Do financial or prosocial incentives reduce over-treatment? Evidence from South Africa

Mylène Lagarde
London School of Economics

Duane Blaauw
University of Witwatersrand

12 January 2020

Preliminary draft – do not cite

Abstract

In many developing countries, scarce resources are often misused on the provision of unnecessary medical treatment. Even in the absence of profit incentives, providers often recommend treatment that are inappropriate in terms of volume or cost. We explore whether financial and prosocial incentives can improve providers' treatment choices and increase efficiency. We conduct an audit study of dispensing general practitioners (GPs) in South Africa, who charge a fixed consultation fee that is inclusive of drugs. As such, they have an incentive to limit over-treatment in order to maximise profit. Each GP was visited by unannounced standardised patients portraying a simple case of respiratory viral infection, whose correct treatment does not require antibiotics. To test the response of providers to financial and prosocial incentives, we exogenously varied two attributes of the standardised patients. First, to vary the financial incentive faced by GPs, two patients were sent to each GP: one who took the drugs dispensed as part of the consultation and one who requested a separate prescription to obtain drugs in a pharmacy. Second, to test whether prosocial concerns limit over-treatment, providers were randomised to receiving patients who were either covered by private insurance or who paid all expenses out-of-pocket. When GPs no longer have an incentive to recommend efficient treatment, they choose inappropriate drugs more frequently, and choose drugs that are more expensive, resulting in significantly more expensive treatments. We find some evidence that doctors prescribe fewer inappropriate drugs to uninsured patients, compared to insured patients, but when they do prescribe these drugs, they choose more expensive drugs for uninsured patients. Hence, overall there is no evidence that doctors respond to prosocial incentives and recommend more efficient treatments for patients who bear those costs. The results highlight both the necessity and potential effectiveness of introducing supply-side cost-sharing financial incentives to alter provider decisions and reduce treatment inefficiency.

1. Introduction

One of the causes of rising health expenditures in many health systems has been linked to over-provision of healthcare, defined as healthcare provided with higher volume or cost than is appropriate (Emanuel and Fuchs 2008, Brownlee, Chalkidou et al. 2017). In the US, evidence suggests that about a third of health care spending comes from unnecessary care (Berwick and Hackbarth 2012). In developing countries such as Kenya or India, where accessing care is often costly to patients, a majority receive inappropriate drugs with little to no benefits for their health (Das, Holla et al. 2012, Daniels, Dolinger et al. 2017). In addition to the waste of resources, over-provision of treatment can also have negative externalities. This is the case of antibiotics whose overconsumption fuels the resistance of bacteria, which poses serious threats to global health by making future treatments of deadly infections less effective (WHO 2014).

In developing countries, a lot of the evidence points to the role of providers in inefficient provision of treatment. In some settings, profit incentives exist that directly encourage providers to recommend unnecessary care (Currie, Lin et al. 2014). Even in the absence of direct financial incentives, over-provision of services by professional healthcare providers is rampant, even when they know the appropriate treatment (Mohanan, Vera-Hernandez et al. 2015). Evidence emerging from audit studies demonstrates that patient demand is not responsible for the large levels of overprovision detected (Das, Holla et al. 2012, Daniels, Dolinger et al. 2017). However, especially in private markets where they are sensitive to competitive pressure, providers may respond to perceived patients' demands (Das, Holla et al. 2016). At a time when many developing countries consider reforms to increase access to health care, including by introducing free care and contracting private providers to deliver services, it is important to explore the potential impact of interventions to reduce over-provision.

One way to alter provider behaviour is to introduce a fixed payment inclusive of all care provided to a patient, which creates an incentive to maximise profit by limiting inefficient care. The typical example of such "supply-side cost-sharing" mechanisms is case-based payment for hospitals where reimbursements rates for hospitalised patients are set prospectively based on groupings of patient diagnoses reflecting resource use (Ellis and McGuire 1993). For individual providers, cost containment mechanisms usually aim to limit the volume of visits and consist of fixed lump sum payment per patient for a certain period, rather than payments per visit or episode of care.

Although they are usually designed to reduce over-provision of healthcare by altering patient behaviour and reducing demand for inappropriate care (patient moral hazard), patient charges may also influence provider treatment decisions. If they are altruistic and care about patients' welfare, providers will be less likely to recommend inappropriate or expensive treatment when patients have to pay for them (Lundin 2000, Crea, Galizzi et al. 2019).

The over-provision of healthcare raises a number of questions about the ways in which providers respond to incentives, and the extent to which these responses can inform the design of policies. First, can supply-side cost sharing create incentives to reduce inefficiency in healthcare provision by doctors? Second, do patient user charges reduce the likelihood of overtreatment by providers, through concerns for patients' welfare? Answers to these questions are crucial to inform effective policies, but they have been limited by the ability to find exogenous variations in financial incentives faced by doctors.

To address this gap, this paper uses data from an audit study conducted in the metropole of Johannesburg (South Africa). We exploit an unusual policy setting, whereby drug pricing reforms introduced in the 2000s to limit over-prescribing of drugs, have effectively introduced supply-side cost sharing for the large proportion of private primary care doctors who dispense drugs. These doctors now charge a flat fee that covers both the consultation and drugs dispensed to patients for their treatment. The audit study relies on the use of standardized patients (SPs) who were trained to accurately present symptoms of a simple medical condition (viral bronchitis) to doctors. After each visit, SPs recorded detailed information about the consultation, including specific information about the type of drugs recommended.

To identify the effect of provider and patient cost-sharing on treatment decisions, we exogenously manipulated the characteristics of the SPs. First, to isolate the impact of provider cost-sharing, each doctor was visited in a random order by a pair of SPs identical in all characteristics except one. Instead of accepting the drugs dispensed by the provider as part of the consultation fee, patients randomly assigned to playing the "prescription" role requested that the doctor write them a separate prescription to be filled in a pharmacy, thereby lifting the cost-sharing incentive faced by doctors in the treatment decisions. The other patient sent to the same GP acted normally and received whatever drugs were dispensed by the doctor as part of treatment.

Second, to test the impact of patient cost-sharing, we partnered with a large private insurer to give full insurance cover to the SPs. When they visited half of the participating doctors chosen randomly, SPs used their insurance cover, thereby signalling to the doctors that they would not bear any cost resulting from the visit. When they visited the other half of doctors, patients pretended not to have any insurance cover and have to pay out-of-pocket. This allows us to test if providers reduce overprovision of healthcare out of prosocial concerns.

We report three main findings. First, we find compelling evidence that financial incentives in the form of provider cost-sharing (a flat rate including consultation and drugs dispensed) can reduce treatment inefficiency. In the absence of such incentives, we find that providers recommend inappropriate drugs (other than antibiotics) more often, choose drugs that are 20 percent more expensive on average, resulting in an increase in the cost of treatment by nearly 90 percent.

However, and this is our second finding, there is no reduction in the volume of unnecessary treatment. Despite evidence that incentives affect provider behaviour through the choice of less expensive drugs, we find very high levels of over-treatment by dispensing GPs, even when dispensing drugs reduces their profit: 71 percent of patients receive some antibiotic. These results do not seem driven by lack of knowledge that the patient suffers from a viral infection that is unlikely to be treated by antibiotics, but rather by norms and habits. However we find that financial incentives seem likely to limit the over-use of other inappropriate drugs (e.g. steroids, antihistamines, etc.).

Finally, we do not find evidence that doctors are altruistic and incorporate the welfare of patients in their treatment decisions. Specifically, when doctors prescribe drugs to patients, we find no consistent evidence that the levels of inefficient treatment decisions are lower for uninsured patients compared to insured patients. There is some suggestive evidence that, compared to insured patients, uninsured patients are more likely to be prescribed antibiotics, but less other inappropriate drugs. However overall there is not enough evidence supporting the altruistic doctor hypothesis that GPs recommend cheaper drugs to uninsured patients compared to insured patients.

This paper provides novel evidence on whether prosocial and financial incentives can reduce over-provision of unnecessary medical treatment by providers, and increase the efficiency of service delivery. As such we contribute to several bodies of literature.

First, this paper is related to the empirical literature on credence goods markets, providing evidence of over-provision of services from caesarean sections, to computer repairs or taxi rides (Balafoutas, Beck et al. 2013, Kerschbamer, Neururer et al. 2016, Balafoutas, Kerschbamer et al. 2017). Together with other studies of healthcare markets (Mohanan, Vera-Hernandez et al. 2015, Das, Holla et al. 2016), we show that high levels of over-provision can exist even in the absence of profit motives of suppliers. A key contribution of our paper is to show that inefficiencies can partly be reduced by the introduction of payment mechanisms that force suppliers to internalise part of the financial consequences of their advice.

This paper also contributes to the literature on the effects of financial incentives to encourage efficiency and limit providers' overtreatment. There is a large literature focusing on hospitals' treatment decisions (Moreno-Serra and Wagstaff

2010, Busse 2012), while studies of individual providers focus on capitation payments compared to fee-for-service arrangements (Lurie, Christianson et al. 1994, Trottmann, Zweifel et al. 2012). Our study is novel in that it looks at the impact of an unusual supply-side cost sharing mechanism that can improve efficiency of treatment decisions made by individual providers in a particular episode of care. It is highly relevant in developing countries where growing evidence points to extreme levels of inefficiency at primary care level.

This paper contributes to the literature exploring the altruistic provider hypothesis. Several studies have looked at the effect of demand-side cost sharing (health insurance cover) on prescribing decisions, especially drug costs, and the share of branded vs. generic drugs (Leibowitz, Manning et al. 1985, Lundin 2000, Mott and Cline 2002, Liu, Yang et al. 2009, Crea, Galizzi et al. 2019). The evidence from this literature is mixed, and plagued by challenges of observational data to control for endogeneity problems of providers and patients' decisions, as well as the inability to disentangle completely providers' from patients' (expressed) preferences and choices. Our study is most closely related to an audit study of Chinese doctors (Lu 2014), where auditors randomly declare to be covered by a public health insurance[1] and request a prescription on behalf of an absent (fictitious) family member. Like Lu (2014), our study clearly identifies the decisions made by providers from patients' characteristics and demands. Our approach adds to this study by looking treatment decisions in addition to drug choices only. It also improves the realism of the audit by sending highly trained standardised patients with real insurance covers.

Finally, this paper adds to the growing literature using audit studies to uncover the determinants of otherwise hard-to-observe behaviours. Building on the long tradition of using simulated or standardised patients[2] (SPs) to train and assess medical students, there is a small but growing literature using SPs to explore the quality of medical advice of practising providers, particularly in low- and middle-income countries (Das, Holla et al. 2012, Mohanan, Vera-Hernandez et al. 2015, Sylvia, Shi et al. 2015, Das, Holla et al. 2016, Daniels, Dolinger et al. 2017). To explore the response of providers to different stimuli, health economists have started to use SPs to undertake audit studies similar to the ones used to study e.g. discrimination (Bertrand and Duflo 2017). These studies have sent SPs matched on all observable characteristics except ethnicity (Planas, García et al. 2015), patient information or requests (Currie, Lin et al. 2011, Currie, Lin et al. 2014), and insurance status (Lu 2014).

[1] In their study, patients never have to give proof of health insurance cover and do not

[2] SPs are lay individuals trained to describe the clinical symptoms and medical history of a particular clinical case, as a normal patient would do with a doctor.

2. Institutional setting

2.1. Primary care in South Africa

Although the public sector is free for all in South Africa, data show that 26% of the population chooses to consult a doctor in the fee-charging private sector, and this proportion rises to 32.2% in urban areas (NDoH, Stats SA et al. 2019). This is due to the perceived higher quality of care received in the private sector where consultations are done by general practitioners (GPs), compared to predominantly nurses in the public sector.

Unlike many other developing countries, the private healthcare market in South Africa is well-regulated and formalised (National Department of Health 2015). To practise medicine in South Africa, doctors must hold a five-year Bachelor's degree in Medicine and Surgery followed by two years doing a clinical internship. Medical practice also requires to be registered with the Health Professionals Council of South Africa (HPCSA) and the Medical and Dental Board and keep their registration up to date through annual licensing by the Board. Medical doctors have the freedom to set up their practice arrangements as they see fit. Although some also operate in small group practices, most doctors work in solo practices. They all operate on a fee-for-service basis, and they are free to set their consultation rates.

Private GPs are used predominantly by the 14.5% of the population who is more affluent and covered by private health insurance (Koch and Omotoso 2017). However, nearly 30% of those choosing to use private GPs rather than free public clinics are poorer patients without health insurance (NDoH, Stats SA et al. 2019), even though this leads to out-of-pocket expenditure (Ataguba and McIntyre 2012, Ataguba and McIntyre 2018).

There are a number of private health insurance providers. Regardless of the provider or insurance plan characteristics, private insurance schemes share a number of similar characteristics. First, while they differ in terms of deductibles and co-pay rates for hospital services, insurance plans in South Africa all offer some 'Medical Savings Account' (MSA) to cover primary care expenditures. In practice, it means patients have full insurance cover for all primary care costs (GP consultation fees, pathology tests and drugs) up to a certain annual threshold, beyond which the patient incurs all the costs[3]. Second, insurance providers have limited oversight over the quality of health care provided by GPs (Competition

[3] Some insurance plans also cover a few extra consultations with GPs even when patients have depleted their annual MSA.

Commission South Africa 2018). Finally, regardless of whether they are treating insured or uninsured patients, GPs are paid on the same fee-for-service basis.

2.2. Dispensing GPs

There are two types of GPs: dispensing and non-dispensing GPs. Approximately 40% of GPs in South Africa are dispensing doctors who are officially licensed to prescribe and dispense pharmaceutical drugs to their patients[4]. While historically the right for GPs to dispense was introduced to address the needs of communities lacking pharmacies, especially in rural areas (Lim, Emery et al. 2009), dispensing GPs can now be found everywhere. In many health systems, dispensing doctors make a profit selling drugs to patients, which creates incentives to prescribe more drugs or drugs with larger mark-ups, (Iizuka 2007, Iizuka 2012, Kaiser and Schmid 2016, Goldacre, Reynolds et al. 2019). This is not the case in South Africa, where changes to drug pricing and dispensing regulations introduced in 2004 to increase transparency in pricing dramatically reduced the possibility to make profits on drugs, to the point that dispensing GPs have regularly complained that they dispense drugs at a loss (Health 24 2015). Since 2004, dispensing GPs purchase drugs at the single exit price (SEP), which is set yearly for all pharmaceutical products by the national authorities[5]. When they sell drugs, GPs are allowed to add a small dispensing fee to cover the costs relating to the management and dispensing of drugs. However, because dispensing fees both vary from one drug to the other, and are regulated each year by the government[6], GPs find this practice complicated and hard to explain to patients. Instead, in practice nearly all dispensing GPs do not charge separately for drugs[7], but offer a flat rate that is inclusive of all drugs dispensed. Some GPs also offer a lower consultation rate when they do not dispense, although this practice is not systematic.

While dispensing GPs charge a flat rate for their consultation, they are obviously free to *not* dispense drugs, or not dispense all of the drugs they recommend for the patient treatment, and write a prescription for (some of) the drugs recommended. For example, doctors will have to prescribe if they do not have the necessary drug in stock. Another example of a dispensing doctor prescribing occurs when they move practice but have not yet moved their dispensing license[8].

[4] To obtain a dispensing licence from the Department of Health, a practitioner has to complete an online dispensing course for R1,000 (about £55). The licence is valid for five years.

[5] The SEP is the price at which a manufacturer must sell to all purchasers (pharmacies or dispensing doctors), irrespective of volume. This transparent pricing was introduced to avoid any perverse incentive, including rebate linked to volume. Thanks to this well-enforced regulation on pharmaceutical pricing, prices are the same everywhere such that the comparison of drug costs is not confounded by GP differences in their ability to influence the purchase or sale price levels.

[6] Since 2017, the dispensing fee for dispensing GPs is set at a maximum of 30% of a drug worth up to R120, and R36 thereafter.

[7] In our study, in the 240 consultations observed, only six patients were charged separately for drugs.

[8] In South Africa, the dispensing license is attached to a specific address of the doctor practice.

3. Study design

3.1. Overview of the audit study with standardised patients

To investigate the efficiency of treatment choices made by GPs, we set up an audit study with standardised patients (SPs) in the greater Johannesburg area.

We recruited and trained 12 individuals to impersonate regular patients seeking medical advice from dispensing GPs. All individuals took part in a 10-day training organised by a multi-disciplinary team of researchers. The training included rehearsals of a detailed script of the clinical case, realistic portrayal of the presentation of the symptoms and patient's attitude, as well as construction of a backstory consistent with the socio-economic status of the individual portrayed (i.e. middle-class person). To create exogenous variation in the incentives faced by doctors, SPs were randomly assigned to different scenarios, discussed in more detail below.

SPs were trained to accurately and consistently present the clinical symptoms and history of a viral respiratory infection (acute bronchitis) in a healthy adult in their early 20s. The case was developed in collaboration with infectious diseases experts with the objective to portray an uncomplicated textbook case of viral bronchitis. This case was chosen for two reasons. First, respiratory tract infections (RTIs) are one of the most frequent causes of deaths and a main reason for primary care use in developing countries (Brink, Van Wyk et al. 2016, Troeger and GBD 2016 Lower Respiratory Infections Collaborators 2018). In South Africa, they are estimated to represent the vast majority of visits during the cold season (July-August), when this study was undertaken. Second, they are arguably the most important cause of over-prescription of antibiotics, with 75 - 80% of antibiotics estimated to be prescribed in primary care, mostly for RTIs such as bronchitis, pharyngitis and sinusitis (Barnett and Linder 2014, Brink, Van Wyk et al. 2016).

The case portrayed was chosen specifically to assess the quality of medical advice provided by the doctor, and specifically whether the doctor recommends inappropriate treatment, especially the recommendation of antibiotics. The age of the patient (young adult) and lack of co-morbidity were chosen to rule out potential concerns of complications that doctors might have for immune-depressed or fragile individuals such as children or elderly, which have been found to increase the likelihood of unnecessary antibiotic prescribing.

To allow us to measure provider effort, the opening statement[9] of the SPs did not provide a lot of information beyond the main complaint (coughing), which could potentially be consistent with a number of illnesses (pneumonia, TB, etc.). Further appropriate questioning and examination would allow the doctor to rule out alternative diagnoses and conclude that the patient suffered from a typical case of viral bronchitis. A full description of the case can be found in Appendix 1.

The validity of the audit study relies on the fact that providers believe that SPs are real patients and do not know they are in fact being observed. To test this, we phoned all GPs who had been visited by our SPs, to determine whether they had suspected any patient to be a SP[10]. Overall, only four GPs indicated some suspicions, in a total of five consultations. Because the description of the patient characteristics and symptoms of one of these doctors did not match our SP, we did not consider this a valid detection. Hence, only four of the 240 SP visits (1.66%) were detected, which is comparable to recent studies conducted in LMICs, and supports the validity of the method.

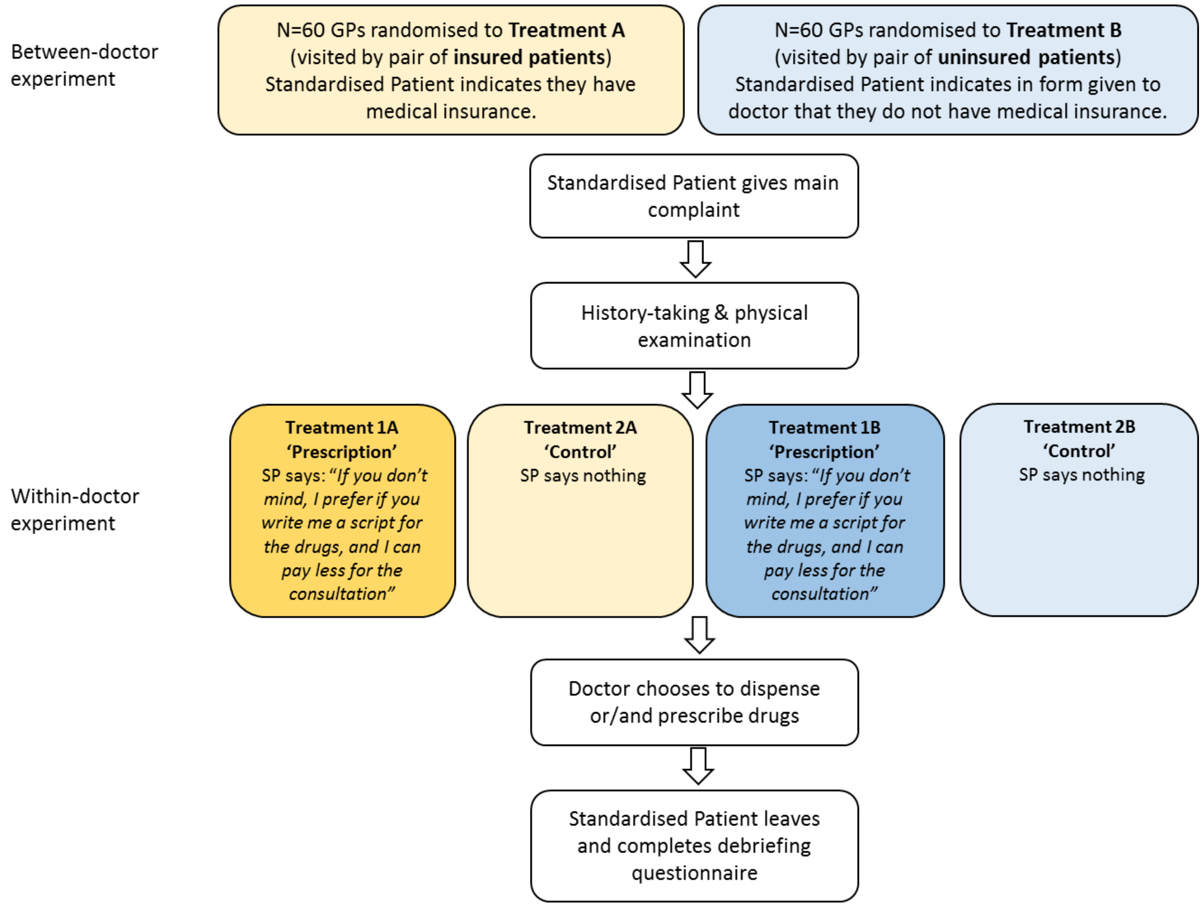
3.2. Experimental design

The audit study was layered with an experimental approach to detect variations in how doctors prescribe depending on the financial and prosocial incentives they face. To maximise the power of the experiment, we employed a mixed design. In the 'within-doctor' encouragement experiment, we send pairs of SPs to doctors, and randomly ask one of the two SPs to request a prescription from the doctor. The other dimension of the experiment is a classic between-subject experiment: we randomly assigned doctors to receive either a pair of insured or uninsured patients. Figure 1 provides a summary of the design and protocol of the two experiment.

[9] SPs were trained to say: "I have been coughing a lot. I had a cold about a week ago. The cold is a bit better now, but the cough is not going away." The statement was translated in several local languages if the doctor addressed a patient in their local language (e.g. based on their name).

[10] When they consented to take part in the study, providers were told to record the name and symptoms of any patient they would suspect to pretend to be a real patient, but in fact be part of our study.

Figure 1: Experimental design



3.2.1. Prescribing encouragement experiment

Our aim is to explore whether the rationing incentive inherent to the way in which dispensing GPs operate in South Africa encourages them to limit inefficient provision of services. Identifying this effect is difficult. Simply comparing dispensing and non-dispensing GPs cannot lead to a robust estimation due to obvious selection problems. It is also not possible to randomise providers to either dispensing or non-dispensing. However, we can use an encouragement design to recover the effect of dispensing by using instrumental variables (Angrist, Imbens et al. 1996).

Each GP in the study was visited by two SPs, identical in all respect except for the fact that one requests a written prescription to obtain drugs from a local pharmacy[11], while the other one does not ask anything. The order of the 'Prescription' and 'Control' patient visits were randomised. To avoid raising suspicions about this unusual request, the 'Prescription' SPs were trained to refer to the fact that some GPs offer a lower consultation rate for consultations without drugs: "*If you don't mind, I prefer if you write me a script for the drugs, and I can pay less for the consultation*." By doing so, the patient also introduces the idea that

[11] This request is similar to the strategy used by Currie, Lin et al (2014). However, in their study, this strategy is used to remove the financial incentives of Chinese doctors who earn a proportion of their income from drug sales. Here, the patient request creates the opposite incentive, and increases the doctor's profit.

they care about the financial consequences of their health-seeking decision, and do not wish to spend more than necessary. If asked further about the motives of their request, the SPs were trained to explain that they would get the drugs from a relative who owned a pharmacy and would not charge them extra for the drugs[12].

This request by the patient introduces a random partial manipulation on the decision made by the provider to prescribe. All doctors can prescribe drugs to patients at any point. In fact, 16.8% of dispensing doctors in our sample choose to prescribe to Control patients. But if the encouragement is successful, some additional proportion of GPs who would have normally chosen to dispense, will prescribe drugs to the patient. By doing so, the rationing constraint they face in their treatment decisions is removed.

Sending both the control and 'prescription' SPs to each doctor potentially alleviates some concerns raised by encouragement designs. Specifically, because we observe whether GPs prescribe or not when they see control patients, we are in a position to identify potential "defiers". We further discuss these issues in the robustness section of the paper.

3.2.2. Patient insurance status

Half of 120 GPs were randomised to receiving a pair of insured patients and the other half was randomised to receiving a pair of uninsured patients. We partnered with one of the largest providers of medical insurance and provided real insurance cover to our SPs, complete with an actual insurance card and electronic file in the insurance company records under the SP's real identity[13]. The insurance plan chosen offered full insurance cover for primary care costs up to R8,316 per year, beyond which a normal patient would have to pay. Our collaboration with the insurance group meant that SPs were never subjected to this limit and always appeared as if they had not used any of their annual primary care allowance, despite the multiple consultations they did. Finally, to increase the similarity between our insured and uninsured patients, and ensure that GPs would always receive their consultation fees immediately for both patients, our insured SPs always paid the consultation fee themselves and asked for a receipt to file for reimbursement.

Panel A in Table A1 in Appendix shows the validity of the randomisation based on the limited data available from the sampling frame. The only observable difference is that a higher proportion of providers who received insured patients were located in the wealthiest areas.

[12] The GP would understand from this statement that the pharmacy would not charge them any dispensing fee on top of the SEP.

[13] Unlike past studies varying the insurance status of standardised patients (eg Lu 2014), in our setting, the insurance status of the patient is systematically verified by providers. Before a consultation, the GP secretary generally asks to see the patient insurance card and checks that they still have enough "savings" to cover the consultation.

4. Data

4.1. Sampling

We carried out the study in the metropolitan area of Johannesburg, the main urban centre of the country. To construct a sampling frame of private GPs, we used a national database of practices registered with the Board of Healthcare Funders, which includes approximately 80% of all registered GPs nationally, and significantly more in urban areas (BHF 2012). The database included the contact details of 1,012 GPs in Johannesburg, 361 of whom (35.7%) were licensed to dispense drugs and eligible to take part in the study[14].

GPs were recruited between March and June 2018, and they were informed that "*enumerators trained to portray real patients*" would visit their practice within the next six months. We called all of eligible GPs to invite them to take part in the study. Of those, 26% (N=94) could not be reached despite several attempts[15] ; 26.3% refused to take part (N=95); 11.6% (N=42) requested further information about the study to make their decision and 36% agreed to take part (N=130). From this final group, we randomly selected 120 to take part in the study, and visits by SPs occurred in August 2018.

4.2. SP role non-compliance

To avoid confounding the fieldworker and the SP roles played, we randomly assigned the four roles to consultations, and then randomly assigned individual fieldworkers to consultations, so that all fieldworkers played all roles. Even though fieldworkers were reminded of their role before each visit, this complexity led to seven cases (out of 240) where SPs did not play the role they had been assigned (Table A2 Appendix X). To preserve the comparison between control and prescription patients, we drop seven pairs of SP visits from the analysis, leading to a final sample of 226 consultations with 113 doctors, 60 who received uninsured pairs of patients and 53 who received pairs of insured patients[16]. While we cannot completely rule out that these seven GPs are different from the others, Panel B in Table A1 in Appendix show the balance between the two groups of GPs is preserved.

4.3. Provider survey and summary statistics

When GPs were contacted about the detection of patients, they were also asked to schedule an appointment to conduct a face-to-face interview. Of the 120 providers in the study, only N=90 (75%) agreed to the interview.

[14] To be eligible to take part in the study, a doctor had (1) to practice general medicine; (2) to work in a private practice.

[15] Either no one responded, or the receptionist refused to pass the communication to the doctor.

[16] Unsurprisingly, failure to portray the role assigned occurred more frequently for the insured SPs because they had to recall more elements.

Table 1 below presents some basic summary statistics about the 113 providers included in the study, based on the information available from the initial database (Panel A) and for the sub-group who took part in the interview data (Panel B). The data show that the GPs are mostly male, 51 years old on average and therefore quite experienced (24 years of practice on average). The doctors work in a wide range of socio-economic areas, even though 57 percent work in the richest 40% areas[17].

Table 1: Provider summary statistics

	Mean	(SD)
Panel A. Medpages data (N=113)		
Male	0.73	(0.44)
Age	51.57	(10.98)
Practice location, by socio-economic quintile of local area		
Practice is located in Q1 (poorest 20%)	0.11	(0.31)
Practice is located in a Q2	0.21	(0.41)
Practice is located in a Q3	0.12	(0.32)
Practice is located in a Q4	0.41	(0.49)
Practice is located in a Q5 (richest 20%)	0.16	(0.37)
Panel B. Interview data (N=86)		
Male	0.73	(0.45)
Age	52.48	(10.78)
Practice location, by socio-economic quintile of local area		
Practice is located in Q1 (poorest 20%) area	0.08	(0.28)
Practice is located in a Q2 area	0.23	(0.42)
Practice is located in a Q3 area	0.12	(0.33)
Practice is located in a Q4 area	0.39	(0.49)
Practice is located in a Q5 (richest 20%) area	0.18	(0.39)
Experience as a doctor (years)	24.35	(9.12)
Ethnicity		
African	0.38	(0.49)
White	0.12	(0.32)
Asian	0.41	(0.49)
Other	0.09	(0.29)
No. of patients per day, previous week	26.28	(13.90)
Works in group practice	0.23	(0.42)
Does sessions in the public sector	0.16	(0.37)
Knows SP case is caused by a virus	0.84	(0.37)
Likelihood AB will help patient recover more quickly (out of 100)	29.33	(29.60)
Likelihood other GPs will give this patient AB (out of 100)	64.59	(29.96)
Likelihood patient won't come back if no AB (out of 100)	61.56	(32.40)

[17] The quintiles were defined based on household data set from the Gauteng City-Region Observatory.

4.4. Treatment choices and cost of treatment

The SPs all presented with a simple case of self-limiting viral infection. According to both local and international clinical guidelines, no medication is necessary to treat the patient, although palliative medications may be considered to relieve the patient's symptoms and discomfort. Such palliative care can include analgesics for pain relief (e.g. paracetamol) and, possibly, some expectorant for the cough xrefx. Any other drug recommended by the provider is considered inappropriate. We consider in particular the prescription of antibiotics which is both unnecessary and potentially harmful in this textbook case of viral respiratory infection.

Following these guidelines, we define three measures of treatment quality: whether the treatment includes palliative drugs; whether it includes antibiotics and whether it includes any other inappropriate drugs[18]. Any treatment that include any inappropriate drugs would be considered inefficient.

We then consider the cost of the treatment. We compute the total cost of the drugs recommended by the doctor, either prescribed and/or dispensed. Using an exhaustive database of all drugs registered in South Africa, we calculated the cost of the treatment by using the SEP of all the drugs included in the treatment[19]. When we could not identify the drug prescribed or dispensed in the drug database, we used one of two strategies to price them. In some cases, with the help of doctors and pharmacists, we could identify a close match in the drug database that would typically be considered as a suitable option by a pharmacist. In this case, we used the price of the cheapest match for computing the drug cost. In other cases, especially for non-regulated drugs such as herbal remedies or probiotics, we used the cheapest available product from two big pharmacy chains which list most of their drugs and prices online. In both strategies the prices used are likely to be conservative estimates.

Our detailed information about the treatment allows us to decompose the total cost of the treatment into the number of drugs dispensed and their average price. We do this for the entire recommended treatment, as well as for the three types of drugs identified above: palliative (potentially useful) drugs; antibiotics and other inappropriate drugs.

[18] This includes the following categories of drugs: steroids, bronchodilator, antihistamines, nasal spray, vitamins, probiotics, and throat preparations.

[19] We had detailed information about the quantity prescribed and the dosage/size of the packet/bottle recommended.

5. Results

5.1. Estimation framework

Table 2 shows that the inducement strategy of the 'prescription' role had a dramatic effect on the probability that the doctor prescribed some drugs, instead of or in addition to dispensing them. When patients asked for a separate prescription, 44% of them received one. In control consultations, fewer than 17% of patients received a prescription. If we add cases where doctors both prescribed and dispensed drugs, the difference is even larger. Just above 60% of patients who asked a prescription received one, while the same doctors wrote a prescription in only 21% of consultations with control patients.

Table 2: Prescribing behaviour

	(1) Control patients	(2) Prescription patients
GP dispensed drugs	78.76	39.82
GP prescribed drugs	16.81	44.25
GP dispensed & prescribed	4.42	15.93
Observations	113	113

Our main objective is to estimate differences in the treatment decisions made by doctors for patients who asked for a separate prescription, as opposed to other control patients. If we find that lifting financial incentive leads to less efficient treatment decisions, then we can conclude that financial incentives could help reduce overprovision of healthcare.

We start by estimating an intention-to-treat effect of prescribing rather than dispensing, i.e. of not facing a cost-sharing incentive. To account for the hierarchical nature of the data, since each doctor sees two patients, we estimate a fixed-effects model of the form:

$$Y_{ij} = \beta_0 + \beta_1 Prescr_{ij} + \gamma_j + order_i + fw_i + \varepsilon_{ij} \qquad (1)$$

where Y_{ij} is the outcome for patient i seen by doctor j, and $Prescr_{ij}$ takes the value 1 if the patient i told doctor j that they wanted to have a separate prescription, and 0 otherwise. The associated coefficient β_1 captures the effect of allowing the provider to release the rationing constraint s-he faces within their dispensing practice. In addition, we control for provider (γ_j), visit order ($order_i$) and fieldworker (fw_i) fixed effects.

Note that this is the intention-to-treat effect because providers may still choose to dispense drugs despite the request made by patients. Because our main interest is to measure more precisely the impact of releasing the cost-sharing incentive when GPs make treatment decisions, which only happens if they *actually* prescribe, we estimate this average treatment on compliers (ATC) by using a 2SLS fixed-effects model in which writing a prescription is an endogenous regressor, and the random assignment of the SP prescription role is used as an instrument to estimate the effect of prescribing:

$$Y_{ij} = \beta_{0j} + \beta_1 \widehat{Prescr}_{ij} + \gamma_j + order_i + fw_i + \varepsilon_{ij} \qquad (2)$$

where $\widehat{Prescr}_{ij}$ is estimated in a first stage using role assignment $Prescr_{ij}$ to predict the decision of provider j to prescribe to patient i.

While β_1 in the second specification provides a useful estimate of the average treatment on the compliers, it is still not an average treatment effect, because, as noted before, nearly 21% of GPs decided to prescribe at least one drug to control patients. In a final estimation, we drop these providers and re-estimate the same 2SLS model with a sample restricted to GPs who only and exclusively dispensed to control patients. By doing so, we seek to estimate an average effect more precisely but we run the risk of introducing some bias, as there might be some selection effect driving the decision to prescribe to control patients, even though we fail to detect any observable ones (see Table A3 in Appendix), and that there are reasons for this decision to be driven by exogenous factors [20]. However, we cannot rule out that this decision is the result of a self-selection process based on unobservable traits: doctors choosing to prescribe might be less concerned about their patients' financial welfare, and/or more concerned about their own profit. If it is the case, excluding these GPs driven by more selfish motives means that the estimated average effect likely under-estimating the true average effect.

To study whether demand-side cost-sharing reduces inefficiency in treatment decisions, we simply disaggregate the estimated effects for insured and uninsured pairs of patients. Mirroring the first estimation, we compute intention-to-treat effects of prescribing for insured and uninsured patients by estimating a fixed-effects model of the form:

$$Y_{ij} = \beta_0 + \beta_{11} Prescr_{ij} \times Ins_i + \beta_{12} Prescr_{ij} \times Unins_i + \gamma_j + order_i + fw_i + \varepsilon_{ij} \quad (3)$$

where Ins_i is a dummy variable taking the value 1 if the patient mentioned they had an insurance and 0 otherwise, and $Unins_i$ takes the value 1 if the patient pretended not to have insurance and 0 otherwise.

Similarly, we then use the random allocation of the prescription role as an instrument to estimate these two interaction effects in a 2SLS model.

[20] Doctors may have to prescribe for quasi-random reasons. For example, they may run out of the chosen drugs, or they may have recently moved practice and not updated their dispensing license.

5.2. Do financial incentives reduce over-provision of healthcare?

Table 3 presents the results of the three estimation strategies looking at the effects of prescribing on the treatment decisions made by doctors. The three outcome variables are whether the treatment included some palliative drugs (Column 1), some antibiotics (Column 2), or any other inappropriate drugs (Column 3).

Table 3: Treatment choices

	(1) Palliative treatment (1=yes)	(2) Antibiotic (1=yes)	(3) Other inappropriate treatment (1=yes)
Panel A. Intention-to-treat estimates			
Patient asked prescription	0.028	0.001	0.091*
	(0.038)	(0.045)	(0.047)
Mean for control patients	0.805	0.708	0.805
Observations	226	226	226
Panel B. IV estimates (AT on compliers)			
Patient asked prescription	0.068	0.002	0.224*
	(0.095)	(0.110)	(0.117)
Observations	226	226	226
Panel C. IV estimates, sub-sample of GPs dispensing to control patients			
Patient asked prescription	-0.042	-0.022	0.233**
	(0.073)	(0.087)	(0.105)
Mean for control patients	0.865	0.753	0.764
Observations	178	178	178

Notes: standard errors are in parentheses. Estimates in Panel A are from a fixed-effects linear model. Estimates from Panel B and C are from a fixed-effects 2SLS model. All regressions include provider, fieldworker and visit order fixed effects, as well as a constant term. Observations are at the SP-provider interaction level.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Looking at the ITT results in Panel A, we find no evidence that patients who asked for a prescription were more likely to receive palliative treatment or some antibiotic. However the probability of receiving some other inappropriate drug increased by about 9.0 percentage points (11 percent more than control patients). Using the IV estimates (Panel B), we can get a more precise estimate of the effect of lifting the supply-side cost sharing incentives on prescribing decisions. As expected, the results are consistent with the ITT estimates but much larger quantitatively. There is still no evidence that a change in the incentives structure faced by providers changes the probability to include palliative or antibiotic drugs in the treatment. However, in the absence of cost-sharing, there is a significant increase by 22.4 percentage points (pp) in the probability that the treatment includes other inappropriate drugs (Column 3). If we consider only providers who only dispensed to control patients (panel C), we find a similar increase in inefficient treatment decisions of 23.3pp.

Results for treatment costs (Table 4) follow the same format as Table 3 but the dependent variables of interest include the total costs of drugs recommended by the GP (Column 1), the number of drugs included (Column 2), the average cost of these drugs (column 3), and the proportion of generic drugs chosen (column 4).

<u>Table 4:</u> Cost of treatment

	(1) Total drug cost	(2) Number of drugs	(3) Average drug cost (cond.)	(4) Proportion of generic drugs
Panel A. Intention-to-treat estimates				
Patient asked prescription	37.602***	0.087	10.135***	-0.063***
	(13.175)	(0.109)	(3.611)	(0.023)
Observations	226	226	226	226
Mean for control patients	101.352	3.708	27.147	0.906
Panel B. IV estimates (AT on compliers)				
Patient asked prescription	92.476***	0.213	24.926***	-0.154***
	(29.759)	(0.270)	(8.050)	(0.056)
Observations	226	226	226	226
Panel C. IV estimates, sub-sample of GPs dispensing to control patients				
Patient asked prescription	99.782***	0.066	29.621***	-0.145***
	(23.392)	(0.218)	(6.691)	(0.041)
Mean for control patients	59.033	3.719	15.620	0.951
Observations	178	178	178	178

Notes: standard errors are in parentheses. In columns 3, the average cost is conditional on a drug being prescribed, which was the case for all patients. Estimates in Panel A are from a fixed-effects linear model. Estimates from Panel B and C are from a fixed-effects 2SLS model. All regressions include provider, fieldworker and visit order fixed effects, as well as a constant term. Observations are at the SP-provider interaction level.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

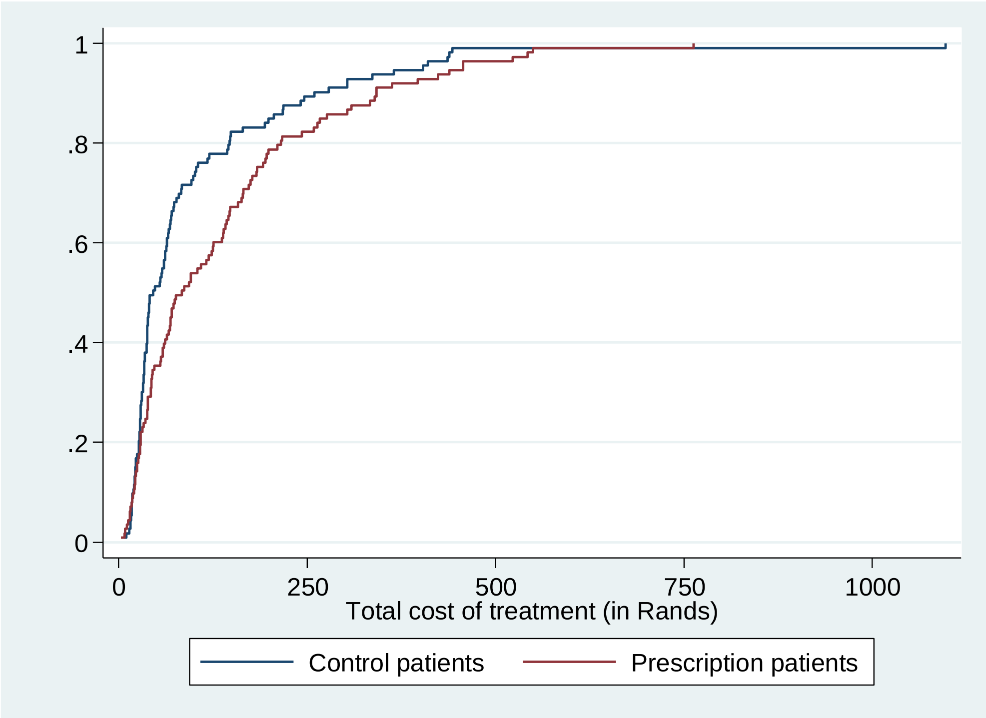
On average, GPs recommended a treatment worth a little over R100 (≈USD 6.9) to control patients, which included 3.7 drugs, the vast majority of which (90 percent) were generic drugs.

The ITT results in Panel A show that for patients who asked for a prescription, the drugs were about R38 higher than that of control patients, an increase of 37 percent. We find that this increase is not driven by a higher number of drugs, but by a higher average cost per drug. On average a drug recommended to a patient who asked for a prescription was about R10 more expensive, or 37 percent more than for control patients. This higher average cost is likely to be partly driven by a 6pp decrease in the share of generic drugs recommended (column 4).

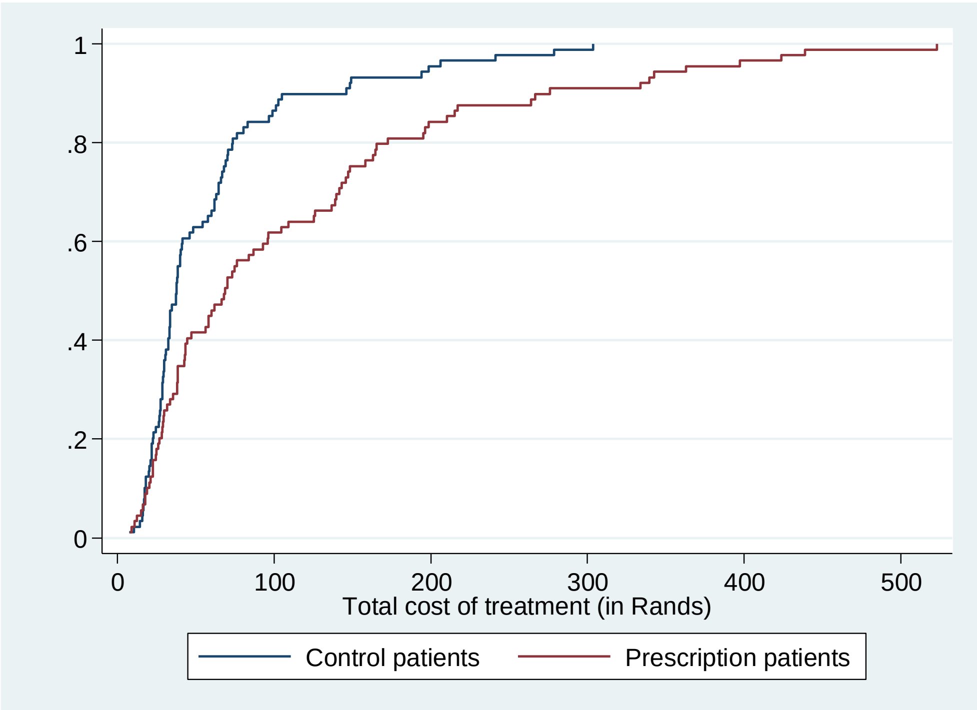
The IV estimates show that lifting the cost-sharing incentives leads to a significant increase in the treatment cost. When they prescribe, the total cost of drugs recommended by doctors is R92 higher (panel B), meaning that the absence of supply-side cost-sharing incentive increases the cost of the drugs recommended by 90 percent. This effect is the same if slightly higher in the restricted sample (Panel C), where we find an increase in cost by nearly R100. As before, we find that this reduction in efficiency is not driven by an increase in the number of drugs but rather by an increase of R29 in the average cost of drugs, representing a doubling of their average cost. Similarly we see a significant decrease by about 15pp in the proportion of generic drugs recommended, both in the whole sample or the restricted one (panel C).

These differences are seen clearly in Figure 1 and 2. Figure 1 plots the cumulative distribution functions (CDF) of the drug costs of control and treatment patients in the whole sample, while Figure 2 is restricted to providers who did not prescribe to control patients. The distribution of drug costs for 'prescription' (Figure 1) first-order stochastically dominates that of the control patients (Kolmogorov-Smirnov test, p=0.008). The corresponding distribution for 'prescription' patients visiting doctors who strictly dispense to control patients (Figure 2) also first-order stochastically dominates that of control patients (Kolmogorov-Smirnov test, p=0.002).

<u>Figure 1:</u> Cost of treatment by patient type, whole sample



<u>Figure 2:</u> Cost of treatment by patient type, excluding doctors who prescribe to control patients



Disaggregating further the cost of drugs across palliative, antibiotic and other unnecessary drugs (Table 5), we find that the choices made by doctors across all drug categories are affected by the incentives they face.

Looking FIRST at palliative drugs (columns 1-3), the ITT results suggest an increase of the average cost chosen by R12.60 representing a 53 percent increase. The average effect estimated on compliers (panel B) is even higher (about R31), and so is the effect in the sub-sample of doctors who only dispense to control patients (Panel C). Across all specifications, this increase in average drug cost results in much higher total cost.

The results for antibiotics are more mixed. All estimates (columns 4-6) point to an increase in the average cost of antibiotics chosen by the doctor for 'prescription' patients. However, the IV estimates also suggest a reduction in the number of antibiotics[21]. Consequently, probably due to the lack of power, the results detect a significant increase in the total cost of antibiotics only in the sub-sample of GPs dispensing to control patients (Panel C), where the effects are sharper and not diluted by the prescribing for control patients.

Finally, looking at the choice of other inappropriate drugs (columns 7-9), both sets of IV estimates suggest a significant increase in the total drug cost by about R30, and about 0.50 more drugs recommended.

[21] A close look at these results seem to be driven by the fact that to reduce the cost of treatment, some doctors prescribe a combination of two antibiotic drugs (amoxicilin and clavulanic acid), instead of simply choosing Co-Amoxiclav which is the combination of the same two drugs into one, and which costs more.

Table 5: Cost and quantity of drugs, by category of drugs

	Palliative drugs			Antibiotics			Other inappropriate drugs		
	(1) Total cost	(2) No. of drugs	(3) Average cost (cond.)	(4) Total cost	(5) No. of drugs	(6) Average cost (cond.)	(7) Total cost	(8) No. of drugs	(9) Average cost (cond.)
Panel A. Intention-to-treat estimates									
Patient asked prescription	12.588***	-0.049	8.373***	13.467	-0.029	25.997**	11.634	0.176	4.063
	(4.595)	(0.109)	(2.696)	(8.797)	(0.051)	(11.675)	(7.554)	(0.106)	(5.549)
Observations	226	226	186	226	226	160	226	226	190
Mean for control patients	23.634	1.478	15.895	37.495	0.761	49.371	40.223	1.46	24.037
Panel B. IV estimates (AT on compliers)									
Patient asked prescription	30.959***	-0.120	20.380***	33.120	-0.071	58.461**	28.612	0.432*	11.815
	(11.593)	(0.267)	(6.912)	(20.809)	(0.126)	(24.300)	(17.853)	(0.262)	(15.438)
Observations	226	226	186	226	226	160	226	226	190
Panel C. IV estimates, sub-sample of GPs only dispensing to control patients									
Patient asked prescription	22.464**	-0.308	17.810***	44.391***	-0.069	58.182***	33.134***	0.451**	24.722**
	(9.690)	(0.225)	(5.581)	(15.018)	(0.104)	(20.239)	(10.777)	(0.229)	(10.689)
Mean for control patients	22.067	1.629	12.672	21.408	0.809	25.149	15.558	1.27	11.873
Observations	178	178	153	178	178	133	178	178	144

Notes: standard errors are in parentheses. In columns 3, 6 and 9 the average cost is conditional on a drug being prescribed. Estimates in Panel A are from a fixed-effects linear model. Estimates from Panel B and C are from a fixed-effects 2SLS model. All regressions include provider, fieldworker and visit order fixed effects, as well as a constant term. Observations are at the SP-provider interaction level.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

5.3. Do pro-social concerns reduce over-treatment?

Using the fact that GPs were randomly allocated to receive a pair of insured or uninsured patients, we can test whether providers internalise patients' welfare, and reduce inefficiency in treatment decisions because of altruistic concerns. If providers care about the financial welfare of their patients, when they prescribe drugs to uninsured patients, they should limit inefficient treatments and/or choose less expensive drugs more than what they do with fully insured patients.

Table 6: Treatment choices, by insurance status

	(1) Palliative treatment (1=yes)	(2) Antibiotic (1=yes)	(3) Other inappropriate treatment (1=yes)
Panel A. Intention-to-treat estimates			
Uninsured patient X asked prescription	0.026	0.037	0.043
	(0.055)	(0.064)	(0.067)
Insured patient X asked prescription	0.030	-0.034	0.137**
	(0.054)	(0.063)	(0.066)
Mean for control uninsured patients	0.850	0.717	0.833
Mean for control insured patients	0.755	0.698	0.774
Observations	226	226	226
Panel B. IV estimates (AT on compliers)			
Uninsured patient X asked prescription	0.096	0.057	0.159
	(0.103)	(0.120)	(0.120)
Insured patient X asked prescription	0.018	-0.075	0.275**
	(0.122)	(0.143)	(0.138)
Observations	226	226	226
Panel C. IV estimates, sub-sample of GPs dispensing to control patients			
Uninsured patient X asked prescription	-0.017	0.063	0.031
	(0.099)	(0.118)	(0.146)
Insured patient X asked prescription	-0.070	-0.114	0.451***
	(0.108)	(0.129)	(0.159)
Mean for control uninsured patients	0.829	0.732	0.707
Mean for control insured patients	0.896	0.771	0.813
Observations	178	178	178

Notes: standard errors are in parentheses. Estimates in Panel A are from a fixed-effects linear model. Estimates from Panel B and C are from a fixed-effects 2SLS model. All regressions include provider, fieldworker and visit order fixed effects, as well as a constant term. Observations are at the SP-provider interaction level.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

The results follow the same format as before, expect that we decompose the effect of prescribing (i.e. removing the supply-side cost-sharing constraint) between insured and uninsured patients. Table 6 suggests some differences in prescribing choices, depending on whether patients are insured or not. While there is no difference in the probability to prescribe palliative drugs (column 1) or antibiotics (column 2), all estimates suggest that providers are likely to prescribe inappropriate treatments to insured patients (column 3), but not to uninsured patients. The ATE estimates in Panel C suggest a 45 percent increase in the probability to receive other inappropriate drugs for insured patients, while providers do not seem to change their prescribing patterns for uninsured patients[22].

Table 7: Quantity and cost of drugs, by insurance status

	(1) Total drug cost	(2) Number of drugs	(3) Average drug cost	(4) Proportion of generic drugs
Panel A. Intention-to-treat estimates				
Uninsured patient X asked prescription	35.513*	0.080	11.522**	-0.067**
	(18.849)	(0.157)	(5.165)	(0.034)
Insured patient X asked prescription	39.804**	0.135	8.765*	-0.058*
	(18.582)	(0.155)	(5.092)	(0.033)
Observations	226	226	226	226
Mean for control uninsured patients	103.165	3.850	25.055	0.897
Mean for control insured patients	99.299	3.547	29.515	0.916
Panel B. IV estimates (AT on compliers)				
Uninsured patient X asked prescription	79.873**	0.453	21.843**	-0.122**
	(31.871)	(0.304)	(8.480)	(0.057)
Insured patient X asked prescription	103.476***	0.133	26.983***	-0.141**
	(38.121)	(0.360)	(10.068)	(0.067)
Observations	226	226	226	226
Panel C. IV estimates, sub-sample of GPs dispensing to control patients				
Uninsured patient X asked prescription	94.767***	-0.088	31.681***	-0.123**
	(31.833)	(0.304)	(9.065)	(0.055)
Insured patient X asked prescription	105.644***	0.338	27.335***	-0.170***
	(34.726)	(0.331)	(9.889)	(0.060)

[22] This difference is significant (p=0.0505) for IV estimates presented in Panel C, but not for the other two specifications, probably because the study is not enough powered to detect it.

Observations	178	178	178	178
Mean for control uninsured patients	60.054	3.833	15.139	0.935
Mean for control insured patients	57.838	3.585	16.183	0.969

Notes: standard errors are in parentheses. Estimates in Panel A are from a fixed-effects linear model. Estimates from Panel B and C are from a fixed-effects 2SLS model. All regressions include provider, fieldworker and visit order fixed effects, as well as a constant term. Observations are at the SP-provider interaction level.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table 7 decomposes the effects of prescribing for insured and uninsured patients on the treatment cost. Unlike results in Table 6, we find no evidence of a difference between uninsured or insured patients. All specifications show that, when they prescribe, GPs recommend more expensive treatments to both insured and uninsured patients (column 1), driven an increase in the average price of drugs chosen (column 3), as they recommend generic drugs less frequently (column 4). Although the size of the estimated increase in total drug cost seems higher for insured patients than for cash patients, these differences are not statistically significant, and we still observe large and significant increases in costs for uninsured patients. In other words, doctors do not show altruistic concerns by choosing less costly treatments for patients who bear the full cost of the treatment.

The apparent discrepancies in the results between Table 6 and 7 is explained by the breakdown of treatment cost by category of drugs (Table AX in Appendix). The results suggest that doctors choose more expensive drugs when they prescribe, recommending more expensive antibiotics to insured patients, while the average cost of other inappropriate drugs prescribed to uninsured patients is much higher than for insured patients.

5.4. Robustness checks

In this section we discuss a number of potential concerns that may question the validity of our results.

First we test the robustness of our findings to alternative econometric specifications. Data with repeated measures for the same subjects are analysed in different ways in different disciplines to account for the correlation structure of the multiple observations (consultation) for the same subject (doctor). Alternative approaches to the fixed-effect model include OLS regression clustering errors at the physician-level; marginal model estimated through the generalised estimating equations (GEE) method and a mixed effects model where individual physicians are treated as random rather than fixed effects. We used these three alternative approaches to check the robustness of our main results presented in sections 5.2 and 5.3. We also use two alternative specifications to retrieve IV estimates: a clustered 2SLS model (the IV equivalent to the OLS model with clustered standard errors), and a random-effects IV model. Tables A4-A5 in Appendix present the results of all of these estimations. The results confirm our main findings and the size of the estimated effects remain very consistent across all specifications.

The validity of the audit study as well as our ability to retrieve a consistent IV estimates of the effect of prescribing through the encouragement design rely on a number of conditions that we discuss here.

First, we estimate the effect of prescribing in a consistent way only if the effect of prescribing is the same whether the doctor is encouraged or not. This begs the question of whether or not the encouragement itself influences the treatment decisions made when prescribing. For example, one issue with the encouragement could be that the doctor interprets the patient's request as a way to obtain more or more expensive drugs. This is unlikely to be the case for uninsured patients who pay for the treatment themselves[23], but it could potentially be a concern when doctors receive patients who are covered by a health insurance. However the absence of difference between uninsured and insured patients suggests that doctors do not interpret the prescription request in relation to the type or quantity of drugs wanted by the patient.

Another question relating to the patient prescription request is whether it influences the decisions of doctors whether or not they decide to prescribe or not. In particular, when a doctor decides to dispense despite the patient's request, does this request still influence his treatment choice? For example, if providers interpret the patient's request as a willingness to obtain more drugs, and respond positively to this perceived demand by *dispensing* more drugs, the prescribing effect would be under-estimated. The doctors we interviewed suggested that this

[23] In interviews undertaken with doctors before the experiment, doctors confirmed that this was unlikely to be what doctors would interpret for uninsured patients. On the contrary, they felt that this could be seen as a sign that cash patients would only want to buy part of the treatment.

would be very unlikely to happen, and that if doctors made this interpretation they would rather write a complementary prescription.

The identification of the prescribing effect in this encouragement design also relies on the assumption that every doctor who prescribes without encouragement, would also prescribe with encouragement. This relates to the issue of defiant behaviour. Because we employ a within-subject encouragement design, we can explore this problem in detail. Table A6 in appendix shows that we there are N=4 defiers. Two doctors prescribed to the control patient but dispensed to the patient requesting a prescription, and two dispensed drugs to the patient requesting a prescription, while they voluntarily prescribed at least part of the drugs to the control patient[24]. When we exclude these four doctors from our sample, we obtain consistent results (Tables A7-A8).

The last issue relates to the question of whether the (additional) group of doctors who responds to the encouragement by prescribing drugs are representative of the broader population of GPs, or are a self-select group of individuals with specific characteristics. This issue does not invalidate our IV estimates, but simply raises the issue of their generalisability. For example, it could be that doctors who complied with the patients' request are better, more attentive doctors in general, who could also prescribe less unnecessary treatment that have no or little benefit to patients. We use the rich data collected by the SPs on quality of care to look at the determinants of the decision to prescribe (Table A9 in Appendix). The results do not suggest that the selection process is based on obvious observable characteristics such as provider's competence or attention to patients as measured by quality of care measures, or by measures capturing altruistic concerns[25]. However it is certainly not possible to totally rule out that there are some unobserved aspects (e.g. concern for profit) that may play a role in this selection process and in the treatment decisions.

[24] Note that these defiant or contrarian behaviours may not necessarily be a reaction to the patient's request. It could also be linked to a change in circumstances (e.g. the doctor runs out of a drug they want to dispense, and therefore has to write a prescription). It could also be linked to a difference in the interpretation of the case: although the patient presents with the same symptoms, the doctor may not arrive to the same diagnosis.

[25] We use two variables to capture altruistic concerns: whether or not the doctor accepts patients with a low-insurance scheme for which doctors are reimbursed at a low rate; and whether they sometimes undertake shifts in the public sector.

6. Conclusion

This study examines whether financial incentives and providers' altruistic concerns can reduce overtreatment and improve efficiency in care. To solve the usual endogeneity problems associated with the decision of doctors to self-select into particular remuneration arrangements, and that of patients choosing certain providers, we conducted an audit study using standardised patients to create exogenous variations in financial and prosocial incentives. The results show that South African doctors prescribe a treatment worth about 25% more in the absence of a cost-sharing incentive. These gains in efficiency are achieved through a choice of cheaper drugs, including more generic drugs. Interviews with dispensing doctors confirmed that to reduce their operating costs, they are mindful to procure only cheaper drug options, including almost exclusively generic drugs.

However when they prescribe drugs, they no longer appear to be concerned with treatment costs for their patients, whether they pay for the drugs out of pocket or not. There is some suggestive evidence that doctors prescribe less inappropriate drugs to uninsured patients, but overall the treatment cost is the same regardless of whether the patient pays for it or not. This finding is at odds with claims made by South African GPs where they indicated that they would provide significantly cheaper treatments for uninsured patients (Chabikuli, Schneider et al. 2002). However, our results echo those of a similar audit study conducted in China (Lu 2014), which also found no evidence that uninsured patients get less expensive drugs[26].

Meanwhile, treatment decisions remain very inefficient, with very high levels of inappropriate drugs prescribed, including when dispensing GPs have an incentive to limit inefficiency to increase their profit. The average cost of the treatment recommended by providers in the presence of supply-side cost sharing is nearly six times more expensive than what can be considered as the cheapest recommended treatment for this case[27]. Data from the survey done with 75% of the GPs suggest that these decisions are not driven by poor knowledge or failure to diagnose the case. When shown a description of the case in a knowledge test, 84 percent of providers know that it is most likely caused by a virus. Focusing specifically on antibiotic prescription, only 20 percent of doctors believe that it would help the patient recovery. However, even if they know they are not necessary, there is suggestive evidence that GPs believe prescribing antibiotics is the norm: 60 percent of doctors think that other GPs would do it, and 57 percent think that the patient would likely not come back if they were not given antibiotics. These results

[26] In the Chinese context, doctors show altruistic concerns only when prescribing increases their income: their willingness to induce demand is lower for uninsured patients than for insured ones.

[27] Based on national guidelines, this treatment was recommended by experts and includes including paracetamol and cough suppressants. Using the cheapest available options for both drugs, it was estimated at a cost of R17.86.

could also be suggestive of habit persistence, which has been shown in several studies to explain prescription patterns by providers (Crea, Galizzi et al. 2019)

The external validity of the results is a potential concern here, since it focuses on a particular segment of providers (dispensing doctors) who agreed to take part in study, assessed with one specific patient case (a viral respiratory infection). The focus on dispensing doctors was necessary to carry out the study, since it was the specificity of these GPs' remuneration that created the opportunity to test its effects on treatment choices. Although these doctors represent about 40% of all primary care providers, they have distinct characteristics. According to data collected as part of a related study, dispensing GPs are older, located in less affluent areas, accept poorer patients, are more altruistic and charge lower rates. These observable characteristics suggest that these dispensing doctors are likely to be less sensitive to profit concerns. Hence the effect of a rationing incentive in this group is probably under-estimating its impact for more profit-oriented non-dispensing doctors. Despite concerns that informed consent from providers could introduce some selection bias in the study, the local ethics committee refused to allow us to waive doctors' informed consent and use a truly random sample. To mitigate the self-selection bias as much as possible, participants were simply told that the study would look at the determinants of "clinical decision making" in primary care, and that the fieldworkers portraying patients would pay the consultation as any normal patient. Yet it is possible that doctors who agreed to participate were more competent and altruistic providers. If that were the case, it would imply that our results may have under-estimated the effects of incentives in the larger population. Respiratory tract infections are the most common reason for primary care consultations in South Africa and in many other developing countries. There is also little reason to believe that doctors would make inefficient treatment decisions for these illnesses, but not for others. If anything, the treatment for ARTIs is simple and does not require any costly investigation or long courses of expensive drugs.

At a time when the government of South Africa is considering ways to contract private providers to expand the offer of free primary care services to the population, the study provides both encouraging and concerning evidence. Our results suggest that bundled payments for consultation and drugs that shift some of the financial risk of treatment decisions onto primary care providers could reduce overprovision due to higher drug costs than appropriate[28]. As the results likely reflect a general pattern of inefficient prescribing by private South African private GPs, the government should consider such payments to encourage savings in the context of a future national health insurance. A potential pitfall of provider

[28] Alternatively, and for GPs who do not dispense drugs, an equivalent solution would be to regulate doctors' choice of drugs by restricting the list of drugs to be reimbursed by a national health insurance fund.

cost-sharing arrangements is that, in the absence of monitoring, they rely on professionalism or providers' altruistic concerns for patient welfare not to skimp on quality and under-provide efficient treatment. In the case of viral infections studied here, it is not a concern. Also, because GPs can still decide to prescribe and release the constraint of cost-sharing, there might enough flexibility in the current arrangement to avoid the main drawbacks.

Our results suggest that plans for expanding coverage should include discussions to introduce incentives to reduce the large levels of efficiency observed. Although the study focused on private providers, there is no reason to believe that the high levels of inefficiency observed were specific to the private sector. In fact, in a study related to this one, we found that inefficiency in treatment decisions were equally rampant in the public sector, despite a greater institutional reference to and use of clinical guidelines. However, the treatment costs in the private sector were much higher than in the public sector. This confirms the need for a future national health insurance as well as private insurers to consider the introduction of cost-containment measures to encourage more rational prescribing.

While a lot of attention has been focused on increasing access to more services to a greater number in low- and middle-income countries, this study cautions against the financial implications of such reforms in the absence of cost-containment measures. At a time when governments consider contracting private GPs to deliver primary care services, this study suggests that innovative supply-side cost-sharing remuneration could generate considerable savings. It remains to be seen whether such payment mechanisms provide enough flexibility to limit the risk that providers skimp on the quality of care and under-provide effective treatment.

References

Angrist, J. D., G. W. Imbens and D. B. Rubin (1996). "Identification of Causal Effects Using Instrumental Variables." J. Am. Stat. Assoc. **91**(434): 444--455.

Ataguba, J. E. and D. McIntyre (2012). "Modelling the affordability and distributional implications of future health care financing options in South Africa." Health Policy and Planning **27**(suppl_1): i101-i112.

Ataguba, J. E. and D. McIntyre (2018). "The incidence of health financing in South Africa: findings from a recent data set." Health Economics, Policy and Law **13**(1): 68-91.

Balafoutas, L., A. Beck, R. Kerschbamer and M. Sutter (2013). "What Drives Taxi Drivers? A Field Experiment on Fraud in a Market for Credence Goods." The Review of Economic Studies **80**(3): 876-891.

Balafoutas, L., R. Kerschbamer and M. Sutter (2017). "Second-Degree Moral Hazard In A Real-World Credence Goods Market." Econ. J. **127**(599): 1--18.

Barnett, M. L. and J. A. Linder (2014). "Antibiotic Prescribing for Adults with Acute Bronchitis in the United States, 1996–2010." JAMA **311**(19): 2020.

Bertrand, M. and E. Duflo (2017). Field Experiments on Discrimination (chapter 8). Handbook of Field Experiments. E. Duflo and A. Banerjee.

Berwick, D. M. and A. D. Hackbarth (2012). "Eliminating Waste in US Health Care." JAMA **307**(14): 1513--1516.

BHF. (2012). "Board of healthcare funders of Southern Africa." Retrieved 01/03/2012, from www.bhfglobal.com.

Brink, A. J., J. Van Wyk, V. M. Moodley, C. Corcoran, P. Ekermans, L. Nutt, T. Boyles, O. Perovic, C. Feldman, G. Richards and M. Mendelson (2016). "The role of appropriate diagnostic testing in acute respiratory tract infections: An antibiotic stewardship strategy to minimise diagnostic uncertainty in primary care." S. Afr. Med. J. **106**(6): 30--37.

Brownlee, S., K. Chalkidou, J. Doust, A. G. Elshaug, P. Glasziou, I. Heath, S. Nagpal, V. Saini, D. Srivastava, K. Chalmers and D. Korenstein (2017). "Evidence for overuse of medical services around the world." The. Lancet **390**(10090): 156--168.

Busse, R. (2012). "Do diagnosis-related groups explain variations in hospital costs and length of stay? – Analyses from the Eurodrg project for 10 episodes of care across 10 European countries." Health Economics **21**(S2): 1--5.

Chabikuli, N., H. Schneider, D. Blaauw, A. B. Zwi and R. Brugha (2002). "Quality and equity of private sector care for sexually transmitted diseases in South Africa." Health Policy Plan. **17**(Suppl): 40--46.

Competition Commission South Africa (2018). Health market inquiry. Provisional findings and recommendations report. Johannesburg, Competition Commission South Africa.

Crea, G., M. M. Galizzi, I. Linnosmaa and M. Miraldo (2019). "Physician altruism and moral hazard: (no) Evidence from Finnish national prescriptions data." Journal of Health Economics **65**: 153-169.

Currie, J., W. Lin and J. Meng (2014). "Addressing antibiotic abuse in China: An experimental audit study." Journal of Development Economics **110**: 39-51.

Currie, J., W. Lin and W. Zhang (2011). "Patient knowledge and antibiotic abuse: Evidence from an audit study in China." Journal of Health Economics **30**(5): 933-949.

Daniels, B., A. Dolinger, G. Bedoya, K. Rogo, A. Goicoechea, J. Coarasa, F. Wafula, N. Mwaura, R. Kimeu and J. Das (2017). "Use of standardised patients to assess quality of healthcare in Nairobi, Kenya: a pilot, cross-sectional study with international comparisons." BMJ Global Health **2**(2).

Das, J., A. Holla, V. Das, M. Mohanan, D. Tabak and B. Chan (2012). "In urban and rural India, a standardized patient study showed low levels of provider training and huge quality gaps." Health Aff (Millwood) **31**(12): 2774-2784.

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

Ellis, R. P. and T. G. McGuire (1993). "Supply-Side and Demand-Side Cost Sharing in Health Care." Journal of Economic Perspectives **7**(4): 135-151.

Emanuel, E. J. and V. R. Fuchs (2008). "The perfect storm of overutilization." JAMA **299**(23): 2789--2791.

Goldacre, B., C. Reynolds, A. Powell-Smith, A. J. Walker, T. A. Yates, R. Croker and L. Smeeth (2019). "Do doctors in dispensing practices with a financial conflict of interest prescribe more expensive drugs? A cross-sectional analysis of English primary care prescribing data." BMJ Open **9**(2): 026886.

Health 24 (2015). Outrage as SA Health Department discourages doctors from dispensing medication. https://www.health24.com/News/Public-Health/Outrage-as-SA-Health-Dept-discourages-doctors-from-dispensing-medication-20150320.

Iizuka, T. (2007). "Experts' Agency Problems: Evidence from the Prescription Drug Market in Japan." The RAND Journal of Economics **38**(3): 844-862.

Iizuka, T. (2012). "Physician Agency and Adoption of Generic Pharmaceuticals." American Economic Review **102**(6): 2826-2858.

Kaiser, B. and C. Schmid (2016). "Does Physician Dispensing Increase Drug Expenditures? Empirical Evidence from Switzerland." Health Econ. **25**(1): 71--90.

Kerschbamer, R., D. Neururer and M. Sutter (2016). "Insurance coverage of customers induces dishonesty of sellers in markets for credence goods." Proceedings of the National Academy of Sciences **113**(27): 7454-7458.

Koch, S. F. and K. Omotoso (2017). "South African trends in medical aid coverage and stated healthcare-seeking preferences: 2004–14." Development Southern Africa **34**(5): 575-592.

Leibowitz, A., W. G. Manning and J. P. Newhouse (1985). "The demand for prescription drugs as a function of cost-sharing." Social Science & Medicine **21**(10): 1063--1069.

Lim, D., J. Emery, J. Lewis and V. B. Sunderland (2009). "A systematic review of the literature comparing the practices of dispensing and non-dispensing doctors." Health Policy **92**(1): 1--9.

Liu, Y.-M., Y.-H. K. Yang and C.-R. Hsieh (2009). "Financial incentives and physicians' prescription decisions on the choice between brand-name and generic drugs: Evidence from Taiwan." Journal of Health Economics **28**(2): 341--349.

Lu, F. (2014). "Insurance coverage and agency problems in doctor prescriptions: Evidence from a field experiment in China." Journal of Development Economics **106**: 156--167.

Lundin, D. (2000). "Moral hazard in physician prescription behavior." Journal of Health Economics **19**(5): 639--662.

Lurie, N., J. Christianson, M. Finch and I. Moscovice (1994). "The effects of capitation on health and functional status of the Medicaid elderly. A randomized trial." Ann. Intern. Med. **120**(6): 506--511.

Mohanan, M., M. Vera-Hernandez, V. Das, S. Giardili, J. D. Goldhaber-Fiebert, T. L. Rabin, S. S. Raj, J. I. Schwartz and A. Seth (2015). "The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural India." JAMA Pediatr **169**(4): 349-357.

Moreno-Serra, R. and A. Wagstaff (2010). "System-wide impacts of hospital payment reforms: Evidence from Central and Eastern Europe and Central Asia." Journal of Health Economics **29**(4): 585--602.

Mott, D. A. and R. R. Cline (2002). "Exploring Generic Drug Use Behavior: The Role of Prescribers and Pharmacists in the Opportunity for Generic Drug Use and Generic Substitution." Medical Care **40**(8): 662--674.

National Department of Health (2015). White paper on national health insurance. Towards universal health coverage. Pretoria, National Department of Health.

NDoH, Stats SA, SAMRC and IC (2019). South Africa Demographic and Health Survey 2016. Pretoria, South Africa and Rockville, Maryland, USA, NDoH, Stats SA, SAMRC, and ICF.

Planas, M.-E., P. J. García, M. Bustelo, C. P. Carcamo, S. Martinez, H. Nopo, J. Rodriguez, M.-F. Merino and A. Morrison (2015). "Effects of Ethnic Attributes on the Quality of Family Planning Services in Lima, Peru: A Randomized Crossover Trial." PLoS One **10**(2): e0115274.

Sylvia, S., Y. Shi, H. Xue, X. Tian, H. Wang, Q. Liu, A. Medina and S. Rozelle (2015). "Survey using incognito standardized patients shows poor quality care in China's rural clinics." Health Policy and Planning **30**(3): 322-333.

Troeger, C. and GBD 2016 Lower Respiratory Infections Collaborators (2018). "Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016." The Lancet Infectious Diseases **18**(11): 1191-1210.

Trottmann, M., P. Zweifel and K. Beck (2012). "Supply-side and demand-side cost sharing in deregulated social health insurance: Which is more effective?" Journal of Health Economics **31**(1): 231--242.

WHO (2014). Antimicrobial Resistance Global Report on Surveillance. Geneva, Switzerlnad.

Appendices

Appendix 1: Standardised patient case

The patient opens the consultation with the following statement: "*I have been coughing a lot. I had a cold about a week ago. I am a bit better now but the cough is not going away*."

Upon relevant questioning, the patient reveals that the cold subsided after 4-5 days, and the only persisting symptoms are the cough, tiredness and a slightly irritated throat. Appropriate history-taking by the doctor should rule out a number of ailments (including TB, pneumonia and asthma or allergies), and a physical examination should further confirm this (in particular the absence of crackling noise characteristic of pneumonia). The likely viral nature of the condition can be inferred from the lack of absence of physical symptoms and fever, in particular the absence of coloured sputum, as well as the short duration of reported symptoms.

Appendix Tables

Table A1: Balance check

	(1) Received uninsured patients		(2) Received insured patients		(3) Difference	
	Mean	SD	Mean	SD	Diff.	p-val.
Panel A: original sample						
Age	51.95	(10.79)	50.75	(11.06)	1.20	0.55
Male	0.75	(0.44)	0.68	(0.47)	0.07	0.42
Practice location						
Practice is located in Q1 area	0.13	(0.34)	0.07	(0.25)	0.07	0.23
Practice is located in Q2 area	0.22	(0.42)	0.20	(0.40)	0.02	0.82
Practice is located in Q3 area	0.10	(0.30)	0.15	(0.36)	-0.05	0.41
Practice is located in Q4 area	0.45	(0.50)	0.35	(0.48)	0.10	0.27
Practice is located in Q5 area	0.10	(0.30)	0.23	(0.43)	-0.13	0.05*
Observations	60		60		120	
Panel B: final sample						
Age	52.12	(10.91)	50.76	(10.85)	1.36	0.52
Male	0.75	(0.44)	0.68	(0.47)	0.07	0.41
Practice location						
Practice is located in Q1 area	0.13	(0.34)	0.08	(0.27)	0.06	0.32
Practice is located in Q2 area	0.22	(0.42)	0.21	(0.41)	0.01	0.91
Practice is located in Q3 area	0.10	(0.30)	0.13	(0.34)	-0.03	0.60
Practice is located in Q4 area	0.45	(0.50)	0.34	(0.48)	0.11	0.24
Practice is located in Q5 area	0.10	(0.30)	0.25	(0.43)	-0.15	0.04**
Observations	60		53		113	

Note: the final sample excludes the 7 providers for whom one of the two SPs did not portray the role that they had been assigned.

Table A2: **Non-compliance of fieldworkers to SP role assigned**

	SP role assigned				
SP role played	Uninsured patient	Uninsured patient, prescription	Insured	Insured, prescription	Total
Uninsured patient	60	0	0	0	60
Uninsured patient, prescription	0	60	0	3	63
Insured	0	0	59	3	62
Insured, prescription	0	0	1	54	55
Total	60	60	60	60	240

Table A3: **predictors of prescribing to control patients**

	(1) Prescribed (1=yes)	(2) Prescribed (1=yes)
No. of items completed	0.014	-0.000
	(0.009)	(0.010)
Consultation duration	-0.003	-0.001
	(0.007)	(0.008)
Age		-0.002
		(0.004)
Accepts low-insurance client		-0.052
		(0.088)
Also does sessions in public sector		-0.054
		(0.129)
Works in a group practice		0.147
		(0.096)
Consultation rate		0.001*
		(0.000)
Observations	113	86

Notes: standard errors are in parentheses. Estimates are marginal effects from a probit regression model.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.