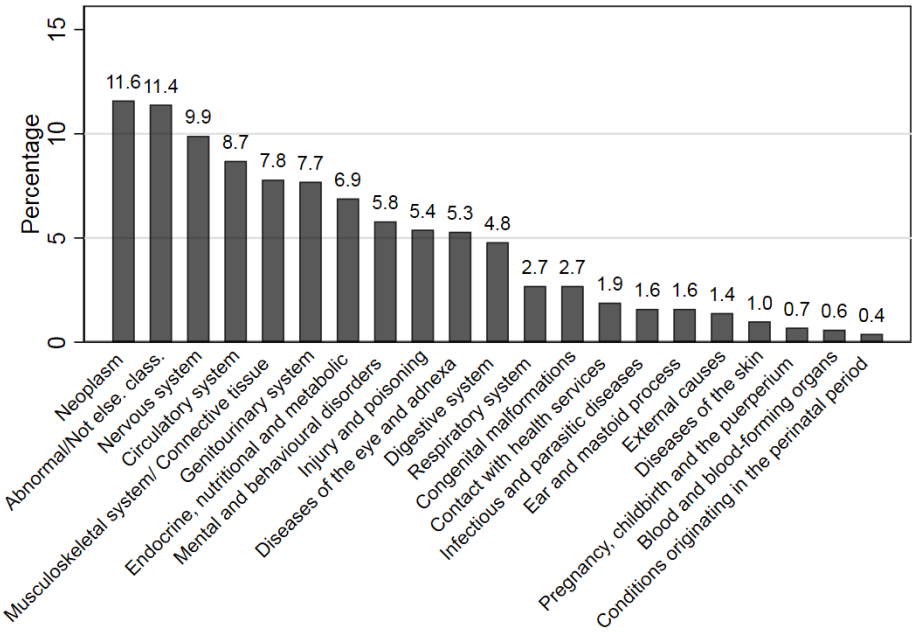
Source: Authors calculation based on Defensoría del Pueblo (2004, 2007, 2009, 2010, 2011, 2012, 2013, 2015). The numbers reported in Panels (c) and (d) include tutelas made against other actors in the system different from insurers (i.e. IPSs, Secretary of Health and Social Protection, local authorities).

Figure B.5: Type of Treatment Requested and Diagnoses Associated with the Tutelas Invoking the Right-to-Health

(a) Type of Treatment Requested in Tutelas Invoking the Right-to-Health



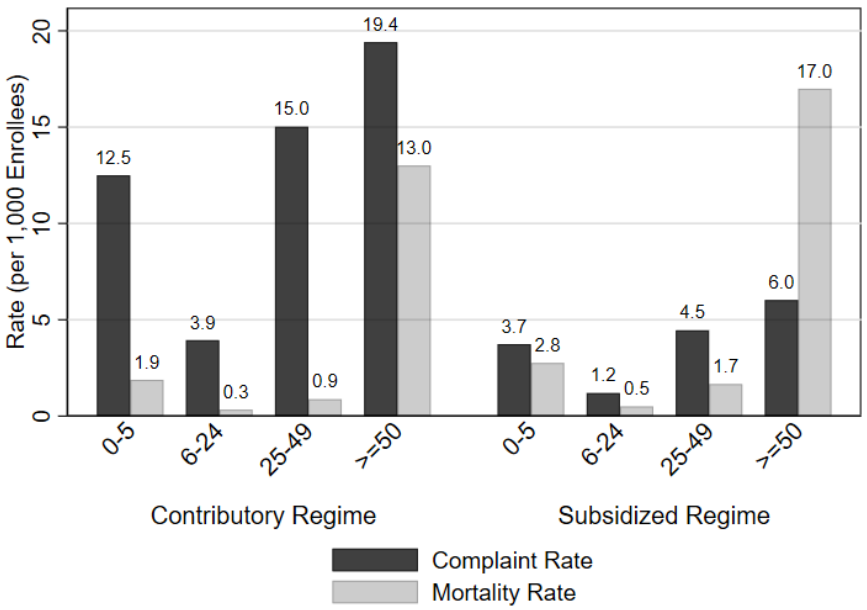
(b) Diagnoses Associated with the Tutelas in 2015



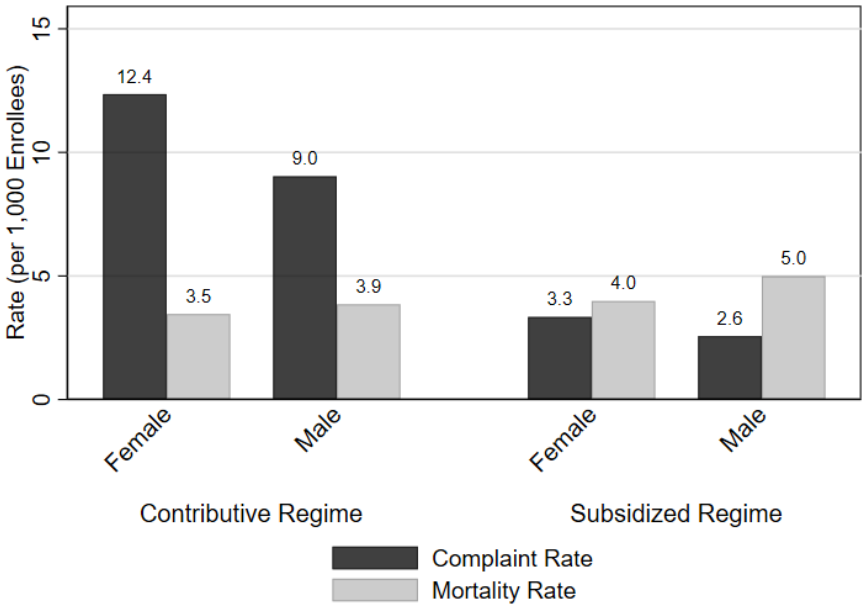
Source for Panel (a): authors calculation based on Defensoría del Pueblo (2004, 2007, 2009, 2010, 2011, 2012, 2013, 2015). Source for Panel (b): Ministerio de Salud y Proteccion Social (2018). The numbers reported in both panels include tutelas made against other actors in the system different from insurers (i.e. IPSs, Secretary of Health and Social Protection, local authorities).

Figure B.6: Complaint and Mortality Rates by Age and Sex Conditional on Regime

(a) By Age



(b) By Sex



Notes: Panel (a) reports the average complaint and mortality rates by age. Panel (b) reports the average complaint and mortality rates by sex. The divisions correspond to the people affected by the complaint, not those that file it

Table B.1: Complaints to the National Health Superintendence Between 2013 and 2017

	Overall		Contributory Regime		Subsidized Regime	
	Number	Share	Number	Share	Number	Share
Restriction of Access to Care Services*	737,785.0	44.1	565,273.0	47.7	119,172.0	32.1
Delays of Authorizations**	305,508.0	18.3	185,351.0	15.6	101,967.0	27.5
Problems with Affiliation	144,362.0	8.6	97,255.0	8.2	41,462.0	11.2
Restriction to Change/Choose Insurer/Provider	84,311.0	5.0	46,828.0	4.0	35,716.0	9.6
Denials of Authorizations**	75,309.0	4.5	51,778.0	4.4	18,445.0	5.0

Notes: The numbers in the overall column include complaints made against other actors in the system different from insurers (i.e. IPSs, Secretary of Health and Social Protection, local authorities). The numbers by regime include only complaints made against insurers. *Includes lack of opportunity to get an appointment with a care specialist, lack of opportunity to program a surgery, lack of opportunity to program an exam, and lack of opportunity to get a medication amongst others. **Includes authorizations for appointments with care specialists, surgeries, exams, and medications.