The London School of Economics and Political Science

Rationalising Health Care Provision under Market Incentives: Experimental Evidence from South Africa

Arthika Sripathy

A thesis submitted to the Department of Health Policy of the London School of Economics for the degree of Doctor of Philosophy

London, May 2020

Declaration of Authorship

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 79,570 words, excluding references and appendices.

Statement of funding sources

I confirm that a part of the fieldwork for Chapter 7 was funded by UK Aid from the UK Department for International Development (DFID), through the Consortium for Resilient and Responsive Health Systems (RESYST). However, the views expressed and information contained in it are not necessarily those of or endorsed by DFID, which can accept no responsibility for such views or information or for any reliance placed on them.

All other fieldwork for this thesis was funded by the Economic and Social Research Council (ESRC) and the Wellcome Trust.

Acknowledgements

The three years of this PhD have been truly transformative, both personally and intellectually. I found love, got married, built a home, and found a job that keeps me learning every day. Intellectually, these years have been an indulgence – they gave me the time and mind space to pursue diverse ideas, to question and grow in the way I think. Such growth would have been impossible without the generosity and support of many.

I must begin by sincerely thanking my supervisor, Mylene Lagarde, whose support made this study both possible and enriching. Thank you for always being generous with your time, good humour, and encouragement; and for calmly entertaining all my totally infeasible research ideas. Your "punchy" advice on my verbose prose was invaluable. Thanks to you, I feel lucky enough to say that my biggest challenge in completing this PhD has been around trimming word count.

I am also grateful to Prof. Elias Mossialos for his generous support and mentorship in starting this PhD journey. Prof. Pierre-Yves Geoffard, at the Paris School of Economics, who first introduced me to health economics and the topics covered in this thesis – thank you for inspiring my interest and setting me on this PhD path.

I gratefully acknowledge financial support from the LSE Department of Health Policy, the Wellcome Trust, the ESRC and DfID. The fieldwork for this thesis was conducted in Johannesburg, and I am indebted to Duane Blaauw and Ling Ting at the Wits Centre for Health Policy for collaborating on this research, and for welcoming and making me feel at home there.

Over the course of this PhD, I have also benefitted from the advice and encouragement of several faculty members and fellow students at the LSE Departments for Health and Social Policy. Special thanks to Prof. Anne West, for her genuine care and advice; to Profs. Alistair McGuire and Andrew Street, for providing valuable feedback; and to my PhD cohort in both departments, for sharing in triumphs and frustrations, and for moral support in the form of an occasional drink at the George. My desk neighbours, Ellie, Laura and Sangjune – thank you for always lending an ear and advice when needed. And to other friends I found at the LSE - especially Sam, Felix, Maria, Priya, Josephine, Michael and Josh – thank you for being a tremendous source of laughter and sanity.

Heartfelt thanks also to long-suffering friends and family in London and further afield, for putting up with my dull PhD rants, and the one-too-many catch-ups I've either cancelled or turned up unfashionably late to (with yet another lame PhD excuse). I promise to be much, much better.

Finally, I dedicate this thesis to my parents and Giancarlo, for their unconditional support, optimism, and sense of humour throughout these years. Thank you for propping me up when I could find no strength of my own – for always being the glass half-full to my glass half-empty.

Abstract

Unnecessary medical treatments place a significant burden on health systems striving for universal health coverage (UHC). This thesis studies inappropriate treatment incentives in the private sector in South Africa, where plans to implement a national health insurance system (NHI) foresee the contracting of private physicians to deliver publicly-funded health care. Private providers are increasingly recognized as necessary partners for UHC success in many low-and-middle-income countries (LMIC). However, aligning the incentives of these actors with UHC and public health goals requires a better understanding of incentive effects in these settings.

I conduct two field experiments with incognito standardized patients (SPs), to both evaluate appropriate care provision and experimentally vary the treatment incentives facing private physicians. First, I run a within-subject experiment with 89 private primary care physicians (GPs) in Johannesburg, to investigate the causal impact of improving patients' financial protection (insurance cover) on physicians' quality of care delivery. The results suggest that more insured patients receive a higher level of visible clinical effort, but a lower level of technical care quality – including a higher likelihood of inappropriate antibiotic treatment. Second, I use data from the same experiment to evaluate the impact of patient insurance on the quantity and costs of care. I find that more insured patients are more likely to receive unnecessary diagnostic tests and treatment procedures, and receive more and more expensive branded drugs, resulting in significantly higher care costs. The results on antibiotic treatment and drug treatment quantity and costs occurred despite the absence of any financial incentives attached to drug prescribing for GPs, which suggests the presence of alternative motives for physicians' treatment decisions that might vary with patient insurance – including intrinsic or altruistic motives. Third, I explore the scope for leveraging such intrinsic motivations to improve physicians' treatment choices. I conduct a randomized (between-subject) experiment with 80 GPs, to evaluate the impact of intrinsic, informational incentives from private performance audit and feedback (A&F) on physicians' antibiotic treatment choices and care costs. The findings suggest that private A&F can significantly reduce the likelihood of inappropriate antibiotic treatment for common viral infections that present in primary care, without simultaneously reducing appropriate antibiotic use for bacterial infections or increasing other inappropriate drug treatments. However, improved performance on antibiotic use does not coincide with significantly lower treatment costs or any improvements in measured diagnostic effort or accuracy. There is indicative evidence that prescribing norms and perceived patient expectations may play an important role in mediating private physicians' treatment choices in all three empirical chapters.

Table of Contents

List of Tables

List of Figures

Abbreviations and Acronyms

ABR	Antibiotic Resistance
A&F	Audit & Feedback
AMR	Antimicrobial Resistance
ANC	African National Congress
API	Active Pharmaceutical Ingredient
CCSA	Competition Commission of South Africa
CoJ	City of Johannesburg
CU	Contracting Unit (for Primary Health Care)
EDL	Essential Drugs List
EHR	Electronic Health Records
ESRC	Economic and Social Research Council
FFS	Fee-for-Service
GDP	Gross Domestic Product
GP	General Practitioner
IRT	Item Response Theory
LMEM	Linear Mixed Effects Model
LMIC	Low and Middle Income Countries
LRTI	Lower Respiratory Tract Infection
MSA	Medical Savings Account
MPR	Medicine Price Registry
MRR	Maximum Reimbursement Rate
NHI	National Health Insurance
OLS	Ordinary Least Squares
OOP	Out-of-Pocket
OTC	Over-the-Counter
P4P	Pay-For-Performance
Q&A	Question & Answer
RCT	Randomised Controlled Trial
RTI	Respiratory Tract Infection
SA	South Africa
SD	Standard Deviation
SEP	Single Exit Price
SID	Supplier-Induced Demand
SP	Standardised Patient

STG	Standard Treatment Guidelines
UHC	Universal Health Coverage
UK	United Kingdom
URTI	Upper Respiratory Tract Infection
US	United States (of America)
USD	United States Dollar
VAT	Value Added Tax
WHO	World Health Organisation
YLL	Years of Life Lost

1 Introduction

Unnecessary medical treatments are avoidable contributors to rising health expenditure around the world. Despite the widespread availability of evidence-based treatment guidelines, scarce resources are frequently misused in the provision of clinically inappropriate and inefficient healthcare (Rosenberg *et al.*, 2015). Recent literature has drawn attention to this problem of "overuse" in health systems striving for universal health coverage (UHC), where money spent on harmful, ineffective or expensive care is money diverted from other essential services (Berwick, 2017).[1] While aggregate global costs are unknown, an estimate from the US is indicative: over USD 270 billion of healthcare spending in 2013 was estimated to be overuse (Brownlee *et al.*, 2017). This type of waste tends to co-exist with the underuse of clinically appropriate yet inexpensive therapies, poor overall quality of care, and significant unmet healthcare need, particularly in low and middle-income countries (LMIC) (Saini *et al.*, 2017; Das *et al.*, 2018).

The inappropriate use of antibiotics is perhaps the best documented form of such overuse globally (Brownlee *et al.*, 2017). The emerging economies of Brazil, Russia, India, China and South Africa accounted for almost 76% of the increase in antibiotic consumption worldwide between 2000 and 2010 (Van Boeckel *et al.*, 2014). In South Africa – the context of this thesis – antibiotic use escalated by 175% over the same period (Republic of South Africa, 2015). Comparable studies have documented inappropriate antibiotic prescriptions for common viral infections in 55% of patient cases in China (Currie *et al.*, 2014), in 62% of cases in India (Dehn Lunn, 2018), and in 72% of cases in South Africa (Lagarde and Blaauw, 2019).[2] The result in South Africa occurred despite the presence of a robust antibiotic regulatory system in line with WHO recommendations (WHO, 2001, 2015; Holloway *et al*., 2016), including published standard treatment guidelines and diagnostic frameworks to rationalise antibiotic use (NDoH, 2014; Brink *et al.*, 2016).

This thesis studies what drives such overtreatment at the level of physician-patient encounters. It explores this question in the context of a renewed global push for universal health coverage (UHC), where improved financial protection for patients and universal access to basic health services have become central aspirations of many health systems – and concurrently, where there

[1] Some stark estimates of overuse include 6.2 million unnecessary caesarean sections each year, at a cost of USD 2.32 billion (50% of these in China and Brazil) (Gibbons *et al.*, 2010); over 22% of percutaneous coronary interventions in Israel; and up to 60% of endoscopies in the US (Brownlee *et al.*, 2017).

[2] In the South African study, the rate of inappropriate antibiotic prescribing was 78% in the public sector and 67% in the private sector.

is scarce evidence that improving access to health services consistently ensures the delivery of necessary, clinically-effective care (Scott and Jha, 2014; Das *et al.*, 2018).

In most countries, the private sector tends to play a prominent role in health service delivery. For instance, private providers account for over 50% of all reported treatment for diarrhoea and cough in children in sub-Saharan Africa (see Appendix A.1) (Wadge *et al.*, 2017).[3] With limited public sector capacity to expand healthcare access, LMIC governments in particular increasingly rely on private providers to deliver publicly financed healthcare. In South Africa, government plans to implement an ambitious national health insurance system by 2026 similarly foresee the contracting of private physicians to expand access to free primary healthcare, and to function as gatekeepers to the national health system (Republic of South Africa, 2019). This creates a context where healthcare provision is increasingly subject to market incentives, and patients (healthcare consumers) are progressively more sheltered from the costs of healthcare overuse. In this setting, the contractual or implicit agency relationships between physicians, patients and third-party payers - how they are structured and incentivised - are likely to have important influences on healthcare provision.

This thesis therefore explores agency problems in the provision of health services in a LMIC setting, where physicians are subject to market incentives. There is comparative evidence that private providers generate more costly care than the public sector (Pongsupap and Lerberghe, 2006; Blaauw and Lagarde, 2019), but little evidence that rates of over-treatment are any worse – particularly with respect to antibiotic treatment – and indeed some evidence that certain aspects of care quality are better, which makes it unclear whether the higher costs simply price in higher quality (Pongsupap and Lerberghe, 2006; Das *et al.*, 2016; Blaauw and Lagarde, 2019). Existing studies on incentives for provider behaviour in LMIC settings tend to focus on the public sector, so little is known about incentive effects in the private sector (Das and Hammer, 2014). Moreover, despite growing third-party payment in healthcare, there is scarce research on how these payers' efforts to regulate healthcare decisions on the supply and demand sides can interact (Bardey and Lesur, 2006) – for instance, how patient cost-sharing may also incentivise provider decisions. By focusing on the microenvironment of individual physician-patient interactions, this thesis investigates the role of private physicians, the incentives they respond to (including possible incentives from patient cost-sharing), and how this can inform policy choices for rationalising healthcare use in increasingly resource-constrained settings.[4]

---

[3] Equivalent figure are much higher in South Asia and South East Asia (around 80% and 65%), respectively. Notably, even in the poorest quintiles of the populations, the majority of these private providers from whom people report seeking care are formal, for-profit providers (not informal or non-profit providers) (Wadge *et al.*, 2017).

[4] I restrict my analysis to the formal, for-profit segment of the private healthcare sector in South Africa. The relative sizes of the informal or non-profit sectors are unknown; however, they are likely to be insignificant in the urban setting of this study (Johannesburg).

The remainder of this introduction is structured as follows. Section 1.1 sets out the definitions and methods employed for analysing "appropriate" healthcare provision in the following chapters. Sections 1.2 and 1.3 outline the two related sources of health market failure that motivate this thesis: *i)* the combination of asymmetric information and externalities in healthcare provision, particularly in antibiotic treatment; and *ii)* the combination of asymmetric information and the dual role of physicians as diagnostic and treatment providers. Section 1.4 discusses a role for physicians' intrinsic motivations in mediating these market failures. Finally, Section 1.5 presents the main research objectives and a roadmap for this thesis.

1.1 Understanding "Appropriate" Healthcare Provision: Definitions and Challenges

1.1.1 Terminology

Defining *appropriate* healthcare is the first step to investigating over-treatment. This study adopts a definition that is synonymous with *"right care"*, as characterised by Vikas Saini and colleagues (Saini *et al.*, 2017). Broadly speaking, it refers to care choices that fulfil three criteria (WHO, 2000): *i) evidence-based* (safe and clinically effective), which addresses the *technical quality* and clinical *necessity* of care; *ii) cost-effective* (delivers targeted outcomes at least cost); and *iii) patient-centred* (responsive to individual patient needs, constraints and preferences).[5] The notion of appropriate healthcare therefore captures the considerable linkages between quality and efficiency in healthcare spending.[6]

Any deviations from this standard of care may be considered *overuse* or *under-use* – in other words, *inappropriate* care. *Overuse* is used synonymously with *over-treatment* and *over-provision* in this study, and refers broadly to the delivery of "any services that are unnecessary in any way" (Brownlee *et al.*, 2017).[7] This includes "the provision of medical services that are more likely to cause harm than good" (Brownlee *et al.*, 2017). It also includes the provision of necessary yet cost-ineffective care - that is, care choices that are clinically indicated, but where there are cheaper alternatives available with equivalent benefit (a prime example being the use of branded drugs over accessible generic equivalents). This latter form of overuse can also be termed *over-*

[5] Note that "appropriate healthcare" is used interchangeably with "appropriate (physician) effort" in this study.

[6] For instance, while quality cannot increase indefinitely without increasing cost, cost-effectiveness necessarily requires that care quality be maximised for a given cost. Moreover, going beyond a single point of care use, poor quality care can lead to higher costs further down the patient care pathway. For example, a physician that fails to take the time to understand a particular patient's ailment and preferences, and tailor a therapeutic regime accordingly, might save time (and cost) during that single encounter. However, higher future costs may be incurred if subsequent treatment failure or patient non-adherence leads to more serious health complications down the line.

[7] While *overuse* can also refer to patients' healthcare choices, this thesis focuses only on physician behavior. Therefore, the term can be used synonymously with *overtreatment* and *overprovision* here.

charging. Conversely, *underuse* indicates "the failure to use effective and affordable medical interventions (or care processes)" (Glasziou *et al.*, 2017), and is used interchangeably with *under-treatment* and *under-provision*.[8]

1.1.2 Challenges in measurement and identification

Studying the drivers of over-treatment under these definitions poses two challenges. The first is a problem of *measurement*. Appropriate care – as defined above - is necessarily patient and case-specific (Das and Hammer, 2014). For the researcher or third-party regulator, systematically determining what is appropriate care in a given clinical context - to identify any unnecessary treatment - is problematic on the basis of direct observations, surveys or administrative data alone.[9] For a defined clinical case, the first two criteria of appropriate care (evidence-based and cost-effective) may be established to some extent from validated national or international clinical guidelines. However, defining the clinical case itself on the basis of what physicians tell patients, or enter in their administrative records, has its limitations – given room for diagnostic error, imprecise clinical coding, and the conflict of interest for the treating physician who also gives the diagnosis (which in turn, determines the assessment of his treatment choices). Moreover, even where the clinical case can be defined with reasonable certainty, commonly-used methods for measuring inappropriate treatment have additional scope for measurement error: direct observation and clinical vignettes can induce Hawthorne effects (Leonard and Masatu, 2010), administrative data can be of very low quality in some LMIC settings, and patient surveys can suffer from recall bias and patient misunderstanding of clinical procedures (Onishi *et al.,* 2010).

A second challenge is one of *identification*. Empirically answering the question of *who* primarily drives healthcare choices at the physician-patient level, and *what* drives them, is a challenge. Healthcare is widely characterised as an expert service, where patients cannot perfectly assess the quality or necessity of the care they receive (Darby and Karni, 1973). In this view, 'expert' physicians inevitably have considerable influence on how healthcare resources are consumed. The theory of supplier-induced demand (SID) (Evans, 1974; Fuchs, 1978) accordingly argues that, under market provision of healthcare, profit-maximising physicians drive over-consumption (given the informational asymmetries and basis of trust in the physician-patient relationship). Nevertheless, there is little robust, causal evidence to support this view. Fundamentally, the

[8] While Saini *et al.* (2017) highlight the concurrent prevalence of overuse and underuse at the health system-level, these issues can also co-exist in the treatment of the same patient. For instance, physicians' underuse of necessary diagnostic processes (including the necessary patient history-taking or examination) can lead to diagnostic uncertainty or inaccuracy, and potential overuse of incorrect treatments (overtreatment) for the same case.

[9] In OECD countries, electronic health records (EHR) and billing data are widely used in quality and cost monitoring. However, in most LMIC settings EHR systems are rare and administrative data quality is often very poor. Therefore alternative methods of quality measurement have been used in these settings, including direct observations, patient exit interviews, and standardized patients (SP) (Das and Hammer, 2014)

healthcare choices observed from individual clinical encounters may be due to the influence of the patient (patient demand, or other factors) or the physician (demand inducement, or otherwise); the problem is in separating the two. This issue has received little empirical attention (Currie *et al.*, 2014; Lu, 2014; Gottschalk *et al.*, 2018). Moreover, evidence of significant over-treatment also in the public sector (with salaried providers) runs counter to the SID characterisation of such behaviour being purely profit-motivated (Mohanan *et al.*, 2015; J. Das *et al.*, 2016; Lagarde and Blaauw, 2019). Most existing studies on SID are observational, and rely on aggregated or administrative data, which makes it impossible to control for all potential confounders of treatment decisions in clinical encounters and isolate the relative influence and motivations of the physician.

To overcome these two challenges, this thesis draws on an audit study approach, using under-cover standardised patients (SP) (Rethans *et al.*, 2007). By standardising all patient and case-specific drivers of treatment choices, this approach allows appropriate care to be determined *ex ante*, and evaluated *ex post*.[10] Moreover, by allowing experimental variation in aspects of the patient and clinical case presentation of research interest, it allows the researcher to not only identify the physician's influence in driving healthcare choices, but also to explore their potential motivations in doing so. The SP method is commonly recognised as the "gold standard" in care quality measurement (Das and Hammer, 2014). This study builds on a growing body of observational and field experimental studies using SPs to evaluate the clinical performance of healthcare providers in LMIC settings (Madden *et al.*, 1997; Kwan *et al.*, 2019). Much of this literature has used SPs to describe and compare the quality of healthcare providers in different settings (Pongsupap and Lerberghe, 2006; Mohanan *et al.*, 2012, 2015; Sylvia *et al.*, 2014; J. Das *et al.*, 2016; Daniels *et al.*, 2017, 2019; Christian *et al.*, 2018). Only a handful of studies have adopted an experimental approach, using SPs to exogenously vary and investigate specific drivers of providers' treatment behaviours (Currie *et al*., 2011; Currie *et al.*, 2014; Lu, 2014; Gottschalk *et al*., 2018). The use of SPs in impact evaluations of randomised provider quality interventions is similarly scarce (Das *et al.*, 2016; Mohanan *et al.*, 2017; Harrison *et al*., 2000; Mathews *et al*., 2009).[11] This thesis specifically contributes to these latter, more nascent applications of SPs in field experimental research.

[10] In this study, the 2014 edition of the South African Standard Treatment Guidelines (STG) and Essential Medicines List (EML) was employed to determine appropriate treatments for the clinical cases presented by SPs (see Section 4.1 for further details on the SP clinical cases used in this study). These guidelines are updated every five years by the national Department of Health, and serve as useful benchmarks for evidence-based and cost-effective treatments in the South African context. Appropriate diagnostic procedures (for the defined clinical cases) were also established in procedural checklists developed by the research team in consultation with clinical experts in the relevant fields (see Section 4.2.1 for further details).

[11] Another two studies either use SPs to evaluate interventions targeting provider attitudes (rather than their care quality performance, per se) (Li, Lin and Guan, 2014) or use non-blinded SP visits where providers are aware of being assessed (which risks a Hawthorne effect confounding the estimated intervention impact) (Sanci *et al.*, 2000).

1.2 Market Provision of Healthcare under Externalities

Informational asymmetries – between the physician and patient, and between the physician and healthcare regulators (government or third-party payer) – are fundamental features of healthcare markets. They result in what Ma and McGuire (1997) consider a missing market in healthcare: physician payment on the basis of clinical effort. While healthcare quantity may be easily observable and (somewhat) contractible,[12] clinical effort is not (McGuire 2000). As neither patients nor third-party regulators can accurately determine or contract for appropriate physician effort, a physician's supply of such effort in any given patient case will entail some private cost and positive externality for which he is not compensated. This suggests that appropriate effort will in general be under-supplied relative to what is in patients and payers' best interests.

Private physicians also have clear incentives to respond to market demand, even where that demand does not reflect appropriate care. For instance, patients are commonly said to demand antibiotics without any clinical indication (Dempsey *et al.*, 2014; Fletcher-Lartey *et al.*, 2016; Farley *et al.*, 2018), and found to prefer branded prescription drugs over generic equivalents despite the higher costs (Himmel *et al.*, 2005; Shrank *et al.*, 2009). Where payment for health services is conditioned on patient or service volume, private physicians have direct incentives to *over-supply* those services that they believe patients demand (or where any negative demand response to over-supply is expected to be low). Conversely, where no payment incentives are attached, physicians may *under-supply* aspects of care where there is little demand response – that is, aspects of care that patients cannot easily assess or respond to, such as the *technical quality* of care. Haas-Wilson (1994) highlights a distinction between the observable, interpersonal aspects and amenities of care, and its less observable technical quality – a distinction that affects the extent to which physicians can 'price in' the quality they supply in patient fees. Das *et al.* (2016) also find that market-determined prices reward observable aspects of care (such as time spent with the patient, the extent of history-taking and examination, and the quantity of medications), but not its unobserved technical quality (the accuracy of diagnoses, or appropriateness of treatments). Studies comparing ambulatory care in the public and private sectors of LMIC concur that, while the private sector performs better in terms of the amenities and observable aspects of care that patients would more readily value (including waiting and consultation times, and patient communication), there is little difference in technical quality (Pongsupap and Lerberghe, 2006; Das *et al.*, 2016; Blaauw and Lagarde, 2019). [13]

[12] Ma and McGuire (1997) argue that even healthcare quantity may not be perfectly contractible due to the possibility of fraud in what is reported to healthcare payers.

[13] Indeed, Blaauw and Lagarde (2019) find that private providers are more likely than public providers to provide *incorrect* diagnoses for the same clinical case – although they are also more likely to communicate *any* diagnosis to the patient.

The non-contractibility of (unobserved) appropriate effort is particularly problematic in the treatment of infectious diseases - including in the rational use of antibiotics. Determining whether antibiotic treatment is necessary for a particular patient case will require some costly clinical effort from the physician,[14] who is not compensated for this - nor for any broader benefits this effort will have on public health, in containing unnecessary growth in antibiotic resistance. The economic theory on externalities argues that market provision of goods that generate benefits (costs) to third-parties, that are unaccounted for in market prices, will lead to under-provision (over-provision).[15] The non-contractibility of appropriate care in general, and the public health spill-overs from appropriate antibiotic use in particular, therefore predict that antibiotics will often be misused relative to the patient's and public's best interests.[16]

Taken together, both theory and evidence support the conclusion that (free) market provision of healthcare is likely to prioritise those aspects of care that the market can observe and reward. This raises the question of how private providers may be incentivised to provide appropriate care - especially those aspects that are unobserved, generate externalities, or conflict with actual or perceived market demand.

1.3 Healthcare as a Credence Good

The potential for inefficiency from market provision of healthcare has long been recognised in the economic literature on 'credence goods' (Darby and Karni, 1973). Credence goods are characterised by consumers' inability to identify the goods that best fit their needs (i.e. to self-diagnose) and to verify the quality of the goods they consume. Consumers must therefore rely on experts – like trained physicians – to diagnose and treat their needs. Another important characteristic of these goods is that there are economies of scope to diagnosis and treatment provision (Dulleck and Kerschbamer, 2006). In other words, it is costly or complicated for consumers to get second opinions, so the expert who gives the diagnosis and treatment recommendation has an advantage in providing the treatment as well.

The information asymmetry between expert sellers and consumers, and the inherent conflict in sellers' diagnostic and treatment functions, raises the possibility of three types of fraudulent behaviour in these markets (Dulleck and Kerschbamer, 2006): *overprovision*, where more services are provided than required for the consumer's needs (service quantity is over-supplied); *under-provision*, where the services provided fail to treat the consumer's needs (service quality

[14] There is an additional effort cost if the patient demands antibiotics in cases where it is inappropriate (as often reported), and the physician needs to take the time to explain to and persuade the patient that antibiotic treatment is unnecessary.
[15] In these cases, some form of government regulation of private provision is generally proposed, such as ('Pigouvian') corrective taxes or subsidies to bring market incentives in line with optimal provision.
[16] This tendency for over-use may be further compounded under market provision if there is significant patient demand for antibiotics.

is under-supplied); and *overcharging*, where appropriate services are provided for the consumer's needs, but the seller overcharges for its quality.[17] Field studies from the SID literature (Gruber and Owings, 1996; Gruber *et al.*, 1999; Clemens and Gottlieb, 2014) initially documented the existence of over-provision for profit in healthcare markets using observational data. More recently, evidence from controlled laboratory and field experiments have supported those earlier findings, contributed new findings on over-charging and under-provision, and proposed certain institutional and incentive mechanisms that might mediate these behaviours (Dulleck *et al.*, 2011; Currie *et al.*, 2014; Lu, 2014; Gottschalk *et al.*, 2018).

1.3.1 Insurance in credence goods markets

The incentives for fraudulent expert behaviour are, in theory, worsened under consumer insurance (or third-party payment, more generally). In healthcare, standard theory on *ex-post moral hazard* stipulates that patients will demand more unnecessary healthcare when they are insured, and do not face the full upfront costs attached (Arrow, 1963; Pauly, 1968; Zeckhauser, 1970). At the same time, the problems of asymmetric information between patients and expert physicians may be worsened: less cost-conscious, insured patients have less incentives to retrospectively verify and hold physicians to account for the quality or necessity of their care, even where this is possible through some search or informational cost to the patient (Pauly, 1978, 1988). The theory of *second-degree moral hazard* argues that profit-oriented physicians will exploit the anticipated moral hazard and lower cost-sensitivity of insured patients to increase their own fraudulent behaviour (i.e. to over-supply or over-charge aspects of care where they can profit, and under-supply those aspects that are uncompensated).

Despite its theoretical appeal, robust field evidence that physicians respond in this way to patient insurance is scarce. A series of observational and quasi-experimental studies have documented a positive relationship between health insurance coverage and treatment choices (and related health expenditure) (Lundin, 2000; Card et *al*., 2008; Wagstaff and Lindelow, 2008; Zhang *et al.*, 2009; Baicker and Goldman, 2011; Martin *et al.*, 2017). However, as with the SID literature discussed previously, these studies fail to isolate the influence of physicians' fraudulent behaviour in response to insurance from patient-side drivers (the fraudulent demands or selective needs of insured patients). Only one study, by Lu (2014), succeeds in fully isolating physician responses to patient insurance. Using an audit study approach with SPs, Lu (2014) randomises both physicians' financial incentives for drug prescribing and the SP's insurance status.[18] The author

[17] Note that, according to the definitions of *overuse* and *underuse* employed in this thesis (see Section 1.1), all three types of fraud (over-provision, under-provision and over-charging) can indicate overuse, whereas only over-provision and under-provision can indicate underuse.

[18] Financial incentives for prescribing are exogenously varied in this study through the SPs' stated preferences on where to buy any prescribed drugs. A financial incentive is provided when the SP asks to buy any prescribed drugs from an affiliated pharmacy, from

finds that, when physicians have a financial incentive to prescribe, they write prescriptions that are 43% more expensive for insured patients compared to uninsured ones. Conversely, physicians do not respond to patient insurance when they have no financial incentives attached. Taken together, these results suggest that physicians' response to patient insurance is driven by profit incentives (*second-degree moral hazard*) rather than any altruistic considerations for patients' financial welfare (*perfect agency*). This is an important finding that merits further investigation, given on-going expansion in formal insurance and for-profit healthcare provision worldwide.

1.3.2 Institutional remedies for expert fraud

In theory and in laboratory settings, it has been shown that imposing the strict condition of *liability* - which holds sellers liable for providing sufficient quality to resolve consumers' needs – may be an effective institutional constraint on expert fraud (Dulleck and Kerschbamer, 2006; Dulleck *et al.*, 2011).[19] However, implementing this condition in healthcare markets is likely to be problematic for two reasons. First, liability requires verifiability of service outcomes.[20] In practice, verifying and attributing health outcomes to specific health services and to specific providers is not straightforward, given the myriad external and patient-related factors that can also influence outcomes. Second, the lab experiments that test this condition do not consider the effects of diagnostic uncertainty or inaccuracy. The credence goods literature generally assumes that experts can diagnose with certainty, and all lab experiments in this area take the diagnosis as given (Kerschbamer and Sutter, 2017). In practice, the considerable uncertainty that is common in clinical decision-making means that strong liability conditions for under-provision can inadvertently lead to more over-provision - a tendency known as 'defensive medicine' (Baicker *et al.*, 2007). The credence goods literature does not distinguish over- or under-provision due to diagnostic uncertainty (or inaccuracy) from fraudulent profiteering.

Pauly (1978, 1988) and others have further argued that some aspects of healthcare may be more characteristic of a 'search' or 'experience' good, rather than a pure credence good. In these cases, through costly search or repeated provider interactions, patients may lower their informational disadvantage (Wolinsky, 1993) to deduce certain qualities of providers (and their services) *ex ante* or infer the quality of services provided *ex post*. In this view, two alternative market

---

which the prescribing physician would receive a profit-share. This incentive is removed when the SP states a preference for purchasing any prescribed drugs at an external pharmacy.

[19] In theory, Dulleck and Kerschbamer (2006) also find that another condition of *verifiability*, which allows consumers to learn the quality of the goods they receive, is effective in preventing over-charging. However, ensuring *verifiability* in a lab setting was far less effective than predicted (Dulleck, Kerschbamer and Sutter, 2011). Kerschbamer *et al.* (2017) put the latter result down to heterogeneous social preferences among sellers, and the presence of 'anti-social' sellers that respond counter to predicted behaviour under verifiability.

[20] In healthcare, the closest institutional solution to implementing liability is a malpractice litigation system, which would rely on the verifiability of health outcomes.

mechanisms may be useful in limiting fraudulent behaviour: *provider competition* and *reputation-building* opportunities.[21]

In general, the results from altering competition and reputational incentives in practice (both in lab and field settings) suggest the obvious: these mechanisms that function through market demand responses are unlikely to bring about improvements to service aspects (price or quality) *unless* those aspects are observable and salient to consumers. While price is easily comparable and highly salient (at least to consumers with little or no insurance, or to third-party payers), appropriate servicing (over- or under-treatment) is not.[22] Therefore, while price competition has been found to drive down market prices, providers may instead compromise on service quality aspects that are less observable, and that cannot trigger the same demand response (Dulleck *et al.*, 2011; Mimra *et al*., 2016).[23] Competition or reputational incentives are then likely to be effective only insofar as they target service aspects where provider performance can be easily observed, measured and compared. As the previous sections highlight, the issues in systematically observing or measuring appropriate healthcare – particularly its technical quality - mean that these regulatory mechanisms have significant limitations in healthcare markets.[24]

To summarise, there are a number of gaps in the literature on credence goods (as applicable to healthcare), and on insurance in these markets in particular, which this thesis seeks to address. First, the literature generally takes experts' diagnostic accuracy as given, and abstracts from the substantial uncertainty that is common in medical scenarios. The inappropriate care choices and unnecessary costs that can stem from diagnostic uncertainty or inaccuracy in real-world settings, as distinct from fraudulent profiteering, have not been investigated in lab or field settings.[25]

---

[21] There is some conceptual overlap in the competition and reputation-building provider incentives evaluated in the literature. Competition effects are commonly tested by investigating provider responses to variations in the number of other competitors in the market (Gottschalk *et al.*, 2018; Rasch and Waibel, 2018), endogenous price-setting (Mimra *et al.*, 2016), consumer search and choice (Dulleck *et al.*, 2011; Gottschalk *et al.*, 2018), and public reporting of comparable competitor information (Dulleck *et al.*, 2011; Mimra *et al*., 2016). Reputation-building incentives are usually tested by evaluating provider responses to variations in the likelihood of repeat customer interactions (Dulleck *et al.*, 2011; Schneider, 2012; Rasch and Waibel, 2018), and to public reporting of provider performance (Mimra *et al.*, 2016).

[22] Repeat business incentives have shown some effect in reducing prices, but no effect on over- or under-provision on average (Dulleck *et al.*, 2011; Schneider, 2012; Rasch and Waibel, 2018). Schneider (2012)'s analysis of reputational incentives in the car repairs market also notes an interesting relationship between customer learning and reputation effects: under the possibility of repeat business, mechanics were less likely to over- or under-provide services that were easier for motorists to verify the quality or necessity of post-service.

[23] For instance, Dulleck *et al.* (2011) find that the combination of consumer choice, endogenous price-setting and public reporting of price information with multiple sellers in a lab setting stimulated price competition and lowered market prices. However, it also led to greater fraud in the form of under-provision (more under-supply in quality). Mimra *et al.* (2016) find that public posting of pricing and quality information in a lab setting lowered prices but exacerbated under-provision and over-charging - once again indicating some trade-off between price competition and quality fraud. Volpp *et al.* (2003) provide some supportive evidence from the field: risk-adjusted AMI mortality rates rose in New Jersey after deregulation of hospital prices, indicating a decline in care quality in response to greater price competition - although no evidence is given that demand elasticity rose and prices actually fell post-deregulation.

[24] Field studies evaluating the quality impact of public reporting in healthcare find little to no effect on targeted technical quality outcomes (Ketelaar *et al.*, 2011), and some perverse effects on non-targeted outcomes, quality reporting fraud, and patient selection (Dranove *et al.*, 2003; Werner and Asch, 2005; Gravelle *et al*., 2010).

[25] The study design in Lu (2014) also negated the need for physicians' to exert much diagnostic effort - as the SPs presented each physician with prior laboratory test results - and minimised the potential for diagnostic uncertainty or inaccuracy. Hence, it does not

Second, how physician experts may trade-off between the *experiential* aspects of healthcare and the purely *credence* aspects has not been explicitly modelled or considered in this literature. Third, it is unclear how different levels of patient cost-sharing (insurance co-payment) affect physician behaviour in these markets. Partial insurance (some patient cost-sharing) may have different incentive effects on physicians than full insurance (no cost-sharing) (Kerschbamer and Sutter, 2017). Lastly, the institutional remedies proposed in the literature for regulating expert fraud – liability, competition or reputational incentives – have informational requirements that are unrealistic in real-world settings. Instead, Kerschbamer and Sutter (2017) conclude that experts' "moral constraints" (or pro-social preferences) may have an important role in regulating fraudulent behaviour where institutional solutions are infeasible. While these authors propose leveraging experts' moral constraints through job selection – implying the constraints are largely fixed, individual attributes – this thesis explores how such intrinsic constraints may also be mobilised through targeted incentives (as detailed further in Section 1.5).

1.4 Role of Intrinsic Motivation

The previous sections highlight a conflicting duality in the physician's role under market incentives - as expert agents in healthcare decision-making and custodians of public health on the one hand, and as profit-maximising economic agents on the other. Where appropriate care conflicts with patients' actual or perceived preferences, private physicians have clear rationale to prioritise and oversupply those services that they believe patients will value – or at the very least, not object to. Moreover, regulators' inability to systematically verify and compensate appropriate care provision means that financial or reputational incentives, tied to imperfect and simplified measures of such provision (such as pay-for-performance or public reporting programs[26] tied to aggregate performance metrics), are likely to be of only limited value in countering these over-treatment incentives (Berenson and Rice, 2015). In this context, finding ways to leverage physicians' *intrinsic motivation* (Bénabou and Tirole, 2003; Galizzi *et al.*, 2015) – be it professionalism, or altruistic concerns for patient health and financial welfare – may be a way forward. While intrinsic motivation has been understood in slightly different ways in different

provide much insight into how physicians may vary their diagnostic effort (and their resulting diagnostic accuracy) in response to patient insurance.

[26] This includes social accountability initiatives, like the community monitoring initiative tested with public providers in Uganda by Björkman and Svensson (2009). While the Hawthorne effect – that is, provider responses to knowingly being monitored – has been proven beneficial in improving monitored aspects of care (Leonard, 2008; Schwartz *et al.*, 2013), the issue remains that it is often difficult to assess whether observed aspects constitute (or lead to) "appropriate" care. For instance, observing that a provider completes monitored aspects of clinical effort does not guarantee that the resulting diagnosis will be accurate and the treatment appropriate. It also does not guarantee that providers will not subsequently compromise (and lower their effort on) non-monitored aspects of care.

literatures (Galizzi *et al.*, 2015; Brock *et al.*, 2016), this thesis broadly defines it as any physician motivation that aligns with patient welfare, independent of financial or reputational gain.[27]

The notion that healthcare providers may be intrinsically motivated is not new (Galizzi *et al.*, 2015) - it has been evidenced both in the lab (Kolstad and Lindkvist, 2013; Hennig-Schmidt and Wiesen, 2014; Lagarde and Blaauw, 2017) and in the field (Lundin, 2000; Granlund, 2009; Ashraf, Bandiera and Lee, 2018).[28] The literature generally identifies two sources of this motivation: *i)* individual-specific endowments, that are fixed and unique to individual providers, and *ii)* environmental factors, that can shape the motivations of all providers (Franco *et al*., 2002; Ashraf and Bandiera, 2017). The discussion on leveraging intrinsic motivations in healthcare has often emphasised the first source, proposing that recruitment practices must be designed to identify and select (highly) intrinsically motivated individuals into health professions (Lagarde, Huicho and Papanicolas, 2019). Another body of work, inspired by behavioural economic theory, has begun exploring whether intrinsic motivations may indeed be malleable (in line with the second source) - and how workplace and job attributes (Tonin and Vlassopoulos, 2010; Ashraf and Bandiera, 2017; DellaVigna and Pope, 2017), social and behavioural reinforcements (Brock et al. 2013), or informational incentives (Kolstad, 2013; Lee, 2018) may be used to improve the performance of the average health worker. A key finding in Leonard *et al.* (2015) is that the first source is useful for improving provider performance only in the public sector in Tanzania, but not in the private sector. Das and Hammer (2014) argue that this may be due to a stronger impact of market incentives in disciplining provider behaviour in the private sector. However, as suggested previously, there may be cases where market incentives do not align with appropriate provider effort – either because market demand is not aligned with appropriate care, or because not all aspects of appropriate provider effort are observable to the market. This thesis therefore investigates both sources of intrinsic motivation among private physicians, and their potential for countering the market failures outlined in the preceding.

[27] Note that "intrinsic motivation" and "altruism" are used interchangeably in this thesis, for simplicity. It is recognised that altruism may be just one form of intrinsic motivation and conceptually distinct from other forms (such as professionalism). This simplification is perhaps more justified in healthcare than in other professions, as the goal of improving patient welfare is consistent with both professionalism and altruism among healthcare workers (professional excellence in the medical profession generally requires providing clinically optimal care to patients).

[28] For instance, it has been used to explain why some workers voluntarily take up low-paying or rural posts (Kolstad and Lindkvist, 2013; Ashraf *et al.*, 2018); why medical students choose to sacrifice their own cash payoffs to benefit unknown patients (Hennig-Schmidt and Wiesen, 2014); and why physicians are more often observed to prescribe cheaper generic drugs to less insured patients, even when they have no financial incentives to do so (Lundin, 2000; Granlund, 2009).

1.5 Thesis Objectives and Roadmap

Building on the findings discussed in previous sections, this thesis uses field experiments with unannounced standardised patients (SP) to examine three research questions in the market for primary healthcare services in South Africa:

1. What is the impact of increasing patient financial protection (insurance) on the appropriateness (quality and efficiency) of care provided by private physicians?
2. How can market pressures and intrinsic motivations mediate the healthcare choices of these physicians?
3. What is the impact of intrinsic informational incentives from private audit and performance feedback (A&F) in lowering inappropriate antibiotic treatment and costs?

The remainder of the thesis is structured as follows. **Chapter 2** provides an overview of the study setting in South Africa, and explains the relevance of this setting for the thesis objectives. **Chapter 3** sets out a theoretical framework that models a typical physician-patient interaction in the private sector, replicating several features of this study setting. The model characterises the decision problem of a representative physician in this study, and the various institutional and incentive elements that shape it. Seven hypotheses on physicians' effort and treatment choices under different incentives are derived for testing in subsequent chapters. **Chapter 4** outlines the experimental methods and main data sources used in this thesis. The chapter also provides a description of the two empirical parts to this study, corresponding to two distinct field experiments.

Chapter 5 evaluates the first (within-subject) field experiment, and tests the impact of varying the level of patients' financial protection (insurance) on physicians' care quality choices - including observable aspects of their clinical effort, as well as the accuracy of diagnoses and appropriateness of treatments given. Using the same experiment, **Chapter 6** evaluates the impact of patient financial protection on the quantity and costs of care provided by physicians. It evaluates this impact both on care aspects with direct financial incentives for physicians (fee-for-service diagnostic tests and treatment procedures) and on aspects without any direct financial incentives attached (drug prescriptions for physicians that cannot dispense drugs). **Chapter 7** evaluates the second, randomised field experiment in this study, to test the effectiveness of private A&F compared to a passive educational intervention alone in lowering inappropriate antibiotic prescribing for a viral respiratory infection. The chapter also evaluates effects on treatment costs, and potential adverse spill-overs on appropriate antibiotic use and other inappropriate drug treatments. Finally, **Chapter 8** concludes with a summary of the main findings from this study, its strengths and limitations, and its research and policy implications.

2 Study Setting

This chapter describes the overarching context of this thesis study in South Africa, and highlights why the country is a particularly relevant setting for exploring the research questions set out in Section 1.5.

With a relatively young population of 54 million, South Africa faces a unique mix of challenges in overcoming the healthcare inequalities inherited from its colonial and Apartheid history, whilst combating a high burden of infectious diseases and high rates of antimicrobial resistance. Central to these efforts is an ambitious agenda to harness a prominent private healthcare sector, to deliver universal healthcare to all citizens within a new National Health Insurance system (NHI). In particular, the NHI seeks to contract private primary care physicians (GPs) to deliver publicly-funded healthcare and function as gatekeepers to the national health system. These trends emphasise an increasingly central role for private GPs in South Africa, in determining the healthcare use and expenditures of a progressively better insured population. Understanding and aligning the incentives of these market-based providers with universal healthcare and public health will be crucial to the success of this agenda.

As South Africa's largest and fastest-growing urban centre - with a large private healthcare market - the city of Johannesburg presents an ideal microcosm for studying market incentives in healthcare provision in this context. The following sections highlight a number of institutional, epidemiological and policy developments that are relevant to understanding the contextual factors that may influence private physicians' clinical decision-making in this setting, as well as the motivations for and timeliness of this study's objectives.

Section 2.1 explains the historical context of the South African health system, which has established a public-private dichotomy in the system today (Section 2.2). Section 2.3 describes the high burden of infectious diseases and antibiotic stewardship in the country. It is noted that, despite the majority of antibiotics being prescribed in primary care, most local antibiotic stewardship initiatives to date have been focused on the public hospital sector. Section 2.4 describes the private market for primary healthcare in South Africa, characterised by a largely insured demand and a supply by independent private GPs. Finally, Section 2.5 outlines the current trend toward universal health coverage (UHC) in South Africa with the planned National Health Insurance system (NHI), and its implications for private GPs.

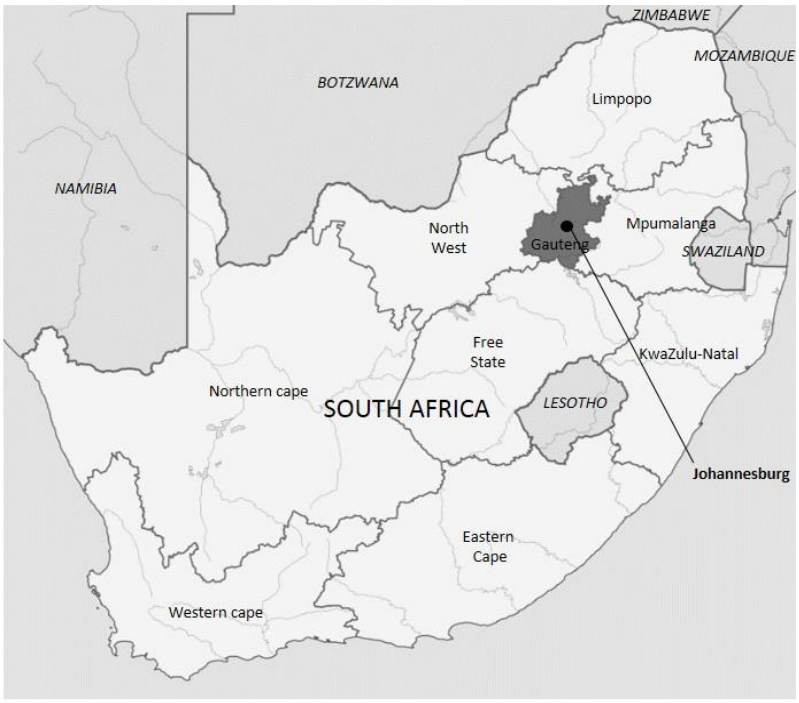
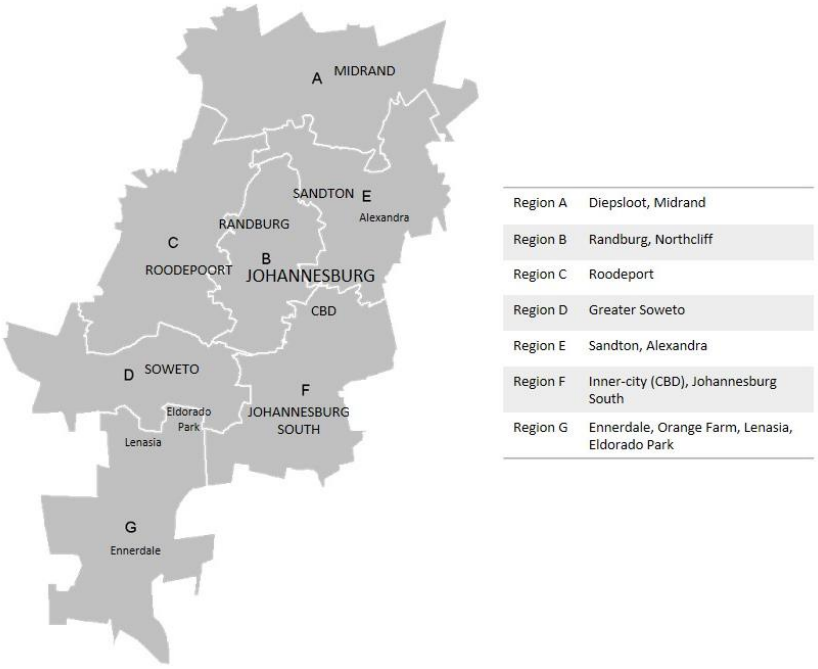


Figure 2.1. Provinces of South Africa



Region A	Diepsloot, Midrand
Region B	Randburg, Northcliff
Region C	Roodeport
Region D	Greater Soweto
Region E	Sandton, Alexandra
Region F	Inner-city (CBD), Johannesburg South
Region G	Ennerdale, Orange Farm, Lenasia, Eldorado Park

Figure 2.2. Regions of the City of Johannesburg

Notes. Apartheid-era racial segregation in Johannesburg had a distinct north-south divide. The northern suburbs around Sandton and Randburg, primarily in Regions C, B and E, were the previously whites-only residential areas. Apartheid policies since 1950 relocated black Africans previously residing in these areas to the newly-created southern suburbs in Regions D, G and (south of) F, and the northern peripheral suburbs in Region A. An exception to this north-south divide is the suburb of Alexandra, bordering Sandton in Region E, which was and still is a predominantly black African neighbourhood. Sandton and Randburg remain two of the wealthiest suburbs in Johannesburg today (and still predominantly resided by white South Africans), and are referred to as the "northern suburbs" in this thesis. See Appendix A.2 for a brief overview of historical spatial segregation in Johannesburg, and how it has translated into spatial inequalities in health and economic opportunity.

2.1 Historical Background

South Africa's history of institutionalised racial segregation has shaped its modern-day health system. Healthcare resources have been historically skewed in favour of the white minority population, urban centres, higher levels of curative care, and a private sector that predominantly serves the white, urban population (Wadee *et al.*, 2003). The Apartheid regime that began in 1948 enforced racial segregation by delineating strict whites-only areas within urban areas; relocating urban-residing blacks to rural homelands and urban peripheries, and regulating their movement into towns and cities. Healthcare facilities were similarly segregated, and urban health services and whites-only facilities were far better resourced than rural or blacks-only facilities. Primary healthcare was severely underfunded in the public system: by 1994, only 11% of public healthcare expenditure was being allocated to non-hospital primary healthcare services (McIntyre *et al.*, 1995). The private healthcare sector expanded rapidly in the 1980s, and the share of doctors working in the private sector increased from 40% to 60% over this decade (Coovadia *et al.*, 2009).[29] At the same time, access to private healthcare was predominantly restricted to the white population, due to both the concentration of private providers in urban centres (where the white population largely resided) and the restriction of private health insurance cover to only this population group until the 1970s.

After the first multi-racial elections in 1994, the new African National Congress (ANC) government ended institutional segregation and introduced a number of reforms to improve the public provision and distribution of basic health services.[30] The public health system was better integrated, and developed into a comprehensive national health service, centred around primary care (which was also made available free of charge to all citizens). Spending on primary healthcare almost doubled as a share of public health expenditure (Coovadia *et al.*, 2009). Nevertheless, government spending on the public health system stagnated after 1997, while spending by private health insurers increased: between 1996 and 1999, annual real expenditure growth per public sector beneficiary was 1%, compared to 10% per private insurance beneficiary.

---

[29] Government funding further subsidised this growth, with tax subsidies for the private health insurance contributions of formal sector workers (amounting to 10-17% of the government health budget in 1994); government (employer) contributions to private health insurance membership for all civil servants and their beneficiaries (amounting to 16% of the health budget in 1992); and the training of all medical personnel (public and private) in the public sector (Wadee *et al.*, 2003) – subsidies that continue to this day.

[30] More generally, since the end of Apartheid, South Africa has enjoyed three decades of positive economic growth, and absolute poverty has declined. Today, it is considered an upper middle-income country and the largest economy in Africa. However, stark socio-economic inequalities exist – particularly along racial lines. Around 29% of the working-age population are unemployed, and this unemployment is largely concentrated among the young (55%), black (17%),[30] and unskilled labour (Stats SA, 2015a, 2019). Only 6% of the black population have some form of higher education, and their median income falls in the income category below R4,800. Moreover, over 64% of black South Africans were classified as living below the poverty line in 2015. On the other hand, equivalent figures for the white population are 30% with higher education, median income in the category R38,400-76,800, and just 1% living below the poverty line (Stats SA, 2015a, 2015b).

2.2 Two-Tier Health System

As a result of historical disparities, South Africa today remains characterised by a two-tier, public-private health system. Private healthcare is highly developed and comprises a substantial share of the national health system, despite a large, parallel public sector that offers free care to most citizens. Compared to other upper middle-income countries, South Africa's total health expenditure as a share of GDP (around 9%) is relatively high. However, more than half of this expenditure comes from private sources (51%), and primarily from voluntary private health insurance schemes (43%) that cater to just 16% of the population.[31] Revenue from general taxes account for less than 48% of total health spending, and yet fund healthcare for over 80% of the population that rely on the public system. This includes 16% of the population that rely on the public sector for hospital services, but choose to pay out-of-pocket (OOP) to access private primary care (Keeton, 2010). These characteristics are symptomatic of the historical under-funding of public primary care in South Africa, and highligh utilisation preferences that are misaligned with health insurance affordability.

Private insurance is increasingly unaffordable for the majority of South Africans,[32] and the likelihood of cover varies by race - and relatedly, by education, location and wealth. In 2016, around 76% of the white population had private insurance, compared to just 11% of black Africans.[33] Private insurance is also twice as likely in urban than rural areas, and positively associated with the level of education and household wealth (NDoH *et al.*, 2019).[34] The distribution of healthcare benefits in the system favours the privately insured, as health insurance schemes spend almost six times more per capita than the public sector (Kula and Fryatt, 2013).

This inequity in financing is reflected in the distribution of healthcare infrastructure and workforce. South Africa faces a general shortage of doctors (with less than 1 doctor per 1000 population) (World Bank, 2017), and retention in the public sector has been particularly problematic. Although all medical training takes place in the public sector, most newly qualified doctors opt to shift into private practice in urban areas, mainly due to poor working conditions and workplace security in the public system (van der Spuyet *et al.*, 2017). Various estimates

[31] These funds are spent almost exclusively on private healthcare services, although they also cover care sought by insured individuals in public facilities – which in practice, tends to be limited to highly-specialised, tertiary care (McIntyre, 2010).

[32] The population covered by private insurance schemes has declined in recent years, from 19.8% of the total population in the late 1990s to just over 16% in 2016 (NDoH *et al.*, 2019), due to dramatic increases in the cost of premiums and co-payments.

[33] Note that, according to the 2011 population census (revised in 2015), 79% of the South African population is black African, 9% is white, 9% is mixed-race, and 3% is Asian, Indian or other race (Stats SA, 2015a).

[34] While overall healthcare financing is progressive (the richest households contribute a higher share of their income to health expenditure), this is primarily driven by private insurance contributions that only benefit scheme members, and does not contribute to income cross-subsidisation within the health system. At the same time, there is some evidence of 'reverse' risk cross-subsidisation, as the distribution of healthcare benefits is regressive (pro-rich) and not in line with health need: the poorest households, that bear a heavier burden of ill-health, incur a smaller share of overall health service benefits within the system (Ataguba and McIntyre, 2012).

suggest that between 70-80% of all doctors in South Africa work in the private sector (including those that work in both sectors) (Keeton, 2010; McIntyre, 2010).[35]

These disparities in healthcare resources translate into significant inequalities in healthcare access, quality, and financial protection (Harris *et al.*, 2011; Lagarde and Blaauw, 2019b). As previously mentioned, despite the small population covered by private insurance, the use of private outpatient services is comparatively high - even in the poorest quintiles of the population. Uninsured individuals frequently opt to pay fully OOP to access private primary care services with lower wait times, greater choice of healthcare professionals, and perceived higher quality than in the public sector (Harris *et al.*, 2011).[36] For instance, it is estimated that 26% of the adult population (aged over 15 years) that utilised outpatient care in 2016 chose to consult private outpatient services – and this figure is higher in urban areas (32%). Of these individuals, 28% paid out-of-pocket (NDoH *et al.*, 2019).

2.3 Burden of Infectious Diseases and Antibiotic Stewardship

South Africa faces a relatively high burden of disease for an upper-middle income country - particularly from infectious diseases such as AIDS/HIV, TB and other lower respiratory infections. It has the highest absolute number of people living with HIV anywhere in the world (19% of the country's adult population (NDoH *et al.*, 2019), and 17% of the world's HIV+ population). [37] While non-communicable diseases are becoming more prevalent, certain communicable infections remain a significant mortality and morbidity burden. According to civil registration data, TB was the leading cause of all mortality in 2015 (accounting for 7.2% of all deaths) (Stats SA, 2017); while AIDS/HIV-related conditions, TB and pneumonia were the three single leading causes of premature mortality (as measured by the number of years of life lost (YLL)) (Groenewald *et al.*, 2017). Around half of all premature mortality was attributed to HIV/AIDS in 2012 (Pillay-van Wyk and Bradshaw, 2017).[38]

The growth in antimicrobial resistance (AMR) is particularly problematic in South Africa, where high rates of AMR (for both Gram-positive and Gram-negative bacteria) co-exist with this high

[35] In 2015, the private sector employed around 1 general practitioner (GP) and 0.9 medical specialists per 1000 of the insured population, compared to equivalent figures of 0.2 GPs and 0.1 specialists per 1000 of the uninsured in the public sector (CCSA, 2018).
[36] Measured clinical quality is also been found to be higher in the private sector (Lagarde and Blaauw, 2019b).
[37] However, the number of deaths attributable to the disease has been falling, and overall life expectancy has increased rapidly since 2005 (from 51.6 to 62.9 years in 2015) due to improved access to antiretroviral treatment and other related health initiatives (WHO, 2018).
[38] The main causes of mortality vary by population group. Non-communicable diseases were the leading causes of mortality among the white and Asian/Indian population groups in 2012, accounting for 80% of deaths; while HIV/AIDS and TB accounted for the majority of deaths among black Africans (38%). The mortality burden of other lower respiratory infections (including pneumonia) is more evenly distributed - as the third and fifth leading causes of mortality among the black and white population groups in 2012, respectively (Pillay-van Wyk and Bradshaw, 2017).

incidence and mortality burden of bacterial diseases (Gelband and Duse, 2011). The prevalence of these conditions makes access to effective antibiotics imperative and, paradoxically, threatens the effectiveness of available ones (DeNegre *et al.*, 2019). AMR is a natural, evolutionary outcome of antibiotic use over time, and it inevitably develops faster in settings where bacterial diseases are more common. Nevertheless, the misuse of antibiotics for viral infections, the use of shorter antibiotic courses than necessary, and the overuse of multiple or broad-spectrum antibiotics unnecessarily exacerbate this problem (Tonkin-Crine *et al.*, 2017).

Policy attention on the issue has grown in recent years, as antibiotic use escalated in South Africa - increasing by 175% between 2000 and 2010 (Republic of South Africa, 2015).[39] The South African Antibiotic Stewardship Programme was initiated in 2012, and brought together a group of multidisciplinary experts to implement stewardship initiatives within healthcare facilities, designed to encourage appropriate antibiotic use. This was followed by the development of a National Strategic Framework on Antimicrobial Resistance (NDoH, 2014a) in 2014, to cohere national efforts to contain AMR. South Africa now has the most active AMR surveillance system in Africa, a drug regulatory system that prohibits the sale of antibiotics without prescription, and an Essential Drugs List (EDL) and evidence-based Standard Treatment Guidelines (STG) (NDoH, 2014b) to regulate and guide antibiotic use in the public health system.

These initiatives have tended to focus on characterising the extent of AMR in the country; and on antibiotic stewardship in the hospital environment (Boyles *et al.*, 2013; Brink, Messina *et al.*, 2016) and in the public sector. South Africa has very little experience or evidence on successful antibiotic stewardship in primary care – particularly in the private sector. This is despite 75-80% of antibiotics in the country being prescribed in primary care (mostly for community-acquired respiratory infections) (Brink, Van Wyk, *et al.*, 2016), and almost 80% of the main prescribers (the doctors) working in the private sector. One exception was a 4-day prescriber training intervention, which was found to be effective in reducing antibiotic prescribing for upper respiratory tract infections in primary care (Meyer *et al*., 2001). This trial was nevertheless conducted in public clinics, and almost 20 years ago. A more recent initiative by Brink, Van Wyk *et al.* (2016) provides a 'diagnostic stewardship' framework to aid the differential diagnosis and management of common respiratory infections in primary care. However, its effectiveness is yet to be evaluated. Besides these educational (or 'decision-aide') initiatives, aimed at improving the knowledge or awareness of physicians and pharmacists, there is still a paucity of active stewardship interventions in this sector (or efforts to evaluate their impacts) to match those seen in South African hospitals (Brink, Messina, *et al.*, 2016; Brink, 2017).

[39] This increase was primarily driven by increases in the use of trimethoprim and broad-spectrum penicillin.

Prescribing practices in the private sector also remain relatively unregulated.[40] The STGs are publicly available to guide the prescription choices of both public and private healthcare providers, but they are primarily directed at the public sector. A recent survey of primary care providers in South Africa (mostly private) also revealed very little reported use of existing guidelines in antibiotic choices (Farley *et al.*, 2018): only 39% of respondents reported using guidelines in prescription choices, and 80% expressed demand for clearer guidelines in hard-copy. While overall antibiotic use has remained fairly stable in the private sector in recent years (Schellack *et al.*, 2017), the sector has seen a large increase in the prescription of newer, more expensive therapies. Moreover, a related study to this thesis (Blaauw and Lagarde, 2019), comparing inappropriate antibiotic use for a viral respiratory infection in the public and private primary care sectors of Johannesburg, found that inappropriate prescribing in the private sector – while lower than in the public sector – was still very high (67%, compared to 78% in the public sector).

There has also been little effort to investigate the *drivers* of inappropriate antibiotic treatment decisions in South Africa, to inform the design of stewardship initiatives. A few recent studies have sought to address this gap (Farley *et al.*, 2018; Lagarde and Blaauw, 2019b; Manderson, 2019). They reveal that, while physicians display relatively good knowledge of AMR and correct antibiotic use, actual or perceived patient demand for antibiotics may be an important influence. Farley *et al*. (2018) find that, of 269 primary care physicians surveyed, 67% reported feeling pressure from patients to prescribe antibiotics. At the same time, respondents had relatively good knowledge of correct antibiotic use and AMR, and better knowledge was associated with more appropriate self-reported prescribing behaviours. Moreover, there was significant demand from respondents for more information on appropriate antibiotic use, both for themselves and for patients. This latter finding suggests that physicians, given their time constraints, may inappropriately prescribe antibiotics because they feel ill equipped or wish to avoid having to explain to expectant patients why antibiotics are not needed. Educational aids may facilitate discussions with patients in this case. In another qualitative study, Manderson (2019) finds that patients did not generally demand antibiotics during direct observations of consultations in primary care; however, physicians reported perceiving patient demand for antibiotics during the same consultations in subsequent interviews. In addition, Lagarde and Blaauw (2019b) report that physicians were less likely to prescribe inappropriate antibiotics in practice (when performance was assessed through unannounced standardised patients) when patients actively expressed reluctance towards receiving unnecessary antibiotics – again, underscoring physicians' sensitivity

[40] The only prescribing constraint imposed on private providers is any restrictions pertaining to the insurance cover of private patients (although this is left to the prerogative of private insurers, and only applies to the 16% of the population with private insurance cover).

to patient demands, whether positive or negative. They also find that better diagnostic and guidelines knowledge are good predictors of appropriate treatment in practice (Blaauw and Lagarde, 2019).

Overall, current evidence on antibiotic use in South Africa suggests that active interventions to address inappropriate prescribing in primary care, informed by rigorous research on its drivers – particularly in the private sector - are urgently needed. To this end, primary care physicians consistently demonstrate a high demand for clearer information and education on AMR and appropriate antibiotic use - both for themselves and for patients. Evidence from other contexts suggests that education-based initiatives may indeed lead to better prescription outcomes (van der Velden *et al.*, 2012). However, the relative effectiveness and cost-effectiveness of such initiatives, in comparison to other possible interventions, must be evaluated more thoroughly in the South African context.

2.4 Private Market for Primary Health Care

South Africa's private healthcare market is particularly prominent in the urban province of Gauteng, where Johannesburg is located. Around 40% of private health insurance beneficiaries, and 45% of all private doctors, dentists and pharmacists are based there (McIntyre, 2010; CMS, 2018). Among all 52 districts in South Africa, the city of Johannesburg has the fourth highest number of private GPs per capita, and the third highest number of private specialists per capita (CCSA, 2018).

Although highly privatised health systems are not uncommon in developing countries (Mackintosh *et al.*, 2016), the South African healthcare market is distinctly more formalised and regulated. Market demand largely comes from privately insured patients, whereas market supply of primary care is provided by self-employed, private GPs.

2.4.1 A largely insured demand

Around 72% of all expenditure on private outpatient services (including primary care services) in South Africa comes from private insurance funds, and 28% from OOP payments (NDoH *et al.*, 2019). Market-based primary care is therefore predominantly funded through private health insurance, provided by around 80 insurance schemes – locally known as "medical schemes." [41] Medical scheme membership is not legally required, but most formal sector employers offer subsidised scheme membership to their staff. All insurance plans must cover a prescribed

[41] Although the schemes are non-profit (and open schemes cannot refuse membership to any applicant) as per the MSA, scheme administration is usually outsourced to for-profit companies.

minimum benefits package, which includes care for certain chronic diseases and hospital-based interventions. Beyond this, plan options can differ in the premiums charged to members (the flat monthly contributions), the benefits covered, and the deductibles and co-payments attached to health service use.

Medical schemes commonly offer Medical Savings Accounts (MSA) to cover primary care expenditures (including drugs), with a certain amount of funds ('savings') allocated per annum. These accounts are funded through a portion of members' monthly premium contributions,[42] and the amount of allocated savings varies by insurance plan. In general, the higher the premiums charged, the higher the annual savings allocation to MSA. Once these savings are depleted, the patient would need to pay for any subsequent primary care OOP, until the account is replenished the following year. Therefore, the higher the savings allocation, the lower the likelihood that the beneficiary incurs any OOP expenditure for primary care in a given year (all else equal).

Medical schemes often have designated or preferred provider networks, comprised of independent private providers with whom they have specific payment or patient volume agreements (contracts).[43] However, there is very little price regulation in the market. Medical schemes set their own prices (maximum reimbursement rates (MRR)) for specific healthcare services, but often allow contracted providers to balance-bill patients above those rates (i.e. to charge patients an additional fee above the MRR). Most medical schemes also allow beneficiaries to consult providers outside their networks, and directly access specialists and higher levels of care without a prior referral ('gate-keeping'). However, external providers tend to charge higher prices and increase OOP expenditure for scheme beneficiaries, as patients are expected to pay these provider fees upfront and claim back from the schemes later (up to the MRR).

The schemes also have very little oversight of the quality or choice of services supplied by providers. Providers have no obligation to report on their care quality performance or costs, so there is no publicly available data for determining the cost-effectiveness of medical interventions or technologies.[44] Quality reporting is voluntarily undertaken by some providers and private

---

[42] The MSA were designed to allow members to take greater ownership of (and rationalise) their own primary care expenditures. However, they necessarily limit risk pooling at the primary care level, and further reduce member contributions to risk pooling at higher levels of care. They may also incentivise members with little or no remaining savings to by-pass primary care in order to avoid OOP costs, and unnecessarily utilise better-covered specialist care. Moreover, as members cannot draw on these account funds for any other uses, there are concerns that generous savings allocations create incentives for overuse of primary care services - both by healthy members seeking to fully benefit from available funds, and from providers expecting patients to be relatively price-insensitive to services paid for through these accounts (CCSA, 2018). Indeed, the average claims-to-savings ratio of beneficiaries has been consistently above 95% for open medical schemes in recent years, showing significant depletion of available funds in general (CMS, 2018).

[43] Providers are free to be contracted by several medical schemes and also see uninsured patients paying OOP.

[44] The costs of health service claims have increased substantially in recent years. Annual claims expenditure rose by 590% in real terms between 1980 and 2016 (CCSA, 2018), largely due to the price-setting power and heavy capacity investments of private hospitals (which have driven over-utilisation and prices). These costs have almost entirely been passed down to scheme members, as reflected in rapidly rising insurance contributions and co-payment rates. For instance, between 2000 and 2017, inflation-adjusted

organisations. However, this tends to be limited to the hospital sector, and the care quality data reported is often not comparable or made publicly available (Carvounes *et al.*, 2017; CCSA, 2018). The schemes primarily rely on claims data submitted by providers for their beneficiaries (which contain diagnostic and procedural data, as recorded by providers) to infer limited aspects of care quality from reported utilisation of specific preventive services or hospitalisations. This administrative data is also very limited, covering only the scheme's beneficiaries (rather than providers' full patient populations), and often reported to be inaccurate and discordant with clinical data (Carvounes *et al.*, 2017).

2.4.2 Supply: private general practitioners (GPs)

GP services account for the majority of all OOP expenditure on private healthcare by the uninsured (almost 50%), highlighting the centrality of private GPs in delivering care to both the insured and uninsured populations (McIntyre, 2010). Individuals frequently forego free healthcare in public clinics to consult a private GP, due to the perceived higher quality of care in the private primary care sector, and the guarantee of seeing a doctor (nurses conduct most primary care consultations in the public sector).

Most GPs are paid fee-for-service (FFS), where a unit fee is paid per service provided. Such volume-based payment has raised concerns around perverse provider incentives to over-supply well-remunerated services relative to what patients need (CCSA, 2018). Evidence of such 'supplier-induced demand' (SID) in the South African primary care sector is relatively scarce, although some indicative evidence can be gleaned from statistics published by the Council for Medical Schemes (CMS) on GP visit claims (CMS, 2018). Comparing the geographical distribution of GP visit claims to medical schemes in four South African provinces, a consistent pattern is clear: areas in the highest quintile in terms of local per capita expenditure on GPs have the highest local GP density[45] and the highest share of total GP visit claims (over 55% of visit claims in all four provinces).[46] While these statistics fail to account for potential differences in local patient characteristics that may contribute to differences in GP utilisation across areas, the striking consistency in the positive association between GP expenditure and GP density across all

---

average monthly contributions per member have increased by 72% (CMS, 2018). The medical schemes themselves have managed to retain stable cash reserves of approximately 32% of gross annual contributions over the last decade - well above the necessary 25% mandated under the MSA (CMS, 2018). This discrepancy between high and rising costs for members, and high and stable cash reserves for the medical schemes, has called into question the statutory role of medical schemes as non-profit entities that act in the sole interest of their beneficiaries.

[45] The relative GP density of an area (postal code) is determined by comparing the area's share of total practicing GPs in the province, and its average patient load per local GP, to other areas in the province.

[46] This pattern is particularly stark in Johannesburg's province of Gauteng, where areas ranking in the top 20% in terms of local per capita GP expenditure have an average GP density that is around 90% higher than in areas ranking in the bottom 20%, and account for 75% of all GP visit claims.

four provinces analysed suggests some level of demand-inducement is plausible (whereby GPs compensate for lower patient loads with a higher number of visits and expenditure per patient).

As all medical practitioners in South Africa, GPs must complete a five-year Bachelor degree in Medicine and a two-year clinical internship in order to practise. They must also be registered with the Medical and Dental Board of the Health Professional Council of South Africa, which is the primary regulatory body for medical practitioners. Beyond this, more active on-going regulation of professional practice and standards is generally absent in the private sector. For instance, there is no systemic requirement for GPs to follow evidence-based treatment protocols, pursue professional development activities, undergo peer review, or adhere to other forms of quality monitoring or reporting.

There are nevertheless two (weak) sources of implicit regulation and cost-control in the private GP market (as detailed below). The first comes from the existence of contracted provider networks for medical schemes (as noted in the previous section), which distinguish 'contracted-in' GPs from those that are 'contracted-out'. The second comes from the existence of drug dispensing licenses for GPs, which create cost-sharing incentives for 'dispensing' GPs that are absent for 'non-dispensing' ones.

2.4.2.1 Contracted-in vs. contracted-out GPs

Private GPs are not obliged to join medical schemes' provider networks in order to see insured patients (these patients are usually also covered for consultations with external providers). However, the majority of GPs in South Africa choose to contract with at least the larger medical schemes, as it can benefit them in two ways. First, they are guaranteed direct reimbursement from schemes up to their maximum rate (MRR) or a negotiated tariff for each service provided, which enables them to take the administrative and OOP burden off their patients (who will then not need to pay up-front and claim back from the schemes themselves). Second, they can benefit from higher patient volumes, as schemes often have incentives for beneficiaries to see providers within their networks. Hence, in areas with a high number of insured patients, particularly from a few large medical schemes, 'contracting-in' can be beneficial to both local GPs and patients. In the city of Johannesburg for instance, 71% of active GPs are reported as contracted-in with medical schemes.[47]

[47] This is calculated from the population of GPs listed as active in Johannesburg on the online Medpages database (which is estimated to list approximately 80% of all medical practitioners in South Africa), with non-missing data on contracting status (N=967).

These provider networks ensure some level of oversight on costs and standards of care,[48] although this oversight is not systemised, and the consequences of suboptimal standards on network GPs are unclear (CCSA, 2018).[49] The prices of 'contracted-in' GPs tend to be lower, as schemes can negotiate cheaper overall tariffs and restrictions on balance-billing in return for preferential channelling of beneficiary volumes and higher MRR's (through preferred payment contracts). As a result, patients are less likely to have to co-pay to cover these GP consultation rates (which are more likely to be on par with their scheme's MRR).

2.4.2.2 Dispensing vs. non-dispensing GPs

The majority of private GPs in South Africa are 'non-dispensing', meaning they do not have a licence to dispense drugs (they can only prescribe). In Johannesburg, non-dispensing GPs account for approximately 60% of all active GPs. Doctors that wish to both prescribe and dispense drugs to their patients must complete a dispensing course and be licensed by the Department of Health. In many parts of the world, where doctors make a profit from drug dispensing, licensing doctors to both prescribe and dispense is often argued to create incentives for doctors to over-prescribe (Chou *et al.*, 2003; Park *et al.*, 2005; Iizuka, 2007, 2012; Rischatsch *et al.*, 2013; Kaiser and Schmid, 2016). On the contrary, in South Africa, dispensing licences for GPs operate in practice as a cost-containment mechanism. Since 2004, new regulations on drug pricing transparency have reduced GPs' ability to profit from dispensing. All GPs must purchase drugs at fixed single exit prices (SEP), set yearly by a national drug-pricing authority. A small dispensing fee may be added to the cost of each drug sold to patients, and these fees are also set at a fixed percentage of each drug's SEP. Nevertheless, dispensing GPs rarely charge patients separately for each drug dispensed, as the pricing formularies are administratively complicated. Most dispensing GPs charge a flat consultation rate, inclusive of all drugs, which creates strong incentives for containing unnecessary prescription costs. Dispensing GPs' consultation rates tend to be higher than those of non-dispensing GPs, and the larger medical schemes also offer higher MRR for dispensing GPs to cover drug costs.[50]

A related study suggests that the cost-containment incentives from GPs' dispensing status can indeed be quite powerful (Lagarde and Blaauw, 2019a): dispensing GPs in Johannesburg prescribed significantly more inappropriate and more expensive treatments when their obligation

[48] For instance, the criteria for entering GP networks often include some quality assessment, supposedly through peer review (by independent GP associations). However, the methods of assessment, their frequency, and the consequences for GPs that do not meet quality standards are unclear.

[49] The strongest form of oversight applies to schemes' Designated Service Provider (DSP) networks, where there are often restrictions on referrals to network-specialists only, and requirements to follow certain treatment protocols and prescribe only formulary-listed medicines. However, these requirements vary between schemes, and the extent to which they are enforced in practice is unclear.

[50] Some dispensing GPs also offer a lower consultation rate in cases where they do not dispense drugs (for instance, where they do not recommend any drug treatment, or where they do not have the necessary drugs in stock and must prescribe instead).

to dispense was removed - that is, when patients asked for a prescription instead. Given the implied cost constraints to dispensing, and that dispensing status is voluntary for GPs in South Africa, some selection bias in GPs' decision to dispense is likely. In Johannesburg, for instance, dispensing GPs were more likely to be located in less affluent areas, accept poorer patients, and display higher levels of measured altruism toward patient welfare than non-dispensing GPs (Blaauw and Lagarde, 2019).

2.5 Universal Health Coverage (UHC) Policies

The South African government has committed itself to progressing universal health coverage (UHC) for all citizens. Central to this effort are plans to dismantle the current public-private dichotomy and establish a single state-run health system; underpinned by national health insurance for the whole population; and centred on universal access to integrated and quality healthcare, initiated at the primary care level. An outline of planned reforms up to 2026 has been set out in two government bills tabled in the South African parliament at the time of writing: the 2019 National Health Insurance (NHI) bill, and the 2018 Medical Schemes Amendment bill. Certain aspects of these proposals are worth noting in the context of this thesis, as detailed below.

The NHI bill proposes to set up a single, state-run insurance fund ('NHI fund') which will pool mandatory tax contributions from all employed citizens (along with the government health budget) and purchase healthcare services on behalf of the whole population.[51] It aims to eliminate OOP expenditure for a comprehensive set of healthcare benefits and ensure equal access for all patients.[52] Accordingly, enrolment in the NHI system will be mandatory for all South Africans, and a universal package of health services and medicines will be made available free at the point of use. The fund will contract with accredited healthcare providers to deliver these services, in line with cost-effective treatment guidelines and at prices determined by a central committee. To gain NHI accreditation, healthcare providers will have to meet certain performance standards and be certified by the Office of Health Standards Compliance. In addition, they will be expected to routinely collect and relay specific care quality performance data, for monitoring and assessment by the fund.

These proposals promise to tackle many of the inequities and inefficiencies in the current two-tier system - in standardising healthcare access, eliminating OOP costs for essential care, reducing

[51] NHI implementation is being phased over a period of 9 years, and the system is planned to be operational by 2026.

[52] Access to NHI-funded services would be conditional on user registration at a contracted primary healthcare provider, which will also function as their point of entry into the health system. All specialist or hospital care may be accessed only through referral by the relevant primary care provider. The fund can refuse reimbursement of health services in cases where it is not included in the universal benefits package, where there is no medical necessity for the service, where the service is not cost-effective according to treatment guidelines, or where the patient does not comply with prescribed referral pathways.

care fragmentation, and enforcing quality and cost-effectiveness in care provision. Nevertheless, serious concerns remain regarding its sustainability[53] and acceptability to all stakeholders. In particular, there are questions on its attractiveness to the private healthcare sector, and the population that rely on it. All current medical scheme beneficiaries will be obliged to contribute to the NHI fund and will no longer receive tax breaks to offset their medical scheme contributions. To avoid payment for duplicative cover, the NHI bill states that medical schemes may only provide complementary cover for health services not covered under the NHI benefits package once the fund is operational.[54]

The Medical Schemes Amendment bill sets out a number of reforms to medical scheme operations, in order to pave the way for full NHI implementation. Perhaps most crucially, the bill states that all co-payments on health services covered by the schemes will be abolished, eliminating all OOP expenditure by beneficiaries. Another amendment states that medical schemes must cover a more comprehensive package of health services (perhaps similar to what the NHI will eventually cover) than just the prescribed minimum benefits that they are currently obliged to cover. The expectation is that this coverage will gradually fall as the NHI comes into operation, at which point the schemes will only provide complimentary cover.[55]

The NHI implications for private GPs are of particular relevance to this study. The public sector lacks the health workforce and capacity to service the NHI system, so the contracting of private providers and facilities to deliver NHI-funded services is an essential part of government proposals. Sustainable contracting of private GPs will require that contractual terms and working conditions are sufficiently attractive to these providers, whilst maintaining incentives for cost-control.[56] A number of contracting models for private GPs are currently being explored, including capped FFS tariffs and sessional contracts. Part-time sessional contracts for private GPs to work in public clinics were trialled in NHI pilot sites between 2012 and 2017.[57] However, uptake of the

---

[53] The system is expected to cost around R265 billion (USD 17 billion) by 2022 (Ngcuka, 2019), and a history or corruption and resource misuse in the public healthcare sector raises questions about the state's capacity to govern such a complex and costly system.

[54] In practice, there is likely to be a significant market for such top-up insurance products – for instance, to cover discretionary medical procedures that are unlikely to be covered by NHI, and to allow direct access to specialist care for people wishing to avoid the wait times and mandatory referral pathways under NHI.

[55] Other proposals in the bill include the abolishment of medical scheme brokers, whose increasingly costly services have shown little return in growing beneficiary numbers; and the set-up of a central beneficiary registry, to gather data on consumer profiles and behaviours in the medical schemes market, for informing the NHI design.

[56] While the exact contractual terms for individual GPs are still to be defined, the latest proposals for organising primary healthcare delivery at the local level are the following (Republic of South Africa, 2019). Contracting Units for Primary Health Care (CU) will be set up at the sub-district level, managed by a horizontal network of local healthcare providers (including a district hospital, primary care facilities or clinics, and individual GPs). These units will be responsible for primary care provision within defined geographic areas (sub-districts), and are expected to contract with the NHI fund on a capitation basis (they will be paid a fixed amount over a defined period of time for each NHI-registered individual in their area). The CU's will then be responsible for contracting and remunerating accredited private GPs in their areas.

[57] The proposed reforms to primary health care were rolled out in 11 pilot districts between 2012 and 2017. The sessional contract trialled with GPs contracts required them to work a portion of their time each week in a designated public healthcare clinic, for which they were paid a fixed hourly rate (sessional rate). The basic rate was set at the maximum public sector rate for doctors.

sessional contracts was very low (only 330 GPs had been contracted by 2017/2018) (NDoH, 2019). GPs flagged several issues with the pilot contract, including a lack of clarity in the contract terms; sessional rates that were not high enough or market-related; limited clinical autonomy and inability to work at their own practice location; and the high workload and poor working conditions in public clinics (Blaauw and Lagarde, 2015; Hongoro *et al.*, 2015). Capped FFS tariffs without any form of cost-sharing by individual GPs can also incentivise unnecessary care provision, if GPs respond to capped prices by increasing service volumes to maintain their profits.[58] While the current FFS model of GP payment is unlikely to be sustainable in the NHI context, in the absence of convincing evidence on the perverse incentives and costs of FFS in a prospective universal health insurance system, ensuring buy-in from GPs and other stakeholders for payment models that move too far from the status quo will be a challenge.

For the purposes of this thesis, it is important to underline the following points from this chapter. The thesis primarily focuses on the population of private GPs that are *i)* contracted-in with medical schemes, *ii)* non-dispensing, and *iii)* paid FFS. These GPs comprise the majority of primary care physicians in South Africa, and they are currently not subject to much cost containment incentives or quality control. Moreover, these GPs will be increasingly exposed to fully financially-protected patients both in the private sector (as OOP payments are eliminated in medical schemes) and in the future NHI system. Understanding how their treatment decisions may change with the level of patient insurance under current contracting terms – and equally, how they may be incentivised to rationalise inappropriate treatments - can therefore inform the future contracting of these providers.

[58] While the CU's will be subject to cost-sharing incentives (through capitated payment methods), it is unclear how and to what extent these incentives will be passed on to individual GPs in their areas.

3 Theoretical Framework

This chapter seeks to model a typical physician-patient interaction in the private primary healthcare sector, replicating certain features of the thesis context in South Africa. The following framework is an extension of those proposed in McGuire (2000) and Bardey & Lesur (2006). It includes two agents: a risk-averse patient and a physician (assumed risk-neutral). As in standard physician agency models with asymmetric information, it is assumed that healthcare consumption is determined by physicians at the point of care access (either objectively, or based on beliefs about patient preferences): patients trust physicians, to some degree, to make healthcare choices on their behalf. The market for physician services is monopolistically competitive, as commonly modelled (Pauly and Satterthwaite, 1981; Dranove and Satterthwaite, 1992; Gaynor, 1994; McGuire, 2000; Gaynor, Ho and Town, 2015), where physicians retain some market power from service differentiation, switching costs, and a high degree of trust in physician-patient relationships, but patients have some choice over the physician they consult.[59] The physician's payment incentives are fixed – I assume he is paid fee-for-service (FFS) and does not dispense drugs ('non-dispensing').[60] However, the insurance status of the patient can vary: the patient can be 'uninsured' (paying fully out-of-pocket (OOP)), 'low-insured' (with limited insurance cover, paying partially OOP), or 'high-insured' (with full insurance cover). A third-party payer (health insurer) fixes the reimbursement rates for all physician services and prescribed drugs in the market.[61] This set-up closely approximates the current institutional context in South Africa, where the majority of private primary care physicians are paid FFS, non-dispensing, and contracted-in with all large insurers that fix reimbursement rates and offer varying levels of patient insurance.

The consultation scenario to be modelled is the following: the patient presents to the physician reporting a set of symptoms, and carrying a single (unknown) underlying illness. Information is incomplete: patient preferences are known to the physician, but the physician must conduct a number of observed and unobserved actions to gather private information about the patient's

[59] The physician's market power in this model stems from both asymmetric information and monopolistic competition.

[60] While he is paid a unit fee for any tests or treatments he administers during a consultation, he is not affected financially from the prescription of drugs.

[61] I abstract from any variations in reimbursement procedures, and assume that all physician services are claimed for directly by the physician rather than the patient (which reduces the scope for balance-billing), while all prescribed drugs must be purchased from pharmacies by patients paying up-front and claiming from the insurer later. As the physicians in this model are assumed to be 'contracted-in' with the third-party payer (as with the majority of private physicians in Johannesburg), *de facto* price regulation can be reasonably assumed.

condition. The problem for the physician is that the patient's symptoms can be compatible with multiple underlying illnesses of varying severity, and an accurate diagnosis greatly increases the likelihood that he prescribes an appropriate treatment to cure the patient. He must first deploy costly clinical effort and the necessary diagnostic procedures to identify the true illness, and then exert further effort to identify a therapeutic recommendation that is both clinically-indicated and aligned with the patient's preferences and financial constraints. Moreover, he must do this whilst satisfying his own objectives, which may not always be aligned with those of the healthcare consumer (the patient) or payer (the insurer; and the patient, if not high-insured). This raises the possibility of *imperfect* agency in the physician-patient and physician-insurer relationships.

The patient cannot observe whether the physician exerts the necessary effort to diagnose and treat his illness appropriately. That is, he cannot observe the *technical quality* of the care he receives. However, where there are cost implications to poor quality care, and the patient is not fully insured, he may infer it from subsequent OOP costs.[62] [63] In essence, the informational disadvantage of the patient with respect to technical quality (and hence his demand elasticity) is determined by the salient costs of poor quality care to that patient:[64] when a patient's OOP share of health expenditure is high, he may be more aware of the cost implications of poor quality care, and more willing to invest in obtaining and responding to better information on his physician's performance.

Unlike McGuire (2000), I allow that physicians may attach some weight to patient welfare in their utility, independent of any related business gains. The framework further builds on Bardey & Lesur (2006) by considering that *i)* physicians not only choose service volume (quantity) but also effort, and that *ii)* they are subject to dynamic (indirect) financial incentives from competition (in addition to immediate payment incentives). In the framework extensions (Section 3.4), I allow that effort is not a homogeneous input, as modelled by McGuire (2000), but that it can have several dimensions – some more easily observable to the patient than others (Chalkley and Malcomson, 1998). A final extension proposed is that physicians' intrinsic motivations (how they value patient welfare for its own sake) may not be fully fixed or exogenous, as commonly modelled. Instead, I consider how certain informational incentives can improve physicians' intrinsic motivations for benefiting patients.

---

[62] For instance, poor quality care can result in additional care costs where poor diagnostic or therapeutic accuracy leaves the underlying illness un-treated, and results in the patient needing to seek additional care elsewhere.

[63] The patient may also infer care quality from subsequent treatment outputs and care outcomes, *if* he is willing to incur some informational cost of his own (for instance, in seeking out publicly available care quality information).

[64] This in turn determines the degree to which he delegates decision-making authority to the physician: the patient is more likely to trust the physician to make decisions on his behalf (and consequently, have a lower demand response to technical care quality) when his informational disadvantage is greater, relative to the physician.

This chapter is organised as follows. Sections 3.1 and 3.2 outline the objectives of the patient and physician, respectively. The physician's decision problem and optimality conditions are also presented in Section 3.2, and their implication are discussed in Section 3.3. Framework extensions are considered in Section 3.4, and a concluding discussion is presented in Section 3.5.

3.1 The Patient

A patient j's benefit from a clinical encounter with physician i (U_{ij}) is an increasing, concave function of two arguments:[65] the quantity of physician services provided in the encounter x,[66] and the effort expended by the physician e (which determines the technical quality and cost-effectiveness of any treatments).[67] The quantity of physician services x is perfectly observed by the patient. Effort e can be both a 'diagnostic' (physical and cognitive effort to arrive at an accurate diagnosis) and 'therapeutic' (cognitive effort to identify clinically-indicated and cost-effective treatments for a specific diagnosis)[68] care input. It can therefore have both easily observable physical components, and more hidden cognitive components. However, it is assumed here for simplicity that e - insofar as it contributes to technical care quality - is not directly observed by the patient.[69] [70] The patient's benefit function is then given by:

$$U_{ij}(e,x) = h_{ij}(e,x) + x_{ij} + m_{ij}(e) \quad (1)$$

where h_{ij} is the marginal health stock attributable to consultation ij, and m_{ij} is the quantity of medications prescribed. h_{ij} is an increasing, concave function of the technical quality of care received (captured by unobserved physician effort e, which in turn affects the appropriateness of any drug treatment m); and to a lesser extent, of the quantity of physician services x.[71] [72] h_{ij} is

[65] The concavity of the benefit function reflects the patient's risk aversion. It may be reasonable to assume that U_j is initially increasing and then eventually decreasing in x, as certain healthcare procedures may actually harm the patient if administered at excessive levels.
[66] This includes all diagnostic tests or procedures ordered during the consultation, which the physician is compensated for.
[67] "Quality" in this context of a physician-patient interaction refers to the *clinical* quality of care, which reflects the care *process* through which the health system's structural inputs are transformed into health outcomes (Donabedian, 1988). As effort is linked to both quality *and* cost-effectiveness, higher effort is also assumed to minimise costs without any tradeoff on necessary quality.
[68] Therapeutic effort can also extend to explaining appropriate treatment choices to patients and ensuring compliance - especially where those choices run counter to patient demands.
[69] The simplifying assumption here is that effort is uni-dimensional, and that *all* effort is "appropriate" in improving the technical quality of care. This assumption is relaxed in Section 3.4, where it is considered that that some types of effort – particularly the observable types - may instead affect interpersonal aspects of care quality (i.e. patient experience) rather than its technical aspects.
[70] Even where it is not directly observed, e may be partially inferred *ex post* from subsequent health and cost outcomes. This assumption is discussed further in Section 3.4.
[71] It is likely that $|h_x(.)| < |h_e(.)|$ at most values of x and e (except in extreme cases), as it is the technical quality of care (rather than quantity) that should primarily determine health impact. As with m, it is also plausible that x is decreasing in e (as greater appropriate physician effort should minimize the need for multiple diagnostic tests or procedures). However, building on other modeling approaches in the literature (McGuire, 2000), I assume here that x and e are independent. Modeling x as a decreasing function of e will not change the main implications of this framework (presented in Section 3.3)
[72] I abstract here from any random factors that can also influence h_{ij}, as this does not change the framework's implications.

not perfectly verifiable by the patient: while he can observe his aggregate health state H_j (and any changes to it), the patient faces some uncertainty in the extent to which any changes are directly attributable to a specific consultation ij. It follows that he cannot perfectly verify how the physician's care changes his health state.

The quantity of medications m is a decreasing (concave) function of e: $m'(e) < 0, m''(e) < 0$. Higher effort improves the physician's diagnostic and therapeutic accuracy (decreases clinical uncertainty). With more effort, physicians are able to minimise diagnostic uncertainty, identify the most appropriate treatments for individual patients, and therefore prescribe fewer medications in total.[73] [74] Where physicians stint on effort, they compensate by prescribing more (and more types of) medications, as a wider range of medications increases the likelihood of addressing the true illness under diagnostic uncertainty. It further appeases the patient, and the (non-dispensing) physician incurs no cost to prescribing.

The observable components of the physician's care (x and m) also enter the patient's benefit function independently of any contribution to health stock. This captures the notion that where patients cannot perfectly verify the necessity or quality of the care they receive, they will derive some independent value from observed care quantity (perhaps inferring that a higher quantity of care should correlate with a larger gain in health stock).[75] An important assumption is that at any level of e, $|h_e| > |m_e|$.[76] The patient is always better off with higher physician effort, despite the countervailing utility effect of fewer prescribed medications. This is plausible given potential adverse reactions and long-term health costs to over-medicating and inappropriate medications, which may limit (or even counter) the positive utility effect of prescribing more medications.

There are also three types of costs from a clinical encounter. First, there is a cost p_x per unit of physician services (x) provided. Second, there is an average unit price p_m for the medications prescribed. As with the quantity of medications (m), this unit price p_m also falls with e: $p'_m(e) < 0$. This can be for two reasons. Physicians that expend more effort are able to identify cheaper drugs, such as cheaper branded or generic substitutes on the market. Moreover, in the absence of sufficient effort for diagnostic or therapeutic accuracy, the physician compensates by prescribing more first-in-class and expensive drugs in the belief that newer, more costly drugs increase the

[73] Precautionary antibiotic use is a common example of where diagnostic uncertainty results in unnecessary drug prescriptions (Leigh *et al.*, 2019), particularly for respiratory infections (Whaley *et al.*, 2013; Brink *et al.*, 2016).

[74] Aside from minimizing diagnostic uncertainty, higher effort can also reduce the number and inappropriateness of medications prescribed by countering patient demands for medications. For instance, physicians report prescribing inappropriate antibiotics in order to avoid lengthy explanations to demanding patients as to why antibiotics are unnecessary (Dempsey *et al.*, 2014).

[75] This may be particularly relevant to medications and other treatments that provide symptomatic relief. As patients can directly observe and attribute any symptomatic effects (unlike health outcomes), they may derive some utility from the quantity of such treatments independent of any effect on health comes (including any negative health effects from over-treatment).

[76] The increase (decrease) in health benefit from a given increase (decrease) in e is always larger than the corresponding decrease (increase) in $m(e)$.

likelihood of addressing the (unknown) underlying illness[77] - and at worst, that it just appeases the patient. However, the physician's ability to lower p_m through exerting more effort falls with his effort level: $p_m''(e) < 0$. Lastly, there is a total consultation cost of $p_R C_R$ for a particular illness episode, where p_R is the physician's flat consultation fee, and C_R is the number of (current and future) consultations required to resolve the illness episode (including any repeat consultations, either with the same physician or other physicians). The total number of consultations required C_R is decreasing in the physician's effort during a given encounter: $C_R'(e) < 0$. The harder the physician works to diagnose and treat a patient appropriately, the less likely it is that the patient will require further consultations to address the same illness.[78] However, the more effort he exerts, the smaller the marginal reduction in repeat consultations for the patient: $C_R''(e) < 0$.

The patient's OOP share of these costs is determined by the co-payment parameter θ_j, which is decreasing in the patient's level of insurance ($\theta_j = 0$ for high-insured, $0 < \theta_j < 1$ for low-insured, and $\theta_j = 1$ for uninsured patients).[79]

A patient j's net benefit (NB) from a consultation ij is then:

$$NB_{ij} = U_{ij}(e,x) - \theta_j(\, p_R C_{R_{ij}}(e) + \, p_{m_{ij}}(e) m_{ij}(e) + \, p_x x_{ij}) \qquad (2)$$

Note that, while the OOP cost components of this function are perfectly observable to the patient, $U_{ij}(e,x)$ is neither fully observable nor perfectly verifiable given the patient's uncertainty around $h_{ij}(e,x)$.

3.2 The Physician

Drawing on a general consensus in the literature (Galizzi *et al.,* 2015), the physician's utility from a clinical interaction is an increasing and separable function of two arguments: the profit generated by his medical choices π, and the welfare (or net benefit) of the patient (NB). The weight (β_i) that physician i attaches to patient welfare (for its own sake) is a measure of his pro-

[77] This relationship between diagnostic uncertainty and more expensive drug choices is indicated in Takemura et al. (2005)'s analysis of antibiotic choices for patients with and without advance diagnostic testing. As with the relationship between uncertainty and the quantity of medications, this reflects the notion of defensive medicine: greater clinical uncertainty encourages physicians to "do more", in order to avoid malpractice liability risks. It follows that $|p_m'(e)|$ and $|m'(e)|$ will likely be greater the higher the expected malpractice liability costs.

[78] Where the physician manages to resolve the illness in a given consultation, and no further consultations are required, C_R=1.

[79] I abstract from common non-linearities in insurance contracts (where the co-pay can vary with the level and types of services claimed) and assume θ_j is fixed here.

social preferences (altruism), and is assumed heterogeneous across physicians and exogenous.[80][81] Although altruism may be considered conceptually different from physicians' professionalism (or intrinsic motivation), the nature of the medical profession is such that significant overlap between intrinsic motivation and altruism can be reasonably assumed: professional excellence is likely to be heavily correlated with consistently delivering benefit to the patient. As such, while acknowledging there may be some conceptual differences, I assume here that intrinsic motivation incorporates altruism as well.[82] Section 3.4.2 extends this discussion on intrinsic motivation and its sources in greater detail.

A physician i's profit from a consultation with patient j is made up of three components:

$$\pi_{ij} = n_{ij}(NB_{ij})p_R + (p_x - c)x_{ij} - C(e_{ij}) \quad (3)$$

The first component is the product of the flat consultation rate p_R[83] and the number of future consultations n that the physician gains from his performance during the interaction. n is an increasing function of the net benefit (NB) he generates for the patient (McGuire, 2000).[84] This captures the dynamic incentives facing the physician in any given consultation: his performance not only determines his immediate payoff from the consultation, but also impacts his reputation in the market and future business.[85] A patient is more likely to return to the same physician when his NB is higher, and may also recommend the physician to friends and family. The rate at which n changes with NB (its elasticity) is a measure of the market competition facing the physician and the informational disadvantage of the patient (asymmetric information). The higher the number of (well-known) competitors in the market, and the more responsive the patient demand with respect to changes in NB, the more business the physician gains for a given increase in

[80] In practice, pro-social preferences can be a combination of individual-specific aspects (i.e. intrinsic, or natural altruism) and common (homogeneous) aspects within a defined physician population (social or professional norms that can be nurtured). This is in line with Ashraf & Bandiera (2017b)'s modelling of "altruistic capital" (Ashraf and Bandiera, 2017). For simplicity, I assume only individual-specific aspects are significant here.

[81] β_i is assumed uniformly-distributed in the interval [0,1]. $\beta_i = 0$ for purely profit-motivated physicians, who act in their patients' interests only so far as it increases their future business. Where $0 < \beta_i < 1$, physicians are partially altruistic, but weight patient welfare (for its own sake) lower than their profits. Where $\beta_i = 1$, physicians give equal weight to patient welfare and profit. I make the assumption here that physicians never weight patient welfare for its own sake higher than profit (which may be more justified in the private sector context of this study than in the public sector).

[82] Altruism is often modeled as a particular form of intrinsic motivation in the literature (Besley and Ghatak, 2018). Other forms of intrinsic motivation relate to reference-based utility, whereby agents are motivated to knowingly perform well relative to professional benchmarks or social standards (Kolstad, 2013). It follows that interventions appealing to physicians' intrinsic motivations – for example, audit and private feedback on clinical performance (see Chapter 7) – can be interpreted as essentially targeting β_i. This is discussed further in Section 3.4.2.

[83] I assume here that the consultation rate p_R is exogenous (fixed by a third-party e.g. the insurer) and constant across all patient types, regardless of their insurance cover and how much they pay OOP.

[84] Where the net benefit (NB) is zero, the physician gains no future consultations from his performance and n=1.

[85] While the total pay-off from any future consultations is the sum of the consultation rate p_R and any FFS profit $(p_x - c)x$, the FFS profit would depend on the physician's future choices of x, and is thus uncertain. I therefore assume for simplicity that the physician only considers the fixed (and guaranteed) component of his future payoffs (the consultation rate p_R) in optimizing his choices of x and e in any given consultation. Considerations of future opportunities for demand inducement are omitted here.

patient NB: the patient (and his friends and family) are then more responsive to performance differentials between physicians.[86]

The physician is paid FFS for all services x at the fixed unit price p_x, which include any diagnostic tests or procedures that can be administered or ordered during consultations. He incurs a fixed unit cost c for these services, so his total FFS profit for providing x services is $(p_x - c)x$. I assume that $p_x > c$, which ensures the participation of physicians in FFS payment contracts, and reflects the imperfect competition in the physician market. The physician also incurs a cost to effort $C_i(e)$. This captures the behavioural (communication, examination), cognitive and time costs involved in delivering high-quality and cost-effective care - including the costs to changing any persistent treatment habits or behavioural norms that run counter to appropriate care.[87] While this is not a monetary cost *per se*, physicians can attach a monetary value to their effort supply. This value is not reimbursed, as effort is difficult to verify and contract for by patients or third-party payers, and therefore enters the physician's profit function as a net cost.[88] The marginal effort cost is given by $C_i'(e) > 0$, and this is assumed constant for all levels e.

The physician's problem is then to choose the optimal quantity of services x and effort e to maximise his objective function, given by:

$$V_{ij}(e,x) = n_{ij}(NB_{ij})p_R + (p_x - c)x_{ij} - C_i(e) + \beta_i(NB_{ij}) \quad (4)$$

From (4), it can be seen that patient net benefit enters the physician's objective function at two points. The physician considers patient welfare in his decision-making due to both its impact on his future business (his *competitive* or *market incentives*) and his concern for patient wellbeing for its own sake (his *altruistic incentives*). Asymmetric information in the physician-patient relationship is characterised by the assumption that - unlike the patient - the physician has perfect information on NB_{ij}, given his medical expertise and knowledge of patient preferences.

The optimal choice of service quantity x^* is then determined by the first-order condition:

$$c - p_x = (\, n_i'(NB)p_R + \beta_i \,)(U_x - \theta_j p_x) \quad (5)$$

---

[86] This requires that both informational and search barriers to patients' demand responses are low (Pauly, 1988); that the costs to obtaining accurate information on NB are significantly lower than the expected benefits to patients, and that patients are aware of alternative, accessible sources of care in the market. For instance, interventions to facilitate public reporting of care quality metrics, or greater patient choice among healthcare providers, essentially seek to improve this demand elasticity.

[87] This can include any behavioural costs involved in changing physicians' default treatment patterns, such as in switching better-known branded drugs for lesser-known generic equivalents.

[88] $C_i(e)$ is assumed heterogeneous across physicians. Although not explicitly modelled here, it is likely to depend on physicians' clinical knowledge, competence, or even age. The more knowledgeable or capable the physician, the smaller the likely cost to effort; and the older the physician, the more likely that treatment habits become engrained and harder to overcome. $C_i(e)$ may also be affected by common environmental factors, including the accessibility and utility of clinical decision aides, such as evidence-based treatment guidelines. Interventions to improve access and periodic reference to such guidelines may lower effort costs for all physicians, and act as effort enablers.

and the optimal effort choice e^* by:

$$C_i'(e) = (\, n_i'(NB)p_R + \beta_i)\left(U_e - \theta_j(p_R C_R'(e) + p_m'(e)m + p_m m'(e)\right) \qquad (6)$$

It is assumed for simplicity that $n_i'(NB)$, the elasticity of physician i's business with respect to the net benefit he generates for a patient, is positive and constant across all patient types (uninsured, low-insured and high-insured) and all components of patient net benefit.[89] [90] However, $n_i'(NB)$ is assumed heterogeneous across physicians, to reflect variations in the level of market competition facing individual physicians (for instance, in the number of local competitors).

3.3 Framework Implications

The two first-order conditions offer a number of hypotheses about physicians' decision-making under different patient insurance statuses, as detailed in the following.

3.3.1 Demand-inducement

From (5), under a constant FFS payment p_x for x, the physician always *induces* demand for his services: at the physician's optimal choice x^*, the marginal benefit for the patient U_x is lower than their marginal cost $\theta_j p_x$ for all insurance types.[91]

> *Hypothesis 1*: The physician *over-supplies* the quantity of healthcare services x, relative to what the patient would demand under perfect information and free choice (where it must be that $U_x = \theta_j p_x$),[92] regardless of patient insurance type.

This demand-inducement behaviour is possible because patients cannot perfectly observe U_x (asymmetric information), which allows the physician to influence demand beyond the patient's optimum. In addition, this result is driven by positive FFS profits ($p_x > c$): in a perfectly

[89] Nevertheless, it may be that certain types of patients are more responsive to their realised net benefit, either through greater effort in understanding the benefits of certain care procedures, or through greater information sharing among peers (which can generate greater business elasticity from these patients). Moreover, patients' ability to observe the different components of their net benefit will also vary: the quantity of care (x and m) will be more observable than technical quality (or physician effort e), and the contribution of unobserved effort e to care costs will be more observable than its contribution to health stock. This can result in variations in the demand response $n'(NB)$ across different patients and components of care. These issues are discussed further in Section 3.4.

[90] A further implicit assumption here is that the physician's business elasticity does not change with his market share (aggregate business). This can be justified if individual physicians' market shares are too small to matter to patients.

[91] This follows from the condition that $c < p_x$, and the assumptions that $n_i'(NB)$ and β_i are non-negative. For equation (5) to hold, it must be that $U_x < \theta_j p_x$ at the physician's optimum.

[92] This is the premise of the *supplier-induced demand* (SID) literature (Arrow, 1963; Evans, 1974; Fuchs, 1978), and is one form of physician moral hazard considered in models of imperfect physician agency (McGuire, 2000).

competitive market where $p_x = c$, the physician has no incentive for demand-inducement, *even* under asymmetric information.

Moreover from (5), since U_x is decreasing in x, the physician's capacity for demand-inducement is greater for more insured patients (all else equal). This is the *second-degree moral hazard* hypothesis (Balafoutas *et al.*, 2017). Less-insured patients suffer larger financial costs to demand inducement and would have a larger (negative) demand response to any suspected unnecessary costs. Physicians therefore have both altruistic and competitive incentives to induce less demand from these patients. Where the patient is fully insured ($\theta_j = 0$) however, the physician is able to induce sufficient demand such that the marginal benefit to the patient U_x is actually negative at the physician's optimal choice x^*.

> *Hypothesis 2*: The more insured the patient (the lower the out-of-pocket price $\theta_j p_x$), the higher the quantity of physician services supplied (x^*).

Note that there are two constraints on the physician's propensity to induce demand: $n_i'(NB)$ and β_i. The competitive (market) pressures facing the physician and his individual altruism compel him to consider patient welfare in his decision-making. These constraints are also the factors that differentiate the extent of demand-inducement by patient insurance type (i.e. that generate the *second-degree moral hazard* behaviour). In the absence of any altruistic or competitive pressure (i.e. $\beta_i = 0$ and $n_i'(NB) = 0$), the physician would induce demand equally and infinitely for all patients (irrespective of insurance status).[93]

3.3.2 Effort-stinting

From (6), we see that - unlike the quantity of physician services, x - physician effort e is always under-supplied at the physician's optimum, for all insurance types. This is another form of physician moral hazard in imperfect agency models (Ma and Mcguire, 1997). Again, this effort-stinting behaviour is made possible by asymmetric information and motivated by an uncompensated cost to physician effort $C_i(e)$.

[93] It is more realistic to assume some upper bound on demand-inducement, due to a de facto budget constraint imposed by the patient (determined by their total income, net of insurance premiums) or the insurer (determined by contractual limits on total claims). This may still result in lower x^* for less-insured patients, who are likely to have a lower overall claims limit and budget for healthcare expenditures.

Hypothesis 3: The physician *under-supplies* effort e, relative to what the patient would demand under perfect information and choice (i.e. where it must be that $U_e = \theta_j(p_R C_R'(e) + p_m'(e)m + p_m m'(e))$), regardless of patient insurance type.

Moreover from (6), the more elastic the patient's OOP payments with respect to effort (i.e. the higher the value of $\theta_j|p_R C_R'(e) + p_m'(e)m + p_m m'(e)|$), the higher the e^* chosen. The choice of e^* is then higher for less insured patients, who have larger OOP payments at all levels of e. Hence, physicians have a greater propensity to stint on effort for high-insured patients - which is another variant of the *second-degree moral hazard* hypothesis (Balafoutas *et al.*, 2017).

Hypothesis 4: The more insured the patient, the lower the overall effort supplied by the physician (e^*), and the lower the technical quality of care outputs (accuracy of diagnoses and clinical-appropriateness of treatment choices).

Moreover, as higher effort increases physicians' diagnostic and therapeutic accuracy, and allows physicians to source more cost-effective treatment alternatives, the lower e for high-insured patients limits physicians' capacity for minimising drug treatment quantity and costs for these patients.[94] High-insured patients therefore have more (and more inappropriate) medications prescribed, and incur a higher overall cost of medications than low-insured ones (since $m'(e) < 0$ and $p_m'(e) < 0$). This result occurs despite the absence of any direct financial incentives attached to drug prescribing for the non-dispensing physician. Therefore, it is not driven by greater demand-inducement for profit – but rather, by poorer clinical effort with high-insured patients.

Hypothesis 5: The more insured the patient, the higher the quantity and costs of medications prescribed ($m(e^*)$ and $p_m(e^*)$).

As lower effort increases the number and cost of medications, and increases the likelihood of the patient needing further care for the same illness, both the financial costs to and (negative) demand response from any degree of effort-stinting by physicians will be greater for less insured patients. Once again, physicians have both altruistic and competitive incentives to supply more effort (stint less) for less insured patients. Therefore, the same two constraints that prevent the physician from inducing infinite demand for his services also prevent him from supplying zero effort: the

[94] Higher effort can also correspond to fewer and more appropriate treatments through another channel. Explaining to patients (who derive some utility from the quantity of medications) why certain popular medications like antibiotics or a higher quantity of medications may not always be beneficial can be time and effort-consuming for the physician. Therefore, unless the physician chooses to expend such effort, he may simply give in to appeasing the patient with more medications.

competitive pressures he faces $n_i'(NB)$ and his altruism β_i again force him to consider the patient's welfare in his choices of effort e. These constraints result in different levels of effort being supplied to uninsured, low-insured and high-insured patients; e^* will be the same for all insurance types and equal to 0 when $\beta_i = 0$ and $n_i'(NB) = 0$.

3.4 Framework Extensions

Two key assumptions of the preceding model are modified in the following. First, it was assumed that patient demand elasticity $n'(NB)$ is the same for physician service quantity x and all types of physician effort e. In practice, how patients respond to different physician inputs can vary depending on the relative visibility of those inputs. Physician effort e may be less recognisable (visible) to patients than service quantity x, resulting in lower demand elasticity with respect to effort. *Types* of physician effort can also be differentiated in a similar way: diagnostic effort that involves physical engagement with the patient (for example, physical examinations or diagnostic tests) will be more visible to the patient than physicians' *cognitive* effort in clinical decision-making, and hence generate higher demand elasticity. The implications of such heterogeneity in physicians' care inputs are discussed in Section 3.4.1.

Second, it was assumed that physicians' intrinsic motivation (which drives their concern for patient welfare for its own safe) is uni-dimensional and exogenous to the regulatory environment in which physicians practice. Section 3.4.2 discusses the implications of relaxing this assumption, and allowing third-party regulators a role in augmenting physicians' intrinsic incentives for appropriate effort.

3.4.1 Observable vs. hidden effort

Clinical effort e was previously modelled as a homogenous care quality input, contributing only to the unobserved technical quality (and subsequent cost-effectiveness) of care. However, it was noted that e may be decomposed into a number of different physician inputs (Chalkley and Malcomson, 1998) - some easily observable (henceforth denoted e^o), and others largely unobservable to the patient (denoted e^u). Effort can also relate to other aspects of care quality, including its interpersonal (verifiable and subjective) dimensions of patient experience and satisfaction (Haas-Wilson, 1994). While some observable physical effort e^o (in the form of history-taking, examination and patient communication) will be necessary for the physician to reach an accurate diagnosis and treatment, it is by no means sufficient. The physician will need to invest additional cognitive effort e^u in determining the most appropriate diagnostic procedures for gathering relevant information on the patient's clinical case, and in processing that information

to arrive at a correct diagnosis and treatment. In this view, it is e^u that ultimately determines (unobserved) technical care quality, while e^o may contribute to patient experience (and enhance interpersonal aspects of care).

An alternative interpretation of e^u is that it is the *appropriateness* of any observed effort e^o,[95] which implies some level of complementarity between e^o and e^u. However, Propper *et al.* (2008) and Dranove and Satterthwaite (2000) suggest that observed and unobserved care quality inputs may indeed be treated as substitutes by providers.[96] [97] In essence, physicians can 'price in' aspects of their costly effort that patients observe and value by stinting instead on unobserved effort and increasing care costs. This trade-off is possible *as long as patients are relatively cost-insensitive*. In other words, physicians' propensity to increase observable (interpersonal) effort at the expense of technical care quality and unnecessary costs should be greater for high-insured patients.

To formally model these considerations within the current framework, consider that patient benefit is now an increasing function of three arguments:

$$U_{ij}(e^o,\ e^u, x) = h_{ij}(e^u, x) + x_{ij} + m_{ij}(e^o, e^u) + e^o \tag{7}$$

Three key features of this function can be noted. As patients directly observe e^o, they are likely to derive utility from it in a similar way to observed x and m, independently of any effect on health outcomes. e^o therefore appears as a separate, independent argument in patient utility. Secondly, the health outcomes attributable to consultation ij are only a function of e^u (and x, as before), reflecting the notion that only *appropriate* physician effort – in other words, effort that determines the technical quality of care - matters for patient health (and subsequent care costs). Observable effort can have a health or cost impact, but only in combination with appropriate cognitive effort. Therefore, while observable effort can be utility-enhancing in itself (patients value their care experience, and would always prefer a more visibly engaged doctor), I assume

---

[95] In essence, e^o is the act of "doing something", while e^u guides these actions to "doing the *right* thing". Whether the observed care (e^o) is necessary or appropriate for a clinical case is not easily verifiable by the patient. Therefore, e^u may be taken as at least partially unobserved.

[96] For instance, in response to competition on quality, UK hospitals were found to prioritise observable and monitored aspects of care quality (wait times) at the expense of unobserved aspects, evidenced in improved wait times but worsened hospital mortality outcomes (Propper *et al*., 2008)

[97] The impact of increasing patient or payer visibility on associated aspects of physician care (be it price or quality) is modelled in Dranove and Satterthwaite (1992). While most empirical evidence on provider competition through public reporting of quality metrics comes from the hospital sector, similar evidence on the effects of direct financial incentives attached to measured and publicly reported quality metrics can be found for instance in literature documenting the effects of the UK's pay-for-performance initiative in primary care (the Quality and Outcomes Framework (QOF)). The QOF contained both competitive profit incentives from public-reporting of provider performance and explicit financial incentives attached to measured indicators, whereas former initiatives in the UK hospital sector relied only on competitive profit incentives from public-reporting. Reviews on the impact of P4P schemes generally suggest that, while there have been modest positive effects on some targeted (incentivised) process indicators of care quality, there is little evidence of effects on health outcomes (Doran, Maurer and Ryan, 2017) and some evidence of detrimental effects on non-incentivised aspects of care quality (Doran *et al.*, 2011). It is nevertheless difficult to separate the effects of the competitive (reputational) incentives from the monitoring and public reporting of care quality indicators from the effects of the financial incentives themselves.

that only e^u matters for a health or cost impact. Thirdly, the quantity of medications m is a decreasing (concave) function of e^u ($m'_{e^u}(.) < 0$, $m''_{e^u}(.) < 0$) and an *increasing* (concave) function of e^o ($m'_{e^o}(.) > 0$, $m''_{e^o}(.) < 0$). Physicians supply a greater number and variety of medications to compensate for less appropriate diagnostic and therapeutic effort (e^u). However, a higher quantity of medications contributes to patients' subjective care experience (and potential satisfaction) and may enable more immediate symptomatic relief (without necessarily addressing the underlying illness, and despite potential longer-term risks), and therefore aligns with physicians' efforts to be seen to be "doing something" (e^o).[98] [99]

The patient's net benefit (NB) is now given as:

$$NB_{ij} = U_{ij}(e^o, e^u, x) - \theta_j(p_R C_{R_{ij}}(e^u) + p_{m_{ij}}(e^o, e^u) m_{ij}(e^o, e^u) + p_x x_{ij}) \quad (8)$$

The number of consultations for a particular illness episode (C_R) and the average price of prescribed medications (p_m) are also decreasing in e^u: with more appropriate effort, physicians are able to diagnose and treat the underlying illness more effectively (and therefore lower the number of repeat consultations necessary to resolve the same illness episode), and also identify more cost-effective treatments on the market. However, p_m is *increasing* in physicians' observable effort e^o. This relationship reflects patients' actual or perceived preferences – *ceteris paribus* - for relatively more expensive options among equivalent treatments, such as branded drugs over generic equivalents. Patients may, for instance, infer a higher quality of treatment from better-known, branded drugs (Himmel *et al.*, 2005; Shrank *et al.*, 2009). Therefore, physicians seeking to improve the patient's experience of their care (through higher e^o) may reasonably prioritise the prescription of relatively more expensive drugs, all else equal.

The physician's objective function is then:

$$V_{ij}(e^o, e^u, x) = n_{ij}(NB_{ij})p_R + (p_x - c)x_{ij} - C_i^o(e^o) - C_i^u(e^u) + \beta_i(NB_{ij}) \quad (9)$$

where $C_i^o(e^o)$ and $C_i^u(e^u)$ are the uncompensated costs to observed and unobserved effort respectively. For simplicity, I assume the marginal costs to both observed and unobserved effort are equal, positive and constant: $C_i^{o\prime}(e^o) = C_i^{u\prime}(e^u)$. Maximising (9) with respect to e^o and e^u, the physician's optimal choices of each effort type are given by the first-order conditions:

$$C_i^{o\prime}(e^o) = (n_i'(NB)p_R + \beta_i)(U_{e^o} - \theta_j(p_{m_{e^o}}(.)m + p_m m_{e^o}(.))) \quad (10)$$

---

[98] This reflects a finding that physicians often prescribe unnecessary medications – particularly antibiotics – purely to satisfy actual or perceived patient expectations, rather than due to poor clinical judgement (Dempsey *et al.*, 2014; Ashworth *et al.*, 2016).

[99] For the same reasons, it is reasonable that x may also increase with e^o (and fall with e^u), somewhat in line with Ma and McGuire (1997)'s modelling of patient demand for medical services (x) being influenced by their observations of physician effort (e^o).

and

$$C_i^{u\prime}(e^u) = (n_i'(NB)p_R + \beta_i)(U_{e^u} - \theta_j(p_R C_R'(e^u) + p_{m_{e^u}}(.)m + p_m m_{e^u}(.))) \quad (11)$$

From (10), it can be seen that at any level of $C_i^{o\prime}(e^o)$, physicians provide a higher level of e^o to high-insured patients (with $\theta_j = 0$) than to low-insured ones (with $\theta_j > 0$) at their optimum.[100] On the other hand, they provide a lower level of e^u to high-insured patients compared to low-insured ones, as indicated in (11).[101] It follows that high-insured patients will also receive a higher quantity and average cost of prescribed medications.[102] [103]

In sum, the cost-sensitivity of the patient (captured by their level of cost-sharing) affects physicians' relative choices of observed and unobserved (appropriate) effort, and subsequent treatment quantity and costs.[104] Physicians exploit the lower cost-consciousness of high-insured patients to supply lower technical quality and less cost-effective care to these patients (captured in a lower e^u). At the same time, patients in monopolistically competitive markets still choose among competing providers on the basis of some observable factor of value – be it price or patient experience (which determine providers' reputation in the market). Hence, physicians' reduced ability to compete on cost (price) for high-insured patients means they are likely to prioritise observable effort e^o for these patients (Haas-Wilson, 1994; Dranove and Satterthwaite, 2000; Gaynor, Ho and Town, 2015).[105] These considerations imply an amendment to *Hypothesis 4* and a new *Hypothesis 6*.[106]

> *Hypothesis 4* (amended): The more insured the patient (the lower the θ_j), the lower the unobserved, appropriate effort supplied by physicians (e^u); and the lower the technical

---

[100] This follows from the assumptions that $p_{m_{e^o}}(.)m > 0$, $p_m m_{e^o}(.) > 0$ and $U_{e^o} > 0$, while U_{e^o} is decreasing in e^o.

[101] Consider that $C_R'(e^u)$, $p_{m_{e^u}}(.)$ and $m_{e^u}(.)$ are all < 0, while as assumed previously, $U_{e^u} > 0$ and decreasing in e^u.

[102] Physicians' optimal choices of x are unchanged from those discussed in Section 3.3.

[103] While these results compare patients with no cost-sharing (fully-insured) to patients with some level of cost-sharing (partially-insured), it can be shown that the same pattern of results also hold for comparisons of patients with different *levels* of cost-sharing: patients with lower levels of cost-sharing (lower θ_j) would receive higher e^o, lower e^u and higher treatment costs.

[104] While demand elasticity $n_i'(NB)$ is assumed to be equal for both observed and unobserved effort in equations (10) and (11), it is reasonable to expect that it may be higher for observed effort for all patients. Nevertheless, this would not change the hypothesised effort choice differences by insurance type (*Hypotheses* 4-6).

[105] From the patient's net benefit function in (8), it can be shown that high-insured patients would indeed demand more e^o at their own optimum than low-insured ones.

[106] These hypotheses rely on the assumption that technical care quality and cost-effectiveness can only be inferred *indirectly* through their OOP cost implications for patients (and hence, only by partially-insured patients). However, the same hypotheses can also be derived if we allow that technical quality (or efficiency) can be inferred more directly, with some search or informational cost to the patient – for instance, in seeking out publicly available care quality information, or patient education tools. This aligns with the conception of healthcare as a search or reputation good, rather than a pure credence good (Pauly, 1978, 1988). In this context, patients' demand for (and willingness to invest in) better care quality information would depend on the costs of poor quality care to them, which in turn would increase in their level of cost-sharing: low-insured patients would invest more in better observing e^u, and exert more competitive pressure on physicians to increase it.

quality of care outputs (accuracy of diagnoses and clinical-appropriateness of treatment choices).

Hypothesis 6: The more insured the patient, the higher the observable effort supplied (e^o) - in terms of the quantity of history-taking, examinations, patient communication and consultation time.[107]

3.4.2 Intrinsic incentives for effort

In previous sections, the objective of the third-party payer (the 'principal' health insurer) was not explicitly discussed. Here, I consider the role of the insurer in regulating the physician's healthcare choices in line with its objective. The insurer's problem is to maximise the patient's health stock ($h(e,x)$) whilst minimising its healthcare payments to the physician (($1-\theta$)($p_R C_R(e) + p_m(e)m(e) + p_x x$)), subject to the physician's participation constraint ($p_x > c$). The model set out in this chapter implies that the insurer can only manipulate extrinsic incentives - which appeal to physicians' financial or reputational concerns - to motivate high quality and cost-effective care. These include the reimbursement structure (the FFS and flat consultation rates, p_x and p_R), the patient insurance status (the cost-sharing parameter, θ_j) and the level of competitive pressure on individual physicians (the business elasticity, $n_i'(NB)$).[108]

Physicians' intrinsic motivations (captured in β_i) were assumed to be uni-dimensional, exogenous and fixed. In this section, I relax that assumption. Specifically, I allow that some types of intrinsic motivation may be endogenous to the regulatory environment, and leveraged through intrinsic incentives for physician effort. First, I distinguish two sources of intrinsic motivation: the ability to complete pro-social or altruistic tasks ('altruism'), and the ability to knowingly perform well relative to a trusted benchmark ('professionalism').[109] I assume that the motivation physicians derive from the former source is fixed (an exogenous endowment). However, I allow the motivation derived from the latter source to vary, depending on the incentive environment. Physicians' intrinsic motivation can then be re-defined as an increasing function of two arguments: $\delta_i(\beta_i, E_i)$. $E_i = \bar{e} - e_i$ captures the current performance of physician i (e_i) relative to some trusted professional (or social) benchmark $\bar{e}$, and reflects the physician's professional motivation for effort (which is increasing (decreasing) in how far the physician is knowingly

[107] Note that this reinforces *Hypothesis 5*: both this higher observed effort (*Hypothesis 6*), and lower unobserved appropriate effort (*Hypothesis 4*), result in a higher quantity and cost of drug treatments for high-insured patients (*Hypothesis 5*).
[108] Insurers may increase competitive pressure on contracted physicians by making comparable care quality and cost information on competitors publicly available, for example.
[109] The latter source of intrinsic motivation aligns with the notion of 'reference-based utility' (Heffetz and Frank, 2008).

performing below (above) the trusted benchmark).[110] β_i remains the fixed, physician-specific endowment of intrinsic motivation (or altruism).

Substituting δ_i for β_i in equation (6),[111] the physician's optimal effort choice is now determined by the condition:

$$C_i'(e) = (\, n_i'(NB)p_R + \delta_i(\beta_i, E_i))(NB_e) \tag{12}$$

In this context, an *intrinsic incentive* for effort is defined as something that solely increases the marginal intrinsic utility from effort (δ_i).[112] If physicians can perfectly self-assess their own relative performance (i.e. perfectly observe E_i), the only way for insurers to increase intrinsic incentives for physician effort is by increasing $\bar{e}$ (the performance benchmark). This is not straightforward (and perhaps infeasible) where such benchmarks are based on evidence-based or established professional standards. Nevertheless, the assumption that physicians can perfectly observe E_i is quite weak, given a common finding that physicians often struggle to self-assess and benchmark their own performance (Davis *et al.*, 2006). This creates a role for the insurer in monitoring and feeding back new information on the physician's relative performance, to generate *intrinsic informational incentives* for effort. Reflecting the usual problems in measuring and contracting for care quality (Pauly, 1980; Ma and Mcguire, 1997), the insurer cannot perfectly monitor or evaluate appropriate physician effort either. The insurer's relative monitoring accuracy (compared to the physician's own self-assessment) depends on the strength of its monitoring technology, which can vary.[113]

Formally, the physician cannot observe E_i,[114] but instead observes $\widehat{E_i} \sim \mathrm{N}(E_i, \sigma)$. If $\widehat{E_i} < E_i$, the physician could benefit from better information on his current performance - more accurate performance information can function as an intrinsic incentive, and improve physician effort. The insurer has access to a performance monitoring technology t, which can be used to (imperfectly) audit and feedback on physician performance. Using t to audit physician i, the insurer observes $\widetilde{E_i} \sim \mathrm{N}(E_i, \alpha(t))$. For simplicity, assume that the insurer's monitoring capacity is always more accurate than the physician's ability to self-assess performance (i.e. $\alpha(t) < \sigma$), and the absolute

---

[110] Assuming that e_i is uniformly distributed within some closed, non-negative interval [0,$\check{e}$] and the benchmark $\bar{e}$ falls within this interval, E_i is then uniformly distributed within the interval [$\bar{e} - \check{e},\ \bar{e}$].

[111] Recall that $NB_e = U_e - \theta_j(p_R C_R'(e) + p_m'(e)m + p_m m'(e))$

[112] Conversely, an *extrinsic incentive* is anything that changes market demand responses ($n_i'(NB)$) or the physicians' payment p_R. This corresponds to reputational or direct financial (payment) incentives.

[113] In addition to quality monitoring with administrative data (including electronic health records and billing data), which tends to be infeasible in many LMIC settings where data quality is often poor, Das and Hammer (2014) provide an overview and comparison of alternative provider quality measurement tools.

[114] Note that both the physician and insurer are assumed to be perfectly aware of $\bar{e}$ (the benchmark standard), but struggle to observe E_i because of their inability to perfectly observe or self-assess individual performance e_i. If this condition does not hold, physicians can also be intrinsically incentivised with better information on benchmark standards (for example, professional best-practice guidelines)

level of the insurer's accuracy depends on the technology t. Following performance feedback from the insurer, the absolute change in the physician's marginal intrinsic utility from effort (and hence, his optimal effort choice) would be proportional to $|\widetilde{E_i} - \widehat{E_i}|$. The larger the deviation in the physician's self-assessment from the insurer's feedback, the greater the change in his effort.[115] As long as $\widetilde{E_i} > \widehat{E_i}$ (i.e. $\widetilde{e_i} < \widehat{e_i}$), the feedback should increase intrinsic incentives for effort.[116]

It is implicitly assumed that physicians have perfect information on the benchmark $\bar{e}$, and only struggle to assess their individual performance e_i relative to it. Where this benchmark is evidence-based clinical guidelines, or professional best-practice standards, keeping up-to-date with or recalling this information can be costly to physicians. These effort costs are captured in $C(e)$. In practice, if this information is not easily accessible or comprehensible, the costs to maintaining perfect knowledge of $\bar{e}$ may be prohibitive. In these circumstances, additional enabling interventions, that seek to improve knowledge of trusted benchmark standards, can both lower these effort costs and complement the informational incentive effects of performance feedback.[117]

Furthermore, to isolate the purely *intrinsic* effect of performance feedback, extrinsic motives (such as financial or reputational concerns) should remain unaltered. This would require that no financial incentives are attached to monitored performance, and that individual performance information is kept private (to avoid any reputational or market demand effects from public performance reporting (Kolstad, 2013)).

To summarise, an additional testable hypothesis from this discussion is as follows:

> *Hypothesis 7:* Where physicians have perfect information on shared best-practice standards, but struggle to self-assess their own relative performance and fall below those standards, private performance feedback can generate intrinsic incentives for increasing clinical effort, and improve audited aspects of care for all patients.[118]

---

[115] The implication is that physicians who are assessed as the worst relative performers should have the greatest intrinsic incentives for improvement.

[116] Where $\widetilde{E_i} < \widehat{E_i}$, the feedback could indeed act as an effort *dis*incentive. However, acknowledging that clinicians (as with individuals in general) tend to have fairly optimistic assessments of their own performance (Baumann *et al.*, 1991; Dunning *et al.*, 2004; Berner and Graber, 2008), objective feedback should serve as an effort incentive in most cases. Formally, this implies that $\widetilde{E_i} > \widehat{E_i}$ on average, which would require that $\widehat{E_i} \sim \mathrm{N}(\overline{\overline{E_i}}, \sigma)$, where $\overline{\overline{E_i}} < E_i$.

[117] An effort *enabler* is something that lowers the marginal cost to effort, including educational materials or clinical decision aides for physicians. Such interventions can function as both effort enablers (by lowering $C_i'(e)$, physicians' marginal cost to effort) *and* effort incentives (by improving physicians' knowledge of $\bar{e}$ if there is a knowledge gap, and altering their marginal intrinsic utility of effort δ_i).

[118] Although this discussion has focused on intrinsic incentives for physician effort, the same result could also apply to physicians' service quantity, if feedback was instead based on $\bar{x} - x_i$.

3.5 Discussion

This framework's results have a number of policy implications. In the presence of positive fee-for-service profit for physicians and an uncompensated cost to clinical effort, the provision of health insurance leads to poorer technical quality and more costly care than financing healthcare out-of-pocket. The physician anticipates a higher tendency for moral hazard from more insured patients, and exploits this to increase his own moral hazard behaviour: over-supplying health service quantity and under-supplying clinical effort (technical quality) relative to what the patient and payer would choose under perfect information. These two forms of physician moral hazard are regulated by the physician's competitive and intrinsic incentives, implying a role for policy levers in shaping these incentives.

In the model, the patient delegates healthcare decisions to the physician, who acts as an imperfect agent: the physician cares about his own profits and effort costs, in addition to patient welfare. Ensuring high-quality and cost-effective care in such a context would require leveraging these physician motives through an appropriate balance of intrinsic (altruistic, professional) and extrinsic (profit, reputational) incentives. First, developing physicians' intrinsic motivations (professionalism, and related altruism), perhaps through regular professional audit and feedback of individual performance, may be one lever. This requires that barriers to clinical effort are relatively low. For instance, where physicians' knowledge and recall of relevant best-practice standards are weak, complementary interventions - for example, interventions that improve the accessibility of clear treatment guidelines and other clinical decision aides – could be beneficial in improving physicians' performance capacity.[119] A similar intervention is evaluated in Chapter 7.

Second, increasing physician competition through competitive market regulation, and increased patient choice and information, can further incentivise appropriate care choices - provided that patients are well-informed and hold correct beliefs about the relative benefits of different care inputs. Relatedly, introducing a higher degree of cost-sharing in the physician payment structure – for example, by increasing fixed payment components (e.g. the flat consultation rate p_R) or including drug costs in physicians' flat consultations rates – may reduce both types of physician moral hazard (and any resulting over-provision of medications). While such interventions are not evaluated in this thesis, evidence from previous studies lend support to these predictions.

[119] As stated previously, where there is a knowledge gap on best-practice standards among physicians, these knowledge-based effort "enablers" can improve physicians' performance capacity, lower their costs to effort, *and* function as complementary intrinsic incentives for effort.

Physician cost-sharing, for instance, has been shown to improve the cost-effectiveness of prescription choices in a related study in South Africa (Lagarde and Blaauw, 2019a).

These interventions can be expected to lower physician moral hazard for the average patient. However, the model is unclear on the extent to which they can bridge the predicted differences in physicians' care choices for high- and low-insured patients. It can be seen from (4) and (5) that any increase in physicians' intrinsic motivation β_i or patient demand elasticity $n_j'(NB)$ should reduce unnecessary care costs and improve care quality for *all* patients, regardless of insurance type. However, whether higher baseline intrinsic motivation or competition (demand elasticity) would reduce or increase the second-degree moral hazard behaviour (i.e. reduce or increase the health service quantity and quality differential between high- and low-insured patients) is not obvious from the physician's optimality conditions (4) and (5). This would require further assumptions, for example on the shape of the patient's marginal utility functions.[120] Chapters 5 and 6 attempt to explore this issue empirically, by testing for potential interactions between measured baseline physician altruism and competition on the one hand, and the level of patient insurance cover on the other.[121]

A further model limitation is the strong assumption that physicians respond only to perfect information on patient net benefit and preferences (which are assumed to be perfectly known to physicians). It may be reasonable to allow that, given the high degree of uncertainty and imperfect information inherent in medical encounters, both physicians and patients must rely to some extent on their *beliefs* about the efficacy and value of different clinical inputs and outputs. The role of physician beliefs is not explicitly accounted for here. However, it is intuitive to think of these beliefs as embedded in how physicians *perceive* patient utility and demand elasticity with respect to specific care choices. A common example in this study's context (see Section 2.3) is physicians' reported beliefs that patients expect antibiotic treatment even for uncomplicated viral infections. Actual or anticipated patient demands are often cited as a reason for excessive prescribing of antibiotics, particularly in primary care (Currie *et al.*, 2014).[122] A role for such beliefs is considered in contextualising the results of the physician incentive experiments evaluated in Chapters 5-7.

[120] If the marginal utility functions are all linear, then an increase in altruism or competition would have no effect on the service quantity or quality difference between high- and low-insured patients.

[121] The subsample analyses in those chapters are nevertheless heavily limited by their small sample sizes.

[122] An important distinction exists between patients' *perceived* net benefit, which is subjective and ultimately what determines patient responses (i.e. patient demand elasticity) and drives the competition effect on physician choices, and patients' *realised* net benefit, which may be more visible to providers (given information asymmetries) and important in driving the physician altruism effect.

4 Study Description, Methods and Data

The empirical component of this thesis study includes two field experiments conducted with private primary care physicians (GPs) in Johannesburg between February 2018 and March 2019. This study is embedded in a larger study investigating antibiotic prescribing behaviour in primary care, funded by the Economic and Social Research Council (ESRC) and co-led by my academic supervisor, Dr. Mylene Lagarde (hereafter referred to as the "ESRC study"). As such, it has shared a number of ESRC study resources, including some of its funding, fieldworkers, clinical advisors, and local research and institutional contacts. Baseline data for the field experiment described in Section 4.3.2 is also drawn from the ESRC study results. The research methods and tools were developed jointly, and adapted to this study purposes where necessary. Certain fieldwork activities were also conducted in parallel, including the training of fieldworkers and recruitment of research participants. Further details on areas of overlap are outlined in the relevant sections below.

This chapter is structured as follows. Section 4.1 gives an overview of the standardised patient (SP) methodology employed in fieldwork. The SP method was used to create experimental variation in patient characteristics and for primary data collection in the two field experiments described in Section 4.3. Section 4.2 gives an overview of the main data sources, outcomes and variables of interest in this study. Finally, Section 4.3 details the design and implementation of the two field experiments and their corresponding SP visits.

4.1 Standardised Patient Methodology

Following Rethans *et al.* (1991) and Das *et al.* (2012, 2016), I employ an audit study approach with unannounced SP visits to primary healthcare providers to evaluate the appropriateness of care delivered during clinical interactions. SPs record rich details on several aspects of each interaction in a questionnaire immediately after the consultation (and retain all dispensed drugs, receipts and prescriptions), allowing an unparalleled insight into both the process and therapeutic outputs of physician-patient interactions. Moreover, the standardisation of patient symptoms, script and characteristics allows control of patient-level heterogeneity, to isolate the effects of variation in physicians' choices alone. To leverage this latter feature, I also follow Lu (2014), Currie *et al.* (2012, 2014) and Gottschalk *et al.* (2017) to employ the SP method within healthcare

field experiments: firstly, to create exogenous variation in the insurance status of the patient in Part 1 of this study (to identify the impact of patient insurance on physician behaviour); and secondly, to vary the severity of the clinical condition in Part 2 (and evaluate the effects of a randomised performance intervention).[123]

4.1.1 Clinical cases: description and relevance

Each SP in this study presented with one of two clinical cases, which differ only in their severity: 'uncomplicated' (viral) acute bronchitis in an otherwise healthy young adult (Clinical Case 1), and 'complicated' (bacterial) acute bronchitis in a HIV+ young adult (Clinical Case 2). In both cases, the SP presents to the physician with visible symptoms (a cough) and an opening statement that can indicate multiple underlying conditions, and that would require further questioning and examination to arrive at an accurate diagnosis and therapeutic decision. This allows assessment of the physician's clinical effort when presented with vague information and a number of possible diagnoses with overlapping symptoms. Brief summaries of the case histories and SP opening statements for the two clinical cases are presented in Table 4.1.[124]

In Clinical Case 1, the patient presents with a persistent cough, which started with a cold a week earlier.[125] 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 should rule out a number of ailments (including TB, HIV, asthma, sinusitis and allergies), whilst relevant examination should support these conclusions and eliminate other serious conditions like pneumonia (for instance, no crackling should be audible in a chest examination). The likely viral nature of the condition can be inferred from the lack of severity and short duration of reported symptoms. While the production of coloured sputum when coughing does not confirm a bacterial pathogen, that the patient reports only the production of a little white sputum (when asked) should further confirm the likely viral infection. As with all self-limiting viral infections, no drug treatment is necessary to cure the patient. Advice to drink plenty of fluids and seek follow-up medical attention only if symptoms persist beyond a few weeks or get worse should suffice. Some over-the-counter (OTC) symptomatic medication may be considered to relieve the patient's symptoms. However, the prescription of antibiotics is both unnecessary and potentially harmful in this case.

---

[123] See Section 4.3 for a description of the two study parts, corresponding to the two field experiments conducted in this study.

[124] In the first field experiment detailed in Section 4.3, the SPs portraying Clinical Case 1 are also subject to variations in their insurance status (low-insured or high-insured). As such, the opening statement for Clinical Case 1 varies slightly according the SP's insurance status in that experiment. These variations are discussed in Section 4.3.1.2, and summarised in Table 4.4.

[125] While this patient states that they had the "flu" (rather than a cold) in their opening statement (see Table 4.1), in South Africa doctors and patients often use "cold" and "flu" interchangeably in colloquial terms. Accordingly, both terms are taken to mean the same thing here: a common cold.

Clinical Case 2 builds on Clinical Case 1, with a HIV+ patient and more severe symptoms. In this case, the patient presents with a cough that has lasted two weeks, beginning with other cold symptoms that have now subsided. In addition, the patient reports that they were feeling feverish the day before, and is HIV+. If questioned appropriately, the patient would reveal that their cough has turned productive of yellow-green sputum over the last 3-4 days, and offer further details regarding their HIV history and on-going treatment. Further questioning and examination should rule out other possible conditions (including TB, asthma, sinusitis, allergies and pneumonia). The combination of the patient's HIV+ status (which indicates greater susceptibility to bacterial lower respiratory tract infections), the duration and worsening of symptoms (recent development of yellow-green sputum), and the reported feverishness should indicate that an underlying bacterial infection is likely. In this case, the doctor would be expected to prescribe a short course of antibiotics.[126] OTC symptomatic medications may also be considered here.

Table 4.1. Description of SP clinical cases and opening statements

	Clinical Case 1 Uncomplicated (viral) acute bronchitis in otherwise healthy young adult	**Clinical Case 2** Complicated (bacterial) acute bronchitis in HIV+ young adult
SP opening statement in consultations	"I had the flu last week but I am still coughing a lot."	"I had a cold. But this cough is not going away. And yesterday I was feeling a bit feverish. I am a bit worried because I am HIV positive."
Description of clinical case	• Young healthy adult (20-35 years) • Had a normal cold (runny nose, sore throat, cough) that started a week ago • Nose and throat symptoms have resolved, but the cough has persisted for 7 days (which is why they have come to the clinic) • Cough is only productive of small amounts of whitish / clear sputum • No fever, shortness of breath, chest pain, or other significant symptoms • No other diseases or family history, and not on any medications • No other signs to find upon examination	• Young adult (20-35 years), HIV+ • Had a normal cold (runny nose, sore throat, cough) that started two weeks ago • Nose and throat symptoms have resolved, but cough has persisted for two weeks and gotten worse • Cough has recently become productive of yellowish / green sputum • Reports feeling feverish the day before; but no shortness of breath, chest pain, or other significant symptoms • On ARV treatment, but no other medications • No other signs to find upon examination

In both cases, SPs were trained to give information beyond their opening statements only when probed by the doctor, to allow the doctor to demonstrate active clinical effort. Detailed question-

[126] The recommendation in South African standard treatment guidelines is a 5-day course of Amoxicillin (500 mg, three times a day). If the patient has a penicillin allergy, the recommendation is a 3-day course of a macrolide like Azithromycin (500mg daily) (NDoH, 2014).

and-answer (Q&A) scripts were developed for each case, to ensure SPs responded to clinical interrogation in a manner consistent with the underlying condition. Case-specific checklists of essential and recommended history-taking and examinations were also developed in consultation with clinical experts (as detailed in Section 4.2.1) to assess clinical effort, and SPs were trained to recall the checklist items completed by the physician during each visit and record them in a debriefing questionnaire immediately after (see Section 4.1.2).

The two clinical cases were chosen for their relevance to the South African context and their compatibility with the SP methodology. Lower respiratory tract infections (LRTIs) are a leading cause of morbidity and premature mortality in South Africa (estimated as the 4th biggest contributor to premature mortality in 2010) (Mayosi and Benatar, 2014), and the most common overlapping symptom across a range of LRTIs is a persistent cough. In a setting where life-threatening infectious diseases such as TB and pneumonia are highly prevalent, diagnostic uncertainty in differentiating less serious viral infections from these bacterial conditions has been identified as a key driver of excessive antibiotic use – 80% of which occurs in primary care, and mostly for acute respiratory infections (Brink *et al.*, 2016). South Africa also suffers the world's highest prevalence of HIV (21% of adults aged 15-49 are estimated to be HIV+) (Republic of South Africa, 2016), and the increased susceptibility of HIV+ patients to more serious respiratory infections is a further consideration in diagnostic decisions (Karim *et al.*, 2009; Jaffer *et al.*, 2017). Potential contra-indications with anti-retroviral therapy is also a factor in therapeutic decisions. These issues have led to calls for better 'diagnostic stewardship', and the national Department of Health and leading medical journals have published clear protocols for the diagnosis and treatment of acute respiratory illnesses in primary care (NDoH, 2014; Brink *et al.*, 2016). These guidelines facilitated the development of essential care checklists and appropriate treatment protocols for the cases in this study, to enable assessment of physicians' clinical effort and care quality performance.[127] The two cases further satisfy essential criteria for use with the SP method (see Section 4.1.4), including posing a low risk of invasive procedures to SPs and involving minimal observable symptoms (to minimise the risk of detection).

4.1.2 SP recruitment and training

SP training was conducted in three stages. The bulk of the training was carried out over seven days (56 hours) in June 2018, where Clinical Case 1 was covered. This was done jointly with the ESRC study, as Clinical Case 1 was used in both studies. A 2-day 'refresher' training was then conducted in mid-July 2018, to refresh Clinical Case 1 and introduce the patient roles relevant to

[127] See Section 4.2.1 for details on how the checklists were developed and the treatments classified, and Appendix B.1 for a full list of checklist items.

the first field experiment in this study (see Section 4.3.1). Finally, Clinical Case 2 was covered over three days in November 2018.[128] For brevity, only the first stage of SP training is described in the following (where most of the training for this study was carried out).[129]

All fieldworkers were black African, educated to university-level, between the ages 20-40 years, and resident in Johannesburg or the surrounding areas of Gauteng Province. Candidate profiles were selected to represent an average young adult patient in Johannesburg, whilst an advanced level of education was considered necessary for certain demands of the SP role (including the need to identify and recall detailed features of the consultation). It was further required that candidates had no prior medical or nursing training that could influence their responses to or behaviour with GPs. Candidates were selected for training following an interview and a basic medical check-up to ensure they had no underlying medical conditions that could bias GPs' clinical decisions in the study. It was important to ensure that all SPs were healthy, and remained so throughout the fieldwork, to maintain the face validity of the SP method. Interviews also assessed candidates on their general communication and presentation skills, and potential biases towards private GPs and health insurance.

42 candidates were interviewed in total, and 23 were eventually selected to undergo the first training program in June 2018. The training was carried out by me and three members of the ESRC study team. It began by introducing the clinical case, the relevant SP opening statement, and related question-and-answer (Q&A) scripts designed to enable SPs to respond to clinical questioning in a manner consistent with the case (see Appendix B.2 for the case-specific Q&A scripts). Certain aspects of the opening statement and Q&A scripts were refined during training, following feedback from fieldworkers on local idioms and so forth.[130] As SPs would be visiting GPs as new patients, the fieldworkers also had to come up with plausible reasons for why they were not visiting their regular doctor.[131]

Important characteristics of SPs were emphasised and rehearsed throughout the training, including the standardisation and consistent presentation of case symptoms, the offering of case history details only when probed, and the accurate observation and recall of all necessary aspects of the consultation (as required in the SP debriefing questionnaire – see Appendix B.4). The fieldworkers initially focused on memorising the SP scripts, and then proceeded to practice through role-play and mock consultations with the research team. They were also trained on how

[128] The training was staggered in this way to align with the fieldwork schedules of the two experiments (see Figures 4.1 and 4.2 for a timeline of fieldwork activities corresponding to each experiment, including SP training).
[129] The 2-day refresher training is discussed briefly in the context of the corresponding experiment in Section 4.3.1. The final stage of training simply introduced Clinical Case 2 in a similar format to Clinical Case 1 in the first stage of training.
[130] The fieldworkers also participated in defining the social background of the patient, including their occupation in case the doctor questioned on this.
[131] Most fieldworkers cited visiting family members in the neighbourhood, or starting a new job in the area.

to improvise the script in responding to certain unexpected questions; and communicate all key pieces of information in a standardised, unaltered format in English, even if the doctor made conversation in a local tribal language.

The tablet-based SP debriefing questionnaire was then introduced, which served as the primary data collection tool in SP visits. Appendix B.4 provides an overview of the questionnaire structure. Fieldworkers were familiarised with its contents and protocols,[132] and questionnaire recording accuracy and submission (via an online app) was practised extensively with further mock consultations.[133] To ensure that fieldworkers could correctly identify the different types of clinical advice and examinations that were covered in the questionnaire, mock examinations were demonstrated and practice consultations were arranged with trained doctors. These doctors also gave valuable feedback to the fieldworkers on portraying a credible patient.

Fieldworkers were further briefed on how to handle potentially dangerous or unforeseen situations during consultations, including doctors insisting on administering invasive procedures like blood tests, receptionists insisting on SPs seeing alternative doctors to the ones they were scheduled to see, or medical emergencies presenting at the clinic. Again, clear protocols for handling these situations were practiced through role-play. At the end of the program, fieldworkers were sent on practice (unannounced) visits to consenting doctors that were not part of the study sample, to build their confidence in the field.

Fieldworkers were continually assessed throughout the training, through informal observation by the research team, feedback from external training facilitators (the trained doctors), and short paper-based quizzes. The research team collated evaluations of individual fieldworkers on a daily basis. At program finish, 10 of these 23 fieldworkers were selected for the second stage of training in July 2018 (and participated in the first field experiment of this study – see Section 4.3.1). 10 of these fieldworkers were also retained for the final training in November 2018 (and participated in the second field experiment – see Section 4.3.2).

4.1.3 Ethical considerations

The unannounced SP method in field research fundamentally relies on participant blinding in gathering data. Method validity relies on doctors believing SPs to be real patients throughout the data collection process (the consultation). This is a necessary condition for claiming the SP method to be any superior to other methods of collecting similar data, such as clinical vignettes

[132] Fieldworkers were advised to find a quiet, private space within or in the vicinity of the GP clinic immediately after each consultation, to complete this questionnaire.

[133] Data was automatically uploaded onto an online server after each questionnaire submission, which facilitated daily monitoring and data quality checks.

or direct observations. Blinding avoids the risk of a Hawthorne effect, whereby doctors may alter their normal behaviour when they are aware of being observed (Beullens *et al.*, 1997; Leonard and Masatu, 2010). This reliance inevitably raises certain ethical concerns, both from the perspective of the research participants (the GPs) and the researchers (SPs), which must be addressed.

First, informed verbal and written consent was obtained from all participating GPs prior to the SP visits in each experiment. Measures were taken to ensure that all participating GPs understood the SP method and its use of blinding prior to consenting to the study.[134] Consent to participation in the first experiment ('Part 1') was obtained during the GP recruitment process (see Section 4.3.1). All recruiters were trained in explaining the SP method to GPs, and in answering any questions or concerns raised. This initial consent was then re-affirmed for participation in the second experiment ('Part 2') for the relevant sample of GPs (see Section 4.3.2). At both stages, a comprehensive study information sheet was shared with GPs to inform them of the study aims and different components.[135] The contact details of a research team member and the ethics committee at Wits University (the local research institution that granted approval for this study) were also shared with all GPs at the time of recruitment, in case of further questions or concerns about the research.[136] While GPs were not told exactly when they would be receiving SPs, they were told that they would receive two to four SP visits over a fixed time period of a few months. After the completion of SP visits in Part 1, a follow-up phone call was conducted with all participating GPs to ensure that no SPs had been detected (see Section 4.3.1.3).

Potential risks to GPs, SPs and the actual patients of GPs were carefully considered, and safeguards were incorporated in the research design where appropriate. Risks to participant GPs were considered minimal, as the SP method involves no procedures or interventions that would require informed consent outside of a research setting. The method is designed to evaluate participants carrying out normal, routine tasks in their usual environment, and therefore entails minimal physical, psychological or behavioural intrusion. GPs also received their usual consultation fees for the SP visits, so they did not incur any financial loss from participation. Moreover, participant confidentiality is strictly protected through anonymisation of all identifiable data, and publication of study results only at an aggregate level. Confidential

[134] While SPs have been used in field research in South Africa before, their use remains relatively unknown to physicians, particularly in the private sector. Previous studies were all conducted in the public sector, and mostly without informed consent from individual providers (Christian *et al*., 2018; Harrison *et al*., 2000; Kohler *et al*., 2017; Mathews *et al.*, 2009).

[135] See Appendix B.7 for an example of the study information sheet.

[136] GPs were further informed about their ability to opt out of the study at any stage. A link to opt out was sent via text message to each GP immediately after their consent to participation.

information on individual performance was fed back privately to some GPs as part of the study (see Section 4.3.2), but this was also subject to their informed consent.

Risks to fieldworkers were minimised by following standard procedures for protecting SPs during their visits (King *et al.*, 2019). The clinical cases used in this study are unlikely to result in any invasive investigations. Nonetheless, SPs were trained to refuse all invasive or risky procedures (for example, blood tests or injections). SPs were also taught exit strategies if they needed to terminate the clinical encounter. If that proved impossible, SPs were told to reveal themselves as researchers to avoid any risks.

Finally, risks to the actual patients of participant GPs – for instance, from longer wait times for medical attention due to SPs booking consultation slots - were considered. However, it was not expected that SPs would add substantially to patient wait times, as a typical SP consultation would not last more than 10-15 minutes. In addition (although private GPs rarely receive emergency cases), in exceptional cases where there was a medical emergency in the clinic, the SPs were briefed to immediately step aside and allow consulting GPs to address the emergency first.

Ethical approval for this study (and the use of SP methods) was obtained from the ethics committees of the London School of Economics (LSE) and the University of Witwatersrand (Wits).[137]

4.1.4 Methodological considerations and limitations

In addition to participant blinding, the validity of the SP method requires that SP presentations are standardised and uniform in all aspects except the experimentally varied feature (in this study, the insurance status of the SP or severity of the clinical case). The use of different fieldworkers poses a challenge to this. As mentioned earlier, the training of SPs plays a crucial part in standardising fieldworker presentations. In addition, as described in the relevant experiment designs in Sections 4.3.1 and 4.3.2, individual fieldworkers or fieldworker-pairs (matched on gender, age and physical appearance) were also randomised to individual GP visits in each experiment, to control for any non-random selection. Fieldworker or fieldworker-pair fixed effects were further included in all regression analyses presented in the empirical Chapters 5-7.

As suggested previously, there are certain limitations to the use of SPs in field research. It restricts the types of clinical cases that can be employed to those that have minimal visible symptoms and do not require any invasive examinations for diagnosis. This limits the generalizability of study results. The use of only two (slightly varying) clinical cases in this study also limits the types of

[137] Institutional approval letters are shown in Appendix B.8.

clinical effort (history-taking and examination) and treatment recommendations that physicians can be assessed on. Where resources permit, the use of a broader range of clinical cases can allow a more holistic assessment. Moreover, SPs can only present to doctors as new, acute patients and have one-off interactions (the risk of detection would be very high with repeat interactions). This prevents participating doctors from demonstrating care continuity, which is an important dimension of care quality in primary care (particularly for chronic patients) and which they may provide to patients with repeated interactions.

4.2 Data and Outcomes of Interest

Before describing the two field experiments in this study, it is useful to first outline the data sources and outcomes of interest from those experiments (which are frequently referenced in the experiment descriptions in Section 4.3). Table 4.3 provides an overview of the primary and secondary data sources in this study, as well as the final outcomes and variables of interest drawn from those sources. Briefly, this study draws on two primary sources of data: *i)* the SP visits (as detailed in Section 4.3) and *ii)* the face-to-face interviews ('GP interviews') conducted with all participant GPs in the two field experiments. The timing of interviews and SP visits in each experiment is shown in Figures 4.1 and 4.2. The GP interviews were designed to collect data on participants' personal and clinical practice characteristics. The interviews also included a dictator game experiment (to elicit a measure of GP altruism) and a short knowledge quiz with clinical vignettes (to measure GPs' diagnostic and therapeutic knowledge). Further details on the interview structure and content are provided in Appendix B.4, and relevant components are discussed alongside variables of interest in Section 4.2.3. The four secondary sources of data (listed in Table 4.3) are also discussed alongside related variables in the following sections.

4.2.1 Care quality outcomes

Similar to Das *et al.*, (2016), I use four measures of care quality comprising both a physician's (observable) care inputs and subsequent output (diagnosis and treatment choices). These measures are constructed for each SP consultation completed in this study, using data collected in the SP debriefing questionnaire and any dispensed drugs or prescriptions. The main portion of the debriefing questionnaire covered the consultation time, the case-specific checklists of history-taking and examinations, any diagnostic tests or investigations ordered, the diagnosis and medical advice offered (if any), and the general quality of communication with the patient.[138] Using this

[138] These sections of the questionnaire were validated by physicians on the ESRC study's clinical advisory panel and lecturers at the Wits University medical school.

data, I first construct proxy measures of GPs' clinical effort during a consultation. These measures gauge how hard a GP works in questioning and examining a patient in order to arrive at an accurate diagnosis and therapeutic recommendation. Second, I construct indicators on whether the GP gave a diagnosis and whether it was correct. Third, I examine the treatment offered: the appropriateness of any drugs prescribed or dispensed for the particular clinical case. Last, I examine the advice given by GPs on if and when to seek further medical attention with the same provider. Descriptive statistics on these outcomes are presented in Chapters 5 and 7 for the relevant samples in each experiment.

4.2.1.1 Clinical effort

Three measures of clinical effort are constructed. At a basic level, the consultation length is taken as one proxy for provider effort: the more time a GP spends with the patient, the more opportunities he allows himself to gather all essential case information. Not all time spent in a consultation is likely to be focused on the patient (or necessary), however. The SPs report whether providers spent any consultation time doing other things like answering phone calls and so on, and I deduct any such reported time from the total consultation length in each case.

Consultation length also provides no information on how well the consultation time was used. Therefore, a second proxy measure captures the proportion of case-specific history-taking questions and examinations completed by the GP. Checklists of both *essential* and *recommended* history-taking and examinations for respiratory conditions, that would allow a GP to differentiate serious conditions like TB and pneumonia from less serious ones like mild bronchitis, were developed together with a panel of clinical experts in respiratory and infectious diseases in South Africa. Appendix B.1 lists these checklist items, and indicates those considered *essential* by the panel. SPs recorded which checklist items were completed by GPs in each consultation, and I use the proportion of all checklist items completed in each case as the second measure of provider effort.[139]

Although the computational ease and transparency of this raw proportion measure is appealing, not all checklist items included are likely to be equally valuable in discriminating between high-effort and low-effort GPs. Some items may require more cognitive effort than others, and some may be more habitual. For instance, asking if the patient is coughing up blood would be much more effective in quickly eliminating a serious TB diagnosis than asking about the patient's TB history. To account for such differences in the discriminating value and difficulty of checklist items, I construct a third measure of clinical effort: a weighted, composite index score of

[139] I further disaggregate this measure into the shares of *essential* history-taking and examinations completed, and analyse these measures separately in Chapter 5.

completed history-taking and examination items using item response theory (IRT). First applied in the context of psychometrics and educational testing, IRT is a model-based approach to developing and scoring tests (or checklists) for estimating latent individual traits (such as ability, or in this case, effort). Using maximum likelihood methods, the IRT score assigns greater weight to items that are more difficult (less likely to be completed) and that discriminate better among GPs in terms of their effort quality (Das and Hammer, 2005).

See Appendix B.6 for a discussion on the key assumptions of IRT, as applicable to this study. Appendices C.2 and E.1 provide a full list of included and omitted checklist items in the IRT analyses of Chapters 5 and 7, respectively. Items that were successfully completed in less than 5% of sample consultations were dropped from the IRT analyses in both cases, as the maximum likelihood estimation procedure is not guaranteed to converge with items that have very low completion rates (Das and Hammer, 2005).

4.2.1.2 Correct diagnosis

After each consultation, the SPs recorded whether a diagnosis was given by the GP and what the diagnosis was. The specific diagnoses were then coded as 'correct', 'partially correct' or 'incorrect' by members of the ESRC study research team and myself, in consultation with two respiratory and infectious diseases experts from the study's clinical advisory panel. A strictly 'correct' diagnosis required the doctor to communicate the diagnosis in exact technical terms to the patient (i.e. "acute bronchitis" or "bronchitis"). The limitation here is that what the physician tells the patient regarding their diagnosis may be simplified (for instance, given expected limitations to the patient's medical knowledge). Hence, a 'partially-correct' diagnosis allowed for some generality in the communicated diagnosis (such as "chest cold") and for select similar diagnoses (such as "tracheitis" or "post nasal drip") where the researchers could agree that the doctor was referring to a correct diagnosis in more colloquial terms or had given a clinically-related or very similar diagnosis, and where the correct pathogenic cause (viral or bacterial, depending on the clinical case) could be established with reasonable confidence.

The SPs' data collection tool also allowed them to record multiple diagnoses, if more than one was given. In these cases, where a correct or partially-correct diagnosis was recorded alongside an incorrect diagnosis, the combined diagnosis was coded as 'incorrect'. The only exception was where a correct or partially-correct diagnosis was combined with a "cold" or "flu" diagnosis. In strict medical terms, the latter constitute incorrect diagnoses. However, acute bronchitis is usually preceded by flu-like symptoms, and SPs presenting with both Clinical Cases 1 and 2 in this study

communicate to doctors that their symptoms began with a "cold".[140] Therefore, a "cold" or "flu" diagnosis was coded as 'correct' (or 'partially-correct') if it was combined with another correct (or partially-correct) diagnosis.[141]

Appendix B.1 lists the diagnoses considered 'correct' or 'partially correct' for both clinical cases presented in this study, as well as a range of 'incorrect' diagnoses that were recorded. Based on this classification, I construct an indicator of a 'correct' or 'partially correct' diagnosis for each consultation where a diagnosis was given.

4.2.1.3 Appropriate treatment

The therapeutic outcome of the consultation is the third measure of care quality examined. All drugs and prescriptions given to SPs were collected after each consultation. Each drug item dispensed or listed in a prescription was recorded, using the name specified in the prescription or on the drug packaging. In prescriptions where GPs specified only the main active pharmaceutical ingredient (API) of a drug rather than a drug name, or where they specified a branded drug but explicitly wrote that generic substitutions were to be used, the name of the cheapest generic-equivalent drug available on the local market was recorded.[142] The recorded drug items were then matched to a corresponding item in a national medicines database managed by the South African Medicine Price Registry.[143] This database contains the Anatomical Therapeutic Chemical (ATC) code for each listed drug - a drug classification system developed by the WHO. After individual drug items were matched to corresponding ATC codes, the resulting ATC groupings (drug categories) were classified as 'appropriate' or 'inappropriate' for each of the two clinical cases. This classification was done in collaboration with the ESRC study team and clinical advisory panel, and informed by evidence from recent systematic reviews (Becker et al. 2015; Smith et al. 2014; Smith et al. 2017; Johnstone et al. 2013) and guidelines stipulated in the South African Standard Treatment Guidelines and Essential Drugs List (STG/EDL) for primary care (NDoH, 2014; Perumal-Pillay and Suleman, 2017).[144] Table 4.2 shows the classification of drug groupings for the two clinical cases considered in this study. Based on the ATC codes, indicators were also

---

[140] Discussions with local medical practitioners highlighted that, in the South African context, doctors often do not make a distinction between "flu" and "cold" in communicating diagnoses to patients. Accordingly, no distinction was made in coding a "flu" and "cold".

[141] The only exception to this rule was where a "Cold" or "Flu" was combined with just "a viral infection". In this case, the diagnosis was considered 'incorrect' as this combination could also indicate just an upper respiratory tract infection, which is an incorrect diagnosis.

[142] The cheapest generic-equivalent drug was found by searching the South African Medicine Price Registry (http://www.mpr.gov.za) for all drugs with the same main API as the one specified in the prescription, and comparing the Single Exit Prices (SEP) per unit of these drugs.

[143] The database was downloaded from the South African Medicine Price Registry (http://www.mpr.gov.za) on 13 June 2018.

[144] The STG/EDL is a set of therapeutic guidelines and essential medicines that satisfy the priority health needs of the population, and determine the drugs available for use in the South African public healthcare sector. The EDL was developed according to WHO guidelines to curb rising pharmaceutical costs and irrational medicine use, and serves as a benchmark for evidence-based, cost-effective therapeutic guidelines in the country (Perumal-Pillay and Suleman, 2017) .

constructed to specify whether a drug item was an antibiotic, a steroid, or other drug types of interest.

Table 4.2. Appropriate drug treatments, by clinical case

	Clinical Case 1 Uncomplicated (viral) acute bronchitis in otherwise healthy young adult	**Clinical Case 2** Complicated (bacterial) acute bronchitis in HIV+ young adult
Appropriate	Cough suppressants/expectorants; Analgesics; Throat preparations (e.g. lozenges)	Antibiotics (short course of Amoxicillin); Cough suppressants/expectorants; Analgesics; Throat preparations (e.g. lozenges)
Inappropriate / harmful	Antibiotics; Steroids; Nasal decongestants; Antihistamines; Bronchodilators; Probiotics; Vitamins; Other	Steroids; Nasal decongestants; Antihistamines; Bronchodilators; Probiotics; Vitamins; Other

Drugs were considered 'appropriate' if they are clinically-indicated for treating or mitigating the underlying condition (i.e. *recommended*), or are *palliative* low-schedule (OTC) symptomatic relievers with no contra-indications. Drugs classified as 'inappropriate' are neither clinically-indicated nor low-schedule symptomatic relievers: *unnecessary* drugs are those with insufficient evidence to be considered either clinically effective or palliative and which are not included in the STG/EDL guidelines, while *harmful* drugs are those with sufficient evidence and expert consensus to be considered harmful or contra-indicated for the particular clinical case. In some cases, it is difficult to clearly distinguish between or build clinical consensus on which drugs are harmful and which are unnecessary. For instance, antibiotics for the uncomplicated acute bronchitis case (Clinical Case 1) may be simply considered unnecessary for the patient, given the underlying viral infection. On the other hand, it is ultimately harmful in the context of growing antibiotic resistance, and there is emerging evidence on the contribution of antibiotics to the incidence of adverse drug events (Linder, 2008). For this reason, my analysis focuses on the distinction between 'appropriate' and 'inappropriate' drugs, without stressing too much the sub-distinctions within these groups.

A key difference in appropriate treatment between the two clinical cases is in the use of antibiotics. While antibiotics are considered inappropriate in the uncomplicated acute bronchitis case, they may be justified for the HIV+ patient with more severe symptoms (longer duration, coloured sputum, feverishness) where a bacterial infection is more likely. Indeed, the 2014 edition of the South African STG/EDL guidelines state that antibiotics may be considered for suspected acute bronchitis in HIV+ patients. These guidelines were recently revised at the end of 2018, and the recent edition of the STG/EDL guidelines published in January 2019 does not contain this

reference anymore. Nevertheless, given the timing of this study (mid 2018 – early 2019), and the likelihood that recent guideline revisions will take some time to be disseminated and adopted into routine clinical practice, the 2014 edition of the STG/EDL guidelines is maintained as a benchmark for evaluating appropriate treatment in this study.

4.2.1.4 Follow-up advice

In addition to classifying the drugs dispensed or prescribed, I also examine whether and under what circumstances the GP advised that the SP return for a repeat consultation (a 'follow-up'). The SPs recorded whether such advice was given, and whether the recommendation was to return in the event of the cough getting worse (or if haemoptysis or shortness of breath develops), or for other / unexplained reasons. Acute bronchitis is most often self-limiting, and should not require any repeat consultation except in the event that it worsens and turns bacterial. The recommendation to return in the event of haemoptysis or shortness of breath is also reasonable: the GP may want to be cautious about a possible missed diagnosis of something more serious (pneumonia or TB). These specific cases were therefore considered 'appropriate' follow-up recommendations, whereas recommendations to return for other or unexplained reasons were considered unnecessary ('inappropriate'). Accordingly, I construct indicators to specify whether any follow-up recommendation was given during the consultation, and whether this recommendation was 'appropriate'.

4.2.2 Care quantity and cost outcomes

Two groups of care quantity and cost measures are also constructed for each SP consultation: the quantity and costs of any dispensed drugs or prescriptions, and the quantity and costs of all fee-for-service (FFS) items provided by GPs.[145] The drug quantity and cost outcomes are calculated by matching data recorded from prescriptions and dispensed drug labels to data available in a national medicines pricing database (as detailed below). The FFS costs are constructed from consultation receipts (where SPs paid in cash), insurance claims (where SPs had insurance cover), and fees recorded in the SP debriefing questionnaire. Summary statistics on these outcomes are presented for the consultation samples in Chapters 6 and 7.

[145] The FFS items include the consultation itself (GPs are paid a unit fee per consultation) and any diagnostic tests or procedures ordered during the consultation that entail an additional fee above the unit consultation fee. Note that drugs are not included in FFS items, as they are either prescribed (and therefore entail no fee to the GP) or dispensed as part of the consultation fee (all dispensing GPs in the sample included drugs in their unit consultation fee, and did not charge extra).

Table 4.3. Summary of study data sources and variables of interest

	Primary sources								Secondary sources			
	SP visits					GP interviews						
	SP Debriefing Questionnaire	Prescriptions	Dispensed drugs	Insurance claims	Consultation receipts	Dictator game	Knowledge quiz	Other questions	Medpages database	SA Medicine Price Registry	WHO ATC codes	2014 South African STG / EDL
Outcomes												
Care Quality												
Consultation length	✓											
Checklist completion / IRT score	✓											
Correct diagnosis	✓											
Drug type (antibiotic, etc.) / Appropriateness of drugs		✓	✓								✓	✓
Follow-up advice	✓											
Care Quantity & Costs												
Quantity & type of diagnostic tests/procedures	✓			✓								
Quantity / dose of drugs		✓	✓									
Branded drug		✓	✓							✓		
Cost of prescription / drugs		✓	✓							✓		
Cost of consultation (excl. drugs)	✓			✓	✓							
GP characteristics												
Age, gender & ethnicity								✓	✓			
Contracting & dispensing status								✓	✓			
Altruism						✓						
Diagnostic & therapeutic knowledge							✓					
AMR knowledge							✓					
Antibiotic beliefs								✓				
Competitor density									✓			
Northern suburb practice									✓			
Group practice								✓	✓			
Patient load								✓				

Notes. The Medpages database lists approximately 80% of all private medical practitioners in South Africa (https://www.medpages.info). The database provides practice details and a few personal characteristics for all listed GPs. The South African (SA) Medicine Price Registry (https://www.mpr.gov.za) contains the regulated Single Exit Prices (SEP) that can be charged for all drug items sold in South Africa. Dispensing doctors and pharmacists can charge patients a small dispensing fee above the SEP; however, these dispensing fees are also regulated at a certain share of the corresponding drug SEP (and are therefore proportional to the SEP). The Anatomical Therapeutic Chemical (ATC) codes are a drug classification system developed by the World Health Organisation (WHO), which allows the grouping of drugs according to the anatomy they act on and their therapeutic, pharmacologic and chemical properties (https://www.whocc.no/atc/). The South African Standard Treatment Guidelines and Essential Drugs List (SA STG/EDL) for primary care are developed and updated by the South African Department of Health, to provide evidence-based treatment guidelines for specific clinical cases (http://www.kznhealth.gov.za/pharmacy/edlphc2014a.pdf). Although they were developed for the public healthcare sector, they provide the most comprehensive set of national clinical guidelines available in South Africa (and therefore serve as a useful benchmark for evidence-based care also in the South African private sector). The 2014 version of the STG/EDL was the most current at the time of study. Note that these guidelines were sometimes complemented with findings from recent Cochrane systematic reviews (where available) in assessing appropriate treatments in this study. For further details on the sources or outcomes listed above, refer to Sections 4.2.1 - 4.2.3.

4.2.2.1 Drug treatments

The name and quantity of every drug item prescribed or dispensed in each SP consultation is recorded. The cost of each drug item is then obtained from the South African Medicine Price Registry (MPR) database, by matching each recorded item and its quantity to a unique identifier (a *nappi* code) in the database.[146] The database contains the regulated Single Exit Prices (SEP) for defined quantities (packs) and for a single unit of each listed item. The total SEP cost of each item is then calculated by multiplying the recorded quantity by the corresponding unit SEP of the drug. Not all drug items recorded in the study are listed on this database. For instance, certain OTC medications and herbal remedies are not listed, so it is not possible to obtain SEP rates and *nappi* codes from the database in these cases. In such cases, the relevant *nappi* codes were sourced from the website of *Medikredit* (the firm that licences this national drug coding system), and retail OTC prices for the recorded quantities were obtained from the websites of *Clicks* and *Dischem* (the two biggest pharmacy chains in South Africa).

In addition to the SEP cost of each prescription item, the 'dispensed' cost is also calculated. This represents the expected retail price of the item (that is, the price paid by the patient or insurer). To calculate this, a dispensing fee is added to the total SEP cost of each prescribed or dispensed item. The relevant dispensing fee is either the doctor's fee (for dispensed drugs) or the pharmacist's fee (for prescription items). These fees are set and regulated by the National Department of Health, and depend on the SEP of the drug: they usually comprise a fixed fee (increasing in the SEP) and a share of the SEP. Finally, a VAT at 15% of the dispensing fee is added to the total drug cost.

The total drug cost is calculated in this way for each individual drug item and aggregated at the consultation-level. In addition to the total drug cost from each consultation, I am also interested in how this cost compares to the cheapest recommended treatment available on the market for the relevant clinical case, as this would give an indication of the cost-effectiveness of the observed treatment. The cheapest recommended treatment for each clinical case was determined by referring to recommendations in the 2014 STG/EDL, and pricing the cheapest options corresponding to those recommendations in the MPR database. For Clinical Case 1, this treatment option was paracetamol and a cough suppressant, which amounted to a total cost of R17.86 at June 2018 prices. For the more complicated Clinical Case 2, the equivalent option was

[146] Drug pricing is highly regulated in South Africa. The Department of Health sets a Single Exit Price (SEP) for all drugs sold in the country, and the SEP determines the cost of drugs to dispensing doctors and pharmacists. It is the regulated price at which drug manufacturers are obligated to sell to dispensing doctors and pharmacists. The SEP does not depend on volume, which maintains a high level of transparency in drug pricing. While retailers can charge patients a dispensing fee above the SEP, these fees are also regulated at levels proportional to the corresponding SEP.

paracetamol, a cough suppressant, and a short (five day) course of Amoxicillin, at a total cost of R32.63.

The MPR database further specifies whether each listed item is a branded drug or a generic. I therefore construct another indicator for each dispensed or prescribed item to specify whether it is a branded drug or not.

4.2.2.2 Fee-for-service procedures and consultation fees

The total cost of each consultation (excluding the cost of drugs) was obtained by collating the receipts retained by SPs after each visit and the itemised claims submitted to the partnering insurer in Part 1 of this study (see Section 4.3.1). These costs included any fees that are paid to the consulting GP: the flat consultation rate, additional fees for any services administered during the consultation (such as in-room nebulisation), and the costs of any diagnostic tests that were ordered (which would require follow-up consultations).

4.2.3 Provider characteristics

In addition to basic socio-demographic characteristics, the following variables were constructed for each participating physician using data from the GP interviews. Descriptive statistics for these variables are presented for the respective GP samples in Chapters 5 and 7.

4.2.3.1 Altruism

The dictator game carried out during the GP interviews yields the measure of individual altruism used in this study. The game is a common economic experiment used to elicit pro-social preferences such as altruism using real monetary incentives (Eckel and Grossman, 1996). The use of monetary incentives and dictator anonymity, whereby GPs must anonymously decide how to split their monetary endowment between themselves and a patient charity of their choice, addresses the potential for 'self-presentation' bias in the use of surveys to measure such preferences: when individuals self-report their social preferences based on hypothetical payoffs rather than real monetary stakes, they may be more prone to over-report their degree of positive and socially-valued characteristics such as altruism (Smith, 1976).

The consensus in the experimental economic literature is that the share of the monetary payoff given up by the dictator for the recipient can be interpreted as a measure of altruism. Accordingly, I compute and employ the proportion of the R300 cash endowment donated to a patient charity by individual GPs in the dictator game as the measure of physician altruism in this study. I examine the distribution of this altruism measure across the sample populations in Chapters 5 and

7, and construct an indicator to categorise 'high altruism' GPs as those that donate more than the sample median share to a patient charity.

4.2.3.2 Provider knowledge

GPs' diagnostic knowledge and awareness of standard treatment guidelines are recognised foundations of clinical competence and performance (Miller, 1990). Following standard knowledge-evaluation methods, I rely on closed, multiple-choice questions from the GP interviews to measure GPs' knowledge of the clinical presentation of different respiratory illnesses, their likely causes, and respective treatment guidelines in South Africa. As detailed in Appendix B.4, the knowledge quiz portion of the interview asks GPs to identify the diagnosis and pathogenic cause of four patient cases: *i)* uncomplicated (viral) acute bronchitis, *ii)* complicated (bacterial) acute bronchitis, *iii)* bacterial sinusitis, and *iv)* bacterial pneumonia. The quiz also queries the recommended drug treatment in national guidelines for three patient cases: *i)* common cold, *ii)* bacterial pneumonia, and *iii)* uncomplicated (viral) acute bronchitis.[147] Answers to these 11 questions are coded as 'correct' or 'incorrect', and one point is awarded for each correct response. For the three treatment guidelines questions, only the choice of whether or not antibiotics are recommended for the patient case is used to determine a correct response.

The first measure of provider knowledge is therefore the total score over all 11 diagnostic and therapeutic questions. While this gives an aggregate indication of differential diagnostic and therapeutic knowledge, I am particularly interested in GPs' knowledge of the two acute bronchitis cases presented by SPs in this study; and also in analysing diagnosis, pathogenic cause and treatment knowledge separately. I therefore construct additional knowledge measures for these two individual cases, with separate indicators for knowledge of correct diagnosis, pathogenic cause and recommended treatment (note that a treatment guidelines question was not included for the complicated (bacterial) acute bronchitis case, so this case does not have a treatment knowledge indicator). This disaggregation is important because, for instance, a GP could successfully diagnose the uncomplicated case as 'acute bronchitis' but fail to identify the viral nature of the underlying condition. This would then alter the treatment he believes is appropriate. Analysing these knowledge components separately can thus allow better explanation of observed care outcomes.

Another measure of provider knowledge that may be relevant to antibiotic treatment choices is GPs' awareness of antimicrobial resistance (AMR) and its causes. This awareness can influence GPs' perceived risks of unnecessary antibiotic prescribing, and thereby mediate how GPs' beliefs

[147] Appendices B.5 (a) and B.5 (b) for a description of these patient case vignettes.

or perceptions about antibiotic prescribing (discussed in the next section) translate into prescribing practices. A set of five questions in the GP interviews accordingly tests for knowledge and awareness of AMR. Answers to these questions are scored as '1' if correct, and '0' otherwise. Aggregate scores over the five questions are calculated for each GP, and a binary indicator is coded to categorise scores above the sample median as indicating relatively 'high knowledge' of antibiotic prescribing risks.

4.2.3.3 Beliefs on antibiotic prescribing

In determining appropriate therapy, regardless of whether the GP holds correct knowledge of what is recommended in national guidelines, he may hold different beliefs about whether what's recommended should be adopted in a particular patient case. Three types of beliefs regarding antibiotic prescribing are examined in the GP interviews: antibiotic efficacy, prescribing norms, and perceived patient expectations. For each of the four patient cases presented to GPs in the diagnostic knowledge questions,[148] GPs are first asked how likely it is that the patient recovers *more* quickly if they are given antibiotics. Responses to this question may be interpreted as *beliefs about the efficacy of antibiotics* for individual patient cases. Secondly, GPs are asked about their *beliefs about peer norms* in prescribing behaviour: how likely it is that other GPs would give antibiotics for each patient case. Thirdly, the interview asks about GPs' *beliefs about patient expectations*: if they do not prescribe antibiotics for a specific patient case, how likely it is that the patient would choose to see another GP next time they are ill. While these beliefs are conceptually different from other behavioural influences, it is interesting to note possible interaction in the behavioural influences of, for instance, beliefs about patient expectations and altruism (more altruistic doctors may be more behaviourally responsive to beliefs about patient expectations), or beliefs about peer norms and competition (doctors under greater competitive pressure may be more likely to allow beliefs about peer norms to shape behaviour).

The GP interview records responses to these questions as a scenario probability between 0 and 100, with 0 reflecting a scenario that is believed to be not likely at all, and 100 indicating a scenario that is believed to be certain. For each question, a binary indicator is constructed to categorise beliefs that are relatively 'positive' on antibiotic prescribing (defined as a stated probability above the sample median for that question). This results in three indicators (one for each belief type) per patient case. I then aggregate these indicators by patient case, to create four categorical variables taking values 0-3 and reflecting the overall level of 'positive' antibiotic prescribing beliefs for each patient case.

---

[148] See Appendix B.5 (a) for vignette descriptions of these patient cases.

4.2.3.4 Competition

In the absence of reliable data on patient flows, the literature on physician market competition has tended to favour locational measures of competition, such as area-level physician density (the number of physicians per 1000 population within defined areas) or average distance to a physician practice (Vallejo-Torres and Morris, 2018). This study similarly utilises two geographic measures of the competitive pressure on sample GPs.

First, I construct an indicator of local competitor density. Given limitations in the data available for this study, it was not possible to estimate local population numbers at the suburban or district level of Johannesburg, to estimate area-level demand for GP services. As such, it is not possible to construct a GP density variable that accounts for local variations in population density within the City of Johannesburg (CoJ). I therefore construct a partial measure which calculates the number of rival GPs within a fixed radius of each sample GP's practice, without adjusting for local population numbers. Using the online *Medpages* database, which contains data on more than 80% of all private GPs active in South Africa (including the location coordinates of their primary practice address), and the *geodist* command on STATA, I calculate the straight-line (geodesic) distances between the practice location of each sample GP and those of all other private GPs in CoJ. This allows me to compute the main proxies for localised GP competition in this study: the numbers of rival GPs practicing within a 1 km, 2km and 5 km radius of each sample GP location. The limitation here is that unaccounted local variations in population density could vary the level of demand that GPs compete for in their area. However, it may be argued that in such an urban setting, area-level (resident) population numbers are less relevant to the *de facto* demand faced by GPs, as many patients may visit practices in areas where they are not resident perhaps because they work in that area or for other reasons of convenience.

Given this limitation (and given mixed findings in the literature as to the competitive effect of physician density (McGuire, 2000; Johnson, 2014)), I also construct a second measure of localised competitive pressure. I use GPs' practice location in the wealthiest suburbs of Johannesburg (the 'northern suburbs' of Sandton and Randburg – see Figure 2.2) as an indicator of potential price (or cost) competition. The price sensitivity of the local patient population in these suburbs is likely to be lower than in other suburbs of Johannesburg (particularly in the south), implying relatively lower price competition for GPs located there. A binary variable is accordingly coded for each GP, taking value '1' if the GP's primary practice is located in the northern suburbs, and '0' otherwise. This alternative measure is not without limitations itself. As with all locational measures of competition, it is hampered by the endogeneity of GPs' practice location decisions. Nevertheless, the two measures together (rather than any one) should give a more balanced indication of the competitive pressure on individual GPs in Johannesburg.

4.3 Study Design

4.3.1 Part 1: Within-subject experiment

The first part of this thesis study involved a within-subject SP field experiment, where each participating GP received unannounced visits from two SPs varying only in their insurance status. See Figure 4.1 for a timeline of fieldwork activities related to the experiment, as described in the following subsections.

The purpose of this experiment was to investigate the effects of patient insurance status on the quality and cost of care provided by primary care physicians (Research Question 1 - see Section 1.5). Data from this experiment is analysed in Chapters 5 and 6. This part of the study partnered with a large private health insurer in South Africa, which contracts with a large majority of all private GPs in Johannesburg. Health insurance cover was provided to the SPs for the purposes and duration of this research. All participating GPs were contracted-in with the partnering health insurer, so they were expected to see any patient with the SPs' insurance cover and to charge within the consultation rates set by the insurer.

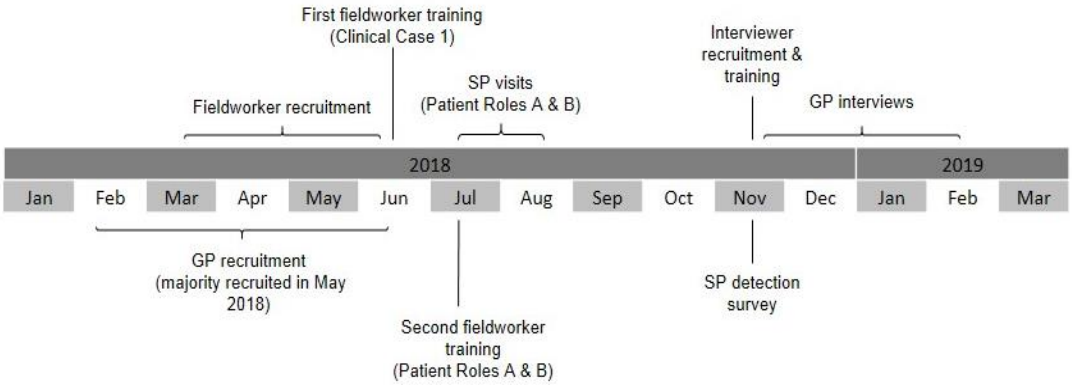


Figure 4.1. Timeline of fieldwork activities for study Part 1

4.3.1.1 Provider sampling and recruitment

The *Medpages* database of private primary care practitioners in the City of Johannesburg (CoJ) was used as the sampling frame for this study.[149] The database consisted of 999 active private GPs within CoJ at the time of recruitment, both of dispensing and non-dispensing status, spread across the 16 townships and seven administrative regions of the city.[150] Column (1) of Table 4.5 provides

[149] The *Medpages* database is a widely used healthcare provider directory in South Africa, and one of the largest and most comprehensive listings of private practitioners in the region. It contains contact, practice and basic demographic details on listed providers, updated on a quarterly basis. The version used in recruitment was downloaded in January 2018.
[150] Figure 2.2 (in Chapter 2) provides a map of these administrative regions in Johannesburg.

basic summary statistics on this GP population. 55% of all listed private GPs are male, with an average age of 51. The majority are non-dispensing (approximately 62%), and located in the northern suburbs of Sandton and Randburg (approximately 35%) or central Johannesburg (26%). Only non-dispensing GPs were targeted for recruitment in this part of the study, as the nature of their payment incentives were of particular interest in addressing the first two research questions outlined in Section 1.5. This is discussed further in Chapters 5 and 6. Non-dispensing GPs totalled 614 listed individuals, and Column (2) of Table 4.5 provides summary statistics for this sub-population. The average age of 50 years is comparable to the full GP population; however, the share of male GPs is slightly lower (47%). Non-dispensing GPs are also more concentrated in the northern suburbs of Sandton and Randburg (44%).

It was necessary to ensure that all participating GPs would accept the SPs' insurance cover, and charge within the fixed consultation rates set by the partnering health insurer. Therefore, a total 279 of these non-dispensing GPs that were not contracted-in with the health insurer were excluded. Of the remaining 335 GPs, a further 64 were deemed ineligible for the study as they were specialist GPs (such as urology specialists) or had begun practising as public health specialists, anaesthetists or other types of medical specialists.

The remaining 271 eligible non-dispensing GPs were then sorted in random order into a recruitment list. Recruitment was done by calling providers at their GP practices, in order of their listing on the recruitment list. I carried out the majority of GP recruitment for this part of the study in May 2018, and some fieldworkers were trained as recruiters to assist.[151] If a GP verbally consented to participating in the study, an automated email was sent to the GP with a study information sheet (summarising the information provided via the call).[152] A text message was also sent with a link to opt-out of the study, allowing GPs the option to withdraw their participation at any stage. Of the 271 GPs on the recruitment list, 208 were successfully reached – meaning, these GPs were successfully contacted and invited to participate in the study. Of these 208 GPs, 112 (54%) were successfully recruited into the study – that is, they consented to taking part and did

[151] Strict recruitment protocols were established, and all recruiters were appropriately briefed and periodically monitored as a quality check. A recruitment call script was developed, covering an explanation of the SP research methodology and information on the broad objectives, structure and timing of the study. GPs were informed that the SP visits would entail no cost to them in terms of time or lost earnings, as each visit (by design) would last the length of a normal consultation and be subject to the usual consultation fee for an insured patient. However, they were told that they would be approached for a brief telephone survey (to check for any SP detections) and a face-to-face interview following the visits. They were also told that they would have the opportunity to receive private and confidential feedback on their performance during the SP visits at a later stage. This feedback was offered to GPs as part of the intervention in Part 2 of this study (see Section 4.3.2), and all GPs not included in that intervention were given the opportunity to request feedback after the study's conclusion. The confidentiality of all identifiable individual-level data, and reporting of study results only at the aggregate level, was also emphasised. Where GPs had more specific questions or concerns that could not be addressed using the call script, recruiters were asked to put them in touch with me or a member of the ESRC study team to follow up. All recruitment calls were made during GPs' working hours, and call-backs were arranged if GPs were busy consulting.

[152] An online system was built using the *REDcap* software, which enabled the tracking of recruitment calls, the recording of GP responses and participation consent, and the programming of automated emails.

not opt-out at a later stage. 34 (30%) of these recruited GPs were also part of the ESRC study sample.

The difference in the number of GPs that were reached and the number that eventually consented raises the possibility of self-selection bias in the study sample: the GPs who consented to participation may be fundamentally different from those who refused. Table 4.5 compares basic socio-demographic characteristics of the final sample and interviewed sub-sample in Columns (3) and (4) (see the following sections for details on these samples) with those of the sampling frame of non-dispensing GPs, although the number of comparative metrics with available data is limited. The two samples are comparable to the sampling frame in the average age of GPs (49 years in both samples, compared to 50 in the sampling frame), and in the number of GPs located in the wealthier northern suburbs of Sandton and Randburg (43% and 45%, relative to 44% in the sampling frame). However, they hold a slightly higher representation of male GPs (52% and 51%, relative to 47% in the sampling frame).

It is difficult to attenuate this self-selection concern, apart from noting that the GPs who refused to participate did so for a number of different reasons: time and work pressures, periods of unavailability (leave) over the study period, multiple practice locations with little foresight on their future rotas, planned relocation, and so on. Some GPs also raised concerns about potential links between the study aims and the planned National Health Insurance (NHI) reform in South Africa, and its implications for future regulation of their private practice. This issue was foreseen by the research team, and recruiters were trained to explain that the study was an independent, academic piece of work investigating the drivers of care quality and cost in primary care, with no links to any government or regulatory body. Despite this, a few GPs still refused to participate due to these concerns.

4.3.1.2 SP visits with health insurance

Every GP in this experiment received two SP visits, each presenting one of two patient roles: a "low-insured" Patient Role A, and a "high-insured" Patient Role B. Both roles presented with Clinical Case 1 (uncomplicated (viral) acute bronchitis). The only distinction between these roles is therefore the patient's insurance status; the clinical case and case-specific Q&A scripts are identical.

The insurance plans assigned to the two roles are further identical in the breadth of their covered primary care services and accessible provider networks. Crucially, they are also identical in the maximum service rates (MRR) that can be charged by contracted primary care providers.[153]

[153] This includes the flat consultation rate and the fee-for-service rates for other healthcare services.

However, they have important differences in the insurance premiums charged (the monthly premium charged for the "high-insured" plan is roughly double that for the "low-insured" plan), and in certain limits on total annual claims. The annual allocation of medical savings per patient, which is used to pay for primary care services such as routine GP consultations and every-day medicines, is substantially lower for the "low-insured" plan. This imposes a tighter budget constraint on the "low-insured" patient's fully-covered primary care expenditures, and increases the likelihood that they incur out-of-pocket (OOP) care expenditure over the course of the year. Therefore, while the breadth of services allowable by their insurance will not differ between the two patient roles in any given consultation, the likelihood that it leads to some level of OOP expenditure for the patient (either from the consultation, or in future consultations due to depletion of the patient's medical savings) is much higher for the "low-insured" role. See Appendix B.3 for details on the insurance plans.

To ensure that treating GPs were aware of the SP's specific insurance cover, and the extent of their medical savings constraints, each patient role had a slightly different consultation opening statement (see Table 4.4). These statements are slight variations on the basic opening statement for Clinical Case 1 (provided in Table 4.1). The key difference in these statements between Patient Roles A and B is the degree of concern the patient communicates regarding their medical savings, to distinguish a relatively more cost-conscious "low-insured" patient. While it is tempting to liken this distinction between "low-insured" and "high-insured" patients to that of 'poor' and 'rich' patients, it is important to note that the insurance plan of the "low-insured" is by no means basic; it is still one of the more comprehensive plans offered by the partnering health insurer. Moreover, in a country where only 16% of the population have any private medical insurance at all, the plan certainly constitutes a 'premium' product for the average South African.

Table 4.4. SP opening statements, by patient role

Patient Role A "Low-insured" & Clinical Case 1	**Patient Role B** "High-insured" & Clinical Case 1
"I've been having this cough that is not going away. I had a cold a week ago, which is better now, but I'm still coughing a lot. My medical savings are a bit low, so I wasn't sure if it's serious enough to come today. But I thought better to make sure anyway."	"I had flu a week ago, which is a bit better now. But I am still coughing a lot - the cough is not going away. I'm not sure if it's that serious, but my medical aid is quite good, so I thought no harm in getting checked out"

Each of the 10 fieldworkers recruited for this part of the study was assigned a specific patient role and corresponding insurance cover – five were assigned the "low-insured" Patient Role A, and

five were assigned the "high-insured" Patient Role B.[154] Certain items were added to the SP Q&A scripts to address insurance-specific questions that may be asked by GPs, such as questions on the level of available medical savings (although it was not expected that many GPs would ask for such details). The second stage of SP training (over two days in July 2018) specifically covered the insurance-specific aspects of these two patient roles. Fieldworkers were briefed on the details of their specific insurance cover, and their new role-specific opening statements and Q&A scripts (see Appendix B.3 for the role-specific scripts and insurance plan details). Fieldworkers also adjusted the occupation and social background of their patient roles, in case of related questioning by doctors, to account for the likelihood that the "high-insured" insurance plan would only be offered in highly-paid jobs or affordable for individuals from wealthy backgrounds.

All SP visits took place over three weeks in July-August 2018, under my supervision and a local fieldwork company's coordination. The five fieldworkers who were assigned the "low-insured" SP role were each paired with one of the five fieldworkers assigned the "high-insured" role. Each recruited GP was then randomly allocated to visits by one fieldworker pair.[155] As the SP method requires that all patient characteristics apart from insurance status are standardised across the two SPs visiting any one GP, it was important to match fieldworkers as closely as possible on the basis of gender, height, physical appearance, and other personal traits. Efforts were also made to match paired visits on the time of day (morning or afternoon), but this proved logistically very difficult to uphold for all visit pairs; as such, this particular constraint was relaxed. To control for learning or order effects, whereby GPs may perform better during the second SP visit due to having seen a similar clinical case in the first one, the sequencing of paired visits was randomised for each GP, such that some GPs saw the "low-insured" SP first whilst others saw the "high-insured" first. Moreover, at least a two-day gap was maintained between paired visits to the same GP, to minimise this learning effect and reduce the likelihood of detection.

Each fieldworker was given a set of items to take to their scheduled consultations. A personal fieldwork journal was provided, detailing their schedule of visits to complete each day. An electronic tablet was provided for completing the SP debriefing questionnaire, along with a hard-paper copy of the questionnaire in case of technical issues with the electronic version. Fieldworkers were also given a sealable plastic bag for each consultation, labelled with their names and the unique identifiers of GPs they were scheduled to see. They were instructed to seal any receipts or prescriptions from each consultation inside these bags.

---

[154] It was not possible to pool fieldworkers and patient roles in this experiment, as specific insurance cover ('high' or 'low' cover) had to be bought for each individual in their own name for the purposes of this research.

[155] Some of these random GP-fieldworker pair combinations had to be manually adjusted, as the overlap in fieldworkers and some participant GPs with the ESRC study meant that some of the fieldworkers had already visited the same doctor or another doctor at the same GP practice during the ESRC study fieldwork.

Table 4.5. Sampling summary statistics (Part 1 of study)

	(1) All GPs	(2) Non-dispensing GPs	(3) Final Sample	(4) Interview Sub-sample
Observations (N)	999	614	89	75
Gender				
Female	45%	53%	48%	49%
Male	55%	47%	52%	51%
Age (mean)	51	50	49	49
Dispensing status				
Non-dispensing	62%	100%	100%	100%
Dispensing	38%	0%	0%	0%
Contracting status				
Contracted-in	71%	60%	100%	100%
Contracted-out	29%	40%	0%	0%
Practice location				
Northern periphery *(Region A & Alexandra)*	8%	9%	4%	3%
Northern Suburbs *(Regions B & E, excl. Alexandra)*	35%	44%	43%	45%
Western Suburbs *(Region C)*	12%	13%	16%	16%
Central Johannesburg *(Region F, excl. Jo'burg South)*	26%	25%	20%	17%
Southern Suburbs *(Regions D & G, & Jo'burg South)*	19%	9%	17%	19%

Source: Medpages database, author's calculations
Note: 'All GPs' included the 999 private GPs listed as active in the City of Johannesburg (CoJ) in the online medical practitioner database, Medpages, which includes approximately 80% of all medical practitioners in South Africa. 'Non-dispensing GPs' are the subsample of all listed GPs that do not have drug dispensing licences. The 'Final Sample' includes the 89 GPs that were successfully visited by both SPs in Part 1 of this study, and the 'Interviewed Sub-sample" includes the 75 GPs of these visited 89 who were also successfully interviewed following the SP visits. 'Contracting status' refers to GPs' contracting terms with private medical schemes. 'Contracted-in' GPs are those who are contracted into schemes' provider networks, and therefore have less discretion over their service pricing for insured patients. The 'Northern Suburbs' include the wealthier suburbs of Sandton and Ranburg in Regions B and E - see figure 2.2 for a map of the administrative Regions A-G in CoJ.

Due to the random allocation of visits to individual fieldworkers and the geographical dispersion of recruited GPs, fieldworkers were often unable to complete their three to four scheduled visits a day. In these cases, visits were rescheduled or swapped with a later visit, as long as these swaps did not violate the minimum two-day gap between paired visits. Consultation appointments were scheduled in advance where possible, and fieldworkers were advised to arrive at least 15 minutes before their scheduled consultation times to avoid any delays. At the end of each fieldwork day, the fieldworkers met with the fieldwork company team for a debriefing, to drop off the sealed plastic bags and contents from the day's consultations, and to flag any issues that arose during the day.

Of the 112 recruited GPs, 89 were successfully visited by both patient roles (both paired visits) and satisfied the eligibility criteria for this part of the study. 13 of the 112 recruited GPs could not be visited for a number of reasons. Some of these GPs were unavailable during the fieldwork period, some could not be contacted for scheduling an appointment, and some could not be visited by any of the five fieldworker-pairs (due to at least one fieldworker in each pair having previously

visited the same GP, or other GPs in the same practice, during the ESRC study fieldwork). A further 10 recruited GPs were visited by SPs, but it was discovered that they were actually dispensing GPs (having dispensed medications during the SP visits) and were therefore not eligible to participate. The visits to these 10 GPs were subsequently dropped from the final sample. This resulted in 89 retained GPs, and 178 completed and valid consultations (at a visit completion rate of 79%) that form the 'final sample' in this part of the study. Summary statistics on basic GP characteristics for this final sample, comparable to those for the sampling frame, are shown in Column (3) of Table 4.5.

4.3.1.3 SP detection survey

In November 2018, the 89 GPs in the final sample were telephoned to conduct a short SP detection survey. Each GP was asked if they recalled having suspected any patient they had seen over the last few months to be a fake patient. If the GP answered yes, they were then asked how many patients they had been suspicious of. 15 of the 89 GPs reported at least one suspected patient (3 of these 15 GPs reported 2 suspected patients, 1 GP reported 3, and another reported 6). For each suspected patient, the GPs were asked for their recollection of the patient's age, gender, symptoms, approximate time of visit (the month, and whether it was the first- or second-half of the month), and any specific reason for their suspicion.

Based on this information, and my records on the fieldworker pair that completed each GP's visits and their timing, it was possible to assess whether each suspecting GP's recollection matched (or approximated) their actual visit experience, and thus whether their detection was valid.[156] This assessment was first done by me and three members of the ESRC study team independently, and then discussed together to come to a consensus on conflicting assessments. It was agreed that only 2 of the 15 GPs reporting a suspicion were likely to have detected an SP during their two visits. While we cannot be fully certain that these detections are valid, we assessed that the likelihood of a correct detection is sufficiently high in these two cases.

[156] To assess these cases systematically, members of the ESRC study team and I first developed a protocol for classifying valid SP detections. This was informed by one member (my academic supervisor)'s previous experience in conducting similar SP studies. We agreed that the reporting of similar symptoms to our clinical cases was the initial screening criteria for valid detections. 5 of the 15 suspecting GPs fulfilled this criteria in at least one of their reported cases. It was also agreed that this was a necessary but insufficient condition for classifying a valid detection, as a lot of patients would be presenting to GPs with similar flu-like symptoms to Clinical Case 1 during that time of year (as it was the flu season). For these five GPs, we then looked at the reported gender and age of the suspected patients, and the timing of their visits, to compare these to their actual SP visits. We also considered the reasons specified for the GP's suspicion. As this was quite a subjective assessment, each of us classified the five cases independently, and then compared our conclusions to come to a consensus.

In the final sample of 89 GPs, this represents a likely detection rate of 2%. In my empirical analyses in Chapters 5 and 6 (which draw on data from this experiment), I check the robustness of all results to excluding these detection cases from the sample.

4.3.1.4 GP interviews

Following the SP detection survey, all GPs in the final sample were approached for a 20-30 minute face-to-face interview, to collect additional data on personal and clinical practice characteristics (as described in Section 4.2). GPs were telephoned at their practices to schedule the interviews, and enumerators followed a set script during these scheduling calls. See Appendix B.4 for an overview of the interview structure and components. All interviews took place at GPs' practices, between November 2018 and February 2019.

Of the 89 GPs in the final sample, interviews were successfully completed with 75, resulting in a non-participation rate of 16%. The majority of the 14 GPs that did not complete an interview cited time constraints as reasons for their inability to continue with the study (30 minutes was required for the interview during their working hours). Four GPs were either not reached or did not confirm an interview time by the end of February 2019, and were not pursued further due to budgetary constraints on the study. Lastly, one GP had moved abroad and was therefore not contacted for the interview. As a result, complete data on all GP characteristics listed in Section 4.2.3 is only available for this sub-sample of 75 GPs (hereafter, the 'interview sub-sample'). While the final sample contains data on all necessary outcomes for answering Research Question 1 (see Section 1.5), the interview sub-sample with additional data on provider-level characteristics (including altruism) is required to fully answer Research Question 2. Both samples are therefore drawn on in Chapters 5 and 6, which seek to address these questions in their analyses.

Summary statistics on basic GP characteristics for this interview sub-sample, taken from the *Medpages* data, are shown in Column (4) of Table 4.5.

4.3.2 Part 2: Between-subject randomised experiment

The second part of this study involved a randomised field experiment to evaluate the impact of a private performance feedback intervention on the appropriateness of care delivered by private GPs – namely, their antibiotic treatment decisions and treatment costs. A timeline of all fieldwork activities in this experiment is presented in Figure 4.2.

The purpose of the experiment was to investigate Research Question 3 (see Section 1.5), and its outcomes are analysed in Chapter 7. The experiment details are provided in the following sections, and a brief overview is as follows. Each sampled GP was first randomised into one of two interventions: *i)* receipt of a written summary of clinical guidelines for common respiratory infections ('control' intervention), or *ii)* receipt of private, individualised feedback on clinical performance during 'baseline' SP visits (conducted as part of the ESRC study) *and* the same written summary of clinical guidelines ('treatment' intervention). These interventions were delivered at the end of the GP interviews for this sample.[157] Following the intervention, each participant GP received 'end-line' SP visits from two SPs varying only in their clinical case (Clinical Case 1 or Clinical Case 2).[158]

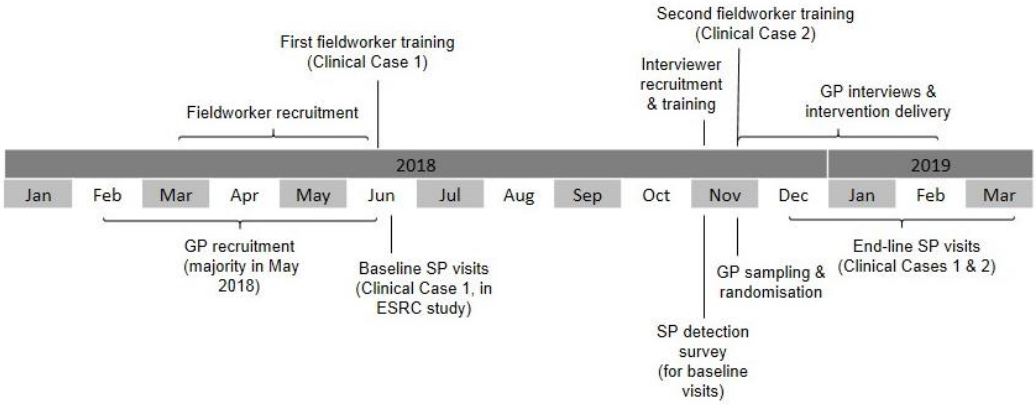


Figure 4.2. Timeline of fieldwork activities for study Part 2

GP participants were drawn from the ESRC study sample, and baseline data was also taken from the ESRC study results. The sampling protocol and stages of the experiment are described below. GP drop-out rates at the randomisation, intervention and end-line visit stages are also detailed there (and later summarised in Figure 7.1 in Chapter 7). Basic summary statistics on the 'initial

[157] The interview itself is described in Appendix B.4.

[158] Note that all SPs in this second experiment presented as uninsured patients (paying in cash), to mirror the 'baseline' SP visits in the ESRC study. The insurance status of SPs was not a variable of interest in this experiment (unlike in Part 1), and it was logistically easier for SPs to be uninsured and pay for consultations in cash. Hence, the only variation in SPs presenting to the same GP in this experiment was the SP clinical case.

sample' of randomised GPs and the 'final sample' of (end-line) visited GPs are presented in Table 4.6, alongside comparable statistics for the *Medpages* GP population and the sampling frame.

4.3.2.1 Provider sampling and randomisation

The sampling frame for this experiment was a sub-set of GPs who participated in the ESRC study and received visits from SPs presenting Clinical Case 1 in that study.[159] The ESRC study sample was randomly drawn from the *Medpages* population of private GPs practicing in the City of Johannesburg. [160]

A total of 122 private GPs received an SP visit with Clinical Case 1 in the ESRC study, and form the sampling frame for this experiment. Results from these visits also provide the baseline data for this part of the study. Of these 122 GPs, efforts were made to exclude those sharing large group practices. This decision was taken partly to facilitate fieldwork, as several fieldworkers had visited GPs as part of the ESRC study, and it was increasingly difficult to find recruited fieldworkers that had not already visited certain group GP practices in the study. It was also done to limit potential spill-over of any treatment effects from treated GPs to control GPs practising in their vicinity. 15 group practice GPs were excluded as a result.[161] In addition, another 8 GPs were excluded either because they could not be contacted for the SP detection survey in the ESRC study or refused to take part in it, or because they reported detections in the survey that were assessed to be valid.

The remaining 99 GPs (the 'initial sample') were then stratified according to their drug dispensing status, contracting status with medical schemes, and baseline performance on the main outcome of interest (whether they had prescribed an inappropriate antibiotic for Clinical Case 1), and randomised to one of two intervention groups: 'control' or 'treatment'. The randomisation resulted in 50 GPs assigned to the treatment group, and 49 to the control group. A comparison of Columns (3) and (4) of Table 4.6 shows that the treatment group has slightly older GPs on average, and also a larger share of male GPs, in this initial sample.

---

[159]Note that there were also experimental variations in some SPs presenting Clinical Case 1 in the ESRC study (similar to the insurance variation in Part 1 of this study). This study considers only those ESRC study SP visits where the 'basic' Clinical Case 1 was presented (by uninsured SPs, with no experimental variation).

[160] That study included both drug-dispensing and non-dispensing GPs, as well as GPs both contracted-in and -out of medical schemes' provider networks (see Section 2.2.4 for a description of these GP practice variations); and the random sampling procedure employed there first stratified the GP population according to dispensing and contracting status. As such, the sample in this experiment also contains both dispensing and non-dispensing doctors, of contracted-in and contracted-out status.

[161] Although it was decided to exclude group practice GPs from the experiment sample, around 6 GPs that were retained in the sample were later found to be working within group practices; this was initially missed in the sampling stage, as these GPs had slightly different practice addresses to other GPs working at the same group practices (so they were assumed to be in solo practices).

4.3.2.2 Randomised interventions

The treatment in this experiment comprised of confidential, written feedback on GPs' individual performance during the baseline SP visits (with Clinical Case 1) in the ESRC study. GPs in the treatment group were handed this feedback in a sealed envelope at the end of their GP interviews, which followed the same format as the interviews in Part 1 of this study (see Appendix B.4). Interviewers were trained to invite the GPs to open and review the feedback, to ensure (where possible) that the GPs actually read the contents and complied with the treatment. See Appendix B.10. (a) for an example of the written feedback sheet, and Appendix B.9 for the interviewer script in delivering the intervention.

Feedback was offered on three areas of GP performance during baseline visits: antibiotic prescribing, clinical effort, and drug treatment costs. As antibiotics are not indicated for Clinical Case 1, GPs were informed whether they correctly abstained from prescribing antibiotics for this case. For clinical effort, the proportion of case-specific history-taking and examination checklist items completed by GPs during their baseline visits was reported.[162] Lastly, GPs were told how much the medications they had dispensed or prescribed for the patient would cost on the local market, and how this cost compared (in multiplier terms) to the cheapest, appropriate treatment available for the clinical case.[163]

Individual performance was reported only relative to evidence-based benchmarks (i.e. how GP performance compared to evidence-based treatment guidelines, clinical effort best-practices, and locally cost-effective treatment choices). It was decided that no peer-comparative performance information would be provided (for example, comparing individual GP performance to the sample average), in order to avoid potentially harmful benchmarking. For instance, if the average antibiotic prescribing rate is high within the sample (as reported in Chapter 7), peer-benchmarking may validate and encourage such prescribing behaviour as being the 'norm' among peers.

In addition to performance feedback, treatment group GPs were also given an educational supplement with a summary of national clinical guidelines for diagnosing and treating common respiratory illnesses that tend to present in primary care (including the clinical cases in this

[162] See Section 4.2.1 for details on how these checklists were devised, and Appendix B.1 for a full list of all case-specific history-taking and examination items included in these checklists.

[163] See Section 4.2.2 for details on how drugs were priced in this study. The cheapest treatment that was considered appropriate for Clinical Case 1 - based on advice from national respiratory and infectious diseases experts, recommendations in the South African Standard Treatment Guidelines and Essential Medicines List (STG/EML), and prices in the South African MPR – was paracetamol and a cough suppressant, priced at a total R17.86.

study),[164] and a summary of the ESRC study from where the feedback performance data was taken.[165]

Control group GPs did not receive the individualised feedback on their baseline performance. They were only given the educational supplement and the summary of the ESRC study. This was also given in a sealed envelope at the end of their GP interviews. See Appendix B.10. (b) for an example of the control group leaflet.[166]

Table 4.6. Sampling summary statistics (Part 2 of study)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All GPs	Sampling Frame	Initial Sample			Final Sample		
			Treat	Control	Total	Treat	Control	Total
Observations (N)	999	122	50	49	99	37	43	80
Gender								
Female	45%	33%	20%	39%	30%	24%	35%	30%
Male	55%	67%	80%	61%	70%	76%	65%	70%
Age (mean)	51	52	54	53	53	55	53	54
Dispensing status								
Non-dispensing	62%	50%	41%	43%	42%	41%	44%	43%
Dispensing	38%	50%	59%	57%	58%	59%	56%	57%
Contracting status								
Contracted-in	71%	79%	84%	82%	83%	81%	81%	81%
Contracted-out	29%	21%	16%	18%	17%	19%	19%	19%
Practice location								
Northern periphery *(Region A & Alexandra)*	8%	4%	6%	4%	5%	3%	5%	4%
Northern Suburbs *(Regions B & E, excl. Alexandra)*	35%	29%	18%	33%	26%	16%	32%	25%
Western Suburbs *(Region C)*	12%	13%	12%	8%	10%	13%	5%	9%
Central Johannesburg *(Region F, excl. Jo'burg South)*	26%	21%	27%	22%	24%	30%	23%	26%
Southern Suburbs *(Regions D & G, & Jo'burg South)*	19%	33%	37%	33%	35%	38%	35%	36%

Source: Medpages database, author's calculations

Note: 'All GPs' included the 999 private GPs listed as active in the City of Johannesburg (CoJ) in the online medical practitioner database, Medpages, which includes approximately 80% of all medical practitioners in South Africa. 'Sampling Frame' includes the subsample of 122 GPs from the ESRC study that were considered for Part 2 of this study. Note that 2 of these 122GPs are not listed in the Medpages data, and are hence exlcuded from the calculated percentages listed here. 'Initial Sample' includes the 99 GPs from the sampling frame that were assessed to be eligible for this study, and randomised to either the treatment or control groups. Note that 1 GP of the 50 randomised to the treatment group is also not included in the Medpages data. The 'Final Sample' includes the 80 GPs that successfully recieved the intervention and were visited by both end-line SPs in Part 2 of this study. "Contracting status' refers to GPs' contracting terms with private medical schemes. 'Contracted-in' GPs are contracted into schemes' provider networks, and therefore have less discretion over their service pricing for insured patients. The 'Northern Suburbs' include the wealthier suburbs of Sandton and Ranburg in Regions B and E - see figure 2.2 for a map of the administrative Regions A-G in CoJ.

[164] These include the common cold, acute bronchitis, influenza, acute bacterial sinusitis and pneumonia.

[165] It was explained that the ESRC study was focused on investigating antibiotic prescribing behaviour in primary care, and was motivated by the high rates of antibiotic prescribing in South Africa and the growing issue of antimicrobial resistance.

[166] All GPs who participated in either Part 1 of this study or the ESRC study (but who were not in the treatment group of this experiment) were also offered individual feedback on their performance in the two studies upon request once the experiment was concluded. Moreover, all GPs were offered feedback on the aggregate results from the two studies.

4.3.2.3 GP interviews and intervention delivery

Following a similar protocol to the scheduling of interviews in study Part 1, the 99 GPs in the initial sample were approached for face-to-face interviews after completion of the SP detection surveys in the ESRC study. The GPs were not informed about the intervention at the time of interview scheduling, to ensure that this knowledge did not influence their decision to interview. At the start of each interview, consent was sought from each GP for continued participation in the study.[167] The GPs were given the study information sheet again at this stage (see Appendix B.7 for the information sheet) and informed that continued participation could entail private performance feedback and further SP visits. The feedback intervention took place at the end of these interviews, subject to GPs' consent to continued participation. Besides being informed about the possibility of receiving private performance feedback and further SP visits, GPs were blinded to the experiment details and their treatment status throughout the whole study period, to prevent this knowledge from influencing their decisions to participate or their responses to the intervention.

83 of the 99 GPs in the initial sample were successfully reached and scheduled for an interview. 12 of the 99 GPs declined the interview, predominantly stating work time constraints. The other four GPs who were not interviewed were either unreachable on their contact numbers over the interview scheduling period, or had moved overseas.

The 83 scheduled interviews were carried out over November 2018-February 2019. This resulted in an intervention non-participation rate of 22% in the treatment group (11 of the 50 treatment group GPs in the initial sample) and 10% in the control group (5 of the 49 control GPs in the initial sample). Other aspects of intervention delivery, and the implications of different non-participation rates in the treatment and control groups, are discussed in Chapter 7.

4.3.2.4 End-line SP visits

As stated previously, the baseline outcomes for this experiment were taken from the SP visits with Clinical Case 1 in the ESRC study. Following the feedback intervention, each of the 83 participating GPs that received the intervention (i.e. those that completed an interview and consented to ongoing study participation) was sent two SPs presenting Clinical Case 1 and Clinical Case 2. The aim of these visits was to collect end-line data to evaluate certain effects of the feedback intervention (as discussed in Chapter 7). Apart from their clinical case, the two end-line SPs were otherwise identical, including in their insurance status: both presented as uninsured

[167] Note that this experiment (study Part 2) was framed as a continuation of the ESRC study, as all participant GPs in this experiment were initially recruited for that study.

patients and paid for the GP consultations in cash. This ensured that any incentives deriving from the insurance status of the patient (as explored in Part 1 of the study) were balanced across the two SPs.

10 fieldworkers were employed in this part of the study. Unlike in Part 1, each fieldworker was trained to play both clinical cases in this experiment. Clinical Case 1 was introduced in the first stage of SP training (as detailed in Section 4.1.2), and Clinical Case 2 was introduced in the third stage in November 2018.[168] This latter training emphasised fieldworkers' ability to differentiate the presentations of the two (similar) clinical cases. Fieldworkers had to memorise a number of details on the history and therapeutic regimen of the HIV+ patient in Clinical Case 2. Extensive role-play, mock consultations and practice in recall (whilst varying the clinical case) ensured that fieldworkers were comfortable with the new clinical case and in switching between the two cases.

As in Part 1, fieldworkers were grouped into five fieldworker-pairs, with pairs matched as closely as possible on age, gender and physical appearance. The gender-matching was relaxed for one pair, where a male was paired with a female (due to uneven numbers of male and female fieldworkers). Each participating GP was then randomly allocated to one fieldworker pair, and each fieldworker within a pair was randomly allocated to one of the two clinical cases to present to each GP. Again, to control for potential learning effects where GPs systematically perform better during the second SP visit, the order of the two visits was randomised; and to reduce the likelihood of detection, at least a two-day gap was maintained between the two SP visits to the same GP. Other fieldwork protocols were also the same as in the SP visits in Part 1 (see Section 4.3.1.2).

The visits in this experiment were completed over a longer period relative to the visits in Part 1, given the need to follow the scheduling of GP interviews where the intervention took place. The interviews were staggered over November 2018-February 2019, to align with GP availability. Therefore, the SP visits in this experiment also spanned a similar period: December 2018-March 2019. The majority of all end-line visits (96%) were completed within 3-4 weeks of intervention delivery (i.e. completion of the corresponding GP interview). However, six visits were delayed and completed between 1-3 months of intervention delivery. These six delayed visits all occurred with control group GPs, and resulting implications for balance in visit characteristics between the treatment and control groups are discussed in Chapter 7.

Of the 83 interviewed GPs, 80 were successfully visited at end-line by both SPs. Of the three GPs that were lost to follow-up, one GP had retired since the interview, one had stopped seeing acute

[168] Some of these 10 fieldworkers also took part in the second stage of training in July 2018, if they were also employed as SPs in Part 1.

patients, and one refused to on-going participation in the study following the GP interview. This resulted in a 'final sample' of 80 GPs (37 in the treatment group, and 43 in the control) and 160 valid consultations, with an attrition rate of 4%. Implications of this attrition are discussed in Chapter 7. Summary statistics on basic GP characteristics for this final sample are presented in Columns (6)-(8) of Table 4.6. Treatment and control group balance along observable GP characteristics and baseline outcomes in this final sample is also evaluated in Chapter 7.

Unlike with the SP visits in Part 1, and with baseline visits in the ESRC study, an SP detection survey was not carried out following these end-line visits. It was decided with the ESRC research team that, given the very low detection rates among visits in Part 1 and in the ESRC study, it would not be cost-effective to carry out a similar survey here.

5 Physicians' Effort Choices and Care Quality under Health Insurance

5.1 Introduction

A rapid expansion in health insurance has underpinned global efforts toward universal health coverage (UHC), characterised by a rising prevalence in third-party purchasing, and reduced financial participation of patients in their own care. To achieve broad coverage and patient choice in health services, many governments have also begun formally engaging the private sector within national insurance schemes (Odendaal *et al.*, 2018; Rao *et al.*, 2018). In South Africa, the government's latest proposals for its planned national health insurance (NHI) reform include contracting private primary care physicians (GPs) to expand access to free care, and abolishing all patient co-payment in private insurance schemes (Republic of South Africa, 2018, 2019). Despite these trends, there is limited theory or evidence on the impact of insurance and third-party financing of market-based healthcare on providers' care quality decisions. Existing literature on the effects of insurance on physician behaviour tend to focus on choices of care *quantity* and costs, rather than quality - in part, due to problems in inferring care quality at the level of physician-patient interactions (Pauly, 1978; Das *et al.*, 2016).

Healthcare quality depends on a number of physician inputs – broadly defined as "effort" - that are not easily verifiable, particularly by third-party payers, and therefore non-contractible (Ma and Mcguire, 1997).[169] This chapter argues that the combination of greater financial protection for patients and third-party purchasing of care from private providers exacerbates a principal-agent problem in incentivising appropriate physician effort, and can worsen the technical quality of care.[170] Where physicians face an uncompensated cost to effort (where there are no immediate financial incentives attached), this effort will in general be undersupplied relative to what is

[169] Here, I distinguish between care quality *inputs*, which includes all forms of clinical effort, and care quality *outputs* (i.e. *actual* or *realised* care quality), which include the accuracy of diagnoses and appropriateness of any treatment.

[170] Technical quality is defined as the *clinical effectiveness* (how well clinical guidelines and medical science are applied to the diagnosis and treatment of a clinical case) and *safety* of care processes, as distinct from interpersonal aspects of care quality (in other words, *patient experience*). Note that all references to physicians' care quality decisions in this chapter refer to process measures of care quality (reflecting what is done to and for the patient during the clinical interaction and any follow-up), as oppose to structural or outcome measures (Donabedian, 1980).

optimal for the patient.[171] The extent of this undersupply ("effort-stinting") may be determined by both the altruistic and market incentives physicians face (Allard et *al.*, 2009) – incentives that can vary in the patient's financial protection, as argued in Chapter 3.

The theoretical framework predicts that, where technical quality coincides with significant *unobserved* physician effort (Chalkley and Malcolmson, 1998; Haas-Wilson, 1994), and has a direct impact in reducing current and future health expenditures (WHO, World Bank and OECD, 2018), its supply will decrease with patient insurance. Private physicians have both altruistic and market incentives to supply more appropriate (and unobserved) effort to low-insured patients, for whom the financial consequences of poor quality care are greater and more tangible. Equally, given their need to compete for patients on *some* observable dimension of care (and their inability to compete on OOP costs with high-insured patients), physicians' relative focus on improving patient experience and satisfying patient demands will increase in the level of patient insurance. This latter prediction also implies that, in situations where optimal care necessitates the refusal of patient demands (such as demand for unnecessary antibiotics), the competitive pressure on physicians to satisfy greater actual or perceived demands from less cost-conscious patients may be an additional channel through which technical care quality can deteriorate in the patient's insurance cover.[172]

To summarise, the primary hypotheses tested in this chapter are the following:[173]

i) Private physicians supply more observable care inputs that enhance patient experience (e.g. physical effort and consultation time) when treating high-insured patients (*Hypothesis 6*).

ii) However, they also supply less unobserved, appropriate care inputs for these patients; resulting in lower quality care outputs (including the accuracy of diagnoses and appropriateness of treatments) for high-insured patients (*Hypothesis 4*).

To test these hypotheses, an experimental audit study with standardised patients (SPs)[174] was carried out in Johannesburg. The experiment involved 89 private GPs, who each received randomly-ordered, incognito visits from two SPs that presented identical clinical cases and varied

[171] This is especially true under regulated prices, as in the context of this study setting, where providers have limited ability to extract compensation for their effort by setting higher fees. In Das *et al.* (2016)'s framework, on the other hand, providers are modeled as charging a piece rate per unit of effort in their consultation fees, which serves to incentivize their effort supply.

[172] These predictions extend Das *et al.* (2016)'s analysis of physicians' effort choices with and without market incentives. Unlike in that study, this chapter considers a context with regulated prices (where physicians have limited capacity to 'price in' their non-contractible effort), and where variations in market incentives may be generated by patient insurance.

[173] These hypotheses correspond to some of the hypotheses derived in Chapter 3. The specific numbered hypotheses are given in parentheses.

[174] See Sections 4.1 and 4.3.1 for a detailed description of the SP method and experiment (comprising Part 1 of this study).

only in their level of insurance cover: high-insured or low-insured. A controlled field experiment was necessary in this case, as the use of observational data presented three primary challenges. First, there are potential endogeneity issues related to patients' selection into insurance and doctors' selection of specific patient types to treat (with specific conditions or insurance). Second, it is difficult to isolate doctors' responses to patient insurance alone, as insurance can also change the behaviour of patients. The SP method allows exogenous variation in the patient insurance status faced by the *same* GPs; and by standardising the script, behaviour and clinical cases of SPs, the influence of additional patient- or case-specific characteristics (including patient demand) can be controlled for. Third, measuring (or inferring) the quality of care provided in clinical interactions is challenging, as optimal care is both patient- and case-specific (Pauly, 1978; Das *et al.*, 2016) and not all determinants can be verified in observational data. Moreover, not all aspects of provider behaviour that influence care quality are easy to measure or systematically reported. Again, by controlling the clinical case presented by SPs, optimal care can be determined *ex ante*, and - through detailed debriefing by SPs - be measured *ex post*.

This chapter's main contribution is in providing novel evidence that isolates the impact of patient insurance status on the quality of care delivery. The results demonstrate that physicians offer lower quality care outputs overall (less accurate diagnoses and appropriate treatment) to high-insured patients, perhaps indicating less appropriate clinical effort with these patients (in line with *Hypothesis 4*). GPs offered a correct or partially correct diagnosis to 46% of high-insured patients, compared to 60% of low-insured ones (a 23% decrease). GPs also prescribed unnecessary antibiotics to 63% of high-insured patients, compared to 51% of low-insured ones (a 24% increase). Notably, this lower quality of care output does not coincide with poorer performance on *observed* aspects of effort for high-insured patients: there were no significant differences in consultation time, and doctors completed slightly more history-taking items and examinations for these patients, as measured by a composite index score and the share of essential physical examinations completed (although the magnitude of these differences are very small). Similarly, doctors were significantly more likely to advise high-insured patients to come back for a repeat visit in case of worsening symptoms (73% vs. 51% for low-insured patients).[175] These latter findings provide some support to *Hypothesis 6*.

[175] While this advice may be considered appropriate, it is not essential for this clinical case, as an accurate diagnosis of an uncomplicated and likely viral respiratory infection in an otherwise healthy young adult patient should give the GP sufficient confidence that the illness is self-limiting and will not require any follow-up. However, the GP may wish to be conservative, reassure the patient, or display more communication effort by providing such advice anyway. It should be noted that there are clear fee-for-service (FFS) payment incentives attached to advising patients to come back for repeat consultations, and GPs may be more inclined to do so when the patient is well-insured.

The remainder of this chapter is organised as follows. Section 5.2 discusses the chapter's contributions to existing literature. The experimental methods and data are outlined in Section 5.3, including a presentation of summary statistics on sample characteristics and the care quality outcomes of interest. The empirical framework is discussed in Section 5.4, followed by a presentation of the main results and robustness checks in Section 5.5. The chapter concludes with a discussion of the main findings and policy implications in Section 5.6.

5.2 Literature Contributions

This chapter builds on several strands of existing literature. Firstly, it draws on and develops the theoretical literature on quality contractibility problems under third-party purchasing. Chalkley and Malcomson (1998)'s analysis of hospital quality contracting is perhaps the most closely related theoretical work. They consider a setting where healthcare is paid for through insurance, and providers' incentives for un-monitored quality rely on patients' demand response. The authors argue that, where patient demand does not fully reflect quality (or where it reflects certain dimensions but not others), market incentives alone will be insufficient for maintaining quality and the optimal contract would depend on the degree of provider altruism ('benevolence'). While they characterise an inevitable trade-off between incentivising quality improvement and cost reduction (unless the hospital is fully benevolent), this chapter argues that higher quality may also coincide with lower costs. The role of altruism and market incentives in ameliorating the quality contractibility problem is also addressed in Allard et *al.* (2009), where uncertainty around the treatment-outcome relationship captures patients' limited ability to observe and verify appropriate physician effort. Building on their framework, this chapter proposes a role for variations in patient insurance (and implied variations in altruistic and market incentives) in determining physicians' effort choices. Related work by Ma and Mcguire (1997) considers physicians' optimal effort choices under different insurance-payment contracts, where quality is observable to the patient (and influences demand for physician services) but unverifiable by the third-party payer. The present chapter relaxes this assumption of complete information in the physician-patient relationship, and considers that physicians' care quality choices may not always be visible (or comprehensible) to patients (Arrow, 1963). McGuire (2000) discusses physicians' effort choices under asymmetric information, regulated prices (as a result of health insurance) and market incentives. However, the focus there is on the effects of different payment structures on care quality choices, rather than on the effects of varying patients' insurance cover. Blomqvist (1991) further underlines the role of the contract structure – fee-for-service reimbursement (under conventional third-party insurance) or salaried payment of physicians (under a HMO-style organisation) – in determining provider choices under health insurance. The author nevertheless

ignores many aspects of care quality by assuming that physicians arrive costlessly at an accurate diagnosis, and by abstracting from any choices on diagnostic or therapeutic effort.

Secondly, this chapter builds on the health policy literature documenting disparities in specific measures of care quality by patient insurance status. Most studies examine associations between patient insurance and three types of care quality metrics: health outcomes (Franks *et al.*, 1993; Hasan *et al.*, 2010; Spencer, Gaskin and Roberts, 2013; Woolhandler and Himmelstein, 2017), utilisation of specific services or treatments (Roetzheim *et al.*, 2000; Churilla *et al.*, 2017), and patient satisfaction (Shi, 2000; Abuosi *et al.*, 2016). While they generally find a positive association between patient insurance cover and quality outcomes, the quality metrics considered are predominantly those that tend to be systematically monitored and publicised (such as mortality outcomes) and that are highly salient (or subjective) to patients. Moreover, these studies do not isolate the role of provider responses to insurance from other confounders. This makes it difficult to draw conclusions on the potential for policies that target provider behaviours in particular, relative to other patient-centred or structural interventions, in improving quality outcomes of interest.

Thirdly, by focusing on the role of the physician, this chapter contributes evidence to an extensive body of empirical work on physician agency in general (see McGuire (2000) and Johnson (2014) for reviews), and a smaller sub-set on physician agency under health insurance (Mort *et al.*, 1996; Lundin, 2000; Iizuka, 2007; Lu, 2014). The bulk of this literature relies on observational data, and where the impact of insurance is estimated, the results include the combined effects of health insurance on the behaviours of both patients and physicians. To my knowledge, only one field experimental study isolates the causal effects of patient insurance status on provider behaviour: Lu (2014) finds that hospital physicians in Beijing prescribe more inappropriate drugs to insured patients (compared to the uninsured) when they have direct financial incentives to do so (that is, when they gain a profit-share from prescribed drugs). Contrary to findings in this chapter, she finds no evidence of an insurance effect on inappropriate drug prescribing when there are no financial incentives attached.[176] A limitation of Lu (2014) is that only one dimension of care quality is evaluated (inappropriate drug treatment for elevated triglycerides). Moreover, studies that examine the effects of patient insurance on provider behaviour generally compare the effects of insured patients to the uninsured. In such cases, other factors besides the insurance status of the patient itself - such as varying consultation rates and the immediacy of reimbursement - may confound physicians' treatment decisions. In this chapter, different *levels* of insurance cover are

[176] GPs recruited for Part 1 of this study do not have a license to dispense drugs, and therefore do not gain (or lose) financially from their drug prescriptions.

compared (high-insured and low-insured),[177] with identical reimbursement rates, reimbursement procedures, and contract terms between the insurer and provider. This is similar to experimentally varying the patient's co-payment rate, as most famously done in the RAND Health Insurance Experiment (HIE) in the 1970s (Newhouse, 1993). Although the HIE primarily examined demand-side responses to variations in the co-payment rate, the technical quality of care received by participants once they accessed care was also analysed, and no co-pay effect was found (Brook *et al.*, 2006). However, the HIE was not designed for evaluating supply-side responses: the co-payment rate was not randomised across or within providers, so the possibility that patients may choose to see different types of providers under different levels of co-payment cannot be controlled for. Additionally, the study's use of claims data limits its ability to infer the actual quality of care provided, for the reasons discussed previously.

In proposing altruism as one explanation for care quality differences by patient insurance, this chapter also relates to literature characterising physicians as 'altruistic' agents who incorporate patient welfare in their behavioural choices (alongside profit motives) (see Galizzi *et al*. 2015 for a detailed review). Empirical evidence of physician altruism largely comes from laboratory experiments with medical students (Hennig-Schmidt *et al.*, 2011; Godager and Wiesen, 2013; Hennig-Schmidt and Wiesen, 2014; Brosig-Koch *et al.*, 2017; Lagarde and Blaauw, 2017), or analyses of prescription data that consider the likelihood of generic-substitution in response to patients' OOP costs (Hellerstein, 1998; Lundin, 2000; Granlund, 2009; Iizuka, 2012; Crea *et al.*, 2019). The former experimental literature is limited in its applicability to field settings with practicing physicians. The latter rely on administrative data, and hence cannot separate physician altruism from other decision drivers, including patient demands (patient moral hazard). Moreover, these observational studies do not consider physician altruism with respect to care quality. Lu (2014) succeeds in experimentally testing for physician altruism in prescribing decisions (separate from profit or market incentives, or patient demands), and finds no supportive evidence. However, as stated previously, this study too gives limited consideration to physician altruism in effort choices (care quality). Admittedly, unlike in Lu (2014), the present chapter cannot isolate the influence of physicians' altruistic incentives from those of market incentives in explaining variations in physicians' choices by health insurance. However, it evaluates a broader range of care quality indicators, and also tests for heterogeneous insurance effects with respect to an experimental measure of baseline physician altruism.

[177] Further details on the insurance covers used in this study are provided in Section 4.3.1 and Appendix B.3. The insurance covers vary in the level of medical savings available to patients for financing their primary care needs in any given year, which is effectively equivalent to variations in their *de facto* annual co-payment rates

This chapter further relates to literature on provider competition on quality under free patient choice. However, it differs from existing works in a number of ways. Most existing evidence relates to competition in the hospital sector, rather than in primary care; and focus almost exclusively on hospital mortality rates as a measure of care quality (Gaynor, 2006; Barros *et al.*, 2016). Moreover, very few studies acknowledge a potential role for patient observability and information in creating heterogeneity in provider competition across different aspects of care quality and patient types. When prices are fixed and providers are unable to compete on cost, theoretical models predict that providers will instead compete on quality, as long as prices are fixed above marginal cost (Gaynor, 2006). Related evidence is mixed. Studies on the UK NHS and the US Medicare systems show that increased competition under fixed prices leads to lower mortality rates for cardiac patients (Kessler and McClellan, 2000; Kessler and Geppert, 2005; Cooper *et al.*, 2011; Gaynor *et al*., 2016) – measures that are commonly monitored and used as a basis for hospital reimbursement. However, there are also studies showing null or opposite effects on cardiac and other outcomes (Mukamel, Zwanziger and Tomaszewski, 2001; Gowrisankaran and Town, 2003; Colla *et al.*, 2016). In general, theoretical models abstract from patients' limited ability to infer care quality objectively, and the potentially high search and switching costs that can prevent patients from responding to poor quality.[178] Acknowledging this, a few papers have noted that provider competition under fixed prices is likely to be higher for aspects of care quality that are more observable to patients (Dranove and Satterthwaite, 1992; Haas-Wilson, 1994; Propper, Burgess and Gossage, 2008). Two studies on competition and quality in primary care show that increased patient choice among providers with fixed prices improved subjective (patient-reported, and therefore observed) measures of care quality, but had a null or smaller effect on objective and less observable measures of technical quality (Dietrichson *et al*., 2016; Gravelle *et al.*, 2018).

The literature is generally ambiguous on the effects of provider competition on quality when there is also price competition (Gaynor, 2006; Gaynor, Ho and Town, 2015). In this chapter's context, patients' OOP prices are not entirely fixed for both patient insurance types. While prices are largely fixed (and near zero) for high-insured patients given their negligible *de facto* co-payment rate, prices may be reduced for low-insured patients with appropriate effort and care quality outputs (implying scope for some level of price competition for these patients).[179] However,

[178] Some exceptions are Dranove and Satterthwaite (1992), Kranton (2003) and Allard et *al.* (2009)

[179] 'Prices' in this study context refer to out-of-pocket costs for patients. Prices for individual services and treatments are capped by the health insurer at the same level for both patient insurance types. Therefore, 'price competition' in this context refers to competition on patients' out-of-pocket care costs, which may be altered by setting prices below the price cap or through the provider's choice of services and medications. Whilst some theoretical models suggest that price competition will lower care quality (as it increases its relative marginal cost) (Gaynor, 2006), this is likely to hold only in contexts where price-cutting alone reduces provider profit, or where the price elasticity of demand is greater than the quality elasticity. In this study context, however, price and quality competition are complementary: a decrease in total costs for the patient is driven by an increase in appropriate provider effort.

causal evidence on the effects of varying the degree of price competition for patients – through variations in patients' price elasticity of demand (through their level of insurance cover, or otherwise) - on providers' care quality choices is currently lacking. One exception is Volpp *et al.* (2003), who provide quasi-experimental evidence that risk-adjusted AMI mortality rates rose in New Jersey after deregulation of hospital prices, indicating a decline in care quality in response to greater price competition - although no evidence is given that demand elasticity rose and prices actually fell post-deregulation. Interestingly, Haas-Wilson (1994) identifies a trade-off between competition on price and *observable* aspects of quality only, with providers supplying higher observable quality at higher prices (and vice versa). On the contrary, no trade-off is found between price and less observable technical care quality. This chapter develops this literature by allowing for both varying degrees of informational asymmetry in care quality and varying levels of *de facto* price competition for patients (by exogenously varying the patient's insurance cover). In doing so, it provides new evidence on the effect of varying providers' cost (price) competition on their choices of both observable and more hidden aspects of care quality inputs and subsequent outputs.

Lastly, this chapter adds to the growing literature using SPs for evaluating the clinical quality of care (Beullens *et al.*, 1997; Glassman *et al.*, 2000; Rethans *et al.*, 2007) - more recently, also within LMIC settings (Das and Gertler, 2007; Mohanan *et al.*, 2012; Das and Hammer, 2014; Daniels *et al.*, 2017). A closely related work to the present one is Das *et al.*(2016)'s use of SPs to study the influence of market incentives on the care quality of dual-practice physicians in India. The authors find that, for the same clinical case, the same doctors deliver better quality care in their private practices relative to their public ones. While this and other studies shed light on the average quality of care in their respective settings, very few exploit the potential for SPs to be used within field experiments, to evaluate the effects of different patient and situational factors on providers' care quality responses. Even in the few experimental SP studies that examine providers' therapeutic choices, corresponding effects on non-contracted aspects of care quality (including clinical effort and diagnostic accuracy) are rarely evaluated.[180] Building on these works, this chapter presents novel field-experimental evidence on the effects of patient insurance on a richer set of care quality measures.

[180] Some exceptions are Currie, Lin and Zhang (2011), Currie, Lin and Meng (2014), and Daniels *et al.* (2019). However, the first two studies consider only a very limited set of clinical effort indicators and do not consider diagnostic accuracy.

5.3 Experimental Methods and Data

5.3.1 Experimental design

To identify the causal effect of patient insurance at the physician-level, the experiment employed a within-subject, audit study approach, where each participating physician received two incognito visits from SPs with different levels of insurance cover. The SPs were either 'high-insured' (with a high level of medical savings to finance their primary care expenditures) or 'low-insured' patients (with a low level of medical savings).[181] These SPs corresponded to either Patient Role A (low-insured) or Patient Role B (high-insured), as described in Section 4.3.1.2. The within-subject experiment design exposes the same physicians to both types of insurance cover, and hence controls for any physician-level heterogeneity that might confound the estimated insurance effect on the average physician. To ensure that the difference in financial protection (likelihood of OOP expenditure) between the two SP types was credible, the SPs were provided with formal insurance cover from an insurance company that contracts with all participating physicians. Physicians would therefore be aware of the type and level of insurance cover attributable to each SP type. The SPs also indicated the extent of their financial protection in their opening statements to physicians, to emphasise this difference (see Table 4.4).

The order in which the two SPs present to physicians may also confound the estimated insurance effect (generating 'order effects'). For one, the likelihood of physicians detecting a fake patient may be higher with the SP that presents second, as both SPs presented with the same clinical case and script (differing only in their insurance cover) within just a few days of each other. Similarly, there may be learning effects, whereby physicians perform better in treating the second SP due to having encountered a very similar patient just a few days before. To minimise these potential effects, the order in which the two SP types presented to each physician was randomised, and at least a two-day gap was maintained between the two SP visits.[182] It was also possible to control for the visit order in the regression analyses of the following sections, to ensure robustness of the main results to these potential confounders.

Similarly, the audit study approach controls for all patient- and case-specific confounders, and the potential for selective matching of certain types of patients with certain physicians. By fixing the clinical case, it also allows better measurement of the quality of care provided during physician-patient interactions. For clean identification of the insurance effect, it was necessary to

[181] The annual medical savings available to the 'high-insured' patient was over 75% higher than that available to the 'low-insured' one. See Appendix B.3 for a brief overview of the insurance covers of the two SPs.

[182] 87% of sample GPs received their two SP visits within 2-7 days of each other. For most remaining GPs, due to consultation scheduling delays, paired visits were conducted with a 7-14 day gap. The maximum 14-day visit gap was exceeded in one exceptional case, where a sample GP received his paired visits with a 26-day gap (due to being on leave within that period).

have otherwise identical SP visits except for the patient's insurance cover. One concern is that there may be individual fieldworker effects, mediating physicians' observed responses to patient insurance. It was not possible to pool individual fieldworkers and insurance covers in this experiment, as formal insurance cover had to be bought individually for each fieldworker. Fieldworkers were therefore assigned to just one insurance cover and SP role each (high- or low-insured). It is possible that the fieldworkers assigned to one insurance cover may be very different from those assigned to the other. Even if fieldworkers assigned to the two insurance covers were balanced *ex ante*, they may behave differently *ex post* (during their visits) conditional on their insurance assignment. To control for both *ex ante* and *ex post* fieldworker variations across the two SP types seen by the same physician, each fieldworker was first matched with another of the opposite insurance cover based on physical characteristics, age and gender, to form fixed fieldworker-pairs. Each physician was then randomly assigned to visits by a unique fieldworker-pair. It is then possible to control for fieldworker-pair fixed effects in the regression models estimating the main insurance effect.[183] In addition, all fieldworkers were trained extensively to follow a standardised script and visit protocol, to minimise variations in presentation (see Section 4.1.2 for details on SP training). A detailed discussion of the SP methodology and visits procedures used in this experiment is provided in Sections 4.1 and 4.3.1.2.

Although the SP method standardises the presentation of patient characteristics that may confound the insurance effect, one limitation is that physicians' may still *infer* those characteristics based on the patient's insurance cover. As patients would usually opt into higher insurance cover, and only specific types of highly-paid jobs would provide the generous cover held by the high-insured patient (if the patient did not pay for it themselves), the physician may infer certain characteristics of individual patients, such as their socioeconomic status or their health-risk preferences, from the comprehensiveness of their insurance cover. While this cannot be controlled for, I ensure to some extent that any inferences made by physicians on patients' socio-economic status will not be too dissimilar across the two insurance covers (e.g. one rich and one poor), as the low-insurance cover used in this experiment is still a highly comprehensive insurance product and a premium option for average South Africans.

5.3.2 Physician sampling and participation

The sampling and recruitment procedures used in this experiment are detailed in Section 4.3.1.1. Briefly, a total of 112 GPs were recruited into the study, of which 89 were successfully visited by

[183] As each fieldworker was assigned to a specific insurance cover, including both individual fieldworker fixed effects and the insurance variable in the statistical model may lead to a high degree of co-linearity. Fieldworker-pair fixed effects are good alternatives to control for any fieldworker bias in the estimated insurance effect, whilst avoiding the co-linearity issue. In the cluster fixed effects model, the inclusion of physician-level fixed effects effectively controls for fieldworker-pair effects that do not vary within-physician.

both SP insurance types. This resulted in a final study sample of 178 consultation visits. GPs were sampled from a population of non-dispensing physicians in Johannesburg (n=614), of whom 271 were deemed eligible for the study. GPs were eligible if they were contracted into private health insurance schemes and were non-specialist physicians. Individual GPs were then contacted at random for participation in the study. Of the 208 GPs successfully contacted, 112 consented to participation. The ethical requirement to obtain informed consent from all participating GPs means the final study sample is not entirely random. Although GPs were randomly sampled for invitation to the study (from the pool of eligible GPs), some level of selection bias is inevitable in those that consented and those that were successfully visited by SPs. While this does not affect the internal validity of study results, it is likely to limit its external validity. As the care quality of GPs that refused participation is not observed, I am unable to verify the extent of this bias. Table 4.5 (Columns (2) and (3)) compares basic socio-demographic characteristics of GPs in the study sample to those of the non-dispensing GP population in Johannesburg. The sample and sampling population are comparable in the average age of GPs (49 and 50 years, respectively), and the share of GPs located in the wealthier northern suburbs of Sandton and Randburg (43%, compared to 44% in the sampling population). However, the study sample has a slightly higher representation of male GPs (52%, compared to 47%).

Interviews with sample GPs were carried out following the experiment, to collect additional data on GPs' socio-demographic and practice characteristics. These interviews were also used to obtain measures of GPs' personal attributes, including individual altruism (through a dictator game), clinical knowledge (through a short quiz), and beliefs pertaining to the prescription of antibiotics. While all 89 GPs successfully visited in the experiment were invited to interview, 9 GPs refused to participate,[184] and 5 GPs were either not reached or did not confirm an interview time before the end of February 2019 when the interviews were concluded. This resulted in a sub-sample of 75 GPs with whom both the experiment and the interviews were successfully completed. The data collected from interviews inform secondary, sub-sample analyses of interest. The validity of these secondary results could be threatened if GPs that refused to be interviewed were systematically different from GPs that did not. Using care quality data from the SP visits, Appendix C.1 shows that participation in interviews was not meaningfully associated with any care quality outcome of interest. The first panel of Appendix C.1 also shows no meaningful association between GPs' basic socio-demographic characteristics and their likelihood of participation.

[184] The GPs that refused to participate in the interviews mainly cited times pressures, as 30 minutes was requested during their work hours to complete the interview.

5.3.3 Sample characteristics

Table 5.1 provides summary statistics on study sample characteristics. As stated earlier, the average age of sample GPs is 49 years, 52% are male, and approximately 43% have practices located in the wealthier northern suburbs. Moreover, the average GP's practice has 7.5 other competing GPs located within a kilometre radius. Table 5.1 also summarises additional characteristics for the interviewed sub-sample of 75 GPs. A large majority of these GPs are white (63%) and only 16% are black-African (or mixed race). GPs of Asian or Indian origin are the second largest racial grouping in the sample (21%). A majority of sample GPs work in group practices (55%), and they report seeing an average 22 patients a day (based on the previous week's patient load). For a measure of GP altruism, the share of a R300 cash endowment that each GP chose to donate to a patient charity during the interview's dictator game was calculated (see Section 4.2.3), and "altruistic" (or "high-altruism") GPs were categorised as those that donated a share at or above the sample median. Notably, a large majority of GPs (63%) chose to donate the full amount. Hence, only the GPs that donated the full R300 were categorised as "altruistic."

As a measure of GPs' clinical knowledge, the average score in the interview's knowledge quiz is shown, indicating diagnostic and therapeutic knowledge on a number of respiratory infections (see Section 4.2.3 for details on the knowledge quiz contents). On average, GPs scored approximately 7 out of a maximum 11 points. On the questions that specifically covered the clinical case presented by SPs in this experiment (an uncomplicated case of acute bronchitis), only 19% of GPs were able to accurately name the case as "acute bronchitis" when presented in vignette-form. However, a large majority (83% of GPs) correctly identified its likely viral cause, which is perhaps more important in determining appropriate treatment (non-prescription of antibiotics). Indeed, this is reflected in a relatively low perceived effectiveness of antibiotics for the vignette case: on average, GPs felt there was only a 25% likelihood that the case would resolve *more* quickly with antibiotic treatment. Contrarily, in a question that specifically asked for the recommended treatment for a *named* uncomplicated acute bronchitis case (revealing the diagnosis), 75% of GPs chose antibiotic treatment. This discrepancy suggests there may be some confusion among GPs on what "acute bronchitis" is, or how it presents.[185] [186] GPs seem relatively

[185] This discrepancy may also present if GPs disagreed with treatment guidelines (as known to them). They may choose antibiotic treatment as the guideline recommendation for acute bronchitis (when the diagnosis is named) because they have poor knowledge of treatment guidelines (even if they understand what "acute bronchitis" is, when named). At the same time, they may report a low perceived effectiveness of antibiotics for the case simply because they disagree with (their incorrect knowledge of) the guidelines.

[186] There are some limitations to the therapeutic knowledge questions in the quiz. Firstly, when asking GPs to identify recommended treatments for specific cases, the quiz named the case (e.g. "acute bronchitis" in an otherwise healthy adult) but did not provide a description of the case presentation (as in the diagnostic knowledge questions, where the case was described in vignette-form). This makes it difficult to understand if GPs were able to correctly link named diagnoses (e.g. "acute bronchitis") to their actual clinical presentations, and therefore to understand if they chose recommended treatments on the basis of known case presentations (which is ultimately what would matter for predicting choices in clinical practice). Secondly, some level of reporting error in the therapeutic

unable to characterise the (vignette-form) case presentation as "acute bronchitis" or identify the recommended treatment for a case so-named. However, this does not affect their ability to identify the presentation's likely viral cause, nor prevent them from holding relatively accurate beliefs about the effectiveness of antibiotics for the case.

Table 5.1. Summary statistics – physician-level characteristics

	Mean	SD
a. Socio-demographic characteristics		
Age*	49	(11.27)
Male*	0.52	(0.50)
Black/Coloured	0.16	(0.33)
Asian/Indian	0.21	(0.41)
White	0.63	(0.49)
b. Personal characteristics		
Altruism: is (highly) altruistic	0.63	(0.49)
Knowledge: correctly named clinical case in vignette	0.19	(0.39)
Knowledge: correctly identified viral cause of clinical case in vignette	0.83	(0.38)
Knowledge: correctly identified treatment guidelines for acute bronchitis	0.25	(0.44)
Knowledge: overall clinical knowledge score (max. 11)	7.35	(1.56)
Knowledge: AMR knowledge score (max. 5)	3.80	(1.02)
Antibiotic beliefs: efficacy for clinical case in vignette[1]	25.16	(30.05)
Antibiotic beliefs: prescribing norms for clinical case in vignette[2]	62	(26.51)
Antibiotic beliefs: patient expectations for clinical case in vignette[3]	53.73	31.74
c. Practice characteristics		
Is in a northern suburb*	0.43	(0.50)
Local GP density (no. competing GPs within 1 km radius)*	7.45	(5.61)
Average daily patient load (previous week)	22.15	(11.75)
Is a group practice	0.55	(0.50)
Observations	75	

Notes: Most data presented here are taken from the GP interviews conducted with participants in Part 1 of this study (see Section 4.2.3 and Appendix B.4). * The average age, share of male GPs, share of GPs located in the northern suburbs of Sandton and Randburg, and the measure of local GP density were calculated using data from the online medical practitioner database *Medpages*, for the full sample of 89 GPs. These four measures therefore have 89 observations (not 75). See Section 4.2.3 for further details on how certain listed GP characteristics were measured. [1] GPs were firstly asked what was the likelihood (between 0 and 100, with 100 indicating certainty) that the patient in the vignette case would recover faster with antibiotics (rather than without). [2] They were then were asked how likely it was that other GPs would prescribe antibiotics to the patient in the vignette case. [3] Lastly, GPs were asked how likely it was that the patient would go to another GP the next time they needed medical attention if they did not receive antibiotic for the case.

Moreover, GPs actual antibiotic prescribing practices may be determined by more than just their clinical knowledge (or beliefs) on indicated treatments. Firstly, their perceptions about the risks of inappropriate antibiotic prescribing, in the context of growing anti-microbial resistance (AMR), may play a role. On this, GPs were generally well informed about AMR and its

knowledge quiz results is possible: these therapeutic knowledge questions were not incentivised monetarily (unlike the diagnostic knowledge questions that preceded them in the quiz), which may distort physicians' motivations in correctly reporting knowledge.

challenges, with the average GP scoring approximately 4 out of 5 points on a short knowledge quiz specifically covering AMR. Secondly, perceived social norms can also play a role. The quiz explored how GPs perceived peer norms and patient expectations in prescribing antibiotics for the vignette case, and revealed relatively *pro*-antibiotic perceived norms: the average GP felt there was a 62% likelihood that other GPs would prescribe antibiotics for the vignette case (of uncomplicated acute bronchitis), and a 54% likelihood that patients would seek care elsewhere next time they fall ill if they were not given antibiotics for the same case.

5.3.4 Care quality outcomes

As described in Section 4.2.1, this study evaluates three aspects of care quality during physician-patient interactions, covering both physicians' (observed) care inputs and subsequent care quality outputs. Table 5.2 shows summary statistics for the care quality outcomes measured in SP visits.

First, observed measures of physicians' clinical effort are evaluated, including the consultation duration, and the raw proportion of the essential and recommended history-taking and examination checklist completed.[187] This checklist was compiled by a panel of clinical experts, and designed to enable physicians to differentiate respiratory conditions of varying severity and overlapping symptoms. The average SP consultation lasted 10 minutes. In this time, physicians completed an average 49% of the essential and recommended care checklist, including 67% of all essential examinations and 45% of all essential history-taking.

Table 5.2. Summary statistics – care quality outcomes

	Mean	SD
a. Diagnostic effort		
Consultation length, in mins.	10.46	(5.48)
Share of history-taking and examination checklist completed	0.49	(0.13)
Share of essential examinations completed	0.67	(0.20)
Share of essential history-taking completed	0.45	(0.17)
IRT score	0.00	(0.89)
b. Diagnosis		
Gave a correct or partially correct diagnosis	0.53	(0.50)
c. Treatment		
Prescribed an antibiotic	0.57	(0.50)
Prescribed other inappropriate drugs	0.94	(0.23)
Gave appropriate follow-up advice	0.61	(0.49)
Observations	178	

Notes: All consultation outcomes were measured using data collected from the post-consultation SP debriefing questionnaire (see Appendix B.4 for an overview of the questionnaire content). For further details on how these outcomes were measured, refer to Section 4.2.1.

[187] See Section 4.2.1 for details on how all care quality outcomes were constructed.

A composite index score using IRT analysis is also computed, weighting completed checklist items by their relative value in differentiating among physicians of different levels and quality of observed effort. See Section 4.2.1 and Appendix B.6 for a brief overview of IRT analysis and its underlying assumptions.[188] Appendix C.2 lists the checklist items included in this chapter's analysis and their corresponding results.[189] The listed parameter estimates are used to weight the checklist items completed and predict a unique IRT score for each consultation. Figure 5.1 shows the distribution of this IRT score in the study sample. By construction, the distribution of the score is centred at a mean of 0, and individual IRT score values are interpreted as standard deviations (SD) from the mean. All IRT scores in this sample fall within approximately 2.25 SD of the mean.

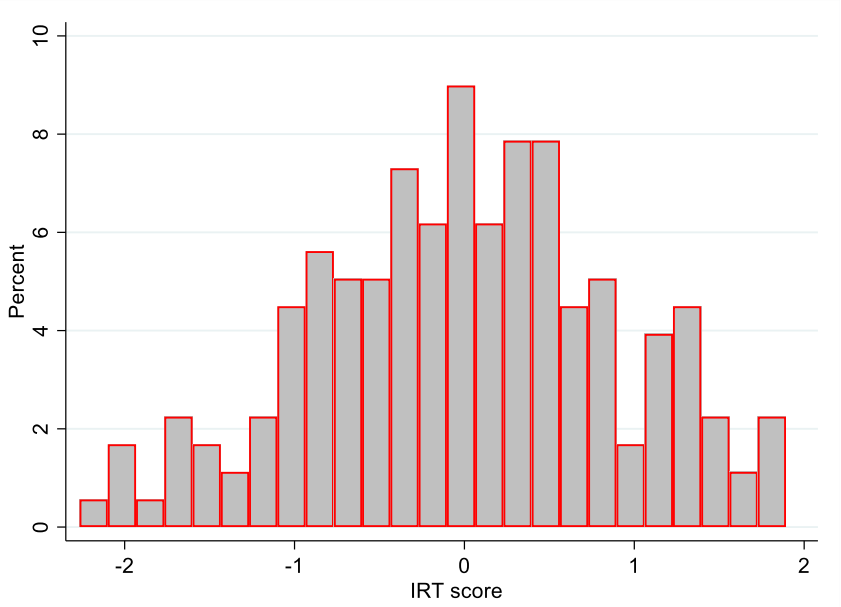


Figure 5.1. Sample distribution of predicted IRT scores

[188] One assumption necessary for IRT analysis is a uni-dimensional latent space. In other words, responses to all items included in the IRT score must reflect only one latent trait: provider effort. To validate this, the eigenvalues from exploratory factor analysis of items included in the score were checked, and the first eigenvalue is almost twice as large as the second one (while the differences between others are significantly smaller). This supports a single factor solution, and the assumption of uni-dimensionality.

[189] Not all items from the essential and recommended care checklists were included in the IRT analysis, for two reasons. First, some items were completed only in a few consultations, and those completed in less than 5% of cases were excluded following Das and Hammer (2005), to ensure convergence of the maximum likelihood function used in the analysis. In particular, history-taking questions about fast breathing (tachypnea) or palpitations were asked in only 2% of consultations, and therefore excluded from the IRT score. Second, the validity of the IRT analysis requires that individual items are conditionally independent of each other. However, some items in the checklist had clear overlap. For instance, whether a physician asks the history-taking question on "coughing up blood" may depend on whether he had already asked the patient about a "productive cough", and the question on "any allergies" may preclude the specific question on "allergy to penicillin." In such cases, the overlapping items were combined into a single item that was indicated as completed if any of the combined items had been completed.

Following physicians' effort inputs, the second aspect of care quality evaluated is their diagnostic output: whether the diagnosis given to the patient was correct or partially-correct.[190] A diagnosis was given in almost all consultations (> 99%); and physicians gave a correct of partially-correct diagnosis in 53% of these cases.

The final aspect of care quality evaluated is the appropriateness of physicians' therapeutic outputs. One indicator of this is whether any unnecessary antibiotic was prescribed for the clinical case. Comparable to other studies in China and India (Das and Hammer, 2007; Currie *et al.,* 2014), I find a high average rate of inappropriate antibiotic prescribing for an uncomplicated viral infection: an antibiotic was prescribed in 57% of all consultations. As shown in Appendix D.1, the most commonly prescribed antibiotic groups were broad-spectrum penicillins (or penicillin-clavulanates) and macrolides. Note that this rate of inappropriate prescribing is much higher than that predicted by GPs' beliefs about the likely effectiveness of antibiotics for this clinical case when presented in vignette-form during interviews (see Table 5.1): the average GP believed that the patient had only a 25% likelihood of recovering quicker with antibiotics. These results appear in line with common findings of a 'know-do gap' in the literature (Das and Gertler, 2007; Leonard and Masatu, 2010; Gertler and Vermeersch, 2013), where physicians' care quality performance in practice is shown to be lower than that indicated by measures of their knowledge. One explanation for this discrepancy, as noted earlier, is GPs' perceptions of relatively strong prescribing norms and patient expectations for antibiotics for the same case (when presented in vignette-form).[191]

Another indicator evaluated is whether any other inappropriate drugs - including steroids, nasal decongestants, bronchodilators, antihistamines or vitamins - were prescribed. The results show that this occurred in a very large proportion of consultations (94%). Appendix D.1 shows that the prescription of steroids was particularly common, occurring in 60% of all consultations. These rates of inappropriate treatment are also much higher than the rate of incorrect diagnoses, suggesting that an accurate (pronounced) diagnosis is unlikely to be sufficient for correct

[190] See Appendix B.1 for a list of diagnoses considered correct or partially correct. Further details on how the diagnosis measures of care quality were constructed are provided in Section 4.2.1. A strictly 'correct' diagnosis required the doctor to communicate the diagnosis in exact technical terms to the patient (i.e. "acute bronchitis" or "bronchitis"). The limitation here is that what the physician tells the patient regarding their diagnosis may be simplified (for instance, given expected limitations to the patient's medical knowledge). Hence, a 'partially-correct' diagnosis allowed for some generality in the communicated diagnosis (such as "chest cold") and for select similar diagnoses (such as "tracheitis" or "post nasal drip"). This chapter accordingly combines both outcomes in its analysis of diagnostic accuracy. Such considerations were taken into account also in coding specific diagnoses, particularly with regards to local colloquial terminology. One example was in coding a "flu" diagnosis. In strict medical terms, this is incorrect. However, acute bronchitis is usually preceded by flu-like symptoms; and discussions with local medical practitioners highlighted that, in the South African context, doctors often do not make a distinction between "flu" and "cold" in communicating diagnoses to patients. Accordingly, no distinction was made in coding "flu" and "cold," and both were coded as partially-correct *if* they were combined with a bronchitis or similar diagnosis (since a cold preceded the acute bronchitis according to the case history).

[191] This mirrors qualitative findings elsewhere that anticipated patient expectations do influence providers' reported treatment choices – particularly under market competition - even where they know those choices are clinically inappropriate (Das and Hammer, 2007).

treatment. This may demonstrate physicians' uncertainty around their pronounced diagnoses, which they compensate for by prescribing different, non-clinically indicated drugs. As discussed before, it may also indicate a 'know-do' gap influenced by other non-clinical factors in treatment choices – such as norms or habits, and anticipated patient demands.[192] Lastly, the appropriateness of any advice to the patient to come back for a repeat consultation is evaluated: in 61% of consultations, physicians appropriately advised patients to return for a repeat consultation if symptoms worsen.

5.4 Empirical Framework

The primary purpose of this chapter is to estimate differences in physicians' care quality performance for low-insured and high-insured patients. To identify the impact of patient insurance, I must estimate the following:

$$Q_{ij} = \beta_0 + \beta_1 Insurance_{ij} + \varepsilon_{ij} \quad (1)$$

where each quality measure Q from consultation i with physician j is regressed on an indicator $Insurance_{ij}$ which equals 1 if the patient in consultation i is high-insured (and 0 otherwise).[193] β_1 measures the insurance effect of interest. It is unnecessary to control for other physician-level covariates to obtain unbiased estimates of β_1: as the same physicians are observed under both levels of $Insurance$, these physician-level covariates are balanced across the two levels. Estimating this equation with pooled OLS requires the assumption that the error terms ε_{ij} are independent and identically-distributed (constant variance). However, the within-subject study design, where each physician is observed twice (under each patient insurance cover), means observations of the same physician (and corresponding error terms) are likely to be correlated. This potential clustering of data at the physician level must be accounted for, to avoid under-estimating the likelihood of type-I errors (false positives) and allow robust inference. Four methods are commonly proposed for handling cluster-correlated data, and the optimal approach is largely study-dependent (and to some extent, subject to disciplinary preferences (McNeish and Kelley, 2019)). These include a paired-sample t-test ('response simplification'); an OLS model with a (physician-level) cluster-robust variance estimator; a (physician-level) cluster fixed effects model; and a linear mixed effects model (LMEM), where individual physicians are treated as random (rather than fixed) effects (Cameron and Miller, 2015).

---

[192] These findings on the relatively low diagnostic accuracy and high prevalence of inappropriate treatment on average align with *Hypothesis 3* in Chapter 3, which posits that appropriate effort will in general be under-supplied (and related inappropriate treatment oversupplied) relative to what is in any patient's best interests (due to the contractibility problem with respect to appropriate effort).

[193] For ease of interpretation, only linear regression models are used to obtain the main results, although robustness checks with logistic regression models are later carried out for all binary outcomes.

Appendix C.3 gives an overview of these four approaches, and their relative strengths and limitations. In general, the three linear regression approaches (OLS with cluster-robust standard errors, cluster fixed effects, and LMEM) are preferred to the paired-sample t-test, given their ability to control for potential confounders of the insurance effect - including any fieldworker-pair effects.[194] [195] I estimate the main results in the following section using the cluster fixed effects model,[196] with and without a cluster-robust variance estimator. Subsequent sub-sample analyses are conducted using linear OLS with the cluster-robust variance estimator (and including fieldworker-pair fixed effects), as the inclusion of physician fixed effects prevents estimation of physician-level covariate effects. To check the sensitivity of main results to the estimation approach, corresponding results from a paired-sample t-test and the LMEM with random intercept and slope are shown in Appendices C.5 and C.6.[197]

5.5 Results

5.5.1 Effects of patient insurance cover on care quality outcomes

5.5.1.1 Observed clinical effort

Table 5.3 shows estimates of the insurance effect on observed measures of clinical effort. Panel A presents results from a fixed effects model with normal standard errors, whilst Panel B shows estimates from the same model with the addition of a cluster-robust variance estimator. All results are shown to be robust across the two specifications.

Column (1) shows that the average consultation time for low-insured patients was 10.34 minutes, and this is not significantly different for high-insured ones. Similarly, there is no difference in the raw proportion of essential and recommended care checklist items completed by insurance cover (Column (2)): for both SP types, physicians completed just under 50% of all checklist items.

[194] The cluster fixed effects model is preferred to the LMEM due to its ease of specification, and the need for fewer assumptions for valid inference. Moreover, it automatically controls for all fixed physician-level confounders of the insurance effect without need for explicit specification of these confounders as model covariates. These include fieldworker-pair effects, as fieldworker-pairs are fixed for each physician. The use of a cluster-robust variance estimator is also feasible with this model, given the relatively large number of clusters (89 sample physicians).

[195] Fieldworker pair effects must be explicitly controlled for in the LMEM and marginal model through the inclusion of fieldworker-pair fixed effects. However, they will be automatically controlled for in the cluster fixed effects model in this study context.

[196] The cluster fixed effects model is estimated through the mean-differencing method (applying linear OLS to the mean-differenced version of (2)), as suggested by Cameron and Miller (2010). The model is also estimated including the cluster-robust variance estimator, to account for any residual within-cluster correlation (Arellano, 1987).

[197] The LMEM specification includes the cluster-robust variance estimator (to correct for potential bias in the variance estimates of β_1 from misspecification of the covariance structure (Liang and Zeger, 1986)) and fieldworker-pair fixed effects. An independent variance-covariance structure is assumed, as with the default case for the *mixed* command in STATA 15.

Table 5.3. Effect of insurance cover on diagnostic effort

	(1) Consultation length	(2) Checklist completion	(3) Essential exams	(4) Essential history	(5) IRT score
Panel A: provider-level fixed effects					
High-insured	0.24	0.02	0.05**	0.03	0.22**
	(0.80)	(0.02)	(0.02)	(0.02)	(0.11)
Obs.	178	178	178	178	178
R-squared	0.00	0.01	0.05	0.02	0.05
Panel B: provider-level fixed effects & cluster-robust standard errors					
High-insured	0.24	0.02	0.05**	0.03	0.22**
	(0.81)	(0.02)	(0.03)	(0.02)	(0.11)
Obs.	178	178	178	178	178
R-squared	0.00	0.01	0.03	0.02	0.04
Mean of low-insured	10.34	0.48	0.65	0.43	-0.11

Notes: Standard errors are in parenthesis. Panel A presents estimates from a fixed effects (mean-differenced) model estimated with linear OLS model. Panel B shows results from a fixed effects model, as in Panel A, with the addition of a cluster-robust variance estimator. *** p<0.01, ** p<0.05, * p<0.1

When evaluated on the IRT score however, physicians scored 0.22 standard deviations (SD) higher when treating high-insured patients compared to low-insured ones (Column (5)), indicating a relatively small positive effect of insurance cover on clinical effort. This effect is still substantial in comparison to those reported for interventions considered to be successful in the literature: Gertler and Vermeersch (2013), for example, find that the introduction of performance pay in Rwanda increased a standardised clinical effort score by 0.13 SD. The effect of patients' insurance cover is found here to be almost 70% larger.[198] Figure 5.2 displays the cumulative distribution functions of the IRT scores for high- and low-insured patients. The two functions are shown to diverge noticeably for IRT scores above the mean: the distribution for high-insured patients dominates that for low-insured patients for all IRT scores in this range.

Isolating only the essential examination items from the full care checklist,[199] Column (3) of Table 5.3 shows that physicians completed a slightly higher share of essential examinations for high-insured patients (70% compared to 65% for low-insured), although the magnitude of the effect is very small and translates into just 0.3 more examinations per patient in absolute terms. No significant differences are detected in essential history-taking.[200]

[198] While consultation time is not found to differ by patient insurance cover, it is still positively correlated with the IRT score (although to a small extent): a one standard deviation increase in consultation time is associated with an increase of 0.06 standard deviations in the IRT score (see Table Column (1) of Appendix C.7).

[199] 'Essential' history-taking and examinations were those categorised by a panel of clinical experts to be crucial among all checklist items for enabling physicians to rule out more severe respiratory illnesses with overlapping symptoms to those of the clinical case in this study.

[200] Looking at individual checklist items, Appendix C.4 shows that physicians were significantly more likely to perform a number of essential history-taking and examination items for high-insured patients compared to low-insured ones. For instance, in conducting essential history-taking on possible night sweats or the patient's TB history, physicians were almost twice as likely to successfully

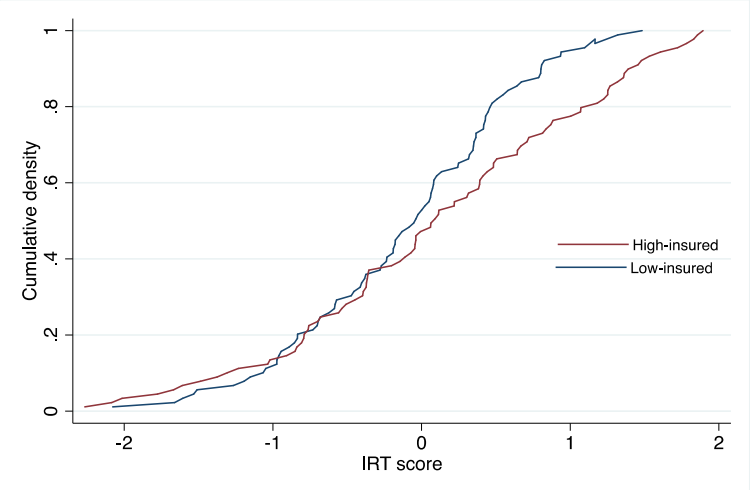


Figure 5.2. Cumulative distribution of checklist completion, by patient insurance type

5.5.1.2 Diagnosis & Treatment

Table 5.4 presents estimates of the insurance effect on diagnosis and treatment outcomes, and follows the same format as Table 5.3. A diagnosis was communicated to patients in almost all consultations in this experiment (177 of 178 consultations). This contrasts with Das *et al.* (2016)'s SP audit study of private providers in India, where only 43% of providers offered a diagnosis to patients presenting with unstable angina or asthma. Conditional on pronouncing a diagnosis, however, physicians' diagnostic accuracy was significantly poorer for high-insured patients. Column (1) of Table 5.4 reports that physicians offered a correct or partially-correct diagnosis to 46% of high-insured patients, compared to 60% of low-insured ones - indicating a 23% lower likelihood of an accurate diagnosis for high-insured patients.

Columns (2)-(4) of Table 5.4 report the estimated insurance effect on treatment outcomes. The main treatment outcome of interest in this study is the prescription of antibiotics. As stated previously, antibiotics are not recommended for the uncomplicated acute bronchitis case presented in this experiment; and yet, an antibiotic was prescribed in 57% of all SP consultations. Moreover, Column (2) indicates that high-insured patients were 24% more likely to receive inappropriate antibiotic treatment compared to low-insured ones (63% vs. 51%). No significant

complete these items with high-insured patients. A notable exception is asking about a penicillin allergy, which is essential when prescribing penicillin-based antibiotics: physicians were 40% less likely to ask about this with high-insured patients when prescribing an antibiotic (despite an equivalent prevalence of penicillin among prescribed antibiotics for both insurance types), indicating poorer safety in prescribing practices for these patients – see Section 5.5.1.2.

differences are detected in the likelihood of prescribing other inappropriate drugs (including steroids, bronchodilators, antihistamines, etc.) by patient insurance type (see Column (3)).

Table 5.4. Effect of insurance cover on diagnoses & treatment outputs

	(1) Correct / partially-correct diagnosis	(2) Antibiotic	(3) Other inappropriate drugs	(4) Appropriate follow-up advice
Panel A: provider-level fixed effects				
High-insured	-0.14*	0.12**	0.04	0.21***
	(0.07)	(0.05)	(0.03)	(0.07)
Obs.	177	178	178	178
R-squared	0.04	0.05	0.03	0.10
Panel B: provider-level fixed effects & cluster-robust standard errors				
High-insured	-0.14*	0.12**	0.04	0.21***
	(0.07)	(0.06)	(0.03)	(0.07)
Obs.	177	178	178	178
R-squared	0.04	0.05	0.03	0.10
Mean of low-insured	0.60	0.51	0.92	0.51

Notes: Standard errors are in parenthesis. Panel A presents estimates from a fixed effects (mean-differenced) model. Panel B shows results from a fixed effects model, as in Panel A, with the addition of a cluster-robust variance estimator.
*** p<0.01, ** p<0.05, * p<0.1

The higher likelihood of antibiotic treatment for high-insured patients coincides with a significantly lower likelihood of being asked about any penicillin allergies by the prescribing physician (see Appendix C.4). This is despite penicillin and penicillin-clavulanates being the most frequently prescribed antibiotic groups (see Appendix D.1) and being no less prevalent in prescriptions for high-insured patients compared to low-insured ones. Conditional on being prescribed an antibiotic, high-insured patients were almost 40% less likely to be asked about penicillin allergies than low-insured ones.[201] This suggests that high-insured patients are more susceptible to both inappropriate antibiotic treatment *and* poorer safety in prescribing practices.

Appendix C.7 indicates that pronouncing a correct or partially-correct diagnosis cannot predict the likelihood of antibiotic prescribing (see Columns (5) and (6)), echoing results from the knowledge quiz in the provider interviews.[202] However, checklist completion has a negative

[201] Among physicians that prescribed an antibiotic to both insurance types, 37% asked high-insured patients about a penicillin allergy, compared to 61% that asked low-insured patients the same question.

[202] In the quiz question that asked about recommended treatment for an uncomplicated case of acute bronchitis, physicians were explicitly told what the diagnosis was, and yet 75% chose inappropriate antibiotic treatment for the case. As stated previously in Section 5.3.3, this may be due to confusion among GPs about the actual presentation of "acute bronchitis" (the quiz question did not provide a description of the case presentation, alongside the diagnosis name). Alternatively, this may lend support to two arguments in Chapter 3: *i)* that, in addition to the (diagnostic) clinical effort necessary to arrive at a correct diagnosis, physicians must exert further therapeutic effort to identify correct, clinically-indicated treatments for that diagnosis; or *ii)* that there are alternative drivers

association: a one SD increase in the IRT score is associated with a 13 percentage point decrease in the likelihood of antibiotic prescribing. This suggests that the slightly higher IRT score for high-insured patients may be attenuating the difference in antibiotic prescribing by insurance type; although, the IRT score difference (0.22 SD) is perhaps too small to fully eliminate the difference in inappropriate treatment.[203]

Lastly, physicians were significantly more likely to give appropriate advice on return visits (i.e. to advise the patient to return for a repeat consultation should the symptoms worsen) to high-insured patients: high-insured patients received this advice in 72% of consultations, compared to 51% for low-insured patients (Column (3) of Table 5.4). Two contextual observations should be noted here. First, while such advice may be beneficial in reassuring the patient, it is by no means necessary given the uncomplicated, self-limiting nature of the clinical case. Viral acute bronchitis in an otherwise healthy young patient should resolve on its own without any treatment. A thorough examination and accurate diagnosis (as well as a physician's confidence in their diagnosis) should therefore negate the need to give such advice. Second, there are clear fee-for-service (FFS) incentives for physicians in recommending repeat consultations. Therefore, communicating such advice may be viewed at least partly in terms of supplier-induced demand (discussed further in Chapter 6), rather than as purely care quality.

5.5.2 Heterogeneous effects

The differences in care quality outcomes by the level of patient insurance are hypothesised to be driven by corresponding differences in physicians' altruistic and competitive incentives for the two patient types. A related question is whether the extent of baseline physician altruism or competition can mediate the insurance effect. The predictions in Chapter 3 are ambiguous on this, as a number of further assumptions would be needed to draw clear predictions on interaction effects. Here, I formally test for interactions between the effects of patient insurance and those of baseline physician altruism and competition, respectively. The measure of altruism in this study was collected for the sub-sample of 75 GPs who were interviewed following the SP visits (see Section 4.3.1.4).[204] The two measures of competition employed here were collected for the full

---

of treatment choices that are unrelated to appropriate effort (including market physicians' sensitivity to perceived patient expectations, perhaps encouraged by prescribing norms).

[203] That GPs are still more likely to prescribe unnecessary antibiotics to high-insured patients (despite the higher IRT score with these patients, and the positive association between IRT scores and antibiotic prescribing) again indicates the presence of additional treatment drivers that are not captured in measured clinical effort.

[204] 14 of the 89 GPs that were successfully visited by SPs were not interviewed, either because they refused to participate due to time pressures or because they could not be reached over the interviewing period. While this non-participation does not threaten the validity of the main results (as the interviews were used only to collect additional data for secondary and sub-sample analyses), it may affect the external validity of secondary analyses if GPs that were interviewed were significantly different from those that were not. However, Appendix C.1 shows that the likelihood of non-participation was not meaningfully associated with any of the care quality outcomes

sample of 89 GPs from the online *Medpages* database. Section 4.2.3 details how these measures were constructed, and Table 5.1 provides summary statistics for each measure (among other measures of GP characteristics). The two locational GP practice characteristics in Table 5.1 are taken as indicators of local GP competition (see Section 4.2.3 for further discussion on this). In addition to the commonly used competition measure of local competitor density, GP practice location in the wealthiest Johannesburg suburbs (the northern suburbs of Sandton and Randburg) is considered another potential indicator of low price (cost) competition, given the likely lower price sensitivity of the local patient population in these suburbs.

Results on interaction effects are presented in Appendices C.9 and C.10. There is some evidence of a mediating effect of physician competition on the impact of insurance on inappropriate prescribing. Appendix C.10 indicates that physicians with a relatively low density of competing physicians in their locality had a significantly larger (positive) insurance effect on the likelihood of antibiotic prescribing. Physicians in the relatively wealthier northern suburbs of Johannesburg were also significantly more likely to prescribe any other inappropriate drugs to high-insured patients compared to low-insured ones.

No significant interaction effects are detected on other care quality outcomes. However, I am unable to rule out the presence of any such effects due to the small sample size (150 observations), and the lack of sufficient power for reliably detecting small-moderate sized interaction effects. Two further limitations to this subsample analyses should be noted. Firstly, the lack of significant heterogeneous effects by physician altruism may be due to limited variation in the degree of baseline altruism in the study sample. In the dictator game during GP interviews (see Appendix B.4), most physicians (63% of the sample) donated their whole cash endowment to a patient charity (and were classified as 'high-altruism' GPs), and only 12% donated nothing. Secondly, there may be some level of measurement error in the indicators of GPs' competitive pressures. Chapter 3 predicts that higher business elasticity (demand response) should affect physicians' treatment quantity and cost decisions. However, the anticipated direction of patients' demand response with respect to different outcomes (in particular, the inappropriate prescription of antibiotics) - and the extent to which its magnitude actually varies with the measures of competition used here - is unclear. The direction and magnitude of demand responses are likely determined by a number of factors, including the corresponding outcome's alignment with patient welfare and preferences, the availability of well-known competitors in the market (alternative sources of care), and patients' ability to recognise care that deviates from their preferences.

---

measured in this study. Appendix C.1 also shows no notable associations between physicians' basic socio-demographic characteristics and their likelihood of attrition. It is therefore unlikely that attrition is a major concern for the external validity of secondary results.

The GP density measure, for instance, indicates the number of alternative sources of care in GPs' localities. Therefore, a higher GP density may be expected to increase demand elasticity. However, the extent to which competing alternatives are *well-known* to patients is unclear. Indeed, Satterthwaite (1979) argues that a larger number of alternatives in the market can actually increase search costs for the patient, and reduce demand elasticity and patient switching. In addition, the direction of any anticipated demand response can vary across different outcomes or patient types. For instance, whether higher competitive pressure on GPs results in more or less inappropriate prescribing for specific patient types would depend on whether GPs anticipate those patients to actually *want* such inappropriate care. In this case, the extent to which a higher GP density captures greater competitive pressure, and how GPs may be expected to respond to such pressure, can vary. This potential for measurement error must be noted in interpreting the evidence on heterogeneous effects by measures of baseline GP competition.[205]

5.5.3 Robustness checks

As outlined in Section 5.4, a cluster fixed effects model (with and without a cluster-robust variance estimator) was employed for estimating the main results, given its relative ease of specification and interpretation. Appendices C.5 and C.6 show that all results are robust to a simple paired-sample t-test and an alternative LMEM specification with a random intercept and slope per physician.

Appendix C.8 presents results from a number of other robustness checks. Potential SP detection during the consultations in this study was a risk, as physicians were informed at recruitment that they would receive 2-4 unannounced SP visits over a period of 6 months. An SP detection survey was therefore carried out with all participating physicians once the SP visits were completed. Section 4.3.1.3 provides further details on the survey and how potential SP detections were categorised. While actual detections cannot be verified (as none of the physicians communicated any suspicions to the SPs during their consultations), two physicians were classified as having a reasonably high likelihood of valid detection based on their responses in the detection survey. Table Column (1) of Appendix C.8 checks the robustness of all main results to the exclusion of visit observations for these two physicians. The estimated results maintain their significance reasonably well for all outcomes.[206]

[205] An additional limitation of these locational measures of GP competition, as noted in Section 4.2.3, is that GPs' practice location decisions are endogenous.

[206] There are marginal changes in the significance of estimated results for the share of essential examinations completed and the IRT score. There results nevertheless maintain p-values < 0.07.

Although the order of the two SP visits was randomised for each physician, Column (2) explicitly controls for the visit order to address any residual order effects driving the main results. All results are shown to be robust to this control. While all SP visits were to be paid for through health insurance, and physicians were expected to bill the costs to the health insurer directly, in 38 consultations the SP was asked to pay the physician in cash and claim back from the insurer themselves. Cash payment is more immediate than insurance claim reimbursement, and this may generate additional treatment incentives for the physician independent of the insurance status of the patient. To ensure that potential cash effects do not confound the estimated insurance effect, the specification in Column (3) explicitly controls for cash consultations. Again, all results are found to be reasonably robust. Lastly, Column (4) shows the robustness of main results for continuous outcomes to the exclusion of extreme values (defined as observed values that are above or below 3 SD of the sample mean).

5.6 Conclusions and Discussion

In utilising an audit study approach and a within-physician experiment design, this chapter provides novel field-experimental evidence on the impact of insurance cover on physicians' care quality choices, controlling for both physician- and patient-level heterogeneity. It finds that physicians exert more observable effort with high-insured patients (in terms of history-taking, examination and communication), but this does not translate into higher quality care outputs: physicians are 23% less likely to give a correct or partially-correct diagnosis and 24% more likely to give inappropriate antibiotic treatment to high-insured patients compared to low-insured ones. The higher rate of antibiotic prescribing also coincides with lower safety in prescribing practices for high-insured patients: physicians are 40% less likely to ask these patients about a penicillin allergy when prescribing an antibiotic.[207] This is despite penicillin being the most frequently prescribed type of antibiotic in this study, and despite no differences in the rates of penicillin-based (versus other) antibiotics prescribed by insurance group.

These findings are generally supportive of the hypotheses tested in this chapter; that physicians prioritise observable effort (patient experience) for high-insured patients who are less cost-sensitive, but this is not always aligned with *appropriate* effort that determines the technical quality of care outputs (including diagnoses and treatments). The difference in antibiotic prescribing is particularly striking, as it occurred despite the absence of any differences in patient demand, patient or provider knowledge, or financial incentives attached to drug prescribing -

[207] Among the sample of GPs that prescribed an antibiotic to both SP types, 37% asked the high-insured SP about penicillin allergies, compared to 61% that asked the low-insured SP.

common factors found to influence physicians' prescribing behaviour in the literature (Currie *et al.*, 2014; Lu, 2014). The likelihood of antibiotic prescribing is not significantly associated with the accuracy of pronounced diagnoses in this study either; suggesting that diagnostic accuracy is also insufficient for appropriate treatment. One explanation for this is that knowing the viral or bacterial cause of a disease is perhaps more important in the decision to prescribe an antibiotic than accurately naming or describing the disease.[208] Unfortunately, data on the believed pathogenic cause of communicated diagnoses was not systematically collected to verify this. A second explanation is that *anticipated* patient demand for antibiotics can play a role – even in the absence of *actual* demand. Qualitative research has shown that providers' often perceive patients to demand unnecessary antibiotics (Das and Hammer, 2007) even in the absence of any (voiced) demand (Farley *et al.*, 2018; Manderson, 2019). Moreover, evidence from interviews conducted in this study demonstrate that physicians perceive competitive pressures to prescribe antibiotics even where they believe it to be likely ineffective. The higher rates of inappropriate prescribing for high-insured patients may then be explained by greater anticipated demand for such treatment from more financially-protected patients.

It is interesting to compare these results to Das *et al.* (2016)'s similar findings of a lack of association between observed clinical effort and appropriate treatment in the public and private primary care sectors in India: while providers in the private sector were found to exhibit higher observed clinical effort than those in the public sector, there was little difference in the quality of treatments given. The authors put this down to differences in provider training (public providers were more likely to be formally qualified) or earning incentives for inappropriate treatments (private providers earn more when they provide inappropriate treatments) off-setting the predicted positive association between observed effort and treatment quality. In the present study, these explanations do not hold: provider training is controlled for through the within-physician experiment design, and physicians have no direct financial incentives attached to drug prescribing (they do not dispense drugs or have any pharmacy affiliations).[209]

---

[208] A related explanation is that what the physician tells the patient (which is where the diagnosis data in this study is collected from) may not perfectly capture what the doctor knows (about the pathogenic cause, or otherwise) which is what would influence treatment choices. Conscious effort (and some degree of leniency) was given to coding the diagnoses, to account for the possibility of simplified patient communication. See Appendix B.1 for a list of communicated diagnoses that were considered correct / partially-correct.

[209] The implication is that separating the prescribing and dispensing functions of physicians (and thereby removing any financial incentives attached to prescribing), which has been shown to reduce inappropriate prescribing elsewhere (Chou *et al.*, 2003; Park *et al.*, 2005), is likely to be insufficient for eliminating the insurance effect on inappropriate treatment choices altogether. This contrasts with Lu (2014)'s finding: physicians in Beijing hospitals were only more likely to prescribe inappropriate treatments to insured patients when they received some financial benefit. A key difference is that the physicians in Lu (2014)'s study were all salaried, whereas physicians in this study are all paid fee-for-service and subject to market (competitive) incentives. Market incentives to prescribe antibiotics may present (even with no direct financial incentives attached) if physicians perceive that it is what patients want and that it would increase the likelihood of return visits. These incentives may be greater the more financially-protected the patient.

The conflicting effects of patient insurance on observable effort and the quality of care outputs may be partly reconciled by considering that not all aspects of physicians' effort can be observed, and that physicians may reasonably infer the cost-sensitivity of patients from their level of insurance cover. First, it may be that some important aspects of clinical effort are unobserved, such as cognitive effort. If GPs supply more appropriate, cognitive effort to low-insured patients, this may offset their lower observable effort, and result in a higher quality of subsequent care outputs. This aligns with the notion of observable (physical) effort being necessary but insufficient for high-quality care outputs. Second, GPs may reasonably assume that low-insured patients are more cost-conscious (given their greater exposure to health expenditure) and more aware that poor-quality care outputs are likely to increase avoidable future costs. While care costs are considered in greater detail in the next chapter, it is worth noting here that an inappropriate antibiotic prescription is associated with an increase of R137 in total drug costs (which represents approximately 51% of the average prescription cost in this study (R268)). Inaccurate diagnoses can equally increase healthcare expenditures, by increasing the likelihood that patients need further diagnostic procedures or care in the future.

That GPs exert more observable effort for high-insured patients but provide higher quality care outputs for low-insured patients is then consistent with *both* provider altruism toward patients' financial welfare *and* providers' competitive incentives to display more observable effort (from which patients more easily infer quality) to patients for whom they are less able to compete on cost. The latter competitive incentives may be driven by providers' *anticipation* of relatively higher demand elasticity with respect to observable effort (patient experience) among high-insured patients, and with respect to cost among low-insured patients. As with the private providers in Das *et al.* (2016), the physicians in this study have clear market incentives to satisfy what they believe patients want – even where they know such care is inappropriate.

Therefore, in a context where physicians are paid FFS, with no payment incentives attached to care quality, greater financial protection for patients results in poorer quality outputs from physicians - leading to care that is more often not only inappropriate, but also potentially harmful. These results have implications for on-going policy efforts in South Africa (SA) and elsewhere, to incorporate the private sector in universal health insurance initiatives. The SA government's plans to abolish co-payments in private health insurance schemes, and roll out an NHI system that would rely on private GPs, must weigh up the welfare benefits of improved financial protection and primary care access for patients (to potentially higher quality care than in the public sector) with potential decreases in the average quality of care in the private sector as a result. The related public health consequences in a country with already high rates of broad-spectrum antibiotic prescribing are equally important (Blaauw and Lagarde, 2019). While this study does not

investigate the effects of different payment incentives on physicians' care quality choices under insurance, it is clear that eliminating patient co-payment without some alteration of supply-side incentives is likely to be inefficient. The competitive incentives that are argued to partly drive the insurance effects estimated in this chapter derive from FFS payment (with no direct incentives for care quality), which remains the predominant form of payment in the SA private sector. In this context, these insurance effects may be attenuated if the SA government directly contracts private GPs in the proposed NHI system under capitation or sessional contracts (rather than capped FFS tariffs, which is one of the current proposals). However, other incentives for poorer care quality (including cost-saving incentives under capitation, or low competitive incentives for effort under sessional contracts) may present instead, and must be controlled for (Zuvekas and Cohen, 2010). [210]

An important limitation of this study, which is common among similarly resource-intensive audit studies, is the small sample size. This was determined primarily by resource constraints, but inevitably results in low statistical power which poses two risks: a higher likelihood of false negatives, and a possibility of inflation in estimated effects sizes (Cohen, 1988; Button *et al.*, 2013). The former issue means I cannot rule out a true effect in some cases where none were detected, particularly in the sub-sample analyses that are more demanding in their total sample size requirements. The latter issue means that large detected effects, such as the insurance effect on antibiotic prescribing, must be interpreted with caution in case of inflation.

External validity of the findings is a further concern. Ethical considerations mandated full informed consent from all research participants, meaning the study sample is not entirely random and will have some level of selection bias. While this does not threaten the internal validity of study results (given the within-subject experiment design), it does limit their generalizability to the full population of private GPs in South Africa. The focus on primary care further limits generalizability to secondary care settings, where lapses in care quality may arguably have more costly consequences. However, given the centrality of primary care in UHC efforts worldwide, as well as in the SA government's plans to reform the national health system (through a proposed gate-keeping system), the focus in this chapter remains highly policy-relevant.

The single clinical case used in this experiment (uncomplicated acute bronchitis) further constrains the generalisability of results to other clinical conditions. This is partially driven by the limitations of the SP method, which cannot support conditions with clear, visible symptoms or those that would require invasive procedures. Appropriate case management in this clinical case

[210] A further concern is the attractiveness of these contracting terms for GPs relative to the current FFS status quo. Preliminary evidence from GP contracting under the NHI pilot schemes suggest that moving too far from the current FFS payment terms will be difficult (NDoH, 2019).

also requires no treatment, whereas many clinical conditions (especially more severe ones) will require some curative treatment. Therefore, while the higher likelihood of little or no treatment for low-insured patients indicates more appropriate care in this setting, a similar result in other case contexts can signal under-provision (poorer quality) of care. Nevertheless, viral acute respiratory infections are some of the most common conditions that present in primary care, and are therefore good candidates for studying the quality of primary care in general. The SP method poses a further limitation here in assessing care continuity, which is a key component of primary care quality. Assessing care continuity would require repeat SP visits to the same doctor, which was not possible due to the increased risk of detection.

Future research in this area should seek to validate these results in other contexts and clinical cases, and with larger sample sizes. Moreover, the present study only evaluates supply-side responses to patient insurance, and proposes one explanation for the results based on physicians' *anticipation* of patient demands with respect to care quality. Further evidence is necessary to fully validate this theory. In particular, empirical validation of the demand-side component of this explanation (that patients' relative demand elasticity with respect to patient experience and inappropriate care varies by insurance status), as well as additional qualitative research on physicians' inferences from patient insurance, would be highly valuable in substantiating the proposed theory. Lastly, while higher insurance cover appears to exacerbate inappropriate prescribing behaviour, the rate of inappropriate antibiotic prescribing among low-insured patients is still very high (51%), and comparable to that found by Currie *et al.* (2014) for a similar clinical case in China (55%). A question for future research, then, is what kinds of interventions (aside from the elimination of positive financial incentives) are needed to reduce inappropriate prescribing for even low-insured or uninsured patients – for whom, in theory, both physicians' altruistic and competitive incentives to minimise inappropriate prescribing should already be high.

6 Second-Degree Moral Hazard in the Quantity and Costs of Care

6.1 Introduction

Rapidly rising healthcare costs have accompanied the expansion of formal health insurance in many parts of the world (Baicker and Goldman, 2011; Xu *et al.*, 2018). Better financial protection for patients has been linked to higher healthcare utilisation, treatment intensity, and costs (Lundin, 2000; Card et *al*., 2008; Wagstaff and Lindelow, 2008; Zhang *et al.*, 2009; Baicker and Goldman, 2011; Martin *et al.*, 2017) – including from wasteful or even harmful care (Long *et al.*, 2012; Zhou *et al.*, 2017). Nevertheless, evidence identifying the causal factors behind these trends is limited (Manning *et al.*, 1987; Lu, 2014). As South Africa shapes its ambitious national health insurance (NHI) reform, which foresees the elimination of most out-of-pocket (OOP) health expenditure in the country, anticipating and offsetting potential effects on wasteful health spending will be crucial to the scheme's sustainability.

Arrow (1963) highlighted a key role for physicians' and patients' agency relationships in driving healthcare spending under insurance. Ma and Mcguire (1997) further underlined the interactions between two missing markets in healthcare: patient insurance conditional on health state, and physician payment conditional on appropriate clinical effort. These missing markets generate incentives for 'hidden action' (or moral hazard) from both insured patients and physicians, which can increase healthcare costs for third-party payers. Patients are more likely to demand unnecessary healthcare (or less likely to seek to minimise care costs) when they do not incur the full financial costs attached (Feldstein, 1973).[211] An emerging body of literature further argues that greater financial protection for consumers (patients) generates additional, supply-side moral hazard in markets for credence goods like healthcare, where consumers cannot perfectly verify the necessity or quality of the services they receive (Wagstaff and Lindelow, 2008; Lu, 2014; Kerschbamer and Sutter, 2016; Balafoutas et *al*., 2017). Insurance in such markets generates

[211] The widespread inclusion of patient co-payments in insurance contracts seeks to limit such *ex post* patient moral hazard (Pauly, 1968; Zeckhauser, 1970). However, it is widely acknowledged that patients cannot always assess the relative value or need for care (Baicker *et al.*, 2015), or drive healthcare decision-making in place of a better-informed, 'expert' physician (Zweifel and Manning, 2000). The high degree of uncertainty in most clinical situations, and the stresses of decision-making under ill-health, mandate an important role for the physician in determining healthcare choices.

profit incentives for 'expert' suppliers to exploit their informational advantage with respect to both the consumer and insurer, and oversupply (or over-charge) less cost-conscious, better insured consumers to a greater degree.[212] In essence, this 'second-degree' moral hazard argument is an extension of theories on supplier-induced demand (Arrow, 1963; Evans, 1974; Fuchs, 1978):[213] suppliers *induce* demand from consumers proportionally to anticipated increases in consumer moral hazard from better insurance.

While this behaviour is usually understood as a response to direct financial (profit) incentives,[214] physicians have been shown to respond to many other stimuli in their treatment decisions. Physicians' concern for patient benefits, costs and preferences (Lundin, 2000; Godager and Wiesen, 2013; Brosig-Koch *et al.*, 2017; Lagarde and Blaauw, 2017); and their tendency to fall back on treatment habits (Hellerstein, 1998; Crea *et al.*, 2019) or defensive medicine under diagnostic or therapeutic uncertainty (Baicker et *al*., 2007) can all play a role. A primary argument of this chapter is that such indirect- or non-financial factors can sustain more and more costly treatment for better-insured patients, *even in the absence of any direct financial incentives* for physicians.

Variations in patient insurance can alter the competitive or altruistic incentives that physicians face - even where direct financial incentives are absent. If physicians are subject to some level of market demand (they compete for patients' future business) or are at least partially altruistic, they can generate higher treatment expenditures for more insured patients through two channels. First, physicians may be simply responding to (correctly or incorrectly) anticipated patient preferences, even in the absence of any explicit patient demand. They may reasonably expect low-insured patients to be more responsive to treatment costs and for high-insured patients to prefer more and more expensive treatments (patient moral hazard).[215] Second, some uncompensated effort may be required of physicians to minimise treatment costs – for example, in improving diagnostic and

---

[212] This is relative to what consumers or insurers would consent to under perfect information.

[213] This refers only to supplier-induced demand motivated by the will to maximizing income, whereas Léonard et *al.* (2009) note that demand-inducement can also be driven by a motive to maximize patient benefit (where the patient does not have full information, and therefore cannot choose the best care themselves). Johnson (2014) argues that the latter form, where demand is influenced in the best interests of the patient, is not supplier-induced demand at all, as "moving demand toward the patient's optimum is a responsibility of physicians".

[214] A study by Lu (2014) finds that although physicians in Beijing hospitals prescribe more and more expensive drugs to insured patients (compared to uninsured) when they have a direct financial incentive to do so (when the patient agrees to purchase the drug in the hospital pharmacy, from which physicians receive a commission), this difference by insurance cover is eliminated when the direct financial incentive is removed (when the patient states they will purchase the drugs elsewhere). The implication is that any oversupply in physician services may not exist in the absence of any positive financial incentives – such as fee-for-service payment systems that reward service volume (Blomqvist, 1991; Bardey and Lesur, 2006), or the ability of physicians to earn mark-ups or commissions from drug prescriptions (Iizuka, 2012; Lu, 2014). This reflects, for instance, policy efforts to separate drug prescribing and dispensing functions in health systems around the world, aiming to minimise wasteful prescribing practices (Morton and Kyle, 2011).

[215] This may be particularly true of demand for prescription drugs: while patients have been shown to prefer more expensive drugs (including branded drugs) in general, believing they are more effective (Cleanthous, 2002; Himmel *et al.*, 2005; Shrank *et al.*, 2009), their demand is also sensitive to out-of-pocket price (Leibowitz *et al.*, 1985).

therapeutic accuracy to minimise unnecessary treatments, in changing costly treatment habits, or in sourcing more cost-effective treatment alternatives on the market. Here, physicians' effort choices are likely to depend both on their anticipated effects on patient welfare and subsequent business opportunities (which rely crucially on the patient's demand response), as well as on the effort costs involved.[216] [217] In either case, physicians are likely to supply more cost-minimising effort and fewer unnecessary or expensive treatments to more cost-conscious, low-insured patients – either because they anticipate these patients' demand to be more responsive to unnecessary care costs, or because they care about patients' avoidable OOP costs (Morton and Kyle, 2011).

The potential for this type of *indirect* second-degree moral hazard from health insurance has received relatively little attention in the physician agency literature. It differs from *conventional* second-degree moral hazard, as the gains to physicians derive *indirectly* - from gains (or cost-savings) to patients, and from effort cost-savings to physicians (effort-stinting) (Ma, 1994).[218]

The two primary hypotheses tested in this chapter are therefore the following:[219]

i) *'Conventional' second-degree moral hazard*: in the presence of direct fee-for-service (FFS) incentives, physicians supply a higher quantity and cost of services to more insured patients, deviating from the best interests of the patient and insurer (*Hypothesis 2*).

ii) *'Indirect' second-degree moral hazard*: even in the absence of direct FFS incentives, physicians supply a higher quantity and cost of services to more insured patients, deviating from the best interests of the insurer (optimal care) *as long as* they *a)* are paid on the basis of patient volume for other services (i.e. are subject to some level of market incentives), or *b)* are altruistic towards patients' OOP costs and anticipated preferences (*Hypothesis 5*). [220]

---

[216] These effort costs can include the costs to higher diagnostic effort, and to sourcing more appropriate and inexpensive treatment options for the patient.

[217] A key assumption here is that physicians view their uncompensated effort and treatment quantity (and costs) as substitutes. For instance, physicians may substitute diagnostic and therapeutic accuracy with treatment quantity. This appeals to the notion of demand-inducement as a form defensive medicine (Johnson, 2014): in the absence of sufficient clinical effort, the physician faces greater diagnostic and therapeutic uncertainty, and therefore prescribes more treatments to increase his chances of addressing the underlying illness. Moreover, physicians may require some effort to break habitual but wasteful practices - for instance, in switching from branded drugs to lesser known generics, or in adopting newer, more cost-effective treatments (Hellerstein, 1998; Frank, 2004; Crea *et al.*, 2019).

[218] Note that this result requires that physicians are paid fee-for-service or on the basis of patient volume for at least some types of care (otherwise they have no competitive incentives to consider patients' demand responses in their treatment decisions), or that they are sufficiently altruistic towards patients' avoidable OOP costs and anticipated preferences.

[219] These correspond to two hypotheses derived in Chapter 3. The specific numbered hypotheses are given in parentheses.

[220] The extent to which this *indirect* physician moral hazard is also a deviation from the patient's best interests (their preferences and willingness-to-pay) will depend on the relative importance of the two underlying behavioural channels proposed: demand-inducement

The real-world extent and relative influence of these two types of physician moral hazard, compared to patient moral hazard, has become a matter of both academic and policy interest (Ma and Mcguire, 1997; Bardey and Lesur, 2006; Kerschbamer and Sutter, 2007). Nevertheless, empirically separating the prevalence of these distinct phenomena (as well as distinguishing either from the effects of adverse selection of sicker patients into health insurance) has eluded most work in this area to date, since the observed outcome is all the same: higher healthcare expenditure for more insured patients (Zweifel and Manning, 2000). To test the hypotheses in this chapter, I therefore utilise a novel, audit study approach to identifying supply-side moral hazard (Lu, 2014; Kerschbamer *et al.*, 2016; Balafoutas *et al.*, 2017). I draw on the same audit experiment as in Chapter 5, which was conducted with 89 private primary care physicians in Johannesburg (as detailed in Section 4.3.1). Each participating physician in the experiment received randomly-ordered, incognito visits from two standardised patients (SP) that varied only in their insurance cover ('high-insured' or 'low-insured'), holding all other aspects of the clinical case and patient presentation fixed.[221] The physicians were all non-dispensing, and therefore had no direct financial incentives attached to drug prescribing. However, they were paid FFS for individual consultations and for any diagnostic tests or procedures they conduct at their clinics. Hence, they had direct financial gains in increasing the number and costs of patient consultations, diagnostic tests and other FFS procedures,[222] but *not* in increasing drug expenditures. This institutional set-up allows simultaneous testing of both primary hypotheses.

In addition to contributing new evidence on conventional second-degree moral hazard, this chapter provides novel field-experimental evidence of its *indirect* manifestation. The results support both hypotheses. Physicians deliver a higher quantity and cost of treatment for high-insured patients, both when they have direct financial incentives to do so, and when they do not. According to national and international clinical guidelines, no diagnostic tests or drug treatments are necessary for the clinical case used in this experiment (an uncomplicated case of acute bronchitis). However, in line with *Hypothesis 2*, physicians were almost twice as likely to provide an unnecessary FFS test or procedure to high-insured patients compared to low-insured ones, with total consultation costs (excluding drugs) that were 7% higher on average (R485, compared to R455 for low-insured patients). Over-treatment with drugs was high in general: the average patient was prescribed 3.3 drugs items, including 2.6 inappropriate drugs, at an average

---

to substitute for an under-supply in cost-minimising effort, or supply responses to (correctly) anticipated patient preferences (in other words, physicians' *perfect agency* (Pauly, 1980).

[221] This creates exogenous variation in the patient insurance cover facing the same physician and overcomes potential issues of insurance endogeneity (patient selection into insurance, or physician selection of insured patients) and identification (the insurance effect on physician behaviour is cleanly isolated from other patient or physician-level confounders). The SP script was standardised to demonstrate the patient's lack of information on their underlying illness and to eliminate any patient demands (or expressed preferences) for treatment – to create the necessary conditions to test for demand-inducement.

[222] Including those they refer to external laboratories, which may require a repeat consultation with the patient to discuss results.

prescription cost of R268.[223] Moreover, in support of *Hypothesis 5*, physicians wrote drug prescriptions that were 17% more expensive for high-insured patients (costing R289 on average, compared to R247 for low-insured patients). This latter result is primarily driven by physicians prescribing more and more branded (originator) drugs to these patients. Taken together, the findings indicate that greater financial protection for patients generates more unnecessary treatment and expenditure by private physicians, even in the absence of any positive demand from patients or direct financial pay-offs for physicians.

The remainder of the chapter is organised as follows. Section 6.2 outlines the chapter's contributions to existing literature. Section 6.3 presents an overview of the field experiment, and summary statistics on the care quantity and cost outcomes of interest. The empirical framework is discussed in Section 6.4, followed by a presentation of the main results and robustness checks in Section 6.5. The chapter concludes with a discussion of main findings and potential policy implications in Section 6.6.

6.2 Literature Contributions

This chapter contributes to four strands of existing literature. Firstly, it develops the literature proposing a role for physicians in explaining the empirical association between insurance cover and healthcare costs (Pauly, 1980; McGuire, 2000). A number of theoretical frameworks predict that the quantity and costs of physicians' treatment decisions will respond to patient insurance (Ma and Mcguire, 1997; Bardey and Lesur, 2006; Rischatsch, Trottmann and Zweifel, 2013). However, robust empirical evidence of this behaviour is scarce. Most existing studies that estimate the effects of insurance on treatment choices are based on observational data (Lundin, 2000; Zweifel and Manning, 2000; Joyce *et al.*, 2002; Card, Dobkin and Maestas, 2007; Wagstaff and Lindelow, 2008; Zhang *et al.*, 2009), and therefore estimate the combined effects of patient and physician responses to health insurance (as well as adverse patient selection into insurance). This prevents them from isolating the relative influence of the physician from the patient, which is crucial from a regulatory perspective for controlling healthcare spending. The RAND and Oregon Health Insurance Experiments isolated the effects of moral hazard (patient or physician responses to insurance) from adverse selection of patients into insurance (Newhouse, 1993; Finkelstein *et al.*, 2012): by randomly assigning patients to different levels of cost-sharing (co-payment rates), they eliminate any systematic self-selection of specific types of patients into

[223] To put this figure in context, the cheapest recommended (symptomatic) treatment for the clinical case presented in this experiment was paracetamol and a cough suppressant, priced at R17.86 (see Section 4.2.2). The average drug cost recorded in this experiment is almost 14 times higher.

specific levels of insurance cover.[224] However, because they do not also randomise the insurance cover facing treating physicians, they cannot fully isolate the role of patient moral hazard from physician moral hazard, or eliminate potential selection of patients into care-seeking from specific types of physicians conditional on assigned insurance status.[225] Mort *et al.* (1996) and McKinlay et *al*. (1996) are two exceptions that identify physicians' independent responses to patient insurance, and find that insurance affects physicians' clinical decisions in ways that can increase care costs. Nevertheless, they rely on physicians' reports of likely choices in specific clinical scenarios, rather than on observations of actual behaviour. This chapter adds to this literature in a number of ways. First, it utilises micro-level data on actual treatment decisions observed during physician-patient interactions. Second, by drawing on a controlled field experiment, it succeeds in creating random variation in the insurance status facing treating physicians, whilst controlling for all other patient-level confounders (including patient demands or adverse selection into insurance). Third, by fixing the clinical case presented by SPs, it can verify the necessity of physicians' treatment choices – which is difficult to do from administrative data, and on the basis of reported diagnoses - and identify the insurance effect on inappropriate health expenditures.

In utilising the SP audit approach, this study also contributes to an emerging literature using 'mystery' shoppers or incognito SPs to identify supply-side moral hazard. These studies have found significant evidence of *conventional* second-degree moral hazard in response to consumer (or patient) insurance in a number of markets with credence goods characteristics, including computer repairs (Kerschbamer *et al.*, 2016), taxi rides (Balafoutas *et al.*, 2017) and drug prescriptions (Lu, 2014). Nevertheless, there is no field-experimental evidence of *indirect* second-degree moral hazard as yet. A closely related study to the present one investigates the impact of patient insurance on physicians' prescribing choices in Beijing hospitals (Lu, 2014). To identify the insurance effect, the author employs a similar controlled field experiment using SPs with varying insurance statuses, and finds that salaried hospital physicians provide prescriptions that are 43% more expensive to insured patients compared to uninsured ones when they expect a financial pay-off.[226] This is a much larger effect than the one found in this study (a 17% increase

[224] However, the RAND experiment was challenged due to substantially higher levels of attrition by participants in the cost-sharing plans (attrition was almost 16 times higher than in the free care plan) (Nyman, 2007). If the likelihood of attrition was linked to the health status of participants or other characteristics that can influence their likelihood of needing care, this would invalidate the randomized assignment of insurance status to avoid adverse selection.

[225] Firstly, the randomly assigned level of cost-sharing may systematically affect where (and with what types of doctors) patients choose to seek care. If the doctors that treat high co-pay patients are systematically different from those that treat low co-pay patients, then differences in treatment choices and costs between high- and low co-pay patients can be either due to differences in the insurance status of the patient or due to systematic differences in the treating doctors. Secondly, even if there were no systematic differences in the treating doctors by the levels of patient co-payment, that patient insurance leads to more medical care and expenditure can be either due to patient demands or doctors' independent responses.

[226] The financial pay-off comes from pharmacy profit-sharing when patients purchase prescribed drugs from the same hospital pharmacy. This pay-off is removed when patients opt to buy their drugs at an external pharmacy.

in prescription costs). However, there are important differences between this study and Lu (2014) that are worth noting. Firstly, unlike in Lu (2014), the physicians in this study do not receive direct financial pay-offs from their prescribing choices (as they are non-dispensing and have no pharmacy affiliations). While *indirect* financial pay-offs (market incentives) from treatment choices is one possible explanation for the insurance effect in this study, these indirect incentives are likely to be far less powerful than direct financial pay-offs (which could explain the smaller estimated effect). Indeed, Lu (2014) finds that when direct financial incentives for drug prescriptions are removed (that is, when the patient opts to buy their drugs at an external pharmacy), the insurance effect on prescription costs is eliminated.[227] It should be noted that Lu (2014)'s sample was composed only of salaried physicians, so market incentives were largely absent in her study.[228] Secondly, while Lu (2014) compares insured patients to fully uninsured patients, the present study compares two *levels* of patient insurance where any differences in physician incentives are likely to be more incremental. Thirdly, drug treatment is largely not recommended for the clinical case used in this study, whereas some treatment *is* clinically indicated for the two cases in Lu (2014).[229] Therefore, appropriate treatment necessitates some positive prescription cost in Lu (2014); whereas *no* drug treatment is clinically indicated in this study, which allows a clearer analysis of the insurance effect on wasteful treatment and costs. Fourthly, in addition to drug prescriptions, this study also investigates the insurance effect on diagnostic tests and other FFS procedures which are not considered in Lu (2014). Finally, this study observes the insurance effect on the *same* physician (the same physicians are visited by both types of insured SPs). While the same hospitals are visited by both types of SPs in Lu (2014), the author cannot verify whether the same physicians were visited within each hospital, and is therefore unable to control for all physician-level characteristics that may confound the estimated insurance effect. As such, this chapter presents a strong methodological contribution to the literature concerned with identifying supply-side moral hazard in healthcare.[230]

Relatedly, this chapter builds on an extensive physician agency literature on supplier-induced demand (SID) (see McGuire (2000) and Johnson (2014) for detailed reviews) – whereby physicians influence healthcare demand away from levels they assess to be in patients' best

---

[227] The author concludes that this counters the 'considerate doctor' hypothesis. In other words, that physicians are not sufficiently sensitive to patients' out-of-pocket costs.

[228] This may partly explain the null insurance effect on prescription costs in the absence of any direct financial incentives, unlike in this study with FFS physicians. The lack of market-based incentives in Lu's sample may limit the tendency for indirect second-degree moral hazard.

[229] The clinical cases used in Lu (2014) were hypertension, and hypertension combined with elevated triglycerides and high blood pressure.

[230] Other experimental audit studies on patient-side influences on physicians' treatment choices evaluate the effects of patient requests for branded drugs (Kravitz et al., 2005), patients' signalled knowledge of appropriate treatments (Currie et al., 2011), and patients' signalling of intent to seek second 'expert' opinions (Gottschalk et al., 2017).

interests for profit. Most empirical work seeking to identify SID in healthcare have relied on exogenous variations in physicians' income levels or fees. For instance, studies have found that the negative income shock implied by an increase in the local physician-to-population ratio increases rates of surgeries and highly profitable Cesarean sections (C-sections) (Fuchs, 1978; Cromwell and Mitchell, 1986; Gruber and Owings, 1996).[231] Similarly, other studies have found increases in health service volumes (also mainly surgeries and C-sections) in response to Medicare fee reductions (Rice, 1983; Yip, 1998; Jacobson *et al.*, 2010), particularly among physicians whose incomes were most affected (Nguyen and Derrick, 1997; Yip, 1998).[232] Currie *et al.* (2014) present field-experimental evidence of SID using a similar SP audit approach to the present study. They exogenously vary the financial pay-offs from drug prescribing for Chinese hospital physicians (in a similar way to Lu (2014)), and find that such pay-offs increase the quantity and costs of physicians' prescriptions.[233] A fundamental assumption that distinguishes SID from other profit-maximising behaviour is that information is asymmetric, and patients are relatively uninformed. Most of this evidence on SID therefore comes from specialist, secondary care contexts, where the asymmetry in information between 'expert' physicians and patients is likely to be highest. More recent studies have sought to test whether physicians' capacity for SID would vary with patient information, with mixed results. The most robust identification strategies in this area are found in two SP field experiments (Currie *et al.*, 2011; Gottschalk *et al.*, 2017). In Currie *et al.* (2011), SPs that signal knowledge of appropriate prescribing outcomes to their consulting physician are prescribed significantly less inappropriate medications (antibiotics) relative to uninformed patients;[234] whereas in Gottschalk *et al.* (2017), patients that signal their likelihood of getting a second diagnostic opinion elsewhere are not treated significantly differently to patients that do not. The present chapter adds to this literature in two ways: it finds evidence of SID *i)* in a non-specialist, primary care context, and *ii)* in response to exogenous variation in patients' insurance cover (and the implied cost-consciousness and care preferences of patients).

Lastly, this chapter relates to a fourth body of literature on alternative, non-financial drivers of physicians' treatment choices (Frank, 2004; Chandra et *al*., 2011) - and how these can vary with the patient's insurance cover. Pauly (1980) argues that physicians have indirect financial

[231] As the local physician density cannot plausibly affect the optimal treatment recommendation for the patient, these results have been interpreted as evidence of demand-inducement.

[232] Gruber *et al.* (1999), on the other hand, find an increase in C-sections in response to a Medicaid fee *increase*, suggesting that the resulting substitution effect dominates the income effect in this case (unlike in the other studies listed).

[233] They exogenously vary physicians' financial incentives for prescribing drugs by sending SPs that express a preference for purchasing prescribed drugs either *i)* at the same hospital pharmacy, which will entail some profit-share for the prescribing physician, or *ii)* at an external pharmacy (which would involve no financial payoff for the prescriber).

[234] However, Currie *et al.* (2014) find that direct financial incentives are a much stronger driver of prescription choices than patients' signalled knowledge.

incentives to minimise patients' OOP costs for services they cannot charge for, as higher patient expenditure on these services reduces what patients can spend on physicians' own (profitable) services. This incentive is of course absent (or very small) with high-insured patients that do not incur OOP costs. Another argument is that physicians' incur effort costs to minimising treatment expenditures for patients (Ma, 1994). First, they may incur some cost to changing habitual behaviours, which deter them from seeking newer, more cost-effective therapies for their patients. For instance, Crea *et al.* (2019) and Hellerstein (1998) find evidence of habit persistence in physicians' prescribing behaviours, which prevent them from switching from branded to cheaper generic drugs. Cutler *et al.* (2019) also argue for persistence in physicians' individual practice styles in explaining regional variations in Medicare expenditure in the US. Second, there is considerable uncertainty involved in clinical decision-making (Chandra et *al*., 2011), and substantial clinical effort may be required of physicians to minimise this. The absence of such effort (or physicians' unwillingness to supply it) may support the tendency for defensive medicine, as noted in some studies (Baicker et *al*, 2007). Johnson (2014) characterises this tendency as an indirect form of SID, where a higher quantity of care is substituted for diagnostic or therapeutic certainty. Physicians' incentives to supply such cost-minimising effort can depend on the OOP costs for the patient. For instance, a large literature considers that physicians are not purely profit-maximising, but are also motivated by patient welfare - including financial welfare (see Galizzi *et al.*, (2015) for a review). In addition, physicians that rely to some extent on patient volume for their income have competitive (market-based) incentives to satisfy the likely greater expectation for cost-effective care from low-insured patients with higher OOP costs, to maintain their repeat business (Allard et *al.*, 2009). Physicians may then supply relatively less cost-minimising effort with more insured patients, resulting in higher costs for these patients even where physicians cannot charge for their treatment choices.

Nevertheless, existing evidence on physicians' treatment responses to patient insurance in the absence of direct financial incentives is mixed. From the literature investigating drug prescribing decisions in settings where physicians cannot dispense drugs, Iizuka (2012) finds that FFS doctors (with similar market-based incentives as the doctors in this study) are not sensitive to patients' OOP costs in their choices of branded or generic drugs – contrary to results in this chapter. Crea *et al.* (2019) and Hellerstein (1998) similarly find no supportive evidence that higher patient insurance increases the prescription costs generated by non-dispensing doctors. By contrast, Lundin (2000), Granlund (2009) and Mott and Cline (2002) find that lower patient insurance increases the likelihood of physicians prescribing the cheaper, generic version of a drug (or allowing generic substitution by pharmacists). These studies are nevertheless limited in their reliance on administrative data, which – as discussed previously - makes it difficult to isolate the insurance effect on treatment choices, and to distinguish physicians' choices from those of

patients. Moreover (with the exception of Iizuka (2012)), although the physicians in these studies do not gain financially from prescription costs, the influence of other market-based incentives (which partly motivate the *indirect* second-degree moral hazard hypothesis in this study) are generally not specified, not controlled for, or non-existent (as with the public sector physicians in Lundin (2000)). A more general limitation of these studies is that, beyond analysing the prescriber choice between a branded and more cost-effective generic version of the same drug, they do not investigate avoidable drug expenditures more holistically in each patient case (which may be determined by choices between equivalent drugs, the overall quantity of drugs prescribed, or even the clinical need for any drug treatment at all in a given patient case). This chapter therefore builds on these works in two ways: *i)* it considers a context where physicians have clear market-based incentives (even where they cannot charge for drug treatments), and *ii)* it draws on field-experimental data to cleanly identify the insurance effect on physicians' choices of both total and unnecessary (or avoidable) treatment quantity and costs.

6.3 Experimental Methods and Data

This chapter draws on the same within-physician, SP audit experiment as Chapter 5. The SP methodology, sample characteristics and experimental procedures are described there and in Sections 4.1 and 4.3.1, and will not be repeated in detail here. Briefly, 89 private primary care physicians in Johannesburg were recruited into the study, and each received unannounced visits from two SPs that varied only their insurance cover (one 'high-insured' with a very low expected co-payment rate, and one 'low-insured' with a higher expected co-payment rate).[235] All other aspects of the SPs' presentation and scripts were standardised - including their portrayal of an uninformed and undemanding patient.[236] All recruited physicians were paid FFS for their consultations and any procedures administered during the visit. However, they were non-dispensing physicians, and therefore did not gain any direct pay-offs from their drug prescription choices. The clinical case presented by all SPs was an uncomplicated case of acute bronchitis in an otherwise healthy, young adult (Clinical Case 1). According to national and international clinical guidelines, no diagnostic tests or treatment are necessary for this case, although some mild palliative treatment may be considered to relieve the patient's symptoms (NDoH, 2014; Brink *et al.*, 2016). After each SP visit, the overall consultation and drug costs attributable to the

[235] The 'high-insured' SP had an insurance plan that allocated a higher level of medical savings to cover primary healthcare costs than that of the 'low-insured' SP (around 75% higher), and therefore had a much lower likelihood of having to co-pay for any care.
[236] All SPs presented to physicians as patients seeking information and expert advice on their underlying illness (*uninformed* patients), without any demands for specific treatments (*undemanding* patients).

visit were calculated using a number of data sources (see Section 4.2.2 for further details).[237] The following section presents descriptive statistics for the main outcomes analysed in this chapter.

6.3.1 Care quantity and cost outcomes

Table 6.1 presents a summary of the care quantity and cost outcomes of interest from the SP visits. A prescription was given in all 178 consultations completed in this experiment. Drug expenditures for each consultation were computed by matching prescription items to corresponding items listed in the South African Medicine Price Registry (MPR), as detailed in Section 4.2.2. The MPR contains the unit Single Exit Price (SEP) for each listed drug, which specifies the maximum charges that can be applied to pharmaceuticals (excluding a regulated dispensing fee) in South Africa. The unit SEP was multiplied by the prescribed quantities to obtain the cost of each prescription item. The database further specifies whether a drug is the originator or a generic equivalent. Where physicians only wrote the active pharmaceutical ingredient (API) of a drug on the prescription (rather than specifying a drug name), the price of the cheapest generic drug with the same main API in the MPR database was chosen to be matched to the prescription item. Some over-the-counter (OTC) prescription items were not included in the database. In these cases, the retail price was obtained from the two leading pharmacies in South Africa (*DisChem* and *Clicks*). For branded (originator) prescription items, a basic search was also conducted in the database to check for the availability of generic substitutes in the market. The costs of all FFS items were taken from itemised insurance claims submitted to the partnering insurer and any receipts provided from the consultations.

Four groups of outcomes are analysed. First, the total drug expenditures per consultation are analysed, including the number of drugs prescribed and the average cost per drug item. The cost of each drug item is the product of the number of units of the drug prescribed (which indicates the treatment *intensity* with that specific drug) and the unit SEP price of that drug (taken from the MPR).[238] As shown in panel *a.* of Table 6.1, the prevalence of over-treatment in the sample is high in general: for a clinical case that requires no drug treatment, the average patient was prescribed 3.31 drugs, at an average cost per drug of almost R80. This amounts to an average cost per prescription of R268.[239] To put this in context, the cost of the cheapest available symptomatic treatment for the clinical case (which is the only treatment recommended for uncomplicated acute

[237] These sources include the debriefing questionnaire completed by SPs immediately after each visit (which noted any tests or procedures administered during the consultation, and any costs that had to be paid upfront in cash and be claimed back), the consultation receipts, the claims submitted to the insurer, the prescriptions from each visit, and a national drug price database that contains regulated pricing information for prescription drugs sold in the country

[238] The average cost per drug item therefore reflects both the average treatment intensity per drug, as well as the average unit cost.

[239] In comparison, Currie *et al.* (2014) find a lower number of 2.4 drugs prescribed for the average patient with flu-like symptoms in their SP audit study of hospital physicians in China.

bronchitis in clinical guidelines) was calculated from the MPR, and amounted to R17.86 - which is just 7% of the average prescription cost realised in this experiment.

Table 6.1. Summary statistics – care intensity and cost outcomes

	Mean	SD
a. All prescription items		
Number of drugs prescribed	3.31	(1.20)
Average cost of prescribed drugs (ZAR)	79.52	(41.26)
Total cost of drugs prescribed (ZAR)	268.46	(166.90)
b. Inappropriate prescription items		
Number of inappropriate drugs	2.57	(1.06)
Average cost of inappropriate drugs (ZAR)	88.22	(52.37)
Total cost of inappropriate drugs (ZAR)	227.42	(162.05)
c. Branded prescription items		
Likelihood of branded drug prescription	0.76	(0.43)
Number of branded drugs	1.28	(1.01)
Average cost of branded drugs (ZAR)	127.91	(72.37)
Total cost of branded drugs (ZAR)	159	(148.87)
d. Fee-for-service items		
Additional fee-for-service tests / procedures billed	0.08	(0.28)
Number of additional fee-for-service tests / procedures billed	0.12	(0.43)
Total cost of consultation, excl. drugs (ZAR)	469.58	(113.33)
Observations (no. consultations)	178	

Notes: All consultation outcomes were measured using data collected from the post-consultation SP debriefing questionnaire, the prescriptions, itemized insurance claim submissions, and a national drug pricing database (the South Africa Medicine Price Registry). All listed figures are on a per prescription or per consultation basis. For further details on how these outcomes were measured and the data sources used, refer to Table 4.3 and Section 4.2.2.

Second, prescribed drugs are classified as either 'appropriate' or 'inappropriate' for the uncomplicated acute bronchitis case, and total inappropriate drug expenditures are analysed in the same way.[240] Panel *b.* of Table 6.1 shows that inappropriate drugs drive a large majority of all drug expenditures: 78% of all drugs prescribed and 85% of total prescription costs were composed of inappropriate drugs. The average patient was prescribed almost 2.6 inappropriate drugs, at an average cost of R88 per item and a total cost of R227 per prescription. Appendix D.1 shows that the most frequently prescribed inappropriate drugs were steroids and antibiotics. The most

[240] This classification was done in consultation with clinical experts in respiratory and infectious diseases in South Africa, and with reference to the South African Department of Health's standard treatment guidelines for primary care (NDoH, 2014).'Appropriate' drugs for this case are classified as those that are recommended for symptomatic relief, but are not clinically-indicated or necessary for case resolution (as with any drug treatment in this case). Among the drugs prescribed, these included OTC analgesics, cough suppressants and expectorants, and throat preparations (see Appendix D.1). 'Inappropriate' drugs, on the other hand, are those that are neither recommended for symptomatic relief nor clinically-indicated (and potentially harmful) for this case. These included antibiotics (prescribed with or without probiotics), nasal decongestants, bronchodilators, antihistamines, steroids, vitamins and other drugs.

expensive prescription items were also antibiotics, at an average cost of almost R148 per item (accounting for more than 55% of the average total prescription cost).

Third, branded (originator) drug costs are analysed as an indicator of avoidable drug expenditures.[241] Over 75% of all prescriptions contained at least one branded drug item, at an average cost of almost R128 per item. In comparison, the average generic drug item prescribed was almost 57% cheaper (costing just R55) – see Appendix D.1. Branded drugs further accounted for almost 40% of all prescribed drugs and 60% of total prescription costs.

Last, the FFS items charged to each consultation are considered (see last panel of Table 6.1). This includes the consultation fee, and the costs of any additional FFS diagnostic tests or procedures that were ordered during the consultation (including any external laboratory tests, that would require follow-up by the referring physician). Figure 6.1 displays the types and frequencies of FFS items charged for in the consultations. Given the uncomplicated nature of the clinical case, and its amenability to a reasonably accurate diagnosis with only history-taking and physical examination, all additional FFS procedures were considered to be unnecessary (over-treatment) in this case. The overall frequency of additional procedures is indeed low: a total 21 procedures were ordered in just 15 consultations (8% of all consultations). This is in contrast to the extent of over-treatment through drug prescriptions, as discussed in the preceding – despite the lack of any FFS incentives attached to the latter. The average consultation cost (excluding drugs, but including the consultation fee and other FFS charges) was approximately R470.

[241] Drugs are classified as either branded or generic using data from the MPR. Not all branded drugs had generic equivalents listed in the MPR. In these cases where generic substitution is not possible, the associated expenditure cannot be considered 'avoidable' (even if it is inappropriate). However, the share of such cases in this experiment is very small: of 228 branded drug items prescribed, only 12 did not have a generic equivalent listed in the MPR.These corresponded to three unique drugs: Fluticasone Furoate (an inhaled corticosteroid), Budesonide-Formoterol Fumorate (a bronchodilator), and Telithromycin (an antibiotic). Excluding these cases does not affect the main results on branded (avoidable) drug expenditure in Section 6.5.1.

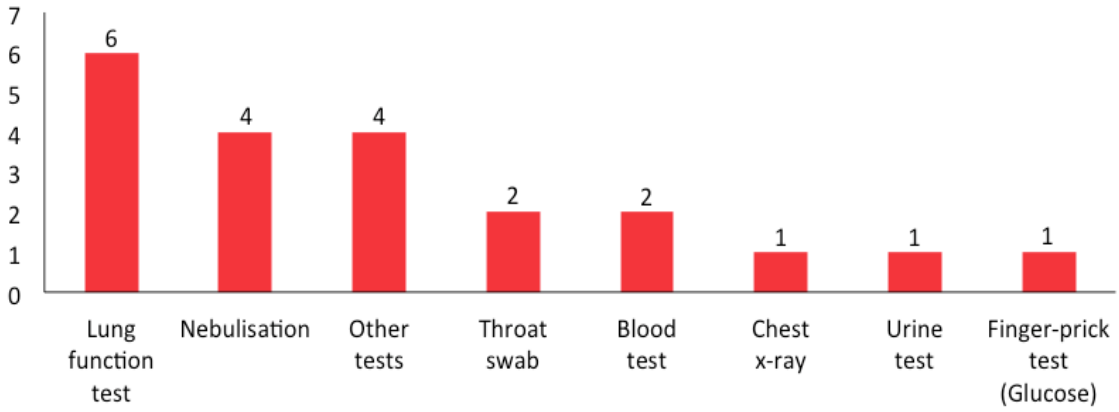


Figure 6.1. Types and frequencies of fee-for-service procedures ordered

6.4 Empirical Framework

The within-subject experiment design and SP method have some implications for the choice of empirical model. As each physician is visited by SPs of both insurance types, the sample is perfectly balanced in terms of physician and clinic characteristics in both 'treatment' arms (high-insured and low-insured SP visits). Moreover, as the SP method allows exogenous variation in the insurance status of patients visiting participating physicians, it avoids systematic selection in the types of insured patients seen by specific types of physicians. These features negate the need to explicitly control for physician or clinic characteristics in the empirical model, in order to obtain unbiased estimates of the insurance effect. However, as each physician is observed twice, the pairs of 'within-physician' observations are likely to be correlated to some extent. This potential for data clustering at the physician-level means statistical inference based on standard ordinary least squares (OLS) estimation is likely to be invalid, and an alternative approach is needed.

In addition, a primary concern with the SP method is that physicians may respond differently to different SPs for reasons other than the insurance cover. For instance, the fieldworkers playing each type of SP may be systematically different, or behave differently conditional on assignment to a certain SP role. To address the former issue, fieldworkers were matched into pairs on the basis of observable characteristics (such as age, gender, and other physical characteristics), and each physician was randomised to visits by a specific SP pair. Behavioural standardisation was also emphasised during SP training (see Section 4.1.2). To address the latter concern, it is necessary to control for fieldworker fixed effects in the empirical model. However, as the SP type played by each individual fieldworker is fixed (each fieldworker was assigned to a specific insurance cover, as formal insurance had to be purchased in their names), including individual fieldworker dummies in the empirical model would create a high degree of co-linearity: the

fieldworker dummies would be perfectly co-linear with the binary insurance variable. An alternative solution is to control for differences across the fixed fieldworker-pairs.

With these considerations in mind, I employ a physician-level fixed effects model of the following form to estimate the main results:

$$Y_{ij} = \beta_0 + \beta_1 Insurance_{ij} + \delta_j + \varepsilon_{ij} \quad (1)$$

where Y_{ij} is the outcome of interest in consultation i with physician j; $Insurance_{ij}$ is an indicator taking value 1 if the SP in consultation i with physician j was high-insured, and 0 otherwise; δ_j are physician-level fixed effects; and β_1 is the parameter to be estimated (the insurance effect).[242] The physician fixed effects automatically control for fieldworker-pair effects, as the fieldworker pairs do not vary within-physician. Following Arellano (1987) and Cameron and Miller (2015), the fixed effects model is also combined with a cluster-robust variance estimator, to account for any residual correlation in the error term from data clustering at the physician level. Main results are presented for both specifications (fixed effects model with and without the cluster-robust variance estimator). All subsequent sub-sample analyses are conducted using linear OLS with the cluster-robust variance estimator only (and explicitly including fieldworker-pair fixed effects), as the physician fixed effects model prevents estimating the effects of physician-level variables.

Robustness of all main results to alternative methods of handling clustered data is also checked in Section 6.5.4, including a paired sample t-test ('response simplification') and a linear mixed effects model (LMEM) with random intercept and slope. See Appendix C.3 for a discussion on these methods.

6.5 Results

6.5.1 Effects of patient insurance cover on drug treatment quantity and costs

Table 6.2 presents estimates of the high-insurance effect on total prescribed drug costs. Panel A shows results from a physician-level fixed effects model with normal standard errors, while Panel B shows results from an equivalent model with physician-level, cluster-robust standard errors. All results are shown to be consistent across the two specifications. Column (3) indicates that total prescription costs for high-insured patients are 17% higher than for low-insured ones (R289

[242] As detailed in Appendix C.3, the fixed effects model is just one of a few alternative approaches to controlling for potential clustering of data at the physician level. Other options include OLS with a cluster-robust variance estimator, response simplification or a mixed effects model with physician-level random effects. The fixed effects model was chosen for the main analysis due to its relative ease of specification, suitability in small sample sizes, and automatic control of fieldworker-pair fixed effects (as the fieldworker-pairs do not vary within-physician, they are implicitly controlled for in the physician-level fixed effects).

compared to R247). In purchasing power parity terms, the R42 difference in total drug costs per patient is equivalent to approximately USD 7.[243]

Table 6.2. Effect of insurance cover on drug treatment intensity and costs – all drugs

	(1) No. drugs	(2) Average drug cost	(3) Total drug cost
Panel A: provider-level fixed effects			
High-insured	0.29***	8.54*	42.21**
	(0.10)	(4.51)	(16.17)
Obs.	178	178	178
R-squared	0.09	0.04	0.07
Panel B: provider-level fixed effects & cluster-robust standard errors			
High-insured	0.29***	8.54*	42.21**
	(0.10)	(4.53)	(16.22)
Obs.	178	178	178
R-squared	0.09	0.04	0.07
Mean of low-insured	3.17	75.25	247.36

Notes: Standard errors are in parenthesis. All costs are in South African Rand (ZAR). Panel A presents estimates from a fixed effects (mean-differenced) model. Panel B shows results from a fixed effects model, as in Panel A, with the addition of a cluster-robust variance estimator. All outcomes are per prescription (per consultation).
*** p<0.01, ** p<0.05, * p<0.1

Higher total prescription costs may be driven by physicians prescribing either a higher number of drugs or more expensive prescription items to high-insured patients. Columns (1) and (2) provide some evidence for both channels. High-insured patients were prescribed 9% more drug items (3.46 items, compared to 3.17 per low-insured patient), and each item was 11% more expensive on average (R83.80, compared to R75.25 for low-insured patients).

As suggested previously, higher prescription costs for more insured patients may be partly explained by physicians' substitution of poorer diagnostic and therapeutic accuracy with higher treatment quantity and costs for these patients.[244] As high-insured patients are less sensitive to drug costs, this tendency for substitution may be higher with these patients. It was shown in Chapter 5 that physicians were 23% less likely to provide a correct or partially-correct diagnosis (conditional on pronouncing any diagnosis) and 24% more likely to prescribe inappropriate antibiotic treatment to high-insured patients, suggesting that diagnostic and therapeutic accuracy was indeed poorer for these patients compared to low-insured ones. Appendix D.2 indicates that a correct or partially-correct diagnosis is not significantly associated with prescription costs.

[243] This is based on the purchasing power parity rate of R6.2/USD for the South African Rand in 2018, published by the OECD and calculated from the OECD National Accounts Statistics – available from: https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm

[244] That is, if physicians believe more and more expensive drugs will compensate for lower clinical accuracy, by increasing the likelihood of successfully treating the unknown underlying illness.

However, the prescription of an antibiotic is associated with an R137 increase in total prescription costs, representing approximately 51% of the mean prescription cost in this experiment. This indicates that some level of substitution in physicians' appropriate effort for higher prescribing intensity and costs is plausible.

Table 6.3 presents estimates of the high-insurance effect on unnecessary drug costs, and follows a similar format to Table 6.2. Although no drug treatment is necessary for resolving the clinical case in this experiment, some OTC (low-schedule) symptomatic treatment may be recommended for reducing the patient's reported symptoms. Therefore, aggregate drug expenditures may be attributed to either appropriate (recommended, low-schedule symptomatic drugs) or inappropriate drugs, with the latter constituting one form of unnecessary drug expenditure. In most cases, physicians also have a choice of prescribing either a branded or cheaper generic version of the same drug. Branded drug prescriptions generate avoidable drug costs in these cases, and constitute a second form of unnecessary drug expenditure.

The results suggest that both types of unnecessary expenditure are higher for high-insured patients. Total inappropriate drug costs per patient are 16% higher for high-insured patients (R244, compared to R211 for low-insured patients). This is driven by both a marginally higher number of inappropriate drug items per prescription (2.65 items, compared to 2.48 for low-insured patients) and a higher average cost per inappropriate item (R93, compared to R83 for low-insured patients) – although the latter results are only weakly significant (at the 10% level). Greater diagnostic accuracy is again not significantly associated with inappropriate drug costs (Appendix D.3). The latter appear to be driven by antibiotic prescriptions (as with total prescription costs).

In terms of branded prescription items, Column (4) of Table 6.3 shows no statistically significant difference in the likelihood of a branded drug being prescribed across insurance types. However, the number of branded drug items per prescription is higher for high-insured patients: these patients were prescribed 1.42 branded drugs on average, compared to 1.13 branded drugs per low-insured patient. This represents a 26% increase in the number of branded prescription items with the higher insurance cover. The absolute difference in the number of branded drugs by insurance cover (0.29 drugs) is comparable to the difference in the number of all prescribed drugs (also 0.29 drugs). Given the lack of an equivalent difference in the number of generics prescribed by insurance cover (see Appendix D.1), this indicates that branded drugs are wholly driving the estimated difference in the quantity of all drugs prescribed. Total branded drug costs are also 21% higher for high-insured patients (R174, compared to R144 for low-insured patients), although this difference is only weakly significant. No significant difference is detected in the average cost of branded drug items. Therapeutic accuracy is negatively associated with branded drug costs (see Columns (5)-(7) of Appendix D.3): an inappropriate antibiotic prescription is associated with a

R78 increase in total branded drug costs and a R43 increase in the average branded drug cost (per prescription), which suggests that physicians may have a preference for originator drugs over generics when prescribing antibiotics.

Table 6.3. Effect of insurance cover on drug treatment intensity and costs – inappropriate and branded drugs

	(1) No. inapprop .drugs	(2) Average inapprop drug cost	(3) Total inapprop drug cost	(4) Any branded drug	(5) No. branded drugs	(6) Average branded drug cost	(7) Total branded drug cost
Panel A: provider-level fixed effects							
High-insured	0.17*	10.09*	33.14**	0.07	0.29***	-0.02	29.81*
	(0.10)	(5.50)	(16.48)	(0.05)	(0.11)	(11.34)	(17.43)
Obs.	178	177	178	178	178	136	178
R-squared	0.03	0.04	0.04	0.02	0.07	0.00	0.03
Panel B: provider-level fixed effects & cluster-robust standard errors							
High-insured	0.17*	10.09*	33.14**	0.07	0.29**	-0.02	29.81*
	(0.10)	(5.52)	(16.52)	(0.06)	(0.11)	(11.35)	(17.48)
Obs.	178	177	178	178	178	136	178
R-squared	0.03	0.04	0.04	0.02	0.07	0.00	0.03
Mean of low-insured	2.48	82.99	210.85	0.73	1.13	128.54	144.09

Notes: Standard errors are in parenthesis. All costs are in South African Rand (ZAR). Panel A presents estimates from a fixed effects (mean-differenced) model. Panel B shows results from a fixed effects model, as in Panel A, with the addition of a cluster-robust variance estimator. All outcomes are per prescription (per consultation).
*** p<0.01, ** p<0.05, * p<0.1

6.5.2 Effects of patient insurance cover on fee-for-service items and costs

Table 6.4 reports estimated high-insurance effects on all FFS items and costs from consultations. Unlike the drug prescription outcomes, this includes all outcomes that present direct financial incentives (gains) for physicians. Total FFS consultation costs (excluding drugs) were 7% higher for high-insured patients compared to low-insured ones (R485, compared to R455). As the consultation rate is generally fixed across the two insurance covers (the insurer sets the same maximum reimbursable consultation rate for both covers), the difference in average per-consultation costs is likely to be primarily driven by additional FFS items ordered in the consultations.[245] Table 6.4 (Columns (1) and (2)) shows that high-insured patients are significantly more likely to receive these additional items: they are twice as likely to be ordered any FFS procedure, and are ordered 2.5 times as many procedures on average. Nevertheless, the

[245] The provision of any such additional services is indeed associated with an increase in total consultation costs of R198 (see Column (3) of Appendix Table D.4).

absolute magnitudes of these differences are small, given the low overall incidence of such additional procedures in this experiment. This might explain the relatively small difference in total (per) consultation costs as well.

Table 6.4. Effect of insurance cover on treatment intensity and care costs – fee-for-service items

	(1) Any FFS tests / procedures	(2) No. FFS tests / procedures	(3) Total consultation cost (excl. drugs)
Panel A: provider-level fixed effects			
High-insured	0.06*	0.10**	29.91**
	(0.03)	(0.05)	(14.86)
Obs.	178	178	178
R-squared	0.03	0.04	0.04
Panel B: provider-level fixed effects & cluster-robust standard errors			
High-insured	0.06*	0.10**	29.91**
	(0.03)	(0.05)	(14.90)
Obs.	178	178	178
R-squared	0.03	0.04	0.04
Mean of low-insured	0.06	0.07	454.62

Notes: Standard errors are in parenthesis. FFS = fee-for-service. All costs are in South African Rand (ZAR). Panel A presents estimates from a fixed effects (mean-differenced) model. Panel B shows results from a fixed effects model, as in Panel A, with the addition of a cluster-robust variance estimator. All outcomes are per consultation.
*** p<0.01, ** p<0.05, * p<0.1

Appendix D.4 also indicates that physicians apply FFS charges that are proportional to their observed clinical effort: a one standard deviation increase in an IRT index of case-specific essential and recommended clinical effort is associated with an increase of R33 in total consultation costs. At the same time, greater diagnostic accuracy (a correct or partially-correct diagnosis) is associated with both lower overall consultation costs *and* a higher likelihood of additional FFS procedures. First, this implies that, unlike the observable aspects of effort reflected in the history-taking and examinations checklist, physicians are less able to price in their (unobserved) technical care quality (partly reflected in their diagnostic accuracy). Second, while diagnostic accuracy appears more likely with some additional costly test being performed, these additional services are ultimately unnecessary: low-insured patients are less likely to receive any additional FFS tests or procedures, and yet (as shown in Chapter 5) they are still more likely to receive a correct or partially-correct diagnosis.

6.5.3 Heterogeneous effects

The observed differences in treatment and cost outcomes by insurance cover are hypothesised to derive from physicians' competitive and altruistic incentives (see Chapter 3 for a more detailed

discussion). The predictions in Chapter 3 are nevertheless ambiguous on potential interactions between the effects of baseline physician altruism and competition, on the one hand, and the insurance effect on the other – in other words, whether a change in physician altruism or competition should affect outcomes for high- and low-insured patients to different degrees. Appendices D.6 – D.8 report results from formal tests for interaction effects between the measures of physician altruism and competition used in this study and patient insurance.[246]

No significant interaction effects are detected for the majority of outcomes. However, the insurance effect on the average cost of branded drugs prescribed is substantially larger in the sub-sample of physicians with high altruism. While Chapter 3 predicts that higher altruism should reduce drug costs for both patient insurance types, these results suggest that high-altruism physicians may exert relatively more cost-minimising effort for low-insured patients.

The insurance effect on the number of inappropriate drugs prescribed is also higher in the subsample of physicians located in the wealthiest northern suburbs of Johannesburg. This indicates that lower expected demand elasticity with respect to care costs (due to wealthier patients in these suburbs on average) increases inappropriate treatments relatively more for high-insured patients. Where the cost elasticity of aggregate demand is low, physicians' ability to compete on cost for high-insured patients is even lower, which generates very little incentive for cost-minimising effort. Physicians may instead try to compete more on treatment quantity for these patients, if they believe that less cost-sensitive patients want more care (i.e. display a higher degree of patient moral hazard).[247]

The same limitations as those noted in Chapter 5 apply to these analyses of heterogeneous insurance effects (see Section 5.5.2). Most importantly, the very small sample sizes (ranging from 113-178 observations) and lack of sufficient power for reliably detecting small-moderate interaction effects means I cannot rule out the presence of such effects where none were detected.

6.5.4 Robustness checks

As discussed in Section 6.4, a physician fixed effects model is just one of a few different approaches to dealing with potential data clustering at the physician level. I test the robustness of main results to two alternative approaches: a simple paired-sample t-test, and a mixed effects

[246] Section 4.2.3 details how these measures of altruism and competition were constructed, and Table 5.1 in Chapter 5 provides summary statistics for each measure (among other measures of GP characteristics). As in Chapter 5, two measures of GP competition are analysed here: a measure of local competitor density and GPs' location in the wealthier northern Johannesburg suburbs of Sandton and Randburg (with a wealthier patient population, and potentially lower price competition).

[247] In the context of the theoretical framework presented in Chapter 3, and in particular with respect to the optimality condition specified in equation (5), these subsample analyses results suggest that patients' marginal net benefit (as anticipated by physicians) is a concave, diminishing function with respect to cost-minimising effort.

model with physician-level random intercepts and slopes. Appendices D.9 - D.12 report the results from these tests, and all results are robust to these alternative specifications (including to the inclusion of fieldworker-pair fixed effects in the mixed effects model).

Appendix D.13 also presents results from a number of other robustness checks, as in Chapter 5. First, in Table Column (1), results are re-estimated excluding the four consultations where an SP detection was likely.[248] All effect sizes pertaining to prescription outcomes become marginally smaller in magnitude when the suspected SP detection cases are excluded. The estimated effects on average drug cost, average inappropriate drug cost, and total branded drug costs are pushed beyond the 10% significance threshold. However, the majority of results maintain their significance with the exclusion of these suspected SP detection cases.

As physicians may react differently to SPs depending on whether they had seen a similar clinical case recently (perhaps due to learning effects), the order of the two insured SP visits can confound the estimated insurance effect. The visit order was therefore randomised for each physician. To check for any residual order (learning) effects, Table Column (2) of Appendix D.13 presents results from a regression specification that explicitly controls for the SP visit order for each physician. All results are shown to be reasonably robust to this control.

Table Column (3) further controls for 38 consultations where SPs were asked to pay the physicians' in cash and claim back from the insurance later themselves.[249] Again, most results are shown to be reasonably robust (only the estimated effect on total consultation costs becomes smaller in magnitude and slightly less significant).[250] Lastly, Table Column (4) of Appendix D.13 checks the robustness of main results for continuous outcomes to the exclusion of extreme values.[251] Interestingly, most estimated effects become much larger in magnitude and more statistically significant. One exception is again the estimated effect on total consultation costs, which maintains its significance but almost halves in size.

---

[248] As previously explained, an SP detection survey was carried out with participating physicians shortly after the completion of all SP visits, where they were asked for details of patients they had suspected to be SPs in previous months. Each reported suspicion was then analysed independently by four researchers (including myself) and compared to the actual SP visits to those physicians. Decisions on valid detections were taken jointly, and it was decided that two physicians in the sample had a relatively high likelihood of a valid SP detection. The data taken from consultations with those two physicians are therefore excluded in the estimates presented in Column (1) of Appendix D.13.

[249] In the majority of cases, the physician's practice claimed from the insurer directly.

[250] One reason for this may be that some of the estimated difference in total consultation costs by insurance cover may be driven by differences in the number of cash consultations: high-insured patients had slightly more cash consultations (two consultations more) than low-insured patients, and while the difference is not large, a cash consultation is associated with a substantial increase in total consultation costs of R351. Explicitly controlling for cash consultations may therefore marginally lower the estimated high-insurance effect on total consultation costs.

[251] Extreme values (outliers) are classified here as values that are more than 3 standard deviations higher or lower than the sample mean. This outlier robustness check was not conducted for results on the number of fee-for-service procedures, as the overall incidence of such procedures was very low, and removing outliers removed all 5 observations where more than one procedure was administered (and left only 10 observations in the sample where any procedure was administered).

6.6 Conclusions and Discussion

This chapter investigates the impact of patient insurance cover on physicians' choices of treatment intensity and costs. In utilising a within-physician audit study approach, it succeeds in isolating physician moral hazard from patient moral hazard and selection, and in distinguishing wasteful from optimal care – problems that have long impeded empirical work in this area (Johnson, 2014; Einav and Finkelstein, 2017). It finds that physicians generate more and more unnecessary care expenditures for high-insured patients both in the presence *and* absence of any direct financial gain. Physicians in this experiment have no financial interest in drug prescribing. Yet, high-insured patients receive prescriptions that are on average 17% more expensive than low-insured patients. This difference is driven by both more over-treatment and more over-charging (in receiving more branded drugs in place of cheaper generic equivalents). High-insured patients also incur higher costs due to physicians' direct financial incentives. They are twice as likely to be ordered unnecessary FFS tests and procedures during their consultations - although this contributes to only marginally higher consultation costs on average (7% higher, given a low incidence of such additional services overall).[252]

While this chapter finds higher unnecessary FFS and prescription costs for high-insured patients, the previous chapter provides supportive evidence that diagnostic and therapeutic accuracy is also poorer for these patients: high-insured patients were 23% less likely to receive a reasonably accurate diagnosis, and 24% more likely to be prescribed an inappropriate antibiotic. While diagnostic accuracy is not significantly associated with prescription costs, an inappropriate antibiotic prescription is associated with an increase of R137 (representing 51% of the mean prescription cost in this experiment). A correct or partially-correct diagnosis is also associated with a small decrease of R25 in total consultation costs (FFS costs). Taken together, these results are consistent with the notion that providers may compensate for poorer diagnostic and therapeutic accuracy (and their under-supply of cost-minimising effort) with more and more costly treatments for high-insured patients – perhaps due to physicians' anticipation of greater demand (or willingness-to-pay) for such treatments from these patients. The results further highlight how under-provision in care quality can co-exist with (and perhaps drive) over-provision and over-charging of treatment quantity for better insured patients. [253]

---

[252] The clinical case used in this experiment (uncomplicated acute bronchitis) is less amenable to over-treatment through diagnostic testing and treatment procedures than through drug prescription, as a fairly accurate diagnosis is possible with only clinical history-taking and examination and without any recourse to diagnostic testing. This may explain the more significant and robust results with respect to drug prescribing.

[253] This reflects the framework predictions in Chapter 3 (*Hypothesis 5*) where the quantity and cost of medications are decreasing in appropriate provider effort, and are hence larger for high-insured patients at the physician's optimum. The results further suggest that physicians' fee-for-service items (denoted x in Chapter 3) should also be a decreasing function of provider effort.

The finding that branded drugs primarily drive the estimated insurance effect on prescription costs is also consistent with costly habits in prescribing practices (Hellerstein, 1998; Rischatsch *et al.*, 2013). Rischatsch *et al*. (2013) suggest that resistance to generic substitution may be overcome if physicians' care for patients' financial welfare and if the cost savings to patients are substantial enough, which could explain the estimated differences in branded drug prescriptions by insurance cover. It may also be that physicians' anticipate patients to prefer branded drugs over generics, especially where their OOP costs are low. This expectation is not unreasonable, given evidence that patients do often prefer more expensive, branded drugs (Himmel *et al.*, 2005; Shrank *et al.*, 2009). At the same time, patient demand for prescription drugs appears highly elastic to OOP costs (Leibowitz *et al.*, 1985). So it is further reasonable that physicians anticipate greater demand for more and more branded drugs from high-insured patients.[254]

Although higher insurance is shown to increase prescription costs, the rate of unnecessary prescribing is still very high for low-insured patients: the average prescription for low-insured patients contained 2.48 inappropriate drugs on average (at a total cost of R211), and 73% of all prescriptions for low-insured patients contained at least one branded drug. This may reflect a general problem of over-prescribing, as noted also in other contexts (Lu, 2014; Das and Hammer, 2007; Currie *et al.*, 2011). The high level of non-incentivised costs in general, even for low-insured patients, is in line with the implications of *Hypothesis 3* in Chapter 3: when some uncompensated effort is required to minimise treatment costs, physicians will in general under-supply this effort and generate higher treatment costs (that entail no cost to them) than what is necessary for the patient.

Interestingly, overall rates of over-treatment – as well as insurance effects on unnecessary treatment costs - are much higher for financially non-incentivised drug prescriptions than for FFS procedures in this experiment.[255] This may be explained by the limited scope for additional FFS procedures within the specific clinical case and primary care context of this experiment. Firstly, given the uncomplicated nature of the acute bronchitis case and the limited capacity available to physicians to provide additional services in-house in primary care clinics (or through external referrals that can guarantee follow-up consultations), physicians' scope for inducing demand and

[254] This chapter's findings are therefore consistent with both the notions of *i)* physicians being *perfect* agents for patients, acting on anticipated patient moral hazard and preferences (that is, providing more and more costly treatments to high-insured patients that can afford it, and lowering out-of-pocket costs for less-insured patients), and of *ii)* physicians being *imperfect* agents, inducing more demand from patients with a higher expected willingness-to-pay, to substitute for lower costly effort in minimising treatment costs. Both cases nevertheless represent imperfect agency with respect to the third-party payer. Unfortunately, the experiment was not designed to discriminate between these two distinct hypotheses in explaining the observed results. Further qualitative work may be useful to validate and better distinguish such underlying drivers.

[255] For example, an unnecessary FFS procedure was administered in only 8% of consultations, whereas a branded drug was prescribed in 76% of consultations, and each prescription contained 2.6 inappropriate drugs on average (at a total cost of inappropriate drugs of R227).

providing unnecessary procedures is very low. Secondly, the potential for repeated interactions with the same patient is arguably higher in the primary care setting of this study, which simultaneously increases physicians' market-based incentives to supply services they believe patients want (such as prescription drugs) and their reputational pressures to limit demand-inducement purely for profit. It is perhaps reasonable to expect that SID and *conventional* second-degree moral hazard would be more prevalent in more severe clinical cases that require curative or emergency treatment (where patients' ability to consider the necessity of clinical decisions would be more limited) and in higher levels of care (where repeated interactions would be less common).[256]

The results further reveal some heterogeneous insurance effects. While Chapter 3 predicts that higher altruism or competitive pressures should reduce physicians' demand-inducement and effort-stinting behaviour for all patients, this chapter finds that these factors may affect certain cost outcomes for high- and low-insured patients to different degrees. For instance, the insurance effect on the average cost of branded drugs is significantly higher in the subsample of high-altruism physicians.[257] The parallel effect on the number of inappropriate drugs prescribed is higher for physicians located in the wealthiest suburbs of Johannesburg, which suggests that lowering physicians' cost pressures may lead to greater increases in unnecessary care costs for high-insured patients (relative to low-insured ones). A possible explanation for this latter result is that the high average level of physician altruism in the study sample may be attenuating the expected effects of low competition on cost outcomes for low-insured patients.

Overall, the findings highlight the potential for exacerbating inefficiencies in healthcare provision under market incentives - in the form of over-provision or over-charging of treatment quantity - with greater financial protection for patients. They build on an earlier finding of another, simultaneous problem of higher insurance: greater under-provision in care quality, as evidenced in Chapter 5. In the context of the South African government's latest proposals to eliminate patient co-payment in private insurance schemes, and contract private primary care physicians as gate-keepers to its prospective NHI system, these results suggest that even incremental increases in patients' financial protection can have significant cost implications by altering physician

[256] This could partly explain why estimated insurance effects under financial incentives are much higher in Lu (2014)'s hospital setting than in the present study (a 43% increase in drug costs for insured patients in Lu (2014), compared to a 7% increase in FFS costs for high-insured patients in this study). It could also explain the relatively larger conventional second-degree moral hazard effects detected in computer repairs (80% higher costs for insured customers in Kerschbamer *et al.* (2016)) and taxi rides (17% higher charges for customers claiming employer travel reimbursement in Balafoutas *et al.* (2017)), where the scope for increasing servicing is perhaps greater - and the scope for repeated interactions perhaps lower - than in primary care.

[257] The insurance effect on the *number* of branded drugs prescribed does not significantly vary by physician altruism, however. It may be that physicians expect all patients (including low-insured ones) to have a preference for branded over generic drugs. Therefore, rather than further lowering the relative number of branded drugs prescribed to low-insured patients, high-altruism physicians may seek to prescribe less expensive branded drug items to these patients as an alternative means to minimising their OOP costs (although no significant heterogeneous effects are detected on total branded prescription costs).

incentives. Presumably, these effects would be larger for patients transitioning from no insurance at all, and where potential demand-side moral hazard (as predicted in standard economic theory) can also occur.

It is important to contextualise these findings with respect to the fixed supply-side incentives in this study. While the chapter says nothing about the relative merits of alternative market or payment conditions, some implications of the non-dispensing status and FFS payment of physicians in this study may be inferred. This chapter finds that increasing patients' financial protection can still worsen unnecessary prescribing and impede generic substitution, even in a context where physicians cannot directly profit from prescribing. The suggestion is that eliminating the positive financial incentives on treatment quantity (FFS payment) may reduce *conventional* second-degree moral hazard, but so long as appropriate physician effort remains uncompensated or un-incentivised, *indirect* second-degree moral hazard may still persist with payment reform.[258]

Inappropriate antibiotic prescriptions and branded drugs are the main drivers of prescription cost differences in this study. Initiatives to rationalise antibiotic prescribing and encourage generic substitution could therefore attenuate unnecessary expenditures, where payment reform proves insufficient. This will require more than just improving physicians' knowledge or clinical competence, which were controlled for in this experiment. Finding ways to incentivise non-contractible effort, especially where both altruistic and competitive incentives for supplying such effort are already high (for instance, with cost-minimising effort for low-insured or uninsured patients), will be key.

There are a number of limitations to this chapter's conclusions. As with Chapter 5, external validity is a concern given the limited scale and scope of the experiment. The focus on the primary care sector and a single, uncomplicated clinical case (where no curative drug treatment is indicated) limits extending conclusions to more serious cases, particularly in secondary or tertiary care where informational asymmetries between doctor and patient are likely to be more severe. This is a key limitation of the SP method, where clinical cases cannot present with observable physical symptoms or invite invasive investigations. The lack of a perfectly random study sample, given the ethical requirement to secure informed consent from all research participants, warrants further caution in extending conclusions to the wider GP population. Future research should test the validity of these findings in broader contexts and samples, and with different clinical cases.

[258] According to the theoretical framework in Chapter 3, this result holds as long as physicians face some level of market competition, or are sufficiently altruistic towards patients' financial welfare.

A more conceptual limitation of this chapter is that, while it finds results consistent with the two physician moral hazard hypotheses, it provides only suggestive evidence on the underlying mechanisms that are hypothesised to drive them. More accurate measurement and greater variation in measures of physician competition and altruism would allow a more robust analysis of the potential influence of these factors. A larger sample size should also improve statistical power to facilitate analysis of heterogeneous effects. The presence and relative importance of the proposed behavioural channels behind these results - of physicians inducing more intensive and expensive treatment when diagnostic and therapeutic accuracy is low, anticipating better-insured patients to expect or accept more care, or maintaining costly treatment habits - must be validated in future qualitative research. In addition, future research testing the effects of varying the over-arching financial or market incentives facing physicians – which are fixed in this study (FFS payment and non-dispensing status) and suggested to partly drive estimated results – would be an insightful extension.

7 Private Performance Feedback and Appropriate Care

7.1 Introduction

A common assertion in the health economics and agency literature is that healthcare providers are, to some extent, intrinsically motivated to fulfil their roles (Bénabou and Tirole, 2003; Galizzi *et al.*, 2015). Unrelated to profit or reputational gains, intrinsic motivation is often associated with providers' self-image, and their ability to knowingly perform well relative to a trusted benchmark (Kolstad, 2013; Eilermann *et al*., 2019) – be it peer norms, or best-practice standards.[259] Previous studies have accordingly shown that such motivation may be leveraged through the provision of information that makes effort more intrinsically rewarding – either in bridging existing knowledge gaps on benchmark standards, or in better informing providers of their own relative performance (Kolstad, 2013; Lee, 2018). In a context of excessive antibiotic use in the South African primary care sector, this chapter tests the incremental effectiveness of a private audit and feedback (A&F) intervention – informing physicians of their prescribing performance relative to evidence-based guidelines – compared to providing educational materials on the guidelines alone.

Earlier findings from this study and others have shown that inappropriate antibiotic prescribing for uncomplicated respiratory infections is in general very high, and can persist even in the absence of any patient demand (Currie *et al.,* 2014; Blaauw and Lagarde, 2019) - highlighting a key role for physicians' decision-making. Despite the availability of clear, evidence-based guidelines on appropriate antibiotic use, the results show significant gaps in their uptake in daily clinical practice. Therefore, identifying physician-level initiatives that can best address this evidence-practice gap is an urgent policy concern.

Such gaps have been commonly linked to providers' lack of guidelines knowledge or diagnostic uncertainty; persistent habits and prescribing norms; and their limited ability to self-assess their own practice (Davis *et al.,* 2006, Brink, Van Wyk, *et al.,* 2016; Tonkin-Crine *et al.,* 2017). Accordingly, educational materials and performance audit and feedback (A&F) – often linked to

[259] It follows that, in order to be effective, intrinsically-motivated providers must have correct knowledge of the relevant benchmark *and* be aware of how they perform relative to it.

published performance scorecards or financial incentives – have been widely used in performance improvement initiatives. Previous evaluations of these initiatives highlight a number of implications and limitations.

Firstly, in comparisons of alternative interventions, physician education has been found frequently effective in lowering antibiotic prescriptions in primary care - often in combination with other intervention elements, such as A&F (van der Velden *et al.*, 2012). Nevertheless, the quality of this evidence base is generally low (Tonkin-Crine *et al.*, 2017); and repeated findings of a "know-do" gap in clinical practice (Das and Gertler, 2007; Leonard and Masatu, 2010)[260] suggest that physician education alone may only be of limited value in improving clinical performance, without additional incentives to raise performance in line with knowledge. Evidence on the incremental incentive effect of A&F - in bolstering educational interventions - is scarce.

Secondly, in designing A&F initiatives, whether audited performance data is kept confidential or disclosed to a wider audience has important implications for associated incentives. Public disclosure of audit data (and its incorporation in financial incentive schemes) has been highly contested, particularly given problems in systematically measuring case-specific and risk-adjusted quality of care through simple, finite metrics. A growing body of evidence further suggests that appealing to such extrinsic motives as financial or reputational concerns alone may be insufficient for improving care quality (Ketelaar *et al.*, 2011; Scott *et al.*, 2011; Berenson and Rice, 2015) and can indeed lead to perverse provider responses (Werner and Asch, 2005) – including up-coding of clinical data (Gravelle, Sutton and Ma, 2010), and 'cream-skimming' of low-risk patients (Schneider and Epstein, 1996; Dranove *et al.*, 2003; Wadhera *et al.*, 2018).[261] Private A&F instead appeals to physicians' intrinsic motivations (or professionalism), and is hypothesised to work by providing physicians new information on their performance relative to a trusted benchmark. In this view, physicians are believed to be intrinsically motivated to improve care, but hampered by their limited ability to self-assess their own performance (Davis *et al.*, 2006). The new, private information from A&F increases the intrinsic returns to effort, by making physicians more aware of their current relative performance and subsequent improvement. Kolstad (2013) finds that such private audit information targeting intrinsic incentives is many times more effective in improving provider performance than public audit information targeting market incentives (market demand).

[260] This refers to instances where physicians' actual clinical performance falls below that predicted by measures of their knowledge.
[261] 'Up-coding' (or 'diagnosis-shifting') refers to manipulation of reported diagnoses in administrative records, to record more severe or antibiotic-appropriate diagnoses than the true (known) diagnosis, in order to justify antibiotic treatment. 'Cream-skimming' refers to the deliberate selection by providers of less complicated clinical cases or low-risk patients for treatment.

Relatedly, in evaluating A&F effects, accurately measuring all intended effects, as well as unintended spill-overs on non-targeted aspects of care, has proved challenging. A common reliance on administrative data limits the scope and accuracy of performance measurement. In general, not all patient and clinical case characteristics that determine appropriate care are measured. Therefore, patient and case selection cannot be fully controlled for, and risk-adjustment in performance measurement is likely to be imperfect (Dranove *et al.*, 2003).[262] The appropriate use of antibiotics further depends on the diagnosis, and the accuracy of the latter cannot be verified from physician reports alone – given the scope for diagnostic error or inaccurate reporting.[263] In addition, targeting aggregate antibiotic prescription rates in A&F interventions can generate unwanted spill-overs if physicians respond by reducing both inappropriate *and* appropriate antibiotic prescribing. The potential for such unintended effects is in general ignored in previous intervention designs and evaluations.[264]

With these considerations in mind, the present chapter evaluates a small, randomised controlled trial (RCT) testing the effects of private, individualised A&F in lowering inappropriate antibiotic treatment – where individual performance was reported relative to standard clinical guidelines - compared to the provision of educational material on the guidelines alone.[265] Standardised patients (SP) were used to measure performance outcomes, to overcome the limitations of administrative data in impact evaluation: all case and patient characteristics were standardised, so it was possible to determine appropriate care *ex ante* and to isolate and measure physicians' care quality performance *ex post.* The RCT was conducted with 80 private, primary care physicians in Johannesburg. At baseline, all participating physicians received incognito visits from SPs presenting with an uncomplicated case of acute bronchitis, where no antibiotic treatment was clinically indicated. Physicians were then randomised to one of two interventions: *i)* receipt of written educational material on AMR and clinical guidelines for the diagnosis and treatment of common respiratory illnesses, or *ii)* receipt of the same educational material *and* a private A&F intervention. The A&F intervention provided written feedback to physicians on key aspects of their individual performance during baseline SP visits, including whether an antibiotic was

[262] "Risk-adjustment" refers to adjustment of performance outcomes (e.g. patient health outcomes) for differences in patient characteristics and the severity of clinical cases.

[263] As such, where pecuniary or reputational incentives are attached to measured performance, the potential for providers to game quality monitoring systems, either through the selection of lower-risk patients or the manipulation of reported diagnoses, cannot be controlled for.

[264] One exception is Meeker *et al*. (2016), who evaluate possible side-effects of peer-benchmarked A&F on diagnosis-shifting (from antibiotic-inappropriate to antibiotic-appropriate diagnoses) and repeat consultations for concerning diagnoses following consultations where antibiotics were not prescribed (indicating negative effects on appropriate antibiotic treatment). However, the authors' reliance on electronic health record (EHR) and billing data is not optimal for ascertaining the true diagnoses and measuring clinical performance.

[265] This corresponds to a test of *Hypothesis 7* in Chapter 3.

inappropriately prescribed.[266] Following the intervention, end-line SP visits were conducted with the same clinical case to evaluate the impact of private A&F on reported aspects of physician performance. Unlike most previous studies, this experiment further tests for potential unintended effects of A&F on *appropriate* prescribing behaviour. A second SP clinical case where antibiotics would be clinically indicated - a more severe acute bronchitis case, in a HIV+ young adult - was also presented to participants at end-line, and rates of appropriate antibiotic prescribing for this case were compared across the two intervention groups.

Three main findings are reported. First, the results suggest that private A&F can significantly reduce inappropriate antibiotic treatments: physicians randomised to the A&F intervention group were 46% less likely to give inappropriate antibiotic treatment at end-line compared to physicians that received educational materials alone.[267] Second, this result does not coincide with any significant differences in observable diagnostic effort or accuracy, suggesting that measured diagnostic performance may not be the primary barrier to appropriate treatment choices. Descriptive evidence points to habit persistence, and social norms in prescribing practices, as alternative barriers to appropriate antibiotic use. If physicians' ability to objectively self-assess and improve their performance is hindered by engrained habits or social norms, it is possible that these barriers can be overcome with the new, objective information provided by A&F. Third, there is no evidence of negative spill-overs from the A&F intervention - either in reducing appropriate antibiotic treatments, or in encouraging the substitution of other inappropriate drugs. Physicians in the A&F treatment group were no less likely to prescribe clinically-indicated antibiotics for the more severe HIV+ patient case, and no more likely to prescribe other inappropriate drugs such as steroids.[268]

While previous studies have evaluated the effects of A&F on antibiotic use, this chapter presents new causal evidence of its incremental effect relative to an educational support alone. The intervention also represents a novel initiative targeting prescribing practices in the primary care sector of South Africa - a country where most (if not all) antibiotic stewardship initiatives are focused on secondary care, despite findings that the majority of antibiotics are being prescribed in primary care (and mostly for respiratory infections) (Brink, Van Wyk, *et al.*, 2016; Tonkin-Crine *et al.*, 2017). While its cost-effectiveness remains to be assessed, the results corroborate

[266] Two other performance outcomes were also reported: the proportion of all case-specific recommended or essential history-taking and examinations that were completed, and the cost of all drugs dispensed or prescribed (relative to the cheapest, recommended treatment on the local market for the case). Nevertheless, inappropriate antibiotic prescribing was the primary component of the A&F intervention, and performance on this aspect was purposefully emphasised in the written feedback form. The stronger emphasis was achieved by listing this performance aspect first, before the other two reported aspects, and using a larger font. See Appendix B.10 (a) for an example of the feedback presentation.

[267] This result is calculated based on an end-line antibiotic treatment rate of 52% in the control group.

[268] It is plausible that the provision of educational materials also to the A&F treatment group, on recommended treatments for a number of respiratory conditions of varying severity, may have acted to minimise the risk of such unintended treatment effects.

similar conclusions elsewhere (Meeker *et al.*, 2016) that the addition of individualised, private A&F to simple educational interventions may be a promising initiative to explore for rationalising antibiotic treatment choices on a larger scale.[269]

The remainder of the chapter is organised as follows. Section 7.2 summaries the chapter's main contributions to related literature. Section 7.3 outlines the RCT design and implementation, and summarises descriptive statistics on participants' baseline characteristics and outcomes of interest. Section 7.4 presents the empirical framework, main results and robustness checks. Finally, Section 7.5 concludes with a discussion on the main findings, policy implications and limitations of the chapter's analyses.

7.2 Literature Contributions

This chapter contributes to four bodies of literature. First, it builds on a large literature in economics that posits that agents – including healthcare providers – can have intrinsic motivations for supplying effort (Bénabou and Tirole, 2003; Berenson and Rice, 2015; Galizzi *et al.*, 2015). "Intrinsic motivation" is broadly defined here as any motivation for effort that is unrelated to profit or reputational concerns. Intrinsically-motivated agents derive utility from performing a task for its own sake ('pure' intrinsic utility), performing well relative to a trusted benchmark (professionalism, or 'reference-based' utility), or performing tasks with positive social externalities (altruism, or 'pro-social' utility). Direct evidence of intrinsic motivation (and relatedly, altruism) among healthcare providers mainly comes from surveys or laboratory experiments (Sicsic, Le Vaillant and Franc, 2012; Godager and Wiesen, 2013; Hennig-Schmidt and Wiesen, 2014). Existing literature often depicts such motivation as an exogenous and fixed trait – one which, for example, explains why some individuals self-select into certain types of jobs (including low-paying or rural jobs) that may be deemed unattractive under standard economic models of utility-maximisation (Kolstad and Lindkvist, 2013; Ashraf, Bandiera and Lee, 2018). In this view, organisations can ensure intrinsically-motivated agents only through selection, rather than through influence (Besley and Ghatak, 2005; Heyes, 2005; Delfgaauw and Dur, 2008; Brekke and Nyborg, 2010; Smith *et al.*, 2013).[270] This chapter instead aligns more closely with a recent literature proposing that agents' intrinsic motivation can indeed be nurtured

[269] Meeker *et al.* (2016) evaluate a similar intervention to the present one with GPs in the US. They provide the same best-practice guidelines information to both treatment and control group GPs (as in this study) and find a marginal A&F effect of similar magnitude (a 47% reduction in inappropriate antibiotic use). Their study differs in two ways to the present one: they report individual performance relative to an anonymised peer performance benchmark (rather than best-practice guidelines) and use electronic health records (EHR) data to measure performance improvement (instead of SPs).

[270] Theories on motivational crowd-out moreover imply that, where such intrinsic motivation can be altered, it may only be reduced (Bénabou and Tirole, 2006). Empirical evidence of reductions in agent effort following the introduction of financial or other extrinsic incentives lend support to these theories (Deci, Koestner and Ryan, 1999; Georgellis, Iossa and Tabvuma, 2010).

or mobilised by organisations (Ashraf and Bandiera, 2017a), through the design of organisational, job or task attributes, and the provision of explicit intrinsic incentives for performance.

In particular, this chapter develops a nascent body of empirical work evaluating intrinsic *informational* incentives for effort. Private performance feedback (A&F) has been proven effective in improving health worker quality and productivity (Ivers *et al.*, 2012; Kolstad, 2013; Lee, 2018; Eilermann *et al*., 2019), particularly where baseline performance is low. The proposed mechanism is that the marginal intrinsic (reference-based) utility from effort is increased when health workers are supplied with new information on their relative performance.[271] Such interventions are distinct from those targeting *extrinsic* information incentives, in the form of public reporting or peer information sharing (Bandiera, Barankay and Rasul, 2005; Mas and Moretti, 2009; Ketelaar *et al.*, 2011), which trigger external reputational or competitive concerns.[272] Private A&F interventions often involve (anonymised) peer comparisons (Kolstad, 2013; Elouafkaoui *et al.*, 2016; Hallsworth *et al.*, 2016; Lee, 2018), where individual performance is reported relative to some aggregate peer performance benchmark. The intervention evaluated here instead benchmarks audited performance to objective, evidence-based clinical guidelines. The benefit of this approach is that it will be appropriate in all contexts, even where – as in this study's context – average peer performance is also low. Moreover, there is little evidence that peer comparison in private A&F is any more effective than comparison to evidence-based clinical standards (Elouafkaoui *et al.*, 2016).

Benchmarking audited performance to clinical guidelines can also improve performance in two distinct ways: by bridging a gap in physicians' knowledge of the guidelines themselves (improving knowledge) and by providing new information to physicians on their relative performance (aiding self-assessment). Although both mechanisms can be considered informational incentives that raise the marginal intrinsic utility to effort (see Section 3.4.2 for a discussion on this), they have very different policy implications.[273] If knowledge of guidelines is the primary effort constraint for physicians, then a less resource-intensive educational intervention than A&F may be just as effective. Other studies evaluating A&F interventions that benchmark performance to best-practice standards do not untangle these two channels (Eilermann

[271] The implication is that a larger gap between baseline performance and the 'best practice' benchmark indicates a greater degree of 'new' information for physicians following performance feedback, thereby stimulating a larger improvement in performance. This supports the idea that A&F interventions work by overcoming physicians' limited ability to self-assess their own performance (Davis *et al.*, 2006).

[272] The latter have proven less effective than intrinsic incentives, and can indeed result in perverse effects (Werner and Asch, 2005; Kolstad, 2013), Public reporting has been found effective in improving quality at the hospital level; however, there is little (if any) evidence of its effectiveness at the physician level (Marshall *et al.*, 2000; Shekelle *et al.*, 2008) or in stimulating the targeted demand response (Epstein, 2006; Ketelaar *et al.*, 2011), and indeed some indication of system 'gaming' through provider selection of low-risk patients and manipulation of clinical case coding (Dranove *et al.*, 2003; Gravelle, Sutton and Ma, 2010).

[273] Only the latter mechanism corresponds to a pure incentive. The former mechanism can also correspond to an effort "enabler," that reduces marginal effort costs by bridging a knowledge gap.

et al., 2019). The intervention evaluated in this chapter therefore presents a novel addition to this literature: by providing basic educational materials on best-practice guidelines to all physicians (in both the A&F treatment and control groups), it ensures that all participating physicians have access to the guidelines information, and thereby isolates the pure incentive effect of private A&F from the knowledge-improvement effect of guidelines information alone.

A third, specific body of related literature evaluates physician-targeted interventions to rationalise antibiotic treatment for respiratory infections (van der Velden *et al.*, 2012; Tonkin-Crine *et al.*, 2017).[274] A review by van der Velden *et al.* (2012) found that multifaceted interventions, that contained at least some educational materials for the physician, were most often effective in lowering overall antibiotic prescription rates. Tonkin-Crine *et al.* (2017)'s review of systematic reviews concludes, however, that available RCT evidence on the effectiveness of multifaceted or single interventions containing educational materials for physicians is generally of low quality. Moreover, the incremental effect of private A&F in improving the impact of physician education alone is unclear. While a few RCTs have evaluated private A&F interventions (Mainous *et al.*, 2000; Gerber *et al.*, 2013; Elouafkaoui *et al.*, 2016; Hallsworth *et al.*, 2016; Meeker *et al.*, 2016), they often include A&F as part of multifaceted interventions which prevent an understanding of its incremental effect (Gerber *et al.*, 2013; Hallsworth *et al.*, 2016); evaluate A&F interventions that feedback on organisation-level performance (rather than physician-level performance) (Mainous *et al.*, 2000; Hallsworth *et al.*, 2016); and evaluate effects on aggregate prescription rates, rather than untangling the desired effects on inappropriate prescriptions alone (Elouafkaoui *et al.*, 2016; Hallsworth *et al.*, 2016). One exception that is closely aligned to the present chapter is Meeker *et al.* (2016), who find that private, individualised A&F on inappropriate prescription rates (prescriptions for antibiotic-inappropriate respiratory diagnoses), combined with guidelines education, lowered inappropriate prescription rates by 47% compared to guidelines education alone. Meeker *et al.* (2016) benchmark individual performance to high-performing peers rather than objective, best-practice guidelines. Moreover, as in other studies in this literature, their reliance on administrative, physician-reported (EHR) data to evaluate prescription performance limits the accurate assessment of prescription appropriateness (which necessarily relies on reported diagnoses).

This chapter therefore builds on a fourth body of literature utilising incognito SPs for evaluating physicians' care quality performance (Beullens *et al.*, 1997; Glassman *et al.*, 2000; Das and Gertler, 2007; Rethans *et al.*, 2007; Mohanan *et al.*, 2012; Das and Hammer, 2014; Daniels *et al.*,

[274] A range of interventions have been tested, including the provision of educational materials, educational meetings, outreach visits, A&F, point-of-care testing, financial incentives, delayed prescription strategies and enhanced patient communication (van der Velden *et al.*, 2012; Tonkin-Crine *et al.*, 2017).

2017). The SP method allows researchers to fix the clinical case (and hence, the diagnosis) and important patient characteristics, and thereby establish optimal care and prescription outcomes *ex ante*. It overcomes any potential for provider selection of low-risk patients, or manipulation of diagnoses coding, which can bias measured performance. Despite its growing use in evaluating provider quality in observational studies, the incognito SP method has rarely been employed in evaluating randomised provider quality interventions (Das, Chowdhury *et al.,* 2016; Harrison *et al.,* 2000, Mathews *et al.,* 2009; Mohanan *et al.,* 2017). One exception is Mohanan *et al.* (2017), who use SPs to evaluate the impact of a social franchising and telemedicine intervention in improving the clinical management of childhood illnesses in Bihar, India. Another exception is Das, Chowdhury *et al.* (2016), who use SPs to evaluate the care quality impact of a training program for informal healthcare providers in West Bengal, India.[275] This chapter further demonstrates this novel application of the incognito SP method for better evaluating randomised physician-targeted interventions.

7.3 Experimental Design, Implementation and Data

The randomised experiment evaluated in this chapter is discussed in detail in Section 4.3.2, and corresponds to Part 2 of this thesis study. An overview of the experiment is provided in the sub-sections below.

7.3.1 Sampling, recruitment and randomisation

The experiment was conducted with private, primary care physicians (GPs) in the City of Johannesburg (CoJ), South Africa. The GP sample was initially recruited as part of the related ESRC study. GPs were contacted for recruitment over February – May 2018, and SP visits presenting an uncomplicated acute bronchitis case (Clinical Case 1)[276] were completed with 122 recruited GPs in June 2018. The baseline data for this experiment was collected from those visits. Of these 122 GPs successfully visited at baseline, 99 were deemed eligible for participation in this experiment.

[275] Li, Lin and Guan (2013) also use SPs to evaluate the efficacy of a randomised hospital-based intervention in China in changing provider behaviour toward HIV-positive patients, although they do not address care quality *per se.* Sanci *et al.* (2000) use SPs to evaluate a randomised educational intervention targeting the quality of adolescent care. However, the SP consultations in that study were not blinded (physicians were aware that the consulting patient was an SP, and that their consultation was being recorded), which raises the risk of observation (Hawthorne) effects biasing measured performance.

[276] See Section 4.1.1, and Appendices B.1 and B.2, for details on the corresponding SP clinical case.

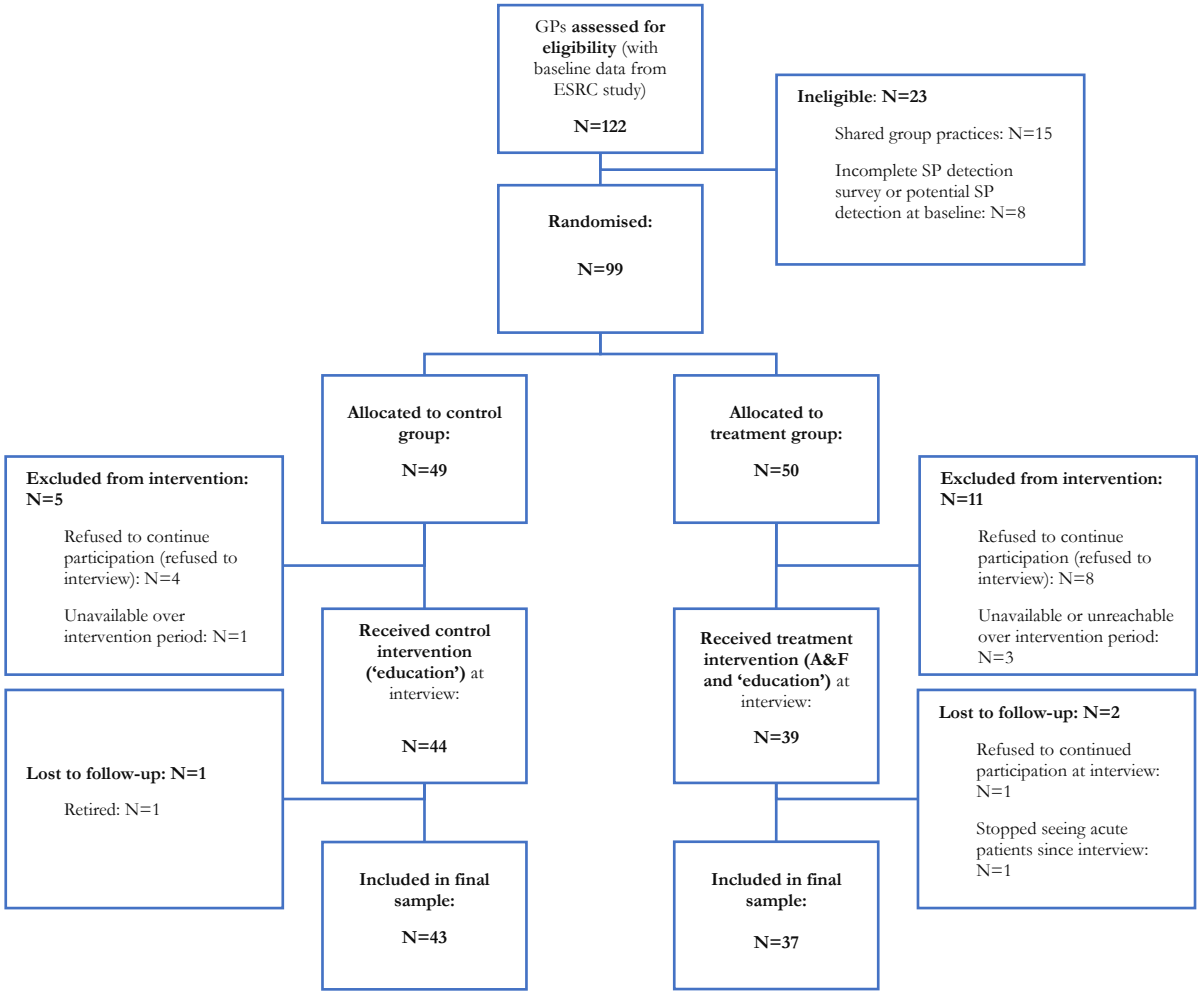


Figure 7.1. Experiment participation – GP participant flow-chart

Eligible GPs were those who were not in shared group practices[277] and not assessed to have detected any SPs during baseline visits (based on an SP detection survey conducted after the visits). The 99 eligible GPs were then stratified according to baseline performance on the main outcome of interest (whether they prescribed an inappropriate antibiotic during the baseline SP visits), as well as their drug dispensing and medical scheme contracting status, and randomised into one of two intervention arms: *i)* receipt of educational materials with private performance feedback ('treatment' group), or *ii)* receipt of educational materials only ('control' group) – see

[277] All GPs that shared a practice with another GP in the 122 assessed for eligibility were excluded. This was due to logistical and fieldworker constraints, as having many participating GPs practicing in the same clinics reduces the number of fieldworkers available to complete all visits (the same fieldworker could not visit the same clinic twice, due to the heightened risk of detection). It was also designed to limit potential treatment spillovers between treated and control group GPs practicing in the same location. 15 group practice GPs were excluded as a result. Despite these efforts to exclude group practice GPs, there were still a few cases of GPs that shared group practices being retained in the final sample (6 GPs), due to errors in GPs' practice address listings.

below. Randomisation was carried out in November 2018, using a random number generator in Excel, resulting in 50 GPs being randomised to the treatment group and 49 to the control group.[278]

See Figure 7.1 for an overview of GP participation, from initial experiment eligibility to final sample composition.

7.3.2 Feedback intervention and implementation

Following random assignment, the 99 GPs were contacted for a face-to-face interview to be conducted at their practices. The experimental intervention was delivered at the end of these interviews. GPs were not informed about the intervention, but they were notified about the general content of the interviews and were informed that they would receive further incognito SP visits following the interviews.

Before each interview, interviewers were given a sealed envelope by the research team, containing either a four-page educational leaflet (the 'control' intervention) or the educational leaflet with an individualised performance feedback sheet (the 'treatment' intervention), depending on the treatment status of the GP they were to interview. At the end of each interview, interviewers were instructed to hand over the sealed envelope to the GP. At this stage, treatment group GPs were informed about the content of the envelope, that the feedback was entirely confidential, and that it was based on previous SP visits that they had consented to.[279]

Following the intervention, each interviewed GPs received two further incognito SP visits ('end-line' visits), where clinical performance data was collected to evaluate the intervention impact. One of these two end-line SPs was identical to the one presented at baseline (Clinical Case 1). The other SP presented with a new, more complicated clinical case: a HIV+ young adult with a more severe case of acute bronchitis (Clinical Case 2). The order of these two SP visits was randomised for each GP, to prevent any order effects from confounding GPs' treatment choices in each case. End-line SP visits were conducted within a few weeks of the corresponding interviews, over the period December 2018-March 2019.[280]

[278] Unfortunately the sample size was restricted by fieldwork and funding constraints, both at baseline (in the ESRC study) and during the experiment.

[279] The specific interviewer scripts when delivering the intervention to control and treatment group GPs are provided in Appendix B.9.

[280] The majority of end-line visits (154 of the 160 visits in the final sample) were conducted within a month of the corresponding GP interview. In 6 visit cases, however, the gap between the end-line visit and the interview were between 1 and 3 months due to issues with GP availability and fieldwork logistics. Potential implications for experiment validity are discussed in Section 7.3.3.

7.3.2.1 'Control' intervention: educational leaflet

A challenge in isolating the pure *incentive* effect of A&F interventions that report performance relative to clinical guidelines is that such feedback can also work through bridging a knowledge gap among physicians on the guidelines themselves.[281] To rule out this knowledge-improvement effect, it was important that both treatment and control groups received the same information on best-practice guidelines. A counterfactual intervention was therefore employed where written guidelines information was shared with control group GPs in the form of an educational leaflet. The same information was also incorporated in the treatment group intervention (see below).

The four-page (A5) educational leaflet had three components (see Appendix B.10 (b) for an example). The front page presented the study motivation, outlining information on AMR and antibiotic consumption in South Africa. The inside two pages summarised both diagnostic and recommended treatment guidelines for five respiratory illnesses that often present in primary care with overlapping symptoms: the common cold, acute bronchitis, flu (influenza), acute bacterial sinusitis and pneumonia. The aim was to inform GPs of the common clinical presentation and any appropriate diagnostic testing in each case, to enable a differential diagnosis; and to summarise the recommended management of each diagnosis (according to national treatment guidelines), emphasising if and when antibiotic treatment should be considered. Control group GPs were also informed that confidential feedback on their performance in SP visits would be made available to them (upon request) following study completion.

7.3.2.2 'Treatment' intervention: educational leaflet and individualised performance feedback

Treatment group GPs also received a four-page leaflet (see Appendix B.10 (a)). The front page ('feedback page') presented individualised feedback on their performance in baseline SP visits, and the following pages detailed all contents of the control group's educational leaflet.[282] The feedback page outlined the clinical case that was presented (an uncomplicated case of acute bronchitis, in an otherwise healthy young adult), and reported three aspects of measured performance. First, it reported whether the GP had prescribed or dispensed an antibiotic, whilst outlining that an antibiotic was not clinically-indicated for the case. Second, it reported the percentage of all case-specific essential and recommended history-taking and examinations that had been completed. It was explained that this checklist was devised in consultation with respiratory and infectious diseases experts, and designed to enable GPs to differentiate respiratory

[281] Where such knowledge gaps exist, the overall effect of a private A&F intervention is the combined effect of improving knowledge of best-practice guidelines *and* providing new information to physicians' on their relative performance. Only the latter effect operates through intrinsic incentives, in theory.

[282] The AMR and ESRC study information was consolidated on the back page, rather than being presented on two separate pages (as for control group GPs).

illnesses with common symptoms.[283] Third, it reported the total cost of the drugs prescribed or dispensed by the GP, and how it compared (in multiplier terms) to the cheapest, appropriate treatment available on the local market for the same case.[284]

A 'traffic-light' colour coding system was used in the text to highlight good or bad performance. If the GP had inappropriately prescribed an antibiotic at baseline, this was reported in red text (whereas green text was used when no antibiotic had been prescribed). Similarly, the colours red, yellow and green were used to report the case-specific checklist completion rates of less than 40%, between 40-60%, and above 60% respectively. Lastly, if the total cost of drugs prescribed or dispensed was *above* the cheapest appropriate treatment cost, this was reported in red text (and in green text, if *below*). To emphasise the main targeted outcome, performance on antibiotic prescribing was presented first.

7.3.3 Experiment validity

7.3.3.1 Non-participation, attrition and balance

The final sample used in this chapter's analysis consists of 80 GPs: 37 in the treatment group and 43 in the control. Of the 99 GPs initially randomised, 16 were dropped from the final sample due to *non-participation* (these GPs received neither the intervention nor the two end-line SP visits)[285] and 3 due to *attrition* (these GPs received the intervention, but not the end-line SP visits).

This non-participation and attrition – particularly the differential rates across the treatment and control groups - may pose a risk to experiment validity. Internal validity is threatened if treatment status influences GPs' decisions to drop-out and biases the comparability of the two groups. However, the majority of sample loss (84%) occurred *before* the intervention was administered (at the invite to interview) and *before* GPs had any knowledge of the intervention or their treatment status. Table 7.1 shows results from binary regressions of non-participation and attrition indicators on all baseline outcomes, as well as on all observable GP characteristics for which data was available prior to the GP interviews.[286] Treatment status is not significantly associated with the likelihood of non-participation (Column (1)) or attrition (Column (4)). However, in the control group, female GPs were more likely to refuse participation (Column (2)) and contracted-in GPs

[283] See Appendix B.1 for a list of these checklist items for Clinical Case 1.
[284] The cheapest treatment that was considered appropriate for this clinical case was a cough suppressant syrup and paracetamol, priced at a total R17.86. This was based on advice from national respiratory and infectious diseases experts, recommendations in the 2014 South African Standard Treatment Guidelines and Essential Medicines List (STG/EML), and regulated prices in the South African Medicine Price Registry.
[285] Most of these GPs declined to interview citing work time constraints (GPs were requested 30 minutes of their work time to complete the interviews).
[286] This data on a limited set of GP characteristics was taken from the online database of medical practitioners used for recruitment (*Medpages*).

were more likely to attrite (Column (5)) – while similar relationships are not observed in the treatment group.[287]

Column (4) of Table 7.2 reports tests of balance across the two groups in the final (retained) sample for a set of baseline GP characteristics measured in the interviews. Column (4) of Table 7.3 reports equivalent tests for all baseline outcomes. The baseline measure of altruism may be conceptually linked to GPs' intrinsic motivations to provide appropriate care to patients, and baseline levels of intrinsic motivation may in turn determine how GPs respond to the intrinsic incentives from performance feedback (Lee, 2018). It is therefore reassuring that this measure of altruism does not significantly vary between the treatment and control groups. Some significant differences are notable, however. Control GPs are less likely to be black African (or mixed-race) than treatment GPs, and more likely to be practising in the wealthy northern suburbs of Sandton and Randburg. They also communicated significantly weaker beliefs that patients would expect antibiotics for Clinical Case 2 (when the case was presented to them in vignette form in the interviews).[288] During the baseline SP visits, control GPs spent almost three minutes less in consultation time than treatment GPs. While not statistically significant, there were further sizable differences in the gender distribution and the total cost of drugs prescribed at baseline: control GPs were almost 50% more likely to be female, and generated total baseline drug costs that were 32% higher on average than treatment GPs. These differences highlight that the treatment and control groups in the final sample cannot be treated as identical. In the following analyses, all significant and sizable observed differences are controlled for as covariates in the main regression specifications.

The external validity of results would also be reduced if GPs that drop out are systematically different from those that remain in the final sample. Column (1) of Table 7.1 shows that GPs that did not participate in the study were slightly younger, and less likely to have prescribed certain inappropriate drugs (excluding antibiotics and steroids) in baseline visits. However, none of the other observed GP characteristics or baseline outcomes are meaningfully associated with study non-participation or attrition.

[287] The latter result is driven by the one GP that dropped out of the control group, who was contracted-out.
[288] See Section 4.2.3 for details on how such beliefs about antibiotics were measured.

Table 7.1. Baseline determinants of non-participation and attrition

	(1)	(2)	(3)	(4)	(5)	(6)
	Non-participation			Attrition		
	Full sample (non-particip =1)	Control group (non-particip. =1)	Treatment group (non-particip. =1)	Full sample (attrition=1)	Control group (attrition=1)	Treatment group (attrition=1)
Group (1 = treatment group)	0.12			0.03		
	(0.07)			(0.04)		
a) GP characteristics						
Dispensing status (1 = dispensing)	0.03	0.10	-0.05	0.01	-0.05	0.08
	(0.08)	(0.09)	(0.12)	(0.04)	(0.05)	(0.07)
Contracting status (1 = contracted-in)	0.12	0.13	0.11	-0.03	-0.11**	0.06
	(0.10)	(0.11)	(0.16)	(0.05)	(0.05)	(0.09)
Age	-0.01**	-0.01*	-0.01	0.00	0.00**	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Gender (1 = female)	0.03	0.18**	-0.13	-0.05	-0.03	-0.07
	(0.08)	(0.09)	(0.14)	(0.05)	(0.05)	(0.09)
Northern suburb practice	-0.00	-0.06	0.14	0.02	0.07	-0.06
	(0.09)	(0.09)	(0.15)	(0.05)	(0.05)	(0.10)
Local GP density (in 1 km radius)	-0.01	-0.00	-0.01	-0.00	-0.00	-0.00
	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
b) Baseline measures of outcomes						
Antibiotic prescription *(stratifying variable)*	0.07	0.07	0.08	0.06	0.04	0.08
	(0.08)	(0.09)	(0.12)	(0.04)	(0.05)	(0.07)
Total drug cost (per consultation), ZAR	0.00	0.00*	-0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of drugs (per consultation)	0.04	0.06	0.03	0.03*	0.01	0.04
	(0.03)	(0.04)	(0.05)	(0.02)	(0.02)	(0.03)
Average drug cost (per consultation), ZAR	-0.00	0.00	-0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Steroid prescription	0.02	-0.01	0.04	0.03	0.05	0.00
	(0.07)	(0.09)	(0.12)	(0.04)	(0.05)	(0.07)
Other inappropriate drug prescription	-0.14*	-0.07	-0.18	0.05	0.03	0.07
	(0.08)	(0.10)	(0.12)	(0.05)	(0.05)	(0.08)
Consultation time (mins.)	-0.01	-0.01	-0.02	-0.00	0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Checklist completion (share)	-0.19	0.20	-0.47	-0.13	-0.03	-0.23
	(0.27)	(0.32)	(0.42)	(0.14)	(0.16)	(0.25)
IRT score	-0.03	0.05	-0.08	-0.02	0.00	-0.04
	(0.04)	(0.05)	(0.06)	(0.02)	(0.03)	(0.04)
Correct / partially-correct diagnosis	-0.07	-0.02	-0.10	0.03	0.05	0.02
	(0.08)	(0.09)	(0.12)	(0.04)	(0.05)	(0.07)
Sample mean	0.16	0.10	0.22	0.04	0.02	0.05
Obs. (no. of GPs)	99	49	50	83	44	39

Notes. Standard errors are in parenthesis. 'Non-participation' indicates those GPs that were randomized into an intervention group, but who refused to the face-to-face interview where the intervention was to be administered (and hence, were neither informed about nor received the intervention) (N=16). 'Attrition' indicates those GPs that received the intervention but either refused to on-going participation in the study or were not successfully followed-up (N=3). 'Northern suburb practice' indicates whether the GP's practice was located in the wealthiest northern suburbs of Sandton and Randburg in Johannesburg. 'Local GP density' indicates the number of competing GPs located within a 1 km radius of the GP's practice. See Section 4.2.2 for details on how drug treatment and cost outcomes were calculated; Appendix B.1 for the checklist items, and for the diagnoses coding; Appendix B.6 for how the IRT score is calculated from checklist items; and Appendix E.1 for a list of all checklist items included and excluded from the IRT analyses in this chapter. *** p<0.01, ** p<0.05, * p<0.1

Table 7.2. T-test of differences in baseline characteristics between treatment & control groups in final sample

	(1) Full Sample		(2) Control		(3) Treatment		(4) Difference (treatment-control)
	Mean	SD	Mean	SD	Mean	SD	
a. GP characteristics							
Dispensing status *(stratifying variable)*	0.57	(0.50)	0.56	(0.50)	0.59	(0.50)	-0.04
Contracted-in status *(stratifying variable)*	0.81	(0.39)	0.81	(0.39)	0.81	(0.40)	0.00
Age	53.51	(11.55)	53.06	(11.03)	54.04	(12.25)	-0.99
Female	0.30	(0.46)	0.35	(0.48)	0.24	(0.43)	0.11
Ethnicity:							
African / Coloured	0.31	(0.47)	0.21	(0.41)	0.43	(0.50)	-0.22**
Asian / Indian / Other	0.38	(0.49)	0.40	(0.49)	0.27	(0.45)	0.13
White	0.31	(0.47)	0.35	(0.48)	0.27	(0.45)	0.08
Northern suburb practice	0.25	(0.44)	0.33	(0.47)	0.16	(0.37)	0.16*
Group practice	0.28	(0.45)	0.30	(0.46)	0.24	(0.43)	0.06
Local GP density (no. competing GPs within 1 km radius of practice)	8.53	(9.13)	7.91	(8.06)	9.24	(10.30)	-1.34
Average daily patient load	24.00	(13.89)	24.49	(13.24)	23.43	(14.77)	1.06
Altruism: is (highly) altruistic	0.60	(0.49)	0.63	(0.49)	0.57	(0.50)	0.06
b. GP knowledge & antibiotic beliefs (based on responses to case vignettes & AMR quiz)							
Knows diagnosis name for *Clinical Case 1 (viral bronchitis)*	0.17	(0.38)	0.21	(0.41)	0.14	(0.35)	0.07
Knows viral cause of *Clinical Case 1*	0.88	(0.33)	0.91	(0.29)	0.84	(0.37)	0.07
Knows treatment guidelines for *Clinical Case 1*	0.24	(0.43)	0.23	(0.43)	0.24	(0.43)	-0.01
Perceived antibio. effectiveness for *Clinical Case 1* [1]	30.10	(31.37)	33.98	(31.00)	25.59	(31.62)	8.38
Perceived antibio. prescribing norms for *Clinical Case 1* [2]	65.38	(28.24)	67.09	(29.81)	63.38	(26.56)	3.71
Perceived patient expectations for antibio. for *Clinical Case 1* [3]	55.56	(32.07)	53.84	(30.35)	57.57	(34.27)	-3.73
Knows diagnosis name for *Clinical Case 2 (bacterial bronchitis)*	0.55	(0.50)	0.56	(0.50)	0.54	(0.51)	0.02
Knows bacterial cause of *Clinical Case 2*	0.68	(0.47)	0.70	(0.46)	0.65	(0.48)	0.05
Perceived antibio. effectiveness for *Clinical Case 2* [1]	71.95	(28.48)	69.56	(29.54)	74.73	(27.33)	-5.17
Perceived antibio. prescribing norms for *Clinical Case 2* [2]	82.56	(23.25)	79.42	(25.87)	86.22	(19.49)	-6.80
Perceived patient expectations for antibio. for *Clinical Case 2* [3]	74.42	(27.94)	68.84	(30.27)	80.92	(23.72)	-12.08*
AMR knowledge score (max 5.)	3.41	(1.00)	3.37	(1.00)	3.46	(1.02)	-0.09
Observations (no. of GPs)	80		43		37		

Notes: Full sample includes randomized GPs with whom both the intervention (interview) and the two end-line SP visits were successfully completed. Most data presented here were collected during the GP interviews where the intervention was delivered. The average age, share of male GPs, share of GPs located in the northern suburbs of Sandton and Randburg, and the measure of local GP density were calculated using data from the online medical practitioner database *Medpages*. See Section 4.2.3 for further details on how certain GP characteristics were measured. [1]GPs were firstly asked what was the likelihood (between 0 and 100, with 100 indicating certainty) that the patient in the vignette case would recover faster with antibiotics (rather than without). [2] They were then were asked how likely it was that other GPs would prescribe antibiotics to the patient in the vignette case. [3] Lastly, GPs were asked how likely it was that the patient would go to another GP next time they needed medical attention if they did not receive antibiotics for the case. *** p<0.01, ** p<0.05, * p<0.1

While these observations (and proposed controls for observed differences) do not rule out all threats to experiment validity from non-participation and attrition (particularly from any unobserved drivers), they do give some confidence that any resulting bias is unlikely to be severe in a per-protocol analysis with appropriate controls. The RCT literature nevertheless proposes various approaches to correcting for any residual bias in the final sample (Glennerster and

Takavarasha, 2013), including intention-to-treat (ITT) analyses that adopt different ways of imputing outcomes for those research participants lost to non-participation or attrition. This ITT approach is carried out as an additional robustness check in Section 7.4.5.

Table 7.3. T-test of differences in baseline measures of outcomes between treatment & control groups in final sample

	(1) Full Sample		(2) Control		(3) Treatment		(4) Difference (treatment-control)
	Mean	SD	Mean	SD	Mean	SD	
Antibiotic prescription *(stratifying variable)*	0.61	(0.49)	0.63	(0.49)	0.59	(0.50)	0.03
Total drug cost (per consultation), ZAR	138.42	(145.04)	156.12	(168.55)	117.84	(110.51)	38.28
Number of drugs (per consultation)	3.42	(1.16)	3.37	(91.02)	3.49	(1.30)	-0.11
Average drug cost (per consultation), ZAR	44.56	(49.73)	51.01	(57.05)	37.05	(39.03)	13.96
Steroid prescription	0.45	(0.50)	0.42	(0.50)	0.49	(0.51)	-0.07
Other inappropriate drug prescription	0.71	(0.46)	0.74	(0.44)	0.68	(0.47)	0.07
Consultation time (mins.)	9.64	(4.79)	8.51	(4.05)	10.95	(5.29)	-2.43**
Checklist completion (share)	0.40	(0.15)	0.41	(0.15)	0.40	(0.15)	0.01
IRT score	0.05	(0.92)	0.07	(0.88)	0.02	(0.98)	0.05
Correct / partially-correct diagnosis	0.43	(0.50)	0.45	(0.50)	0.41	(0.50)	0.14
Observations (no. of GPs)	80		43		37		

Notes: Full sample includes randomized GPs with whom both the intervention (interview) and the two end-line SP visits were successfully completed. All listed outcomes were measured using data collected from the SP debriefing questionnaire, and any dispensed drugs or drug prescriptions, from the SP consultations conducted at baseline (in the ESRC study, with Clinical Case 1). For further details on how there outcomes were measured, refer to Sections 4.2.1 and 4.2.2. *** p<0.01, ** p<0.05, * p<0.1

7.3.3.2 Treatment compliance and spill-overs

Two further threats to experiment validity are partial intervention compliance within the treatment and control groups, and potential exposure of control group GPs to the treatment mechanism (treatment spill-over). Partial compliance in this case would amount to some GPs not receiving the A&F or educational materials they were assigned to (because of interviewers not complying with interview protocols), or simply not reading the materials they receive. To minimise this risk, interviewers were carefully trained on correct and timely delivery of the intervention envelopes at the end of each interview, and their tablet-based interview guide incorporated the intervention delivery script and instructions. Random audio audits were also programmed into the tablet-based interview guide as a quality check to ensure that interviewers followed their scripts and interview protocols. During interviewer debriefings, there were no cases reported of interviewers failing to deliver the intervention envelopes as instructed or of GPs refusing to receive them.

While it is impossible to know if and when GPs reviewed the contents of their envelopes, interviewers were instructed to invite treatment group GPs to review their feedback and ask any initial questions. A specific question in the interview guide and questionnaire asked interviewers

whether the GP actually opened and reviewed the feedback envelopes in front of them. Interviewers reported that 35 of the 37 treatment group GPs reviewed the feedback immediately in their presence, suggesting that non-compliance among the treatment group is unlikely to be a concern. One weakness of the intervention design is that control group GPs were not similarly invited to review the educational materials they were given by interviewers (see corresponding interviewer scripts in Appendix B.9), so I cannot obtain a similar indication of likely compliance in the control group.

The spill-over of A&F incentives to the control group can also confound estimated treatment effects. This can occur if some control GPs accidentally received the A&F intervention; or if there is close interaction between control and treated GPs, and the former observe and mimic the behavioural responses of the latter. To address the former risk, before each interview, the research team ensured that each interviewer took with them only the intervention envelope addressed to the specific GP that was to be interviewed (to avoid mix-up with other envelopes). The interview guide also prompted interviewers to check that they had the envelope labelled with the correct GP name before entering each interview. As mentioned earlier, efforts to exclude GPs working together in group practices from the experiment sample was designed to avoid potential interactions and behavioural spill-overs between treated and control GPs working in the same practices.[289]

7.3.3.3 Evaluation and evaluator effects

The experiment's use of incognito SPs for gathering performance data avoids a common evaluation-driven effect (the 'Hawthorne effect') whereby research participants adjust their normal behaviour under observation. As in all SP studies, possible SP detection by research participants is a risk. Studies often run SP detection surveys with participants after study conclusion, to identify and control for potential cases of valid detections. A survey run by the ESRC study found a very low detection rate: less than 1% of SP visits were reported as suspicious by GP participants in that study. The present study sample includes only those GPs (and corresponding baseline data) that did not have a valid detection in the ESRC study. Given the low detection rate in the ESRC study, the research team decided that it would not be cost-effective to run another detection survey after the end-line SP visits in this study. Debriefings with fieldworkers that carried out the end-line visits also did not flag any cases of doctors raising suspicions about fake patients.

[289] Six GPs that shared a group practice with another sample GP were mistakenly retained in the final sample due to errors in their address listings.

The validity of the SP method for intervention evaluation further relies on all SP visits at baseline and end-line, across both intervention groups, being identical (apart from the variation in clinical case at end-line). One concern is that there may be individual fieldworker effects influencing GPs' intervention responses, and these effects may vary systematically across intervention groups. To address this, SP fieldworkers were randomly allocated to GP visits across both intervention groups and clinical cases. No fieldworker was restricted to visits in any one clinical case or intervention group. In addition, fieldworkers were blinded to the experiment details and the treatment status of GPs, to prevent their behaviour in consultations and subsequent recall of GP performance being affected by the experiment aims or GPs' treatment status.[290] Fieldworker fixed effects are nevertheless included in all results analyses conducted in this chapter, to control for any residual evaluator effects. GPs were similarly blinded to the experiment details and their relative treatment status during the study period, to prevent any knowledge of this influencing their behaviour. Control GPs were further informed that they could request feedback on their individual performance in SP visits after study completion, to avoid any demoralisation effects from not receiving feedback from their participation.

Table 7.4. T-test of differences in end-line SP visit characteristics between treatment & control groups in final sample

	(1) Full Sample		(2) Control		(3) Treatment		(4) Difference (treatment-control)
	Mean	SD	Mean	SD	Mean	SD	
Visit order (1=Uncomplicated SP visit first)	0.41	(0.50)	0.37	(0.49)	0.46	(0.51)	-0.09
Hour of day							
Uncomplicated visit	12.01	(2.43)	11.84	(2.21)	12.22	(2.68)	-0.38
HIV+ visit	12.23	(2.31)	11.98	(2.02)	12.51	(2.61)	-0.54
No. days since intervention (interview)							
Uncomplicated visit	17.44	(11.58)	19.7	(14.23)	14.81	(6.67)	4.89*
HIV+ visit	15.94	(9.96)	17.42	(13.14)	14.22	(3.27)	3.20
Observations (no. of GPs)	80		43		37		

Notes: Full sample includes randomized GPs with whom both the intervention (interview) and the two end-line SP visits were successfully completed. All listed visit characteristics were measured using data collected from the SP debriefing questionnaire following end-line visits and from the interviews. The "uncomplicated visit" corresponds to the SP visit with Clinical Case 1, and the "HIV+ visit" corresponds to the visit with Clinical Case 2. *** p<0.01, ** p<0.05, * p<0.1

Table 7.4 tests for balance between the treatment and control groups in a set of end-line visit characteristics. As each GP received two end-line SP visits, the order of these visits - if it varies systematically across the treatment and control groups - may confound estimated treatment effects. The second SP visit to each GP may have a higher risk of SP detection, and potential learning effects (whereby GPs perform better at treating a patient when they have seen a similar

[290] The interviewers that delivered the intervention were also not informed about the experiment details or aims.

clinical case recently), as both clinical cases are quite similar. To minimise such order effects, the visit order of the two clinical cases was randomised for each GP in both intervention groups, and at least a two-day gap was maintained in between the two visits.[291] Table 7.4 accordingly shows no significant differences in visit order between the two intervention groups. The time of day of the SP visits can also influence GPs' treatment decisions. For instance, physicians' have been shown to be more prone to prescribing inappropriate antibiotics later in the day, perhaps due to decision fatigue (Linder *et al.*, 2014). Unfortunately, due to logistical reasons and the large distances that fieldworkers often had to travel between scheduled GP visits, it was not possible to strictly control this during fieldwork. Nevertheless, there are again no significant group differences detected in Table 7.4 for the hour of day in which the two end-line SP visits were completed.

Another limitation of the intervention design is that the time lapse between intervention delivery (in the GP interviews) and impact measurement (in end-line SP visits) was not perfectly controlled, due to GP availability and logistical constraints. While the majority of all end-line visits (154 of the final 160 sample visits) were completed within 3-4 weeks of intervention delivery, 6 visits were delayed and completed between 1-3 months. These 6 anomalies all occurred with control group GPs. Table 7.4 indeed shows a significant difference in the number of days since intervention delivery between treatment and control groups for the uncomplicated SP visit. To check the sensitivity of main results to such imbalances in end-line visit characteristics, these characteristics are explicitly controlled for in the results analyses as additional robustness checks in Section 7.4.5.

7.3.4 Final sample characteristics

Column (1) of Tables 7.2 summarises some basic characteristics of the final sample. The average age of GPs is 54 years, 30% are female, and the ethnic distribution is approximately balanced across the three main ethnic groups (black African or mixed-race (31%), Asian or Indian (38%), and white GPs (31%)). Around a quarter of the sample work in group practices, and a similar share are located in the wealthier northern Johannesburg suburbs (Sandton and Randburg). The average GP has around nine competing GPs located within a kilometre of their practice, and reported seeing an average of 24 patients per day.

A large majority of sample GPs are contracted in with medical insurance schemes (81%), which is reflective of the broader GP population in Johannesburg. However, a higher proportion of

[291] The majority of sample GPs received their two end-line SP visits with a 2-7 day gap. Seven GPs had longer visit gaps: three GPs had an 8-14 day gap, one GP had a 30-day gap, and three GPs had a 30-60 day gap.

sample GPs have a drug dispensing licence (57%) compared to the sample population (38.5%). Measured altruism in the sample is also relatively high. To measure altruism in this study, the share of a R300 cash endowment that each GP chose to donate to a patient charity during a dictator game in the GP interviews was calculated (see Appendix B.4 for a description of the game), and "altruistic" (or "high-altruism") GPs were classified as those who donated the full endowment. Around 60% of the sample donated the full endowment.

To measure GPs' clinical knowledge, a vignette-based knowledge quiz was included in the interviews. When presented with the clinical cases used in this experiment in vignette-form, GPs were more able to correctly name the diagnosis as "acute bronchitis" for Clinical Case 2 (55%) than for the uncomplicated Clinical Case 1 (17%). However, a high proportion of all GPs correctly identified the likely pathogenic cause in both cases (88% in Clinical Case 1, and 68% in Clinical Case 2), which is perhaps more important in informing appropriate antibiotic treatment choices. This is reflected in a relatively low perceived effectiveness of antibiotics for the (viral) Clinical Case 1 and a high perceived effectiveness for the (bacterial) Clinical Case 2: on average, GPs felt that there was only a 30% likelihood that Clinical Case 1 would resolve more quickly with antibiotics, and a 72% equivalent likelihood for Clinical Case 2. Overall, this suggests that GPs' clinical knowledge pertaining to appropriate antibiotic use for the experiment's two clinical cases is quite high.

However, actual prescribing practices may be determined by more than just GPs' knowledge (or perceptions) of clinically-indicated treatments. Firstly, their knowledge or beliefs about the risks of inappropriate antibiotic prescribing may play a role. Again, the sample demonstrates relatively good knowledge of these risks: on a short quiz to assess GPs' knowledge of AMR, the average GP answered more than three out of five questions correctly. Secondly, perceived peer norms and patient expectations ('social pressures') around prescribing practices can influence treatment choices. Sample GPs reported relatively strong social pressures to prescribe antibiotics for both clinical cases. This is concerning for Clinical Case 1, where no antibiotics are indicated: despite GPs correctly believing that antibiotics would be of low effectiveness in that case, they reported a 65% likelihood that other GPs would prescribe antibiotics for the same case, and a 56% likelihood that patients would seek care elsewhere next time they are ill if they did not receive antibiotics for the case.

7.4 Results

Most outcomes evaluated in this section were measured in end-line SP visits presenting the uncomplicated acute bronchitis case (Clinical Case 1).[292] These visits mirror those conducted at baseline. End-line SP visits presenting the more severe acute bronchitis case (Clinical Case 2) are used to measure only one outcome: spill-over effects on appropriate antibiotic treatment (as detailed in Section 7.4.3).

7.4.1 Empirical framework

The impact of the A&F intervention can be measured by comparing end-line outcomes across the treatment and control groups in a simple regression framework. For each physician-level outcome, I estimate the following:

$$Y_j = \beta_0 + \beta_1 Treat_j + \beta_2 Y_j^{base} + \boldsymbol{Strat'}_j\beta_3 + \boldsymbol{X'}_j\beta_4 + \delta_f + \varepsilon_j \quad (1)$$

where Y_j is the outcome for physician j at end-line; $Treat_j$ is a dummy variable equal to 1 if physician j is in the treatment group, and 0 otherwise; Y_j^{base} is the value of the same outcome for physician j at baseline;[293] $\boldsymbol{Strat'}_j$ is the set of stratification variables used in randomising treatment assignment (GPs' dispensing and contracting status, and baseline antibiotic prescribing); $\boldsymbol{X'}_j$ is a set of controls for all significantly (or sizeably) unbalanced GP characteristics and baseline outcomes in Tables 7.2 and 7.3 (baseline consultation duration, baseline drug costs, ethnicity, gender, northern suburb practice location, and perceived patient expectations for antibiotics in the complicated bronchitis case); and δ_f is a fieldworker fixed effect (for the corresponding end-line SP visit). The error term ε_j is assumed independent and identically distributed (i.i.d) across sample GPs.

β_1 measures the effect of interest: the marginal effect of the private, individualised A&F, compared to the educational intervention alone.[294] The main results present estimates of β_1 from a per-protocol analysis, including only those 80 GPs that successfully received both the intervention and end-line SP visits.[295] In the following results tables, three sets of estimates are reported for each outcome. In the basic specification in Column (1), only the treatment group

[292] See Section 4.1.1 and Appendices B.1 and B.2 for details on the SP clinical cases.

[293] Note that Y_j^{base} is excluded from the specification in estimating spill-over effects on appropriate antibiotic treatment (for the complicated HIV+ patient case). There is no baseline data for this particular outcome, since the complicated acute bronchitis case was only used in end-line visits.

[294] Strictly speaking, β_1 captures an intention-to-treat (ITT) effect, as it is not certain that all GPs fully adhered to the intervention (that they read all intervention materials provided to them).

[295] An ITT analysis including all 99 randomised GPs, imputing outcomes for the 19 that were dropped due to non-participation or attrition, is carried out as a robustness check in Section 7.4.5.

dummy $Treat_j$, the baseline outcome Y_j^{base}, and the fieldworker fixed effect δ_f are included; in Column (2), the controls $\boldsymbol{X'}_j$ are added, to adjust for unbalanced observables; and finally in Column (3), the stratification variables $\boldsymbol{Strat'}_j$ are also included to improve estimate precision. The specification in Column (3), which corresponds to the full equation in (1), is taken as the main specification in each case. All estimations are carried out using linear OLS.

7.4.2 Impact of A&F on treatment quality and costs

7.4.2.1 Inappropriate antibiotic treatment

The primary outcome evaluated in this chapter is inappropriate antibiotic treatment for the uncomplicated acute bronchitis case (Clinical Case 1). Systematic reviews of existing clinical evidence, clinical expert opinions, and the South African national treatment guidelines all concur that antibiotics are not recommended for this case.

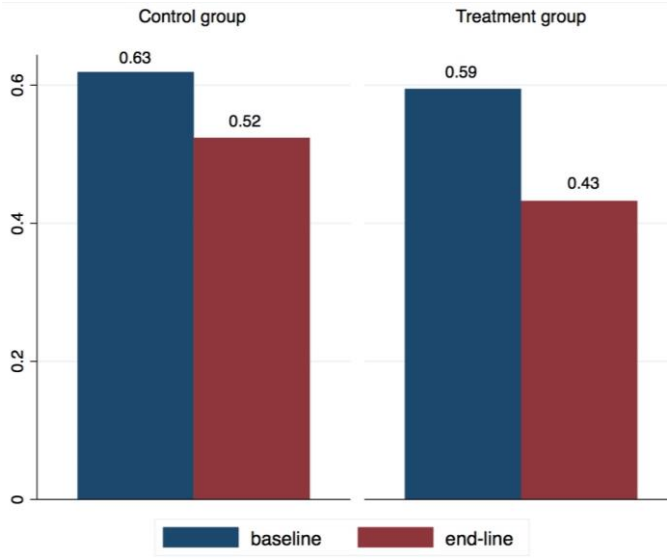


Figure 7.2. Share of inappropriate antibiotic treatment at baseline and end-line, by intervention group

Nevertheless, echoing previous study results (see Chapter 5), baseline antibiotic treatment for the case was very high: an antibiotic was inappropriately prescribed in 61% of all baseline SP visits in the sample (see Column (1) of Table 7.3).[296] This rate of inappropriate treatment is much higher than that predicted by GPs' relatively good knowledge of the viral cause and low effectiveness of antibiotics for the same clinical case (when presented in vignette-form). This may again be reflective of a 'know-do' gap in the sample, where physicians' performance in practice is shown

[296] The most commonly prescribed antibiotic groups were penicillin or penicillin-clavulanate, and macrolides.

to be poorer than that predicted by indictors of their knowledge (Das and Gertler, 2007; Leonard and Masatu, 2010). Insufficient diagnostic effort or accuracy – which is explored further in Section 7.4.4 - may be one explanation for this gap. Perceived social pressures to prescribe antibiotics, as noted in the interviews, may be another.[297] [298] In either case, where such 'know-do' gaps exists – and physician knowledge does not translate into equivalent effort – it is reasonable to expect that the effort incentives from A&F should improve upon the effectiveness of educational interventions alone.

Table 7.5. Effect of private performance feedback on likelihood of inappropriate antibiotic treatment

	(1) Antibiotic	(2) Antibiotic	(3) Antibiotic
Treatment (1 = feedback given)	-0.12	-0.22**	-0.24**
	(0.11)	(0.10)	(0.10)
Baseline antibiotic treatment, for Clinical Case 1 (1 = antibiotic given)	0.45***	0.50***	0.52***
	(0.11)	(0.10)	(0.10)
Fieldworker fixed effects	Yes	Yes	Yes
Unbalanced covariates	No	Yes	Yes
Stratification variables	No	No	Yes
Mean of control group	0.52	0.52	0.52
Mean of sample	0.48	0.48	0.48
Obs.	79	79	79
R-squared	0.36	0.52	0.55

Notes. All specifications are estimated using linear OLS. Standard errors are in parenthesis. Column (1) presents the treatment effect estimate from a basic specification, controlling only for baseline antibiotic treatment and fieldworker fixed effects. Column (2) additionally controls for all significantly or sizably unbalanced observables (baseline consultation duration, baseline drug costs, ethnicity, gender, northern suburb practice location, and perceived patient expectations for antibiotics in Clinical Case 2). Column (3) presents the main specification, controlling also for stratifying variables (GPs drug dispensing and contracting status, in addition to baseline antibiotic treatment). Multicollinearity is unlikely to be an issue in these three specifications: mean variance inflation factors (VIF) for each specification was around 2, and individual VIF for all included variables were all under 4. See notes in Table 7.2 and Section 4.2 for details on how certain covariates were measured. *** p<0.01, ** p<0.05, * p<0.1

Table 7.5 presents estimates of the impact of private A&F on inappropriate antibiotic treatment. The binary outcome variable indicates whether an antibiotic was prescribed for Clinical Case 1 in end-line SP visits. Focusing on the main specification in Column (3), the results suggest that private A&F has a substantial effect in lowering inappropriate antibiotic treatment. While 52% of control group GPs gave an antibiotic for the viral clinical case, GPs that received private A&F were 24 percentage points less likely to do so (95% CI -0.44 to -0.03) – a 46% reduction in the likelihood of inappropriate treatment, significant at the 5% level. For comparison, van der Velden *et al.* (2012) find that antibiotic prescriptions were reduced by 11.6% on average in their review

[297] Beliefs about patient expectations do influence physicians' treatment choices – particularly under market competition – even where physicians' are aware that those choices are not clinically-indicated (Das and Hammer, 2007).

[298] While not evidenced in this study, the Hawthorne effect – whereby physicians perform better whilst under direct observation during the knowledge quiz than when being assessed incognito by SPs – is a further potential explanation.

of 87 interventions (and 58 studies) targeting antibiotic use for respiratory infections in primary care.

While the base rate of antibiotic treatment is still high in the control group, this has also fallen substantially since baseline (from 63% to 52% at end-line – a 17% decline). This may be indicative of the educational intervention effect. It may also be due to seasonal effects: the baseline visits occurred during the cold season in Johannesburg (June-August 2018) when respiratory infections are likely to present more frequently to GPs, whereas end-line visits occurred during the summer (December 2018-March 2019). However, given the lack of a pure control ('no-intervention') group, I cannot isolate the effect of the educational intervention from general time trend or seasonal effects.

Lastly, there is some indication of habit persistence in prescribing behaviour: inappropriate antibiotic treatment at baseline is associated with an increase in treatment likelihood at end-line of over 50 percentage points.

7.4.2.2 Drug treatment costs

As secondary indicators of over-treatment, I examine A&F effects on treatment costs for Clinical Case 1. Baseline performance on this metric, relative to the cheapest appropriate treatment on the market, was also reported in the feedback sheets shared with treatment GPs. Table 7.3 (Column (1)) shows that treatment efficiency at baseline was low in general. GPs gave 3.4 drugs per baseline consultation on average, at a cost of R45 per item – although no drug treatment is clinically-indicated for the case.[299] This amounted to a total drug cost per consultation of R138 - almost 7 times the cost of the cheapest recommended treatment on the market for the same case.[300]Among the SP visits analysed in previous chapters, antibiotics were the most expensive drug items prescribed by GPs. An inappropriate antibiotic prescription was associated with significantly higher total drug costs. Hence, all else equal, a lower likelihood of antibiotic treatment should also coincide with lower overall drug costs for the treatment group in this chapter.

Table 7.6 presents estimates of the A&F effect on total drug costs, the number of drugs given, and the average cost per drug. While private A&F is estimated to lower total drug costs as expected (by R41 compared to a control group mean of R142), this result is not statistically

[299] Some over-the-counter (OTC) symptomatic treatment was still considered appropriate for the case, although not clinically-indicated for case resolution.

[300] The cheapest appropriate (low-schedule, symptomatic) treatment for the uncomplicated acute bronchitis case was some cough syrup and paracetamol, which amounted to a total cost of R17.86.

significant. Similarly for the number of drugs and average drug costs, effect estimates are in the anticipated direction, but none are statistically significant.

One explanation for the absence of significant cost effects may be the small sample size and a lack of sufficient statistical power. An effort to improve the precision of effect estimates, by including additional covariates, is carried out as an additional analysis in Section 7.4.5. While this reduces the magnitude of estimated effects, it does not improve statistical significance (see Appendix E.2). A further explanation may be that, although the rate of antibiotic prescribing is reduced in the treatment group, those GPs may substitute other inappropriate and equally costly drugs in place of antibiotics. This latter channel is examined in the next subsection.

Table 7.6. Effect of private performance feedback on treatment costs

	(1) Total drug costs	(2) Total drug costs	(3) Total drug costs	(4) No. drugs	(5) No. drugs	(6) No. drugs	(7) Av. drug cost	(8) Av. drug cost	(9) Av. drug cost
Treatment (1 = feedback given)	-33.36	-37.98	-40.67	-0.23	-0.39	-0.41	-6.89	-7.64	-8.46
	(24.19)	(25.40)	(25.43)	(0.24)	(0.25)	(0.25)	(7.41)	(7.14)	(7.20)
Baseline outcome measure, for Clinical Case 1	0.35***	0.24**	0.19*	0.62***	0.69***	0.75***	0.34***	0.20**	0.15*
	(0.08)	(0.09)	(0.10)	(0.12)	(0.13)	(0.13)	(0.07)	(0.08)	(0.08)
Fieldworker fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unbalanced covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Stratification variables	No	No	Yes	No	No	Yes	No	No	Yes
Mean of control group	142.39	142.39	142.39	3.29	3.29	3.29	45.78	45.78	45.78
Mean of sample	124.75	124.75	124.75	3.33	3.33	3.33	40.29	40.29	40.29
Obs.	79	79	79	79	79	79	79	79	79
R-squared	0.36	0.44	0.48	0.45	0.56	0.58	0.37	0.54	0.56

Notes. All specifications are estimated using linear OLS, and all costs are in South African rand (ZAR). Standard errors are in parenthesis. Columns (1), (4) and (7) present treatment effect estimates from a basic specification, controlling only for baseline outcome values and fieldworker fixed effects. Columns (2), (5) and (8) additionally control for all significantly or sizably unbalanced observables (baseline consultation duration, baseline drug costs, ethnicity, gender, northern suburb practice location, and perceived patient expectations for antibiotics in Clinical Case 2). Columns (3), (6) and (9) present the main specifications, controlling also for the stratification variables (GPs drug dispensing and contracting status, and baseline antibiotic treatment). Multicollinearity is unlikely to be an issue in these specifications: mean variance inflation factors (VIF) for each specification was around 2 or lower, and individual VIF for all included variables were all around 4 or under. See notes in Table 7.2 and Section 4.2 for details on how certain covariates and cost outcomes were measured. *** p<0.01, ** p<0.05, * p<0.1

7.4.3 Unintended effects

7.4.3.1 Appropriate antibiotic prescribing: HIV+ patient case management

The A&F intervention was primarily designed to target inappropriate antibiotic prescribing for an uncomplicated viral respiratory infection. The feedback (and its combination with the educational leaflet) accordingly emphasised the case-specific context in which antibiotic treatment was considered inappropriate. The intervention should therefore have no impact on *appropriate* antibiotic treatment - that is, in cases where antibiotics *are* clinically-indicated. However, it is still

a concern that GPs may misinterpret the feedback message and try to cut back on all antibiotic prescribing – both inappropriate and appropriate. To ensure there were no such unintended spill-overs, I also test for treatment effects on the likelihood of antibiotic treatment for Clinical Case 2, where a short course of antibiotics was recommended in national treatment guidelines (NDoH, 2014) given the patient's HIV+ status and the more severe case presentation (that increase the likelihood of a bacterial infection).

Table 7.7 shows no evidence of such unintended A&F effects on appropriate antibiotic treatment. Appropriate treatment in the control group was reasonably high, perhaps in part due to the educational intervention: 62% of control GPs chose some antibiotic treatment for the HIV+ patient case. Moreover, treatment group GPs were no less likely to choose antibiotic treatment for this case than control GPs.

Table 7.7. Effect of private performance feedback on likelihood of appropriate antibiotic treatment for the HIV+ patient case

	(1) Antibiotic	(2) Antibiotic	(3) Antibiotic
Treatment (1 = feedback given)	-0.06	0.05	0.05
	(0.11)	(0.12)	(0.12)
Baseline antibiotic treatment, for Clinical Case 1 (1=antibiotic given)	0.46***	0.43***	0.44***
	(0.11)	(0.12)	(0.12)
Fieldworker fixed effects	Yes	Yes	Yes
Unbalanced covariates	No	Yes	Yes
Stratification variables	No	No	Yes
Mean of control group	0.62	0.62	0.62
Mean of sample	0.59	0.59	0.59
Obs.	79	79	79
R-squared	0.29	0.37	0.37

Notes. All specifications are estimated using linear OLS. Standard errors are in parenthesis. Column (1) presents the treatment effect estimate from a basic specification, controlling only for baseline antibiotic treatment (for the uncomplicated acute bronchitis case that was presented in baseline SP visits) and fieldworker fixed effects. Column (2) additionally controls for all significantly or sizably unbalanced observables (baseline consultation duration, baseline drug costs, ethnicity, gender, northern suburb practice location, and perceived patient expectations for antibiotics in Clinical Case 2). Column (3) presents the main specification, controlling also for the stratifying variables (GPs drug dispensing and contracting status, in addition to baseline antibiotic treatment). Multicollinearity is unlikely to be an issue in these specifications: mean variance inflation factors (VIF) for each specification was around 2 or lower, and individual VIF for all included variables were all around 4 or under. See notes in Table 7.2 and Section 4.2 for details on how certain covariates and outcomes were measured. *** p<0.01, ** p<0.05, * p<0.1

7.4.3.2 Substitution of other inappropriate drugs

Another potential side-effect of A&F targeting a reduction in antibiotic treatment alone may be an increase in other inappropriate drug treatments, if GPs substitute other drugs in place of antibiotics. In a previous experiment (see Chapter 5), a number of other inappropriate drugs were frequently given for the same uncomplicated bronchitis case (Clinical Case 1). Treatment with

steroids was particularly common. Therefore, unintended A&F effects on the likelihood of other inappropriate drug treatments are also tested for here.

Table 7.8 shows estimated treatment effects on the likelihood of steroid or other inappropriate drug treatments. Overall, there is little indication of such unintended substitution effects. Inappropriate drug treatment was high in general: 45% of control group GPs gave inappropriate steroids for Clinical Case 1, and 69% gave other inappropriate drugs. Treatment group GPs were 16 percentage points (36%) less likely to give steroids for the same case, contrary to the substitution effect hypothesis (and indicative of positive spill-overs instead); however, this result is not statistically significant. Similar, no statistically significant differences are detected in the likelihood of other inappropriate drug treatment.

Once again, there is some evidence of habit persistence in inappropriate prescribing behaviours: GPs that prescribed a steroid for Clinical Case 1 at baseline were 50 percentage points more likely to prescribe a steroid for the same case at end-line.

Table 7.8. Effect of private performance feedback on likelihood of other inappropriate drug treatments

	(1) Steroids	(2) Steroids	(3) Steroids	(5) Other inapprop. drugs	(6) Other inapprop. drug	(7) Other inapprop. drug
Treatment (1 = feedback given)	-0.16	-0.16	-0.16	0.04	0.08	0.08
	(0.10)	(0.11)	(0.11)	(0.11)	(0.12)	(0.12)
Baseline outcome measure, for Clinical Case 1 (1 = drug given)	0.54***	0.50***	0.50***	0.19	0.26*	0.23
	(0.10)	(0.11)	(0.11)	(0.12)	(0.14)	(0.14)
Fieldworker fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Unbalanced covariates	No	Yes	Yes	No	Yes	Yes
Stratification variables	No	No	Yes	No	No	Yes
Mean of control group	0.45	0.45	0.45	0.69	0.69	0.69
Mean of sample	0.42	0.42	0.42	0.72	0.72	0.72
Obs.	79	79	79	79	79	79
R-squared	0.40	0.43	0.47	0.17	0.26	0.28

Notes. All specifications are estimated using linear OLS. Standard errors are in parenthesis. Columns (1) and (4) present treatment effect estimates from a basic specification, controlling only for baseline outcome values and fieldworker fixed effects. Columns (2) and (5) additionally control for all significantly or sizably unbalanced observables (baseline consultation duration, baseline drug costs, ethnicity, gender, northern suburb practice location, and perceived patient expectations for antibiotics in Clinical Case 2). Columns (3) and (6) present the main specifications, controlling also for the stratification variables (GPs drug dispensing and contracting status, and baseline antibiotic treatment). Multicollinearity is unlikely to be an issue in these specifications: mean variance inflation factors (VIF) for each specification was around 2 or lower, and individual VIF for all included variables were all around 4 or under. See notes in Table 7.2 and Section 4.2 for details on how certain covariates and treatment outcomes were measured. *** p<0.01, ** p<0.05, * p<0.1

7.4.4 Potential channels: effects on diagnostic effort and accuracy

A possible mechanism behind the estimated A&F effect on inappropriate antibiotic treatment is that GPs receiving the A&F intervention exert more diagnostic effort with their patients as a result - thereby improving their diagnostic accuracy, and lowering the likelihood of antibiotic treatment

due to diagnostic uncertainty. To test this, I estimate treatment effects on three measures of diagnostic effort (Table 7.9) and a measure of diagnostic accuracy (Table 7.10) for Clinical Case 1. Diagnostic effort is measured by consultation duration, and the raw proportion of all case-specific essential and recommended history-taking and examinations (checklist) completed.[301] At baseline, the average SP consultation lasted 10 minutes, where just 40% of the case-specific checklist was completed. IRT analysis is also used to compute a composite index score of checklist completion, which weights completed checklist items by their relative value in discriminating among GPs of varying effort quality.[302]

Table 7.9. Effect of private performance feedback on diagnostic effort

	(1) Cons. time	(2) Cons. time	(3) Cons. time	(4) Share check list	(5) Share check list	(6) Share check list	(7) IRT score	(8) IRT score	(9) IRT score
Treatment (1 = feedback given)	0.19	1.62	1.85	-0.04	-0.02	-0.02	-0.33*	-0.22	-0.19
	(1.59)	(1.66)	(1.69)	(0.03)	(0.03)	(0.03)	(0.17)	(0.18)	(0.18)
Baseline outcome measure, for Clinical Case 1	0.31*	0.19	0.22	0.61***	0.54***	0.54***	0.64***	0.55***	0.57***
	(0.17)	(0.17)	(0.17)	(0.10)	(0.12)	(0.12)	(0.09)	(0.11)	(0.11)
Fieldworker fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unbalanced covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Stratification variables	No	No	Yes	No	No	Yes	No	No	Yes
Mean of control group	8.93	8.93	8.93	0.44	0.44	0.44	0.16	0.16	0.16
Mean of sample	9.80	9.80	9.80	0.42	0.42	0.42	0.00	0.00	0.00
Obs.	80	80	80	80	80	80	80	80	80
R-squared	0.24	0.35	0.36	0.43	0.49	0.53	0.50	0.56	0.59

Notes. All specifications are estimated using linear OLS. Standard errors are in parenthesis. Columns (1), (4) and (7) present treatment effect estimates from a basic specification, controlling only for baseline outcome values and fieldworker fixed effects. Columns (2), (5) and (8) additionally control for all significantly or sizably unbalanced observables (baseline consultation duration, baseline drug costs, ethnicity, gender, northern suburb practice location, and perceived patient expectations for antibiotics in Clinical Case 2). Columns (3), (6) and (9) present the main specifications, controlling also for the stratification variables (GPs drug dispensing and contracting status, and baseline antibiotic treatment). Multicollinearity is unlikely to be an issue in these specifications: mean variance inflation factors (VIF) for each specification was around 2 or lower, and individual VIF for all included variables were all around 4 or under. See notes in Table 7.2 and Section 4.2 for details on how certain covariates and the effort outcomes were measured.
*** p<0.01, ** p<0.05, * p<0.1

To gauge GPs' diagnostic accuracy, I evaluate whether the diagnoses communicated to SPs for Clinical Case 1 were correct or partially-correct.[303] During baseline SP visits, a diagnosis was

[301] See Appendix B.1 for the case-specific checklists, and Section 4.2.1 for a description of how the effort measures were constructed. The checklist items were compiled by a panel of clinical experts, and designed to enable physicians to differentiate respiratory conditions with varying severities and overlapping symptoms.

[302] Section 4.2.1 provides a brief overview of how the IRT score is calculated, and Appendix B.6 gives further details on the assumptions underlying IRT. Appendix E.1 lists the checklist items included in this chapter's IRT analysis at baseline and end-line, and the corresponding weight estimates.

[303] See Appendix B.1 for a list of diagnoses considered correct or partially correct for each clinical case. Further details on how the diagnosis measures were constructed are provided in Section 4.2.1. A strictly 'correct' diagnosis required the doctor to communicate the diagnosis in exact technical terms to the patient (i.e. "acute bronchitis" or "bronchitis"). The limitation here is that what the physician tells the patient regarding their diagnosis may be simplified (for instance, given expected limitations to the patient's medical knowledge). Hence, a 'partially-correct' diagnosis allowed for some generality in the communicated diagnosis (such as "chest cold") and for select similar diagnoses (such as "tracheitis" or "post nasal drip"). This chapter accordingly combines both outcomes in its

communicated to SPs by almost all GPs in the sample (79 of 80). Of these, only 43% communicated a correct or partially-correct diagnosis. These observations suggest that insufficient diagnostic effort and accuracy may partly contribute to the high rate of inappropriate antibiotic treatment observed at baseline.

Table 7.10. Effect of private performance feedback on diagnostic accuracy

	(1) Correct / partially-correct diagnosis	(2) Correct / partially-correct diagnosis	(3) Correct / partially-correct diagnosis
Treatment (1 = feedback given)	-0.01	0.00	0.02
	(0.12)	(0.13)	(0.14)
Baseline correct/partially-correct diagnosis, for Clinical Case 1	-0.03	-0.03	-0.00
	(0.12)	(0.13)	(0.14)
Fieldworker fixed effects	Yes	Yes	Yes
Unbalanced covariates	No	Yes	Yes
Stratification variables	No	No	Yes
Mean of control group	0.46	0.46	0.46
Mean of sample	0.49	0.49	0.49
Obs.	75	75	75
R-squared	0.27	0.30	0.33

Notes. All specifications are estimated using linear OLS. Standard errors are in parenthesis. The outcome variable indicates whether the diagnosis communicated to SPs was correct or partially-correct, conditional on a diagnosis being communicated. This analysis therefore excludes the 5 GPs who did not communicate any diagnosis to SPs either at baseline or end-line. Column (1) presents the treatment effect estimate from a basic specification, controlling only for baseline diagnostic accuracy and fieldworker fixed effects. Column (2) additionally controls for all significantly or sizably unbalanced observables (baseline consultation duration, baseline drug costs, ethnicity, gender, northern suburb practice location, and perceived patient expectations for antibiotics in Clinical Case 2). Column (3) presents the main specification, controlling also for the stratification variables (GPs drug dispensing and contracting status, and baseline antibiotic treatment). Multicollinearity is unlikely to be an issue in these three specifications: mean variance inflation factors (VIF) for each specification was around 2 or lower, and individual VIF for all included variables were all around 4 or under. See notes in Table 7.2 and Section 4.2 for details on how certain covariates and the diagnostic accuracy outcome were measured.
*** p<0.01, ** p<0.05, * p<0.1

Nevertheless, results in Tables 7.9 show no evidence of an impact of private A&F in improving diagnostic effort: there are no significant positive effects detected on consultation duration, the proportion of all case-specific checklist items completed, or the IRT score. Similarly, Table 7.10 shows no evidence of an A&F effect in improving diagnostic accuracy (the likelihood of a correct or partially-correct diagnosis) relative to the educational intervention alone. One reason for these null effects may be that the feedback was not optimally designed to incentivise improved diagnostic effort. Although the share of all case-specific checklist items completed at baseline was reported in the feedback sheet, the individual checklist items (or GPs' disaggregated

analysis of diagnostic accuracy. Such considerations were also taken into account in coding specific diagnoses, particularly with regards to local colloquialisms. One example was in coding a "flu" diagnosis. In strict medical terms, this is incorrect. However, acute bronchitis is usually preceded by flu-like symptoms; and discussions with local medical practitioners highlighted that, in the South African context, doctors often do not make a distinction between "flu" and "cold" in communicating diagnoses to patients. Accordingly, no distinction was made in coding "flu" and "cold," and both were coded as partially-correct *if* they were combined with a bronchitis or similar diagnosis (since a cold preceded the acute bronchitis according to the case history).

performance on each item) were not specified. Therefore, it is unlikely to have been particularly effective in informing GPs on their effort performance. [304]

Taken together, these results indicate that increased diagnostic effort or accuracy cannot explain the A&F effect on inappropriate antibiotic treatment. This suggests that alternative drivers, including habit persistence or social pressures in treatment choices – behavioural barriers that may be countered with intrinsic incentives - could instead play a role.

7.4.5 Robustness checks

Appendix E.2 displays output from six robustness checks on estimated results from the main specifications. Firstly, Columns (1) and (2) show output from two approaches to correcting for non-participation and attrition bias in the main results. Column (1) presents estimates from an ITT analysis that includes all 99 randomised GPs, where missing outcomes for those lost to non-participation or attrition are imputed by carrying forward their baseline values.[305] The estimated A&F effect on the main outcome (inappropriate antibiotic treatment) is reduced in magnitude and statistically insignificant. This reduction in effect magnitude is to be expected, as baseline antibiotic prescribing was used as a stratification variable in the randomisation process. Using those baseline values to then impute missing outcomes for almost 20% of the total sample should skew the result toward a null effect to some extent. Nevertheless, the absolute magnitude of the point estimate remains large and the effect is still negative. A further limitation of this ITT analysis is that the 19 GPs lost to non-participation or attrition have very limited covariate data available (as interviews were not conducted with 16 of these GPs). This prevents tests and controls for any imbalance along observable covariates between the treatment and control groups in the regression specification (as in the main specifications in Sections 7.4.2 - 7.4.4).

Column (2) estimates upper and lower bounds for the true effect (in the absence of any non-participation or attrition bias) using the Manski-Horowitz method (Horowitz and Manski, 2000). The upper bound is estimated by assigning the best possible outcome to treatment group GPs with missing outcomes, and the worst possible outcome to control GPs with missing outcomes. The lower bound is estimated using the reverse assumption: the worst outcome is assigned to treatment group GPs, and the best outcome to control GPs, with missing outcomes. This method is suitable when the outcome is binary and attrition rates are low, as continuous outcomes and high levels of

[304] Only marginal improvements in diagnostic effort and accuracy are noted even in the control group, compared to their baseline performance (see Table 7.2). This may indicate that the educational intervention alone also had only limited effect on these measures. Again, I am unable to isolate the educational (control) intervention effect from general time trend, due to the absence of a 'no-intervention' group.

[305] This 'last-value carried forward' approach assumes that the last (baseline) observation is representative of subsequent missing observations. A clear limitation here is that this amounts to assuming that the intervention would have no effect on attriters and non-participators.

missing data can yield very large bounds for the estimate of interest and become uninformative (Glennerster and Takavarasha, 2013). As such, it is only used here to estimate upper and lower bounds for the main binary outcome of interest (inappropriate antibiotic treatment). Columns (2a) and (2b) suggest the true effect could fall within a large interval: somewhere between a small and insignificant negative effect at the lower end (-0.03, p=0.781) and a strong negative effect at the upper end (-0.32, p=0.002). The main estimate from this chapter (from the per-protocol analysis in Section 7.4.2) falls within this bound (-0.24, p=0.027). Considering the small sample size in this study, these results provide some assurance to the chapter's main conclusion on the effectiveness of private A&F in reducing inappropriate antibiotic treatment.

Columns (3) – (5) adjust for other characteristics of study design and implementation. Although the order of the two end-line SP visits was randomised for each GP in both intervention groups, Table Column (3) explicitly controls for the visit order to adjust for any residual order effects. Column (4) controls for the hour of day of each SP visit, to adjust for any time-of-day effects; and Column (5) controls for the number of days in between intervention delivery and end-line outcome measurement (the corresponding SP visit) for each GP, to account for the previously noted imbalance in this visit characteristic between the treatment and control groups. All results are shown to be robust to these three controls.

Finally, given the small sample size of the experiment, the precision of estimated results may be improved with the addition of explanatory covariates that account for more of the unexplained variance in outcomes. Focusing on the main outcome, a number of explanatory factors behind inappropriate antibiotic treatment have been identified in the literature. The specification in Column (6) accordingly controls for additional GP-level covariates (as listed in Table 7.2), that proxy commonly cited factors: indicators of GPs' perceived prescribing norms and patient expectations for antibiotics for the two clinical cases in the experiment, to gauge social prescribing pressures (Altiner *et al.*, 2004; Dempsey *et al.*, 2014; Farley *et al.*, 2018);[306] [307] practice location and local competitor density, to gauge local market competition (Butler *et al.*, 1998; Bennett *et al.*, 2015); indicators of diagnostic competence and perceived effectiveness of antibiotics for the two clinical cases, to proxy knowledge and diagnostic uncertainty (Brink *et al.*, 2016); and reported daily patient load, to capture time constraints (Dempsey *et al.*, 2014).[308]

[306] Perhaps the most commonly-cited driver of inappropriate treatment choices by physicians is actual or perceived patient demand (Butler *et al.*, 1998; Dempsey *et al.*, 2014; Farley *et al.*, 2018). While actual patient demand is controlled for using scripted SPs, it is possible that GPs still *believe* that patients will expect antibiotics due to prescribing norms or culture (even if patients do not voice it). Whether GPs respond to these perceived pressures can also depend on market competition (i.e. how likely it is that patients can find alternative willing providers, and how costly it is to lose individual patient business).

[307] Two of the available indicators on perceived social pressures listed in Table 7.2 are excluded, as they are highly correlated with some of the unbalanced covariates included in the main specifications. These are the two variables on perceived patient expectations for antibiotics in Clinical Case 1 and perceived antibiotic prescribing norms for Clinical Case 2.

[308] See Section 4.2.3 for details on how these GP-level covariates were measured or constructed.

While the magnitude of some effects change under this specification (notably, for the drug cost and diagnostic accuracy outcomes), the statistical significance of all estimates remains largely unchanged.

7.5 Conclusions and Discussion

This chapter evaluates the impact of intrinsic informational incentives for physicians in lowering inappropriate antibiotic treatments and subsequent costs, in a context where physicians are informed about correct treatment guidelines as well. To solve common measurement and patient selection problems associated with the use of administrative data in assessing physician performance, it demonstrates a novel application of the SP audit approach to evaluating physician-targeted interventions. In doing so, it provides highly granular, causal evidence on the impact of private A&F on physicians' antibiotic treatment choices.

Unlike previous studies, this chapter not only isolates the intrinsic incentive effects of A&F, but also investigates potential causal channels, and tests for unintended spillovers from the intervention. The results indicate that giving private, targeted performance feedback to primary care physicians reduces the likelihood of inappropriate antibiotic treatment for a common, viral respiratory infection by 46%, compared to providing educational materials on clinical guidelines alone. There is also suggestive evidence of lower treatment costs following A&F (although this result is not statistically significant).[309] Reassuringly, the private A&F had no unintended spill-overs on physicians' ability to differentiate and appropriately manage more severe presentations of the same clinical case, or on their likelihood of giving other inappropriate drugs (such as steroids).[310]

The main finding on inappropriate antibiotic treatment lends support to two hypotheses from previous literature. First, that physicians respond positively to private performance feedback, despite the lack of any financial, reputational or market demand consequences, aligns with a common characterization of healthcare providers as being – to some extent – intrinsically-motivated (see Galizzi *et al.* (2015) for a review). Physicians derive some intrinsic utility from effort, independent of any profit or reputational motive. Moreover, this utility need not be peer-

[309] However, treatment costs remain inefficiently high even after feedback: the average cost of drugs given by physicians who received A&F is still almost 6 times that of the cheapest, appropriate treatment available on the local market for the clinical case (based on expert advice and treatment guidelines, this included paracetamol and a cough suppressant syrup, at a total cost of R17.86).

[310] The absence of such unintended spill-over effects may be down to two features of the intervention design. First, the feedback content was very specific about the particular clinical case that it applied to, where antibiotics were considered inappropriate. The aim was to minimize the risk of physicians misinterpreting the feedback message as applying more broadly to antibiotic treatment in general. In addition, the educational materials that were provided to all physicians (including those that received A&F) were designed to inform them of appropriate treatments for a number of respiratory infections of varying severity and drug indications. This baseline educational intervention may have also acted to offset any adverse effects of A&F on appropriate drug use.

referenced, as often argued (Kolstad, 2013; Ashraf and Bandiera, 2017b): the feedback in this experiment gave no information on peer performance, and instead benchmarked individual performance to evidence-based guidelines. Second, it implies that a lack of knowledge is not the primary barrier to appropriate treatment choices, reflecting previous findings of a "know-do" gap among healthcare providers in low-and-middle-income countries (LMIC) (Das and Gertler, 2007; Leonard and Masatu, 2010; Mohanan *et al.*, 2015), including among this study's sample at baseline. All physicians (in treatment and control groups) received the same educational materials on appropriate management of the clinical cases presented by SPs. Yet the control group still gave inappropriate antibiotics in 52% of cases. This suggests that the A&F intervention cannot have worked by improving a knowledge gap, but rather by bridging a 'know-do' gap in appropriate treatment choices.

Visibly higher diagnostic effort or accuracy cannot explain the main A&F effect either. Physicians that received A&F did not spend more consultation time with the patient, conduct any more thorough history-taking or examinations, or give more accurate diagnoses than those in the control group. This suggests that measured diagnostic performance cannot fully explain treatment choices, and raises the question of alternative channels through which the A&F intervention might work. There is some indication that habit persistence, aligned with social prescribing pressures, may influence treatment decisions - echoing results in previous studies on prescription drivers (Crea *et al.*, 2019). Physicians who chose inappropriate antibiotic treatment at baseline were 50 percentage points more likely to choose the same treatment for the same case at end-line. Moreover, in interviews conducted in this study, physicians reported peer norms and patient expectations for prescribing antibiotics, in cases where they know antibiotics are unlikely to be clinically effective.[311] Treatment habits and perceived social pressures are costly to overcome (Hellerstein, 1998; Frank, 2004; Crea *et al.*, 2019), even where physicians are intrinsically-motivated to provide optimal care. Thus, the observed A&F effect on inappropriate antibiotic treatment may demonstrate the relative effectiveness of intrinsic informational incentives in helping to counter such barriers to behavior change – barriers that are, perhaps to some extent, sustained by physicians' limited awareness of their own performance.

While much of the policy discourse on clinical performance improvement has focused on external incentives – including via pay-for-performance (P4P) schemes or public reporting – this chapter finds that intrinsic incentive interventions may be an alternative, effective and less contestable

[311] When an uncomplicated bronchitis case was presented to physicians in vignette-form, they believed there was a 56% likelihood that patients would seek care elsewhere if they did not give antibiotics for the case, and a 65% likelihood that other GPs would give antibiotics for the same case - despite also believing there was only a 30% likelihood that the case would resolve more quickly with antibiotics.

means to achieving targeted improvements – *without* worsening performance in non-targeted aspects of care, or generating incentives for providers to 'game' performance auditing systems. Among studies that have found a positive impact of P4P on clinical performance, the effect size is very moderate at best (a 5% average improvement in targeted measures) (Eijkenaar *et al.*, 2013). Evidence on public reporting similarly concludes little to no effect on provider performance (Metcalfe *et al.*, 2018). While this is a relatively small study, the 46% improvement in targeted performance reported here represents a sizable effect in comparison.[312] Perhaps the most comparable existing study to this present one revealed an effect of similar magnitude: Meeker *et al.* (2016) find that private A&F lowered inappropriate antibiotic treatment for respiratory infections by GPs in the US by around 47% (compared to guidelines education alone). That study differed in its use of peer-benchmarking in performance reporting, and its reliance on electronic health records (EHR) for measuring clinical performance.[313] The present study therefore validates this earlier finding in an LMIC setting where baseline performance was much poorer,[314] using standardised patients to improve performance measurement, and benchmarking performance to best-practice standards.

As with any intervention, successful implementation relies on its acceptability to participants and other stakeholders. With the feedback intervention evaluated in this chapter, the confidentiality of feedback data, and use of SPs for accurate performance measurement, overcomes some of the common measurement and reputational issues tied to public reporting that have made these latter initiatives highly contested among healthcare providers and policy-makers. While scaling the use of SPs in performance measurement may not be feasible in resource-constrained settings, the present study nonetheless offers a proof-of-concept for evaluating its cost-effectiveness and feasibility on a larger scale.

As the South African government moves forward with its plans to contract private GPs to expand access to free primary care, addressing the inefficiencies and norms of over-treatment noted in this study will be paramount. Current discussions have focused on reforming payment incentives; for instance, through the introduction of cost-containment measures such as capitation or bundled payments. However, these latter measures raise the risk that physicians instead under-treat

---

[312] It is substantial even in comparison to average effect sizes reported for other physician-level interventions targeting antibiotic treatment (an average 11.6% improvement in targeted outcomes (van der Velden *et al.*, 2012)). One explanation for this is that patient demands for antibiotics, which can hinder appropriate treatment choices in practice, cannot be controlled for in studies using administrative data. In the present study, SPs were trained not to ask for any drugs, so the influence of actual patient demand (in tempering A&F effects) is completely eliminated. The effect size difference may also be partly due to the small sample size in this study (which can inflate estimated effects) and the single end-line observation per sample GP for each clinical case (rather than multiple observations of the same GP, which is most often the case in studies using administrative data).

[313] Meeker *et al.* (2016) further evaluate outcomes at the practice level, rather than at the individual physician level (although feedback in their intervention is provided at the physician level, as in this study).

[314] The mean inappropriate antibiotic prescription rate at baseline was around 22% in Meeker *et al.* (2016)'s treatment and control groups, compared to 61% in this study sample.

patients and supply poorer quality care, under pressure to cut costs. Without a perfect means to monitor such risks, finding ways to leverage physicians' intrinsic motivations to provide high quality care will be necessary. This chapter provides preliminary evidence that physicians respond well to intrinsic informational incentives; and that periodic private performance feedback, perhaps utilizing the SP audit approach, can be an effective way to rationalize treatment choices.

A number of limitations to this chapter's findings must be noted. As mentioned previously, a weakness of the experiment design is that intervention compliance – particularly in the control group – cannot be verified. As I am more certain of full intervention compliance in the treatment group,[315] the estimated ITT effect may be over-estimating the true marginal effect of A&F relative to the educational intervention alone (that is, if all control group GPs had also complied). Moreover, this study does not evaluate the persistence of A&F effects over time. Each GP was assessed only at one post-intervention time point for each outcome of interest. End-line SP visits were mostly conducted within a month of intervention delivery, and it is plausible that a later scheduling of these visits may have recorded diminished effects.

The external validity of results is a further concern, given the clinical cases used and the ethical requirement for informed consent from all GP participants. Only two variants of an acute bronchitis case are employed here in assessing GP performance, which limits generalizability of findings to other clinical scenarios. Nevertheless, in a country where almost 20% of adults aged 15-49 are HIV+, and where respiratory infections are the most common reason for consultations and antibiotic treatment in primary care, the cases remain highly pertinent to investigating appropriate and differential antibiotic treatment choices by GPs.

There is also likely to be some self-selection bias in the GPs that consented to participation in this study. In the ESRC study, from which this study sample was drawn, the non-participation rate was quite high among GPs approached for recruitment (63%).[316] It is possible that GPs who choose to participate in research studies on care quality are more intrinsically-motivated, or competent, than those who do not.[317] If this is the case, and baseline intrinsic motivation (or competency) increases the likelihood that physicians respond well to private A&F (as suggested in Lee (2018)), then the estimated results in this chapter would represent upper bounds of the true effects in the population. It is difficult to verify the extent of this bias, given the limited data

[315] The interviewers that delivered the intervention explicitly recorded whether treatment group GPs read the intervention materials immediately in front of them (and almost all treatment group GPs were reported to have done so (35 of 37)). This procedure was not followed for control group GPs, however.

[316] Non-participation was either because GPs could not be reached (despite several contact attempts by the research team), were reached but refused to take part, or were reached but did not give a final decision to participate.

[317] This may partly explain the high level of measured altruism in the study sample – a trait that is conceptually related to intrinsic motivation.

available on the sample population. From available data, it can be shown that the study sample has a lower proportion of female GPs (30% vs. 45%) and a higher proportion of dispensing GPs (57% vs. 39%) than the GP population of Johannesburg (see Table 4.6 in Chapter 4). Based on data collected in the ESRC study, dispensing GPs tend to be located in less affluent areas, accept poorer patients, charge lower rates, and display higher levels of measured altruism than non-dispensing GPs. Female GPs are also more likely to be non-dispensing, and be located in wealthier areas (northern suburbs of Johannesburg). This suggests that altruism in the study sample (and relatedly, intrinsic motivation) may indeed be positively skewed, relative to the broader population.

The study's small sample size presents three further limitations. First, there is a risk that the magnitude of the estimated treatment effect on inappropriate antibiotic use is inflated, as is sometimes the case in small sample studies. The large effect size should therefore be interpreted with caution. Second, there is a higher risk of obtaining false negatives, meaning I cannot rule out a true effect where a statistically-significant effect is not detected (for instance, on total drug costs) (Cohen, 1988; Button *et al.*, 2013). Third, it prevents testing for heterogeneous treatment effects, which could yield valuable insights on underlying causal channels for improving the intervention design and targeting. Relatedly, the small sample size prevented the inclusion of an additional, 'no-intervention' comparison group in the experiment - to isolate the effects of the educational intervention alone (which currently forms the control intervention).

These limitations present opportunities for future research. A larger sample study could test for interactions between the A&F mechanism and GP characteristics of interest. For instance, Ivers *et al.* (2012)'s review concludes that A&F is more effective in contexts where baseline performance is low, whereas Lee (2018) finds that health workers' response to intrinsic informational incentives is greater when baseline intrinsic motivation is high.[318] It would be interesting to validate these findings in the South African or similar LMIC context, and test whether the A&F effect differs between physicians with high and low altruism, or between those with high and low baseline performance. In addition, although A&F has proven effective in this study, more evidence is needed to evaluate its cost-effectiveness relative to less resource-intensive interventions (such as the educational intervention in this experiment), to understand its feasibility in resource-constrained settings. Lastly, while inappropriate antibiotic use was the main target in this study, the SP method can be equally applied to evaluating A&F interventions designed to

[318] Both these findings may also explain why the main treatment effect estimated in this chapter is so large: poor baseline performance in rational antibiotic use and relatively high baseline altruism (or intrinsic motivation) in the study sample on average.

prioritize other care quality or cost dimensions (such as diagnostic effort or treatment costs, which are treated as only secondary feedback components in this study)[319] to test for similar effects.

[319] The lack of emphasis or actionable detail given to GPs on these other performance feedback dimensions in this study may partly explain the absence of any significant A&F effects on these dimensions. It remains to be seen if similar future interventions, designed to prioritise these alternative dimensions with more detailed feedback, can yield significant effects.

8 Discussion and Conclusion

This thesis began with the observation that healthcare overuse places a significant burden on health systems around the world, at a time when many low- and middle-income countries (LMIC) are stretching public budgets to engage private providers and improve care access for their populations. Private healthcare providers are increasingly recognised as necessary partners for universal health coverage (UHC) success (McPake and Hanson, 2016; Clarke *et al.*, 2018). However, valid concerns remain regarding their limited accountability to patients (Montagu and Goodman, 2016), and subsequent effects on the quality and efficiency of care delivery. Robust evidence on the drivers of inappropriate care in the private sector in LMIC settings - to inform the regulation of these trends – is limited.

Focusing on inappropriate antibiotic use in particular, this thesis sought to explore the behaviour of private physicians, and their incentives to over-treat - particularly as patients become more financially protected under UHC. In the context of South Africa - where private physicians are set to play an integral role in the planned national health insurance (NHI) reform - it set out to investigate the following:

1. The impact of increasing patient insurance (financial protection) on the quality and efficiency of care provided by private physicians.
2. The mediating role of market competition and intrinsic motivation on physicians' healthcare choices.
3. The impact of intrinsic informational incentives from private performance feedback (A&F) in lowering inappropriate antibiotic treatment and costs.

To explore these questions, I designed and conducted two field experiments with private primary care physicians (GPs) in Johannesburg, South Africa. This chapter pulls together and summarises the key findings of this study (Section 8.1); highlights its strengths, limitations and future research implications (Section 8.2); and concludes with the main policy implications (Section 8.3).

8.1 Summary of Main Findings

The study reveals that improved financial protection for patients significantly increases the likelihood of over-treatment by private physicians, even in the absence of any patient demand or influence. Notably, this finding holds *both* when physicians have direct (fee-for-service) financial

incentives to over-treat *and* when they do not - in line with the theoretical predictions in Chapter 3 (*Hypotheses 2* and *5*). Patients with more generous insurance are almost twice as likely to be ordered unnecessary fee-for-service (FFS) tests or procedures, and are charged more in FFS consultation costs (Chapter 6). This aligns with findings elsewhere - in markets for healthcare, computer repairs and taxi rides – that support the conventional *second-degree moral hazard* hypothesis: expert suppliers will exploit the lower cost-consciousness of more insured patients to oversupply services when they have direct opportunities to profit (Lu, 2014; Kerschbamer *et al.*, 2016; Balafoutas *et al.*, 2017). However, Chapter 5 finds that inappropriate antibiotic prescribing by physicians who have no opportunity to profit from drug prescriptions is also almost 25% higher for more insured patients. These patients further receive more, and more expensive branded drugs (Chapter 6), despite the absence of any clinical need for drug treatment or any financial incentives for the prescribing physician. This latter result runs counter to most existing literature on physicians' prescription choices, which conclude that physicians are unresponsive to patient insurance when they have no financial gain (Hellerstein, 1998; Iizuka, 2012; Lu, 2014; Crea *et al.*, 2019). These patterns of over-treatment coincide with significant reductions in the technical quality and efficiency of care with better insurance.

What drives this greater tendency for over-treatment with more insured patients? While the conventional theory on *second-degree moral hazard* can explain the result for treatments incentivised with FFS payment, it cannot explain the same result for drug prescriptions. The study design eliminated all differences in patient demand or characteristics (apart from insurance status), which rules out the potential influence of patient moral hazard or adverse selection into insurance. The within-subject experiment design also rules out the influence of fixed physician-level characteristics, such as clinical knowledge or experience.

Instead, the theoretical framework and results point to a more nuanced effect of provider effort. The model developed in Chapter 3 predicts that less appropriate clinical effort and greater diagnostic uncertainty may be one reason for over-prescribing. Physicians are not observed to exert any less clinical effort for more insured patients (Chapter 5) - indeed, they exert marginally higher observable effort with these patients. Yet, diagnostic accuracy is still poorer for high-insured patients, suggesting that there may indeed be some (unobserved) elements of appropriate clinical effort that are lower for these patients. At the same time, the study finds no significant association between diagnostic accuracy and the likelihood of antibiotic treatment (or the quantity and costs of drug treatment) on average - implying that measured diagnostic performance cannot be the only driver of drug treatment choices either.

Chapter 3 further proposes a role for physician tendencies to satisfy perceived patient expectations, in driving over-treatment. It predicts that private physicians will prioritise aspects

of care that are observed and valued by patients, irrespective of clinical need. Taken together, the empirical findings in this study corroborate this argument that physicians provide care in line with the expected value and cost salience to patients, as anticipated from patients' insurance status. They reduce overtreatment for more cost-sensitive, low-insured patients; and provide relatively more observable yet costly aspects of care (including aspects of clinical effort and treatments that patients value) to less cost-conscious, high-insured patients. Physicians may also 'price-in' their observable effort in their FFS costs: the results show a significant, positive association between measured clinical effort and FFS consultation costs – both of which are higher for more insured patients.

The theoretical framework proposes that these concerns for perceived patient preferences are rooted in private physicians' competitive pressures to secure repeat business, and their altruism towards patients. While the framework is silent on how changes in physician competition or altruism should alter the observed treatment differentials by patient insurance cover,[320] the results provide some evidence on mediating effects. More altruistic physicians are found to generate relatively lower avoidable drug costs for less insured patients; whereas physicians located in wealthier suburbs (with presumably lower localised cost competition) prescribe relatively more inappropriate drugs to more insured patients. Physicians with a lower density of competing physicians in their locality are also relatively more likely to prescribe inappropriate antibiotics to high-insured patients.[321]

Two further findings from this analysis are notable. Firstly, although physicians had a lower tendency to over-treat less insured patients, the average level of over-treatment with drugs was still very high in this patient group (and in the study samples overall) – suggesting a general problem of over-prescribing, as implied in *Hypothesis 3.* Secondly, there is some evidence of a 'know-do' gap with respect to antibiotic treatment for the average physician: while a large proportion of physicians were able to correctly identify the viral cause of an uncomplicated acute bronchitis case in a knowledge quiz (and report that antibiotics would not be very effective for the case), many of the same physicians also prescribed antibiotics for the same case in their clinical practice. Indicative evidence that physicians perceive strong prescribing norms and patient expectations for antibiotics for that clinical case may partly explain this gap.

These findings raise the question of how physicians may be incentivised to reduce over-treatment for the average patient. Chapter 7 finds that a private audit and feedback (A&F) intervention –

[320] The framework merely predicts how a change in these factors should change care for the *average* patient (not how it should affect care differences between high- and low-insured patients).

[321] It is nevertheless noted that the net demand elasticity effect of competitor density is unclear. Although a higher physician density is commonly associated with a higher level of physician competition, it may also increase search costs for patients (Satterthwaite, 1979) and therefore lower demand elasticity.

informing physicians of their performance relative to professional best-practice guidelines – can substantially reduce inappropriate antibiotic treatment for uninsured patients (for whom both physicians' competitive and altruistic incentives to reduce over-treatment should already be high) compared to a passive educational intervention alone. Once again, there is little evidence of a consistent link between diagnostic accuracy and antibiotic treatment decisions: despite reducing inappropriate antibiotic treatment, private A&F had no corresponding effect on diagnostic effort or accuracy. Together, the results underscore the implications of a 'know-do' gap: simply ensuring that physicians have sufficient diagnostic or therapeutic (guidelines) knowledge is not enough to minimise over-treatment. Additional incentives are required to bring actual performance in line with knowledge.

8.2 Strengths, Limitations and Future Research Implications

Each empirical chapter (Chapters 5-7) discussed the literature contributions and limitations of its own analysis. This section summarises those discussions, reflects on the overall strengths and limitations of this study, and draws implications for future research.

8.2.1 Theoretical validity

This thesis began with a conceptual framework that brought together several theoretical arguments and empirical observations on physician behaviour, to guide the empirical investigation and interpretation of results. The findings lend support to many of the hypotheses derived in that framework. While the framework's assumptions were mostly grounded in findings from previous literature, a few key assumptions could not be fully validated in this study and invite further research.

Firstly, the framework assumes that physicians perfectly anticipate how patient demand would vary with insurance cover, and perceive more insured patients to be less cost-sensitive. In this study setting, it is not known how actual patient demand responds to insurance cover or how physicians perceive this demand response in practice. The most robust evidence in support of how the framework models demand responses to insurance comes from the US, from the RAND and Oregon health insurance experiments (Newhouse and Insurance Experiment Group, 1993; Finkelstein *et al.*, 2012). However, equivalent evidence from LMIC is limited;[322] and moreover, even if patient demand responds in this way, it is not clear if physicians correctly anticipate those responses as predicted in the framework. A qualitative investigation into how physicians perceive

[322] One exception is a recent study by Hausofer *et al.* (2020), which randomly allocated private health insurance to patients in Kenya and found no effect on healthcare utilisation. This result may be due to other demand-side constraints.

patients' healthcare expectations to vary with insurance would be an interesting extension to this study.

Relatedly, that physicians supply more profitable services and less appropriate effort to more insured patients (as evidenced in this study) is consistent with both the theory of *second-degree moral hazard* – as modelled in the framework - and *perfect agency* (Pauly, 1980). While Chapters 5 and 6 argue in favour of the former, to fully validate this theory it is necessary to show evidence of demand-inducement (SID) and effort-stinting in the study sample – in other words, that both high- and low-insured patients in practice demand less services and more clinical effort than what physicians are observed to provide. As this study does not evaluate demand-side responses to insurance, it cannot verify this. Although Chapters 5 and 6 show evidence of unnecessary treatment (more frequently for more insured patients) without any actual demand from SPs, the theory of *perfect agency* could argue that this observed behaviour is just physicians trying to perfectly satisfy *anticipated* patient demand – without any selfish, or profit-driven motivation. Nevertheless, there is supportive quasi-experimental evidence of SID under FFS incentives from other contexts (Fuchs, 1978; Gruber *et al*., 1999). One experimental study with SPs also succeeds in separating these two hypotheses (by experimentally varying both the financial incentives and insurance status' facing physicians), and provides evidence in support of *second-degree moral hazard*, but not in support of *perfect agency* (Lu, 2014). Taken together, this evidence suggests that *second-degree moral hazard* is likely to explain the study results. Nevertheless, a 2x2 experimental design like in Lu (2014), where both the GPs' financial incentives and patient insurance status are simultaneously and exogenously varied, would be necessary to fully verify this. This was not feasible in this study context, and is left to future research.

Thirdly, the two factors that are predicted to influence physicians' sensitivity to patient demand are their competitive (market-based) and intrinsic motivations. The findings in Chapters 5 and 6 lend some support to this.[323] However, the results in these chapters cannot perfectly separate the influence of competition from that of intrinsic motivation, as both factors are predicted to influence physician choices in the same way (less overtreatment for less insured patients). Again, complementary qualitative research on the relative influences of competitive pressures and intrinsic motives in physicians' care choices would be insightful.

Lastly, the framework predicts that physicians' intrinsic motivations can be increased by providing them with new information on their performance relative to an accepted 'best practice' benchmark (be it professional guidelines or peer norms). The results from the private A&F

[323] In interviews, physicians reported strong perceptions of patient demand elasticity with respect to their antibiotic treatment choices, for example. Moreover, the average level of measured altruism in the study samples were very high, and there is limited evidence that measures of competition and altruism mediate physician responses to patient insurance in this study.

intervention evaluated in Chapter 7 provide some support to this. However, the framework also models this intrinsic incentive effect as being proportional to the baseline performance of physicians: it is implicit that those performing worst relative to the benchmark will receive the largest incentives to improve from the 'new' A&F information. This assumption is drawn from Tonkin-Crine *et al.* (2017)'s evidence review, which concludes that A&F interventions are likely to be more effective when baseline performance is low. It was not possible to test this assumption in the present study, however. The small sample size (as discussed further in the next section) impeded the testing of heterogeneous treatment effects by baseline performance. Future studies with larger sample sizes could seek to validate this assumption – by investigating how estimated A&F effects vary by baseline performance – and provide further insights into how intrinsic incentives might operate through A&F.

8.2.2 Methodological considerations

The experimental methods used in this study enabled the generation of novel, field-experimental evidence on the impact of patient insurance on physicians' treatment decisions, and the scope for intrinsic incentives in regulating these decisions. The SP audit methodology allowed the study to make two broad contributions. Firstly, it builds on existing evidence on the effects of patient insurance on treatment outcomes in two important ways: *i)* unlike almost all other empirical studies evaluating insurance effects (which rely on observational data), this study's SP method enabled the isolation of physician-level drivers of treatment outcomes from patient-level factors; and *ii)* unlike the one experimental study that also uses the SP method to study insurance effects on treatment outcomes (Lu (2014)), this study's within-subject experiment design enabled observation of the insurance effect on the *same* physician (which controls for all physician-level characteristics that can confound the estimated effect). Secondly, this study develops a nascent application of SPs to the evaluation of randomised provider quality interventions. To my knowledge, very few studies have utilised SPs in this way (Das *et al.*, 2016; Harrison *et al.*, 2000; Mathews *et al.*, 2009; Mohanan *et al.*, 2017). This approach also offers a promising means to validate findings from evaluations of similar interventions conducted with routine, administrative data – for instance, Meeker *et al.* (2016)'s evaluation of a similar A&F intervention as in this study, which finds a comparable effect.

Despite these contributions, there are certain limitations to the use of SP methods in this study. First, SP audit studies are very resource-intensive. As such, the sample sizes in this study are small, and there are only two observations per GP in each experiment. This is a common limitation of SP audit studies in general, and this study benefits from slightly larger samples than some others (Lu, 2014). The small samples constrain the study's statistical power for detecting small-medium true effects, and for investigating heterogeneous treatment effects; and increase the risk

of estimating inflated effect sizes. These limitations must be kept in mind when interpreting study results. The small sample constraint is a relative weakness of using SPs to measure treatment outcomes compared to large administrative datasets, which could offer many more observations per participant. Again, this highlights the value of cross-validation using both approaches where possible, and invites future research that uses larger sample sizes and perhaps insurance claims data to replicate and validate this study results.

Second, only two variants of an acute bronchitis case are used in the experiments to evaluate physicians' clinical performance. These cases are appropriate for evaluating inappropriate antibiotic treatments, particularly in the South African primary care setting. However, there are other clinical cases that are also compatible with the SP methodology, and that may be used to test the generalizability of this study's conclusions in future research.

Third, patient care continuity - an important dimension of primary care quality - cannot be assessed through the SP method given the heightened risk of detection from repeated SP interactions. The SP method is only amenable to evaluating physicians' performance in one-off, acute clinical cases (as in this study). Future studies could pursue a more holistic evaluation of physician performance by combining SP assessments with the analysis of clinical records that can track how the same physicians care for patients over time.

Relatedly, administrative data could be combined with SP assessments to evaluate the persistence of treatment outcomes following a performance improvement intervention, like the A&F in this study. A limitation of the A&F evaluation in Chapter 7 is that intervention outcomes are only observed at a single point in time. Repeated assessments with SPs over time would be very resource-intensive and was infeasible in this study. Although the two methods of quality evaluation are not equivalent, where resources are constrained and comparable electronic health records (EHR) are made available (unlike in this study), EHR could be a useful complement to SPs for evaluating the persistence of certain outcomes over time.

Further to the SP method constraints, four limitations to this study design must be noted. First, the two field experiments draw on two different GP samples. This was done partly for logistical reasons, and partly because different GP characteristics were of importance in the two experiments. Baseline data for the second experiment (which was taken from the ESRC study) was not available for the majority of GPs in the first experiment. To avoid funding additional baseline visits, it was decided to limit the second experiment sample to only those GPs with baseline data from the ESRC study. Moreover, all GPs in the first experiment are non-dispensing and contracted-in with large health insurance schemes (by design, to satisfy the research objectives of the first experiment); whereas GPs in the second experiment vary in their contracting and dispensing statuses, to better represent the distribution of these characteristics in the broader

GP population. As GPs voluntarily choose their dispensing and contracting status, this could imply systematic differences in the average GP across the two samples. For instance, in Johannesburg, dispensing GPs are more likely to be located in less affluent areas, accept poorer patients, and display higher levels of measured altruism toward patient welfare than non-dispensing GPs (Lagarde and Blaauw, 2019b). These sample differences necessarily limit the extent to which the findings from one experiment sample can be generalised to the other. The second experiment findings are perhaps more generalizable in this respect, as the main results are estimated with controls for GPs' contracting and dispensing status.

Second, the SPs used in the two experiments have different insurance statuses. All SPs have some form of insurance cover in the first experiment, while all SPs are uninsured in the second. Again, this was done to simplify logistics and satisfy the different experiment objectives. The first experiment was designed to evaluate the effects of different levels of patient insurance, so it was necessary that all SPs had some insurance cover there. In the second experiment, insurance was not a research concern and it was logistically simpler to have cash-paying SPs, to avoid having to purchase health insurance for fieldworkers and to ensure the SPs would be seen by all sample GPs (irrespective of their contracting status with insurance schemes). This difference may limit the extent to which the findings on intrinsic incentives from the second experiment (in Chapter 7) may be applied to a context where patients are insured, for example. However, the theoretical framework (in Chapter 3) would suggest that estimated effects with uninsured patients in Chapter 7 represent a lower bound of what could be expected with insured patients. Baseline intrinsic incentives for minimising overtreatment – as modelled in Chapter 3 - should be at their highest for uninsured patients. Therefore, the marginal effect of an A&F intervention targeting an increase in these incentives should be relatively lower for these patients than insured ones. This is a hypothesis left to future empirical investigation. A similar conclusion could also apply to findings from the first experiment (in Chapters 5 and 6). The estimated effects of increasing the level of patient insurance cover are likely to be lower-bound effects of moving from a patient with no insurance (SPs in the second experiment) to a patient with some insurance (SPs in the first experiment). A comparison of estimated insurance effects in Chapter 6 to those in Lu (2014) - where the author compares treatment choices for an uninsured patient and an insured one - lends some support to this hypothesis: the magnitude of the insurance effect on drug costs (for example) are much larger in Lu (2014) than in this study. Nevertheless, physicians' financial incentives were different in Lu's study (the doctors there were salaried hospital physicians and had positive financial incentives to prescribe). It remains to be investigated if a similar result can hold with private, non-dispensing GPs.

Third, GPs in both experiment samples were recruited with informed consent, which implies some level of selection bias in those that opted to participate in this research. In the first experiment, 46% of GPs contacted for recruitment refused to participate. The equivalent rate was 63% in the ESRC study (from which the second experiment sample is drawn). A further 16% of recruited GPs in the second experiment refused to on-going study participation following randomisation (prior to intervention delivery).[324] No monetary incentives were offered at the recruitment stage, and it was further explained to GPs at recruitment that the study was investigating the quality of care provision. It is therefore plausible that GPs who agreed to participate may be more altruistic (intrinsically motivated) or more clinically confident than those who refused. It is difficult to verify this with study data. It was shown in Chapter 4 that both experiment samples have a lower share of female GPs, and the second experiment sample has a higher share of dispensing GPs, than the GP population in Johannesburg. The ESRC study found that dispensing GPs are likely to be more altruistic than non-dispensing GPs, whereas female GPs are more likely to be non-dispensing and located in the wealthier areas of Johannesburg (the northern suburbs) (Lagarde and Blaauw, 2019b). This suggests that GPs in the study samples may indeed be more altruistic (and intrinsically motivated) than the broader GP population.

If baseline intrinsic motivation compounds how GPs respond to patient insurance or to A&F, then estimated effects in this study may be overestimating true population effects. The theoretical framework assumes that any changes in baseline intrinsic motivation will not alter the insurance effect;[325] and moreover, that baseline intrinsic motivation will not affect the intrinsic incentive effect. However, limited evidence from the first experiment (Chapter 6) suggests that altruism may mediate the insurance effect on certain care cost outcomes. Lee (2018) also finds that the baseline intrinsic motivations of rural health workers can mediate their performance responses to A&F. Nevertheless, more robust empirical research is needed to conclude how baseline intrinsic motivation mediates physician responses to the altruistic incentives from patient insurance, or the intrinsic informational incentives from A&F.

A last limitation to be noted relates to the first experiment design (in study Part 1). The SPs in that experiment had different levels of insurance cover, which was hypothesised to signal different levels of patient cost-sensitivity to the physician. The SPs also explicitly signalled their different cost-sensitivities (with reference to their respective insurance covers) in their opening statements. These opening statements were specifically designed to ensure that physicians were aware of the

[324] GPs who refused to participate did so for a number of different reasons: time and work pressures, periods of unavailability (leave) over the study period, multiple practice locations with little foresight on their future rotas, planned relocation, and so on. Some also raised concerns about study implications for the NHI design, and their engagement in it.

[325] An increase in baseline altruism is modelled to affect care choices for high- and low-insured patients in the same way, thereby leaving the treatment differential unaltered

different likelihoods of OOP payment for the two SPs. However, this explicit signalling makes it difficult to separate the pure insurance effect (the *implied* patient cost-sensitivity) from the effect of patients explicitly voicing their relative cost concern (in some ways, different explicit demands for cost-effective care). This distinction is unlikely to be of much consequence to treatment differences, since the SPs do not make differential treatment or treatment cost demands in their opening statements. Nevertheless, this could be investigated with an alternative 2x2 experiment design, where both the patient's insurance status and their explicit signalling of cost awareness (or explicit demand for cost-effective care, such as generic substitution in drug treatment) is experimentally varied. Such an experiment would yield further insights into the potential for patients to influence cost-effective treatment choices.

8.3 Policy Implications

Findings in Chapters 5 and 6 show that healthcare overuse related to patient insurance is not just a demand-side issue, as commonly implied in the literature on *ex post* moral hazard. The optimal choice of patient cost-sharing in insurance contracts must also consider subsequent incentives on the supply-side. The findings contribute to the current debate on how private GPs should be contracted under the NHI system in South Africa. It is clear that the most prevalent financial incentive structure for these GPs (fee-for-service, non-dispensing) will not be optimal, given the incentives to over-treat - and over-treat more as patients become better insured. Eliminating all patient co-payment (as proposed in the 2018 Medical Schemes Amendment Bill and the 2019 NHI Bill) without altering this incentive structure is likely to be inefficient. Given the already high rates of inappropriate, broad-spectrum antibiotic prescribing in South Africa, the potential public health consequences are considerable. A related study shows that the dispensing status of GPs ensures some level of supply-side cost-sharing (given the drug pricing regulations in South Africa) (Lagarde and Blaauw, 2019a) and could be a means of cost-containment as patients become more insured. Alternative cost-sharing proposals include some form of capitation payment for GPs, in place of (or combined with) FFS. Nevertheless, further evidence is needed to understand how these cost-sharing incentives will interact with patient insurance, and whether they instead create perverse incentives for *under*-treatment. The extent to which these alternative payment arrangements will be attractive to private GPs is another concern.

In the absence of a perfect means to contract for optimal patient care, Chapter 7 shows that appealing to physicians' intrinsic motivations instead may be an alternative (or complementary) strategy – even under the current FFS payment structure. It presents preliminary evidence that private A&F, combined with clear information on treatment guidelines, can encourage appropriate treatment choices. While the A&F intervention tested in this study specifically

targeted antibiotic treatment, a similar intervention could also target generic substitution (as branded drugs were another important driver of avoidable treatment costs in Chapter 6). However, any proposal to roll out a similar intervention at scale must first address two things.

First, the cost-effectiveness of such an intervention must be considered. As stated earlier, SP audits are highly resource-intensive, so ways of combining these audits with more readily available data sources could be investigated. To facilitate this, the routine collection of standardised and comparable data on contracted provider performance and costs should be a priority in the NHI system. Although such administrative data collection and monitoring is vulnerable to provider manipulation (for example, through up-coding of recorded diagnoses), the scope for this may be minimised by limiting the ties between monitored provider performance and pecuniary incentives. Moreover, the intervention effects detected in this study should ideally be validated in a larger sample study, and with both insured and uninsured patients. Economic analyses of such antibiotic stewardship interventions must also find ways to value the broader public health benefits of intervention effects, in relation to antibiotic resistance, in evaluating their cost-effectiveness (Leal *et al.*, 2017).

Second, although the scope of this study has been restricted to the private sector, the effectiveness of such an intervention in the public primary care clinics of South Africa should also be assessed. Blaauw and Lagarde (2019) found that rates of inappropriate antibiotic treatment in public primary care clinics (where more than two thirds of urban South Africans are estimated to seek primary care (NDoH *et al.*, 2019)) were almost 17% higher than in the private GP practices of Johannesburg, highlighting a considerable need for antibiotic stewardship also in those settings. The responses of salaried public health workers to private A&F may differ from those of private GPs – particularly if baseline intrinsic motivation is associated with the decision to work in the public sector (Kolstad and Lindkvist, 2013).

A final consideration for policy is the influence of patient demand. This study has primarily focused on evaluating supply-side incentives for over-treatment. However, a repeated argument throughout the study (and others) is that perceived patient demand may have an important influence on physicians' treatment choices, even in the absence of actual demand. This highlights an important role for patients, and patient education, in encouraging appropriate antibiotic use and treatment choices. Other studies have shown that when patients signal appropriate treatment preferences, this can result in better choices by physicians (Currie *et al.*, 2011; Lagarde and Blaauw, 2019). It seems reasonable to expect that if physicians respond to the implied cost-sensitivity of patients from their insurance, they would be just as responsive when patients vocalise this sensitivity. This suggests an important role for informed patient demand in influencing not only clinically-appropriate treatment choices (as evidenced in Currie *et al.* (2011)

and Lagarde and Blaauw (2019)), but also cost-effective ones like generic drug equivalents. Patient-side interventions – to educate and encourage patients to question physicians' antibiotic or branded drug choices, for example – could therefore be important complements to altering supply-side incentives.

References

Abuosi, A. A., Domfeh, K. A., Abor, J. Y. and Nketiah-Amponsah, E. (2016) 'Health insurance and quality of care: Comparing perceptions of quality between insured and uninsured patients in Ghana's hospitals', *International Journal for Equity in Health*, 15.

Aguiar, F., Brañas-Garza, P. and Miller, L. M. (2008) 'Moral distance in dictator games', *Judgment and Decision Making*, 3(4), pp. 344–354.

Allard, M., Thomas Léger, P. and Rochaix, L. (2009) 'Provider Competition in a Dynamic Setting', *Journal of Economics & Management Strategy*, 18(2), pp. 457–486.

Altiner, A., Knauf, A., Moebes, J., Sielk, M. and Wilm, S. (2004) 'Acute cough: a qualitative analysis of how GPs manage the consultation when patients explicitly or implicitly expect antibiotic prescriptions', *Family Practice*, 21(5), pp. 500–506.

Arellano, M. (1987) 'Practitioners' Corner: Computing Robust Standard Errors for Within-groups Estimators', *Oxford Bulletin of Economics and Statistics*, 49(4), pp. 431–434.

Arrow, K. J. (1963) 'Uncertainty and the welfare economics of medical care', *American Economic Review*, 53(5), pp. 941-973.

Ashraf, N. and Bandiera, O. (2017a) 'Altruistic capital', *American Economic Review*, 107(5), pp. 70–75.

Ashraf, N. and Bandiera, O. (2017b) 'Social Incentives in Organizations', *In prepartion for the The Annual Review of Economics, June 2017*.

Ashraf, N., Bandiera, O. and Lee, S. (2018) 'Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the delivery of Public Services', *Development Discussion Papers (65)*. London: STICERD, London School of Economics.

Ashworth, M., White, P., Jongsma, H., Schofield, P. and Armstrong, D. (2016) 'Antibiotic prescribing and patient satisfaction in primary care in England: cross-sectional analysis of national patient survey data and prescribing data', *The British journal of general practice*, 66(642).

Ataguba, J. E. and McIntyre, D. (2012) 'Paying for and receiving benefits from health services in South Africa: is the health system equitable?', *Health Policy and Planning*, 27(1), pp. 35–45.

Baicker, K., Fisher, E. S. and Chandra, A. (2007) 'Malpractice liability costs and the practice of medicine in the Medicare program', *Health Affairs*, 26(3), pp. 841–852.

Baicker, K. and Goldman, D. (2011) 'Patient Cost-Sharing and Healthcare Spending Growth', *The Journal of Economic Perspectives*. American Economic Association, 25(2), pp. 47–68.

Baicker, K., Mullainathan, S. and Schwartzstein, J. (2015) 'Behavioral Hazard in health insurance', *Quarterly Journal of Economics*, 130(4), pp. 1623–1667.

Balafoutas, L., Kerschbamer, R. and Sutter, M. (2017) 'Second-Degree Moral Hazard In A Real-World Credence Goods Market', *The Economic Journal*. Oxford, 127(599).

Bandiera, O., Barankay, I. and Rasul, I. (2005) 'Social Preferences and the Response To Incentives: Evidence From Personnel Data', *Quarterly Journal of Economics*, 120(3), pp. 917–962.

Bardey, D. and Lesur, R. (2006) 'Optimal Regulation of Health System with Induced Demand and Ex post Moral Hazard', *Annales d'économie et de statistiques*, (83/84), pp. 279–294.

Barros, P. P., Brouwer, W. B. F., Thomson, S. and Varkevisser, M. (2016) 'Competition among health care providers: helpful or harmful?', *The European Journal of Health Economics*, 17(3), pp. 229–233.

Baumann, A. O., Deber, R. B. and Thompson, G. G. (1991) 'Overconfidence among physicians and nurses: The "micro-certainty, macro-uncertainty" phenomenon', *Social Science & Medicine*, 32(2), pp. 167–174.

Becker, L. A., Hom, J., Villasis-Keever, M. and van der Wouden, J. C. (2015) 'Beta2-agonists for acute cough or a clinical diagnosis of acute bronchitis', *Cochrane Database of Systematic Reviews*, (9).

Bénabou, R. and Tirole, J. (2003) 'Intrinsic and Extrinsic Motivation', *The Review of Economic Studies*, 70(3), pp. 489–520.

Bénabou, R. and Tirole, J. (2006) ‘Incentives and Prosocial Behavior’, *American Economic Review*, 96(5), pp. 1652–1678.

Bennett, D., Hung, C.-L. and Lauderdale, T.-L. (2015) ‘Health Care Competition and Antibiotic Use in Taiwan’, *The Journal of Industrial Economics*, 63(2), pp. 371–393.

Berenson, R. A. and Rice, T. (2015) ‘Beyond Measurement and Reward: Methods of Motivating Quality Improvement and Accountability’, *Health Services Research*, 50(2), pp. 2155–2186.

Berner, E. S. and Graber, M. L. (2008) ‘Overconfidence as a Cause of Diagnostic Error in Medicine’, *The American Journal of Medicine*, 121(5), pp. S2–S23.

Berwick, D. M. (2017) ‘Avoiding overuse—the next quality frontier’, *The Lancet*, 390(10090), pp. 102-104.

Besley, T. and Ghatak, M. (2005) ‘Competition and Incentives with Motivated Agents’, *American Economic Review*, 95(3), pp. 616–636.

Besley, T. and Ghatak, M. (2018) ‘Prosocial Motivation and Incentives’, *Annual Review of Economics*. Annual Reviews, 10(1), pp. 411–438.

Beullens, J., Rethans, J. J., Goedhuys, J. and Buntinx, F. (1997) ‘The use of standardized patients in research in general practice’, *Family Practice*.

Birnbaum, A. (1968) ‘Some latent trait models and their use in inferring an examinee’s ability.’, *Statistical theories of mental test scores*. Addison-Wesley, pp. 397–472.

Björkman, M. and Svensson, J. (2009) ‘Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda’, *The Quarterly Journal of Economics*, 124(2), pp. 735–769.

Blaauw, D. and Lagarde, M. (2015) 'Uptake of public sector sessional contracts by private general practitioners in South Africa', *RESYST Working Paper 8.*

Blaauw, D. and Lagarde, M. (2019) 'Better, but at what cost? Quality and efficiency of public and private care providers in South Africa.' Johannesburg: Centre for Health Policy, Wits University.

Blomqvist, Å. (1991) ‘The doctor as double agent: Information asymmetry, health insurance, and medical care’, *Journal of Health Economics*, 10(4), pp. 411–432.

van Boeckel, T. P., Gandra, S., Ashok, A., Caudron, Q., Grenfell, B. T., Levin, S. A. and Laxminarayan, R. (2014) ‘Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data’, *The Lancet Infectious Diseases*, 14(8), pp. 742–750.

Boyles, T. H., Whitelaw, A., Bamford, C., Moodley, M., Bonorchis, K., Morris, V., Rawoot, N., Naicker, V., Lusakiewicz, I., Black, J., Stead, D., Lesosky, M., Raubenheimer, P., Dlamini, S. and Mendelson, M. (2013) ‘Antibiotic stewardship ward rounds and a dedicated prescription chart reduce antibiotic consumption and pharmacy costs without affecting inpatient mortality or re-admission rates’, *PloS one*, 8(12).

Brekke, K. A. and Nyborg, K. (2010) ‘Selfish bakers, caring nurses? A model of work motivation’, *Journal of Economic Behavior & Organization*, 75(3), pp. 377–394.

Brink, A. (2017) 'Antibiotic Stewardship: From Theory to Practice', *2017 Antibiotic Guide.* Pretoria: AMPATH.

Brink, A., Messina, A., Feldman, C., Richards, G. A., Becker, P. J., Goff, D. A., Bauer, K. A., Nathwani, D. and van den Bergh, D. (2016) ‘Antimicrobial stewardship across 47 South African hospitals: an implementation study’, *The Lancet Infectious Diseases*, 16(9), pp. 1017–1025.

Brink, A., Van Wyk, J., M Moodley, V., Corcoran, C., Ekermans, P., Nutt, L., Boyles, T., Perovic, O., Feldman, C., Richards, G. and Mendelson, M. (2016) 'The role of appropriate diagnostic testing in acute respiratory tract infections: An antibiotic stewardship strategy to minimise diagnostic uncertainty in primary care', *South African Medical Journal*, 106(6).

Brock, J. M., Lange, A. and Leonard, K. (2016) 'Generosity and Prosocial Behavior in Healthcare Provision: Evidence from the Laboratory and Field,' *Journal of Human Resources*, 51(1), pp. 133-162.

Brook, R., Emmett, K., Lohr, K., Newhouse, J., Ware, J., Rogers, W., Davies, A., Sherbourne, C., Goldberg,

G., Camp, P., Kamberg, C., Leibowitz, A., Keesey, J. and Reboussin, D. (2006) 'The Health Insurance Experiment: A Classic RAND Study Speaks to the Current Health Care Reform Debate.' Santa Monica, CA: RAND Corporation.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N. and Wiesen, D. (2017) 'The Effects of Introducing Mixed Payment Systems for Physicians: Experimental Evidence', *Health Economics*, 26(2), pp. 243–262.

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

Butler, C., Hood, K., Verheij, T., Little, P., Melbye, H., Nuttall, J., Kelly, M., Mölstad, S., Godycki-Cwirko, M., Almirall, J., Torres, A., Gillespie, D., Rautakorpi, U., Coenen, S. and Goossens, H. (2009) 'Variation in antibiotic prescribing and its impact on recovery in patients with acute cough in primary care: prospective study in 13 countries', *BMJ*, 338.

Butler, C., Rollnick, S., Pill, R., Maggs-Rapport, F. and Stott, N. (1998) 'Understanding the culture of prescribing: qualitative study of general practitioners' and patients' perceptions of antibiotics for sore throats', *BMJ,* 317(7159), pp. 637–642.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. and Munafò, M. R. (2013) 'Power failure: why small sample size undermines the reliability of neuroscience', *Nature Reviews Neuroscience*, 14.

Cameron, A. C. and Miller, D. (2010) 'Robust Inference with Clustered Data', in Ullah, A. and Giles, D. E. (eds) *Handbook of Empirical Economics and Finance*. 1st edn. Chapman and Hall.

Cameron, A. C. and Miller, D. (2015) 'A Practitioner's Guide to Cluster-Robust Inference,' *Journal of Human Resources*, 50(2), pp. 317-372.

Card, D., Dobkin, C. and Maestas, N. (2007) 'The Impact of Health Insurance Status on Treatment Intensity and Health Outcomes.' RAND Working Paper, Santa Monica, CA: RAND Corporation.

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), pp. 2242–2258.

Carpenter, J., Verhoogen, E. and Burks, S. (2005) 'The effect of stakes in distribution experiments', *Economics Letters*, 86(3), pp. 393–398.

Carvounes, A., Childs, B., Dreyer, K., Kantor, G. and Ranchod, S. (2017) 'Quality Measurement and Reporting in the South African Private Hospital Industry.' Johannesburg: Hospital Association of South Africa.

CCSA (2018) 'Health Market Inquiry: Provisional Findings and Recommendations'. Pretoria: Competition Commission of South Africa.

Chalkley, M. and Malcomson, J. M. (1998) 'Contracting for health services when patient demand does not reflect quality', *Journal of Health Economics*, 17(1), pp. 1–19.

Chandra, A., Cutler, D. and Song, Z. (2011) 'Who Ordered That? The Economics of Treatment Choices in Medical Care', in Pauly, M., McGuire, T. and Barros, P. (eds) *Handbook of Health Economics*, vol 2. Elsevier.

Chou, Y. J., Yip, W. C., Lee, C.-H., Huang, N., Sun, Y.-P. and Chang, H.-J. (2003) 'Impact of separating drug prescribing and dispensing on provider behaviour: Taiwan's experience', *Health Policy and Planning*, 18(3), pp. 316–329.

Christian, S. C., Gerdtham, U.-G., Hompashe, D., Smith, A. and Burger, R. (2018) 'Measuring Quality Gaps in TB Screening in South Africa Using Standardised Patient Analysis', *International Journal of Environmental Research and Public Health*, 15(4).

Churilla, T. M., Egleston, B., Bleicher, R., Dong, Y., Meyer, J. and Anderson, P. (2017) 'Disparities in the Local Management of Breast Cancer in the US according to Health Insurance Status', *The Breast Journal*. John Wiley & Sons, Ltd (10.1111), 23(2), pp. 169–176.

Clarke, D., Doerr, D., Hunter, M., Schmets, G., Soucat, A. and Pavizaa, A. (2018) 'The private sector and universal health coverage', *Bulletin of the World Health Organization*, 97, pp. 434–435.

Cleanthous, P. (2002) 'Patient Welfare Implications of Innovation in the U.S. Antidepressant Market.' New Haven.

Clemens, J. and Gottlieb, J. D. (2014) 'Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?', *American Economic Review*, 104(4), pp. 1320–1349.

CMS (2018) 'CMS Annual Report 2017/2018'. Johannesburg: Council of Medical Schemes.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences: Second Edition*. 2nd edn. Routledge.

Colla, C., Bynum, J., Austin, A. and Skinner, J. (2016) 'Hospital Competition, Quality, and Expenditures in the U.S. Medicare Population', *NBER Working Paper Series,* No. 22826.

Cooper, Z., Gibbons, S., Jones, S. and McGuire, A. (2011) 'Does Hospital Competition Save Lives? Evidence from the English NHS Patient Choice Reforms', *Economic journal*, 121(554), pp. 228–260.

Coovadia, H., Jewkes, R., Barron, P., Sanders, D. and McIntyre, D. (2009) 'The health and health system of South Africa: historical roots of current public health challenges', *The Lancet*, 374(9692), pp. 817–834.

Crankshaw, O. (2008) 'Race, Space and the Post-Fordist Spatial Order of Johannesburg', *Urban Studies*, 45(8), pp. 1692–1711.

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

Cromwell, J. and Mitchell, J. B. (1986) 'Physician-induced demand for surgery', *Journal of Health Economics*, 5(4), pp. 293–313.

Crosson, F. (2009) 'Change the microenvironment: delivery system reform essential to controlling costs'. New York: The Commonwealth Fund.

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

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

Cutler, D. M., Skinner, J., Stern, A. D. and Wennberg, D. (2019) 'Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending', *American Economic Journal: Economic Policy*, 11(1), pp. 192–221.

Daniels, B., Dolinger, A., Bedoya, G., Rogo, K., Goicoechea, A., Coarasa, J., Wafula, F., Mwaura, N., Kimeu, R. and Das, J. (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).

Daniels, B., Kwan, A., Pai, M. and Das, J. (2019) 'Lessons on the quality of tuberculosis diagnosis from standardized patients in China, India, Kenya, and South Africa', *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, 16.

Daniels, B., Kwan, A., Satyanarayana, S., Subbaraman, R., Das, R. K., Das, V., Das, J. and Pai, M. (2019) 'Use of standardised patients to assess gender differences in quality of tuberculosis care in urban India: a two-city, cross-sectional study', *The Lancet Global Health*, 7(5).

Darby, M. R. and Karni, E. (1973) 'Free Competition and the Optimal Amount of Fraud', *The Journal of Law and Economics*, 16(1), pp. 67–88.

Das, J., Chowdhury, A., Hussam, R. and Banerjee, A. V (2016) 'The impact of training informal health care providers in India: A randomized controlled trial', *Science*, 354(6308).

Das, J. and Gertler, P. J. (2007) 'Variations In Practice Quality In Five Low-Income Countries: A Conceptual Overview', *Health Affairs*. Health Affairs, 26(3), pp. 296–309.

Das, J. and Hammer, J. (2005) 'Which doctor? Combining vignettes and item response to measure clinical competence', *Journal of Development Economics*, 78(2), pp. 348–383.

Das, J. and Hammer, J. (2007) 'Money for nothing: The dire straits of medical practice in Delhi, India', *Journal of Development Economics*, 83(1), pp. 1–36.

Das, J. and Hammer, J. (2014) 'Quality of Primary Care in Low-Income Countries: Facts and Economics',

Annual Review of Economics. Annual Reviews, 6(1), pp. 525–553.

Das, J., Holla, A., Mohpal, A. and Muralidharan, K. (2016) 'Quality and accountability in health care delivery: Audit-study evidence from primary care in India', *American Economic Review*, 106(12), pp. 3765-99.

Das, J., Woskie, L., Rajbhandari, R., Abbasi, K. and Jha, A. (2018) 'Rethinking assumptions about delivery of healthcare: implications for universal health coverage', *BMJ*, 361.

Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E. and Perrier, L. (2006) 'Accuracy of Physician Self-assessment Compared With Observed Measures of CompetenceA Systematic Review', *JAMA*, 296(9), pp. 1094–1102.

Deci, E. L., Koestner, R. and Ryan, R. M. (1999) 'A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation.', *Psychological Bulletin*, 125(6), pp. 627–668.

Dehn Lunn, A. (2018) 'Reducing inappropriate antibiotic prescribing in upper respiratory tract infection in a primary care setting in Kolkata, India', *BMJ Open Quality*, 7(4).

Delfgaauw, J. and Dur, R. (2008) 'Incentives and Workers' Motivation in the Public Sector', *The Economic Journal*, 118(525), pp. 171–191.

DellaVigna, S. and Pope, D. (2017) 'What Motivates Effort? Evidence and Expert Forecasts', *The Review of Economic Studies*, 85(2), pp. 1029–1069.

Dempsey, P. P., Businger, A. C., Whaley, L. E., Gagne, J. J. and Linder, J. A. (2014) 'Primary care clinicians' perceptions about antibiotic prescribing for acute bronchitis: a qualitative study', *BMC Family Practice*, 15(1).

DeNegre, A. A., Ndeffo Mbah, M. L., Myers, K. and Fefferman, N. H. (2019) 'Emergence of antibiotic resistance in immunocompromised host populations: A case study of emerging antibiotic resistant tuberculosis in AIDS patients', *PLOS ONE*, 14(2).

Dietrichson, J., Ellegård, L. M. and Kjellsson, G. (2016) 'Patient Choice, Entry, and the Quality of Primary Care: Evidence from Swedish Reforms', Working Paper No. 36. Lund: Lund University, Department of Economics.

Djulbegovic, B. and Elqayam, S. (2017) 'Many faces of rationality: Implications of the great rationality debate for clinical decision-making', *Journal of Evaluation in Clinical Practice*, 23(5), pp. 915–922.

Donabedian, A. (1980) *Explorations in Quality Assessment and Monitoring: The definition of quality and approaches to its assessment*. Chicago, IL: Health Administration Press.

Donabedian, A. (1988) 'The Quality of Care: How Can It Be Assessed?', *JAMA,* 260(12), pp. 1743-48.

Doran, T., Kontopantelis, E., Valderas, J. M., Campbell, S., Roland, M., Salisbury, C. and Reeves, D. (2011) 'Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework', *BMJ*, 342.

Doran, T., Maurer, K. A. and Ryan, A. M. (2017) 'Impact of Provider Incentives on Quality and Value of Health Care', *Annual Review of Public Health*, 38(1), pp. 449–465.

Dranove, D., Kessler, D., McClellan, M. and Satterthwaite, M. (2003) 'Is More Information Better? The Effects of "Report Cards" on Health Care Providers', *Journal of Political Economy*, 11(3), pp. 555–588.

Dranove, D. and Satterthwaite, M. A. (1992) 'Monopolistic Competition When Price and Quality are Imperfectly Observable', *RAND Journal of Economics*, 23(4), pp. 518–534.

Dranove, D. and Satterthwaite, M. (2000) 'The industrial organization of health care markets', in Culyer, A. and Newhouse, J. (eds) *Handbook of Health Economics*, vol 1, Part B, pp. 1093–1139. Elsevier.

Drasgow, F. and Lissak, R. I. (1983) 'Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses.', *Journal of Applied Psychology*, 68(3), pp. 363–373.

Dulleck, U. and Kerschbamer, R. (2006) 'On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods', *Journal of Economic Literature*, 44(1), pp. 5–42.

Dulleck, U., Kerschbamer, R. and Sutter, M. (2011) 'The Economics of Credence Goods: An Experiment

on the Role of Liability, Verifiability, Reputation, and Competition', *American Economic Review*, 101(2), pp. 526–555.

Dunning, D., Heath, C. and Suls, J. M. (2004) 'Flawed Self-Assessment: Implications for Health, Education, and the Workplace', *Psychological Science in the Public Interest*, 5(3), pp. 69–106.

Eckel, C. C. and Grossman, P. J. (1996) 'Altruism in Anonymous Dictator Games', *Games and Economic Behavior*, 16(2), pp. 181–191.

Eijkenaar, F., Emmert, M., Scheppach, M. and Schöffski, O. (2013) 'Effects of pay for performance in health care: A systematic review of systematic reviews', *Health Policy*, 110(2), pp. 115–130.

Eilermann, K., Halstenberg, K., Kuntz, L., Martakis, K., Roth, B. and Wiesen, D. (2019) 'The Effect of Expert Feedback on Antibiotic Prescribing in Pediatrics: Experimental Evidence', *Medical Decision Making*, 39(7), pp. 781–795.

Einav, L. and Finkelstein, A. (2017) 'Moral Hazard in Health Insurance: What We Know and How We Know It', *Journal of the European Economic Association*, 16(4), pp. 957-982.

Elouafkaoui, P., Young, L., Newlands, R., Duncan, E. M., Elders, A., Clarkson, J. E. and Ramsay, C. R. (2016) 'An Audit and Feedback Intervention for Reducing Antibiotic Prescribing in General Dental Practice: The RAPiD Cluster Randomised Controlled Trial', *PLOS Medicine*, 13(8).

Epstein, A. J. (2006) 'Do Cardiac Surgery Report Cards Reduce Mortality? Assessing the Evidence', *Medical Care Research and Review*, 63(4), pp. 403–426.

Evans, R. G. (1974) 'Supplier-Induced Demand: Some Empirical Evidence and Implications', in Pelman, M. (ed.) *The Economics of Health and Medical Care*. London: Palgrave Macmillan.

Farley, E., Stewart, A., Davies, M.-A., Govind, M., van den Bergh, D. and Boyles, T. H. (2018) 'Antibiotic use and resistance: Knowledge, attitudes and perceptions among primary care prescribers in South Africa', *South African Medical Journal*, 108(9).

Feldstein, M. S. (1973) 'The Welfare Loss of Excess Health Insurance', *Journal of Political Economy*, 81(2), pp. 251–280.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J., Allen, H. and Baicker, K. (2012) 'The Oregon Health Insurance Experiment: Evidence from the First Year', *The Quarterly Journal of Economics*, 127(3).

Fletcher-Lartey, S., Yee, M., Gaarslev, C. and Khan, R. (2016) 'Why do general practitioners prescribe antibiotics for upper respiratory tract infections to meet patient expectations: a mixed methods study', *BMJ Open*, 6(10).

Franco, L. M., Bennett, S. and Kanfer, R. (2002) 'Health sector reform and public sector health worker motivation: a conceptual framework', *Social Science & Medicine*, 54(8), pp. 1255–1266.

Frank, R. G. (2004) 'Behavioral Economics and Health Economics', *NBER Working Paper Series*, No. 10881.

Franks, P., Clancy, C. M. and Gold, M. R. (1993) 'Health Insurance and Mortality', *JAMA*, 270(6).

Fuchs, V. (1978) 'The Supply of Surgeons and the Demand for Operations', *The Journal of Human Resources*, 13, pp. 35-56.

Galizzi, M. M., Tammi, T., Godager, G., Linnosmaa, I. and Wiesen, D. (2015) 'Provider altruism in health economics', *National Institute for Health and Welfare: Discussion Paper* , 4/2015.

Gaynor, M. (1994) 'Issues in the Industrial Organization of the Market for Physician Services', *Journal of Economics & Management Strategy*, 3(1), pp. 211–255.

Gaynor, M. (2006) "What Do We Know about Competition and Quality in Health Care Markets?," *Foundations and Trends in Microeconomics*, 2(6), pp 441-508.

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

Gaynor, M., Propper, C. and Seiler, S. (2016) 'Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service', *American Economic Review*, 106(11), pp. 3521–3557.

Gelband, H. and Duse, A. G. (2011) 'GARP: Executive summary', *South African Medical Journal,* 101(8).

Georgellis, Y., Iossa, E. and Tabvuma, V. (2010) 'Crowding Out Intrinsic Motivation in the Public Sector', *Journal of Public Administration Research and Theory*, 21(3), pp. 473–493.

Gerber, J. S., Prasad, P. A., Fiks, A. G., Localio, A. R., Grundmeier, R. W., Bell, L. M., Wasserman, R. C., Keren, R. and Zaoutis, T. E. (2013) 'Effect of an Outpatient Antimicrobial Stewardship Intervention on Broad-Spectrum Antibiotic Prescribing by Primary Care Pediatricians: A Randomized Trial', *JAMA*, 309(22), pp. 2345–2352.

Gertler, P. and Vermeersch, C. (2013) 'Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes', *NBER Working Paper Series,* No. 19046.

Gibbons, L., Belizán, J., Lauer, J., Betrán, A., Merialdi, M. and Althabe, F. (2010) 'The Global Numbers and Costs of Additionally Needed and Unnecessary Caesarean Sections Performed per Year: Overuse as a Barrier to Universal Coverage', *World Health Report Background Paper*, No. 30. Geneva: World Health Organisation.

Gideon, U., Alfonso, E. and Diaz, Y. (2013) 'The Impact of Universal Coverage Schemes in the Developing World: A Review of the Existing Evidence', *Universal Health Coverage (UNICO) Studies Series,* No. 25. Washington DC: The World Bank.

Glassman, P., Luck, J., M O'Gara, E. and W. Peabody, J. (2000) 'Using Standardized Patients to Measure Quality: Evidence from the Literature and a Prospective Study,' *The Joint Commission Journal on Quality Improvement*, 26(11), pp. 644-653.

Glasziou, P., Straus, S., Brownlee, S., Trevena, L., Dans, L., Guyatt, G., Elshaug, A. G., Janett, R. and Saini, V. (2017) 'Evidence for underuse of effective medical services around the world', *The Lancet*, 390(10090), 169-177.

Glennerster, R. and Takavarasha, K. (2013) *Running randomised evaluations: a practical guide*. New Jersey: Princeton University Press.

Godager, G. and Wiesen, D. (2013) 'Profit or patients' health benefit? Exploring the heterogeneity in physician altruism', *Journal of Health Economics*, 32(6), pp. 1105–1116.

Gottschalk, F., Mimra, W. and Waibel, C. (2017) 'Health Services as Credence Goods: A Field Experiment', *SSRN Electronic Journal*.

Gowrisankaran, G. and Town, R. J. (2003) 'Competition, payers, and hospital quality', *Health Services Research*, 38(6), pp. 1403–1421.

Granlund, D. (2009) 'Are private physicians more likely to veto generic substitution of prescribed pharmaceuticals?', *Social Science & Medicine*, 69(11), pp. 1643–1650.

Gravelle, H., Liu, D., Propper, C. and Santos, R. (2019) 'Spatial Competition and Quality: Evidence from the English Family Doctor Market', *Journal of Health Economics*, 68.

Gravelle, H., Sutton, M. and Ma, A. (2010) 'Doctor Behaviour under a Pay for Performance Contract: Treating, Cheating and Case Finding?', *The Economic Journal*, 120(542), pp. 129–156.

Groenewald, P., Bradshaw, D., Day, C. and Laubscher, R. (2017) 'Burden of Disease', in *District Health Barometer 2016/17*, pp. 206–226. Durban: Health Systems Trust.

Gruber, J., Kim, J. and Mayzlin, D. (1999) 'Physician fees and procedure intensity: the case of cesarean delivery', *Journal of Health Economics*, 18(4), pp. 473–490.

Gruber, J. and Owings, M. (1996) 'Physician financial incentives and cesarean section delivery', *RAND Journal of Economics*, 27(1), pp. 99–123.

Haas-Wilson, D. (1994) 'The Relationships between the Dimensions of Health Care Quality and Price: The Case of Eye Care', *Medical Care*, 32(2), pp. 175–182.

Hallsworth, M., Chadborn, T., Sallis, A., Sanders, M., Berry, D., Greaves, F., Clements, L. and Davies, S. C. (2016) 'Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial', *The Lancet*, 387(10029), pp. 1743–1752.

Harris, B., Goudge, J., Ataguba, J. E., McIntyre, D., Nxumalo, N., Jikwana, S. and Chersich, M. (2011) 'Inequities in access to health care in South Africa', *Journal of Public Health Policy*, 32(1), pp. 102–123.

Harrison, A., Karim, S. A., Floyd, K., Lombard, C., Lurie, M., Ntuli, N. and Wilkinson, D. (2000) 'Syndrome packets and health worker training improve sexually transmitted disease case management in rural South Africa: randomized controlled trial', *AIDS*, 14(17).

Hasan, O., Orav, E. J., Hicks, L. S. and Orav, E. J. (2010) 'Insurance status and hospital care for myocardial infarction, stroke, and pneumonia', *Journal of Hospital Medicine*, 5(8), pp. 452–459.

Haushofer, J., Chemin, M., Jang, C. and Abraham, J. (2020) 'Economic and psychological effects of health insurance and cash transfers: Evidence from a randomized experiment in Kenya', *Journal of Development Economics*, 144.

Heffetz, O. and Frank, R. (2008) 'Preferences for Status: Evidence and Economic Implications', in Benhabib, J., Bisin, A., and Jackson, M. (eds) *Handbook of Social Economics*. Elsevier.

Hellerstein, J. K. (1998) 'The Importance of the Physician in the Generic versus Trade-Name Prescription Decision', *The RAND Journal of Economics*, 29(1), pp. 108–136.

Hennig-Schmidt, H., Selten, R. and Wiesen, D. (2011) 'How payment systems affect physicians' provision behaviour—An experimental investigation', *Journal of Health Economics*, 30(4), pp. 637–646.

Hennig-Schmidt, H. and Wiesen, D. (2014) 'Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students', *Social Science & Medicine*, 108, pp. 156–165.

Heyes, A. (2005) 'The economics of vocation or "why is a badly paid nurse a good nurse"?', *Journal of Health Economics*, 24(3), pp. 561–569.

Himmel, W., Simmenroth-Nayda, A., Niebling, W., Ledig, T., Jansen, R., Kochen, M., Gleiter, C. and Hummers-Pradier, E. (2005) 'What do primary care patients think about generic drugs?', *International Journal of Clinical Pharmacology and Therapeutics*, 43(10), pp. 472–479.

Hoffman, E., McCabe, K., Shachat, K. and Smith, V. (1994) 'Preferences, Property Rights, and Anonymity in Bargaining Games', *Games and Economic Behavior*, 7(3), pp. 346–380.

Hoffman, E., McCabe, K. and Smith, V. L. (1996) 'Social Distance and Other-Regarding Behavior in Dictator Games', *American Economic Review*, 86(3), pp. 653–660.

Holloway, K. A., Rosella, L. and Henry, D. (2016) 'The Impact of WHO Essential Medicines Policies on Inappropriate Use of Antibiotics', *PLOS ONE*, 11(3).

Hongoro, C., Funani, I. I. N., Chitha, W. and Godlimpi, L. (2015) 'An Assessment of Private General Practitioners Contracting for Public Health Services Delivery in O.R. Tambo District, South Africa', *Journal of Public Health in Africa*, 6(2).

Horowitz, J. L. and Manski, C. F. (2000) 'Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data', *Journal of the American Statistical Association*, 95(449), pp. 77-84

HSRC (2013) 'The City of Johannesburg Economic Overview: 2013.' Cape Town: Human Sciences Research Council.

Iizuka, T. (2007) 'Experts' agency problems: evidence from the prescription drug market in Japan', *The RAND Journal of Economics*, 38(3), pp. 844–862.

Iizuka, T. (2012) 'Physician Agency and Adoption of Generic Pharmaceuticals', *American Economic Review*, 102(6), pp. 2826–2858.

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., O'Brien, M. A., Johansen, M., Grimshaw, J. and Oxman, A. D. (2012) 'Audit and feedback: effects on professional practice and healthcare outcomes', *Cochrane Database of Systematic Reviews*, (6).

Jacobson, M., Earle, C. C., Price, M. and Newhouse, J. P. (2010) 'How Medicare's Payment Cuts For Cancer Chemotherapy Drugs Changed Patterns Of Treatment', *Health Affairs*. Health Affairs, 29(7), pp. 1391–1399.

Jacqmin-Gadda, H., Sibillot, S., Proust, C., Molina, J.-M. and Thiébaut, R. (2007) 'Robustness of the linear mixed model to misspecified error distribution', *Computational Statistics & Data Analysis*, 51(10), pp. 5142–5154.

Jaffer, A., Devani, N., Brown, J., Mandal, S. and Lipman, M. (2017) 'HIV and respiratory illness in the antiretroviral therapy era', *Annals of Research Hospitals,* 1(6).

Johnson, E. M. (2014) 'Physician-Induced Demand', in Culyer, A. J. (ed.) *Encyclopedia of Health Economics*, pp. 77–82. Elsevier.

Johnstone, K. J., Chang, A. B., Fong, K. M., Bowman, R. V and Yang, I. A. (2013) 'Inhaled corticosteroids for subacute and chronic cough in adults', *Cochrane Database of Systematic Reviews*, (3).

Joyce, G. F., Escarce, J. J., Solomon, M. D. and Goldman, D. P. (2002) 'Employer Drug Benefit Plans and Spending on Prescription Drugs', *JAMA*, 288(14), pp. 1733–1739.

Kaiser, B. and Schmid, C. (2016) 'Does Physician Dispensing Increase Drug Expenditures? Empirical Evidence from Switzerland', *Health Economics*, 25(1), pp. 71–90.

Karim, S. S. A., Churchyard, G. J., Karim, Q. A. and Lawn, S. D. (2009) 'HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response', *The Lancet*, 374(9693), pp. 921–933.

Keeton, C. (2010) 'Bridging the gap in South Africa', *Bulletin of the World Health Organization*, 88, pp. 803–804.

Kerschbamer, R., Neururer, D. and Sutter, M. (2016) 'Insurance coverage of customers induces dishonesty of sellers in markets for credence goods', *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), pp. 7454–7458.

Kerschbamer, R. and Sutter, M. (2017) 'The Economics of Credence Goods – a Survey of Recent Lab and Field Experiments', *CESifo Economic Studies*, 63(1), pp. 1–23.

Kerschbamer, R., Sutter, M. and Dulleck, U. (2017) 'How Social Preferences Shape Incentives in (Experimental) Markets for Credence Goods', *The Economic Journal*, 127(600), pp. 393–416.

Kessler, D. P. and Geppert, J. J. (2005) 'The Effects of Competition on Variation in the Quality and Cost of Medical Care', *Journal of Economics & Management Strategy*, 14(3), pp. 575–589.

Kessler, D. P. and McClellan, M. B. (2000) 'Is Hospital Competition Socially Wasteful?', *The Quarterly Journal of Economics*, 115(2), pp. 577–615.

Ketelaar, N. A., Faber, M. J., Flottorp, S., Rygh, L. H., Deane, K. H. and Eccles, M. P. (2011) 'Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations', *Cochrane Database of Systematic Reviews*, (11).

King, J. J. C., Das, J., Kwan, A., Daniels, B., Powell-Jackson, T., Makungu, C. and Goodman, C. (2019) 'How to do (or not to do) … using the standardized patient method to measure clinical quality of care in LMIC health facilities', *Health Policy and Planning*, 34(8), pp. 625–634.

Kohler, P. K., Marumo, E., Jed, S. L., Mema, G., Galagan, S., Tapia, K., Pillay, E., DeKadt, J., Naidoo, E., Dombrowski, J. C. and Holmes, K. K. (2017) 'A national evaluation using standardised patient actors to assess STI services in public sector clinical sentinel surveillance facilities in South Africa', *Sexually Transmitted Infections*, 93(4), pp. 247-252.

Kolstad, J. R. and Lindkvist, I. (2013) 'Pro-social preferences and self-selection into the public health sector: evidence from an economic experiment', *Health Policy and Planning*, 28(3), pp. 320–327.

Kolstad, J. T. (2013) 'Information and Quality When Motivation is Intrinsic: Evidence from Surgeon Report Cards', *American Economic Review*, 103(7), pp. 2875–2910.

Kranton, R. E. (2003) 'Competition and the Incentive to Produce High Quality', *Economica*, 70(279), pp. 385–404.

Kravitz, R. L., Epstein, R. M., Feldman, M. D., Franz, C. E., Azari, R., Wilkes, M. S., Hinton, L. and Franks, P. (2005) 'Influence of Patients' Requests for Direct-to-Consumer Advertised AntidepressantsA Randomized Controlled Trial', *JAMA*, 293(16), pp. 1995–2002.

Kula, N. and Fryatt, R. J. (2013) 'Public–private interactions on health in South Africa: opportunities for scaling up', *Health Policy and Planning*, 29(5), pp. 560–569.

Kwan, A., Daniels, B., Bergkvist, S., Das, V., Pai, M. and Das, J. (2019) 'Use of standardised patients for healthcare quality research in low- and middle-income countries', *BMJ Global Health*, 4(5).

Lagarde, M. and Blaauw, D. (2017) 'Physicians' responses to financial and social incentives: A medically framed real effort experiment', *Social Science and Medicine*, 179, pp. 147-159.

Lagarde, M. and Blaauw, D. (2019a) 'Can financial or prosocial incentives reduce over-treatment? Evidence from South Africa', *Paper presented at the American Society of Health Economists (ASHEcon) Conference*. Washington DC.

Lagarde, M. and Blaauw, D. (2019b) 'Unnecessary prescription of antibiotics in South Africa.'

Lagarde, M., Huicho, L. and Papanicolas, I. (2019) 'Motivating provision of high quality care: it is not all about the money', *BMJ,* 366: l5210.

Laird, N. M. and Ware, J. H. (1982) 'Random-Effects Models for Longitudinal Data', *Biometrics*, 38(4), pp. 963–974.

Laxminarayan, R., Duse, A., Wattal, C., Zaidi, A. K. M., Wertheim, H. F. L., Sumpradit, N., Vlieghe, E., Hara, G. L., Gould, I. M., Goossens, H., Greko, C., So, A. D., Bigdeli, M., Tomson, G., Woodhouse, W., Ombaka, E., Peralta, A. Q., Qamar, F. N., Mir, F., Kariuki, S., Bhutta, Z. A., Coates, A., Bergstrom, R., Wright, G. D., Brown, E. D. and Cars, O. (2013) 'Antibiotic resistance—the need for global solutions', *The Lancet Infectious Diseases*, 13(12), pp. 1057–1098.

Leal, J. R., Conly, J., Henderson, E. A. and Manns, B. J. (2017) 'How externalities impact an evaluation of strategies to prevent antimicrobial resistance in health care organizations', *Antimicrobial Resistance & Infection Control*, 6(1).

Lee, S. (2018) 'Intrinsic Incentives: A Field Experiment on Leveraging Intrinsic Motivation in Public Service Delivery', *Unpublished PhD Job Market Paper.* Boston: Harvard Medical School and Brigham and Women's Hospital.

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

Leigh, S., Grant, A., Murray, N., Faragher, B., Desai, H., Dolan, S., Cabdi, N., Murray, J. B., Rejaei, Y., Stewart, S., Edwardson, K., Dean, J., Mehta, B., Yeung, S., Coenen, F., Niessen, L. W. and Carrol, E. D. (2019) 'The cost of diagnostic uncertainty: a prospective economic analysis of febrile children attending an NHS emergency department', *BMC Medicine*, 17(1).

Léonard, C., Stordeur, S. and Roberfroid, D. (2009) 'Association between physician density and health care consumption: A systematic review of the evidence', *Health Policy*, 91(2), pp. 121–134.

Leonard, K. (2008) 'Is patient satisfaction sensitive to changes in the quality of care? An exploitation of the Hawthorne effect', *Journal of Health Economics*, 27(2), pp. 444–459.

Leonard, K. and Masatu, M. (2010) 'Using the Hawthorne effect to examine the gap between a doctor's best possible practice and actual performance', *Journal of Development Economics*, 93(2), pp. 226–234.

Leonard, K., Masatu, M., Herbst, C. and Lemiere, C. (2015) 'The Systematic Assessment of Health Worker Performance: A Framework for Analysis and its Application in Tanzania', *HNP Discussion Paper Series.* Washington DC: The World Bank.

Li, L., Lin, C. and Guan, J. (2014) 'Using standardized patients to evaluate hospital-based intervention outcomes', *International Journal of Epidemiology*, 43(3), pp. 897–903.

Liang, K.-Y. and Zeger, S. L. (1986) 'Longitudinal data analysis using generalized linear models', *Biometrika*, 73(1), pp. 13–22.

Linder, J. A. (2008) 'Antibiotics for Treatment of Acute Respiratory Tract Infections: Decreasing Benefit, Increasing Risk, and the Irrelevance of Antimicrobial Resistance', *Clinical Infectious Diseases*, 47(6), pp. 744–746.

Linder, J. A., Doctor, J. N., Friedberg, M. W., Reyes Nieva, H., Birks, C., Meeker, D. and Fox, C. R. (2014) 'Time of day and the decision to prescribe antibiotics', *JAMA internal medicine*, 174(12), pp. 2029–2031.

Long, Q., Klemetti, R., Wang, Y., Tao, F., Yan, H. and Hemminki, E. (2012) 'High caesarean section rate in rural China: Is it related to health insurance (New Co-operative Medical Scheme)?', *Social Science & Medicine*, 75(4), pp. 733–737.

Lu, F. (2014) 'Insurance coverage and agency problems in doctor prescriptions: Evidence from a field

experiment in China', *Journal of Development Economics*, 106, pp. 156–167.

Lundin, D. (2000) 'Moral hazard in physician prescription behavior', *Journal of Health Economics*, 19(5), pp. 639–662.

Ma, C.-T. A. (1994) 'Health care payment systems: cost and quality incentives', *Journal of Economics & Management Strategy*, 3(1), pp. 93–112.

Ma, C.-T. A. and Mcguire, T. G. (1997) 'Optimal Health Insurance and Provider Payment', *American Economic Review*, 87(4), pp. 685–704.

Mackintosh, M., Channon, A., Karan, A., Selvaraj, S., Cavagnero, E. and Zhao, H. (2016) 'What is the private sector? Understanding private provision in the health systems of low-income and middle-income countries', *The Lancet*, 388(10044), pp. 596–605.

Madden, J. M., Quick, J. D., Ross-Degnan, D. and Kafle, K. K. (1997) 'Undercover careseekers: Simulated clients in the study of health provider behavior in developing countries', *Social Science & Medicine*, 45(10), pp. 1465–1482.

Mainous, A., Hueston, W., Love, M., Evans, M. and Finger, R. (2000) 'An evaluation of statewide strategies to reduce antibiotic overuse', *Family Medicine*, 32(1), pp. 22–29.

Manderson, L. (2019) ' Prescribing, resistance and care: antibiotic use in urban South Africa.' Johannesburg: Centre for Health Policy, Wits University.

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B. and Leibowitz, A. (1987) 'Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment', *American Economic Review*, 77(3), pp. 251–277.

Marshall, M. N., Shekelle, P. G., Leatherman, S. and Brook, R. H. (2000) 'The Public Release of Performance DataWhat Do We Expect to Gain? A Review of the Evidence', *JAMA*, 283(14), pp. 1866–1874.

Martin, A. B., Hartman, M., Washington, B. and Catlin, A. (2017) 'National Health Spending: Faster Growth In 2015 As Coverage Expands And Utilization Increases', *Health Affairs*, 36(1), pp. 166–176.

Mas, A. and Moretti, E. (2009) 'Peers at work', *American Economic Review*, 9(1), pp. 112–145.

Mathews, C., Guttmacher, S. J., Flisher, A. J., Mtshizana, Y. Y., Nelson, T., McCarthy, J. and Daries, V. (2009) 'The Quality of HIV Testing Services for Adolescents in Cape Town, South Africa: Do Adolescent-Friendly Services Make a Difference?', *Journal of Adolescent Health*, 44(2), pp. 188–190.

Mayosi, B. M. and Benatar, S. R. (2014) 'Health and Health Care in South Africa — 20 Years after Mandela', *New England Journal of Medicine*, 371(14), pp. 1344–1353.

McGuire, T. G. (2000) 'Physician agency', in Culyer, A. & Newhouse, J. (eds) *Handbook of Health Economics*, vol 1, Part A, pp. 461-536. Elsevier.

McIntyre, D. (2010) 'Private sector involvement in funding and providing health services in South Africa: Implications for equity and access to health care.' Harare: Network for Equity in Health in Southern Africa (EQUINET).

McIntyre, D., Bloom, G., Doherty, J. and Brijlal, P. (1995) 'Health expenditure and finance in South Africa.' Durban: Health Systems Trust.

McKinlay, J. B., Potter, D. A. and Feldman, H. A. (1996) 'Non-medical influences on medical decision-making', *Social Science & Medicine*, 42(5), pp. 769–776.

McNeish, D. and Kelley, K. (2019) 'Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations.', *Psychological Methods*, 24(1), pp. 20–35.

McPake, B. and Hanson, K. (2016) 'Managing the public-private mix to achieve universal health coverage', *The Lancet*, 388(10044), pp. 622–630.

Meeker, D., Linder, J. A., Fox, C. R., Friedberg, M. W., Persell, S. D., Goldstein, N. J., Knight, T. K., Hay, J. W. and Doctor, J. N. (2016) 'Effect of Behavioral Interventions on Inappropriate Antibiotic Prescribing Among Primary Care Practices: A Randomized Clinical Trial', *JAMA*, 315(6), pp. 562–570.

Metcalfe, D., Rios Diaz, A. J., Olufajo, O. A., Massa, M. S., Ketelaar, N. A., Flottorp, S. A. and Perry, D. C. (2018) 'Impact of public release of performance data on the behaviour of healthcare consumers and providers', *Cochrane Database of Systematic Reviews*, (9).

Meyer, J. C., Summers, R. S. and Möller, H. (2001) 'Randomized, controlled trial of prescribing training in a South African province', *Medical Education*, 35(9), pp. 833–840.

Miller, G. E. (1990) 'The assessment of clinical skills/competence/performance', *Academic Medicine*, 65(9).

Mimra, W., Rasch, A. and Waibel, C. (2016) 'Price competition and reputation in credence goods markets: Experimental evidence', *Games and Economic Behavior*, 100, pp. 337–352.

Mohanan, M., Das, V., Tabak, D., Chan, B., Holla, A. and Das, J. (2012) 'Standardised Patients and the Measurement of healthcare quality', *MAQARI Project.*

Mohanan, M., Giardili, S., Das, V., Rabin, T. L., Raj, S. S., Schwartz, J. I., Seth, A., Goldhaber-Fiebert, J. D., Miller, G. and Vera-Hernández, M. (2017) 'Evaluation of a social franchising and telemedicine programme and the care provided for childhood diarrhoea and pneumonia, Bihar, India', *Bulletin of the World Health Organization*, 95(5), p. 343–352.

Mohanan, M., Vera-Hernández, M., Das, V., Giardili, S., Goldhaber-Fiebert, J. D., Rabin, T. L., Raj, S. S., Schwartz, J. I. and Seth, A. (2015) 'The Know-Do Gap in Quality of Health Care for Childhood Diarrhea and Pneumonia in Rural India', *JAMA Pediatrics*, 169(4), pp. 349–357.

Montagu, D. and Goodman, C. (2016) 'Prohibit, constrain, encourage, or purchase: how should we engage with the private health-care sector?', *The Lancet*, 388(10044), pp. 613–621.

Mort, E. A., Edwards, J. N., Emmons, D. W., Convery, K. and Blumenthal, D. (1996) 'Physician Response to Patient Insurance Status in Ambulatory Care Clinical Decision-Making: Implications for Quality of Care', *Medical Care*, 34(8), pp. 783–797.

Morton, F. and Kyle, M. (2011) 'Markets for Pharmaceutical Products', in Pauly, M., Mcguire, T., and Barros, P. (eds) *Handbook of Health Economics*, pp. 763–823. Elsevier.

Mott, D. A. and Cline, R. R. (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), pp. 662–674.

Mukamel, D. B., Zwanziger, J. and Tomaszewski, K. J. (2001) 'HMO penetration, competition, and risk-adjusted hospital mortality', *Health Services Research*, 36(6), pp. 1019–1035.

NDoH (2014a) 'Antimicrobial Resistance National Strategy Framework 2014-2024.' Pretoria: National Department of Health, Republic of South Africa.

NDoH (2014b) *Primary Healthcare Standard Treatment Guidelines and Essential Medicines List*. 5th edn. Pretoria: National Department of Health, Republic of South Africa.

NDoH (2019) 'Evaluation of Phase 1 implementation of interventions in the National Health Insurance (NHI) pilot districts in South Africa*,' Final Evaluation Report*. Pretoria: National Department of Health, Republic of South Africa.

NDoH, Stats SA, SAMRC and ICF (2019) *South Africa Demographic and Health Survey 2016*. Pretoria: National Department of Health, Republic of South Africa.

Newhouse, J. and Insurance Experiment Group (1993) *Free for All? Lessons from the RAND Health Insurance Experiment*. Harvard University Press.

Ngcuka, O. (2019) 'South Africa puts initial universal healthcare cost at $17 billion', *Reuters*, 8 August.

Nguyen, N. X. and Derrick, F. W. (1997) 'Physician behavioral response to a Medicare price reduction', *Health Services Research*, 32(3), pp. 283–298.

Nyman, J. A. (2007) 'American Health Policy: Cracks in the Foundation', *Journal of Health Politics, Policy and Law*, 32(5), pp. 759–783.

Odendaal, W. A., Ward, K., Uneke, J., Uro-Chukwu, H., Chitama, D., Balakrishna, Y. and Kredo, T. (2018) 'Contracting out to improve the use of clinical health services and health outcomes in low- and middle-income countries', *Cochrane Database of Systematic Reviews*, (4).

Onishi, J., Gupta, S. and Peters, D. H. (2010) 'Comparative analysis of exit interviews and direct clinical observations in Pediatric Ambulatory Care Services in Afghanistan', *International Journal for Quality in Health Care*, 23(1), pp. 76–82.

Park, S., Soumerai, S. B., Adams, A. S., Finkelstein, J. A., Jang, S. and Ross-Degnan, D. (2005) 'Antibiotic use following a Korean national policy to prohibit medication dispensing by physicians', *Health Policy and Planning*, 20(5), pp. 302–309.

Pauly, M. V (1988) 'Is Medical Care Different? Old Questions, New Answers', *Journal of Health Politics, Policy and Law*, 13(2), pp. 227–237.

Pauly, M. V. (1968) 'The Economics of Moral Hazard: Comment', *American Economic Review*, 58(3), pp. 531–537.

Pauly, M. V. (1978) 'Is Medical Care Different?', in Greenberg, W. (ed.) *Competition in the Health Care Sector: Past, Present, and Future*. Gaithersburg, MD: Aspen.

Pauly, M. V. (1980) *Doctors and Their Workshops: Economic Models of Physician Behaviour*. University of Chicago Press.

Pauly, M. V and Satterthwaite, M. A. (1981) 'The Pricing of Primary Care Physicians Services: A Test of the Role of Consumer Information', *The Bell Journal of Economics*, 12(2), pp. 488–506.

Perumal-Pillay, V. A. and Suleman, F. (2017) 'Selection of essential medicines for South Africa - an analysis of in-depth interviews with national essential medicines list committee members.', *BMC Health Services Research*, 17(1).

Pieterse, E. and Owen, K. (2018) 'Johannesburg: Confronting Spatial Inequality', *World Resources Report Case Study*. Washington DC: World Resources Institute.

Pillay-van Wyk, V. and Bradshaw, D. (2017) 'Mortality and socioeconomic status: the vicious cycle between poverty and ill health', *The Lancet Global Health*, 5(9).

Pongsupap, Y. and Lerberghe, W. Van (2006) 'Choosing between public and private or between hospital and primary care: responsiveness, patient-centredness and prescribing patterns in outpatient consultations in Bangkok', *Tropical Medicine & International Health*, 11(1), pp. 81–89.

Propper, C. (2010) 'The operation of choice and competition in healthcare: a review of the evidence.' London: 2020 Public Services Trust.

Propper, C., Burgess, S. and Gossage, D. (2008) 'Competition and Quality: Evidence from the NHS Internal Market 1991–9', *The Economic Journal*, 118(525), pp. 138–170.

Radyowijati, A. and Haak, H. (2003) 'Improving antibiotic use in low-income countries: an overview of evidence on determinants', *Social Science & Medicine*, 57(4), pp. 733–744.

Rao, K., Paina, L., Ingabire, M. and Shroff, Z. (2018) 'Contracting non-state providers for universal health coverage: learnings from Africa, Asia, and Eastern Europe', *International Journal for Equity in Health*, 17(1).

Rasch, A. and Waibel, C. (2018) 'What Drives Fraud in a Credence Goods Market? – Evidence from a Field Study', *Oxford Bulletin of Economics and Statistics*, 80(3).

Republic of South Africa (2015) 'Antimicrobial Stewardship: The South African Perspective.' Pretoria: National Department of Health.

Republic of South Africa (2016) 'South Africa Demographic and Health Survey 2016 - Report.' Pretoria: National Department of Health.

Republic of South Africa (2018) *Medical Schemes Amendment Bill*. Pretoria.

Republic of South Africa (2019) *National Health Insurance Bill*. Pretoria.

Rethans, J., Gorter, S., Bokken, L. and Morrison, L. (2007) 'Unannounced standardised patients in real practice: A systematic literature review', *Medical Education*, 41(6), pp. 537-49.

Rice, T. (1983) 'The Impact of Changing Medicare Reimbursement Rates on Physician-Induced Demand', *Medical Care*, 21(8), pp. 803–815.

Rischatsch, M., Trottmann, M. and Zweifel, P. (2013) 'Generic substitution, financial interests, and imperfect agency', *International Journal of Health Care Finance and Economics*, 13(2), pp. 115–138.

Roetzheim, R. G., Gonzalez, E. C., Ferrante, J. M., Pal, N., Van Durme, D. J. and Krischer, J. P. (2000) 'Effects of health insurance and race on breast carcinoma treatments and outcomes', *Cancer*, 89(11), pp. 2202–2213.

Rosenberg, A., Agiro, A., Gottlieb, M., Barron, J., Brady, P., Liu, Y., Li, C. and DeVries, A. (2015) 'Early Trends Among Seven Recommendations From the Choosing Wisely Campaign', *JAMA Internal Medicine*, 175(12), pp. 1913–1920.

Saini, V., Brownlee, S., Elshaug, A. G., Glasziou, P. and Heath, I. (2017) 'Addressing overuse and underuse around the world', *The Lancet*, 390(10090), pp. 105-107.

Sanci, L. A., Coffey, C. M., Veit, F. C., Carr-Gregg, M., Patton, G. C., Day, N. and Bowes, G. (2000) 'Evaluation of the effectiveness of an educational intervention for general practitioners in adolescent health care: a randomised controlled trial', *BMJ*, 320(7229), pp. 224–230.

Satterthwaite, M. A. (1979) 'Consumer Information, Equilibrium Industry Price, and the Number of Sellers', *The Bell Journal of Economics*, 10(2), pp. 483–502.

Schellack, N., Benjamin, D., Brink, A., Duse, A., Faure, K., Goff, D., Mendelson, M., Meyer, J., Miot, J., Perovic, O., Pople, T., Suleman, F., van Vuuren, M. and Essack, S. (2017) 'A situational analysis of current antimicrobial governance, regulation, and utilization in South Africa', *International Journal of Infectious Diseases*, 64, pp. 100–106.

Schneider, E. C. and Epstein, A. M. (1996) 'Influence of Cardiac-Surgery Performance Reports on Referral Practices and Access to Care — A Survey of Cardiovascular Specialists', *New England Journal of Medicine*, 335(4), pp. 251–256.

Schneider, H. S. (2012) 'Agency Problems and Reputation in Expert Services: Evidence from Auto Repair', *Journal of Industrial Economics*, 60(3), pp. 406–433.

Schwartz, A., Weiner, S. J. and Binns-Calvey, A. (2013) 'Comparing Announced with Unannounced Standardized Patients in Performance Assessment', *Joint Commission Journal on Quality and Patient Safety*, 39(2), pp. 83–88.

Schwarz, N. and Sudman (Ed.), S. (1992) *Context Effects in Social and Psychological Research*. Springer-Verlag New York.

Scott, A., Sivey, P., Ait Ouakrim, D., Willenberg, L., Naccarella, L., Furler, J. and Young, D. (2011) 'The effect of financial incentives on the quality of health care provided by primary care physicians', *Cochrane Database of Systematic Reviews*, (9).

Scott, K. W. and Jha, A. K. (2014) 'Putting Quality on the Global Health Agenda', *New England Journal of Medicine*, 371(1), pp. 3–5.

Shekelle, P., Lim, Y., Mattke, S. and Damberg, C. (2008) *Does public release of performance results improve quality of care? A systematic review*. London: The Health Foundation.

Shi, L. (2000) 'Type of health insurance and the quality of primary care experience', *American Journal of Public Health*, 90(12), pp. 1848–1855.

Shrank, W. H., Cox, E. R., Fischer, M. A., Mehta, J. and Choudhry, N. K. (2009) 'Patients' Perceptions Of Generic Medications', *Health Affairs*, 28(2), pp. 546–556.

Sicsic, J., Le Vaillant, M. and Franc, C. (2012) 'Intrinsic and extrinsic motivations in primary care: An explanatory study among French general practitioners', *Health Policy*, 108(2), pp. 140–148.

Singer, E. and Ye, C. (2013) 'The Use and Effects of Incentives in Surveys', *The Annals of the American Academy of Political and Social Science*, 645, pp. 112–141.

Smith, R., Lagarde, M., Blaauw, D., Goodman, C., English, M., Mullei, K., Pagaiya, N., Tangcharoensathien, V., Erasmus, E. and Hanson, K. (2013) 'Appealing to altruism: an alternative strategy to address the health workforce crisis in developing countries?', *Journal of Public Health*, 35(1), pp. 164–170.

Smith, S. M., Fahey, T., Smucny, J. and Becker, L. A. (2017) 'Antibiotics for acute bronchitis', *Cochrane*

Database of Systematic Reviews, (6).

Smith, S. M., Schroeder, K. and Fahey, T. (2014) 'Over-the-counter (OTC) medications for acute cough in children and adults in community settings', *Cochrane Database of Systematic Reviews*, (11).

Smith, V. L. (1976) 'Experimental Economics: Induced Value Theory', *American Economic Review*, 66(2), pp. 274–279.

Spencer, C. S., Gaskin, D. J. and Roberts, E. T. (2013) 'The Quality Of Care Delivered To Patients Within The Same Hospital Varies By Insurance Type', *Health Affairs*, 32(10), pp. 1731–1739.

van der Spuy, Z., Zabow, T. and Good, A. (2017) 'Money isn't everything – CMSA doctor survey shows some noteworthy results', *South African Medical Journal*, 107(7), pp. 550–551.

Stats SA (2015a) 'Census 2011 Persons (10% sample) dataset, revised 2015'. Pretoria: Statistics South Africa.

Stats SA (2015b) 'Poverty Trends in South Africa: An examination of absolute poverty between 2006 and 2015.' Pretoria: Statistics South Africa.

Stats SA (2017) 'Mortality and causes of death in South Africa, 2015: Findings from death notification.' Pretoria: Statistics South Africa.

Stats SA (2019) 'Quarterly Labour Force Survey (QLFS), 1st Quarter 2019.' Pretoria: Statistics South Africa.

Stuckler, D., Basu, S. and McKee, M. (2011) 'Health care capacity and allocations among South Africa's provinces: infrastructure-inequality traps after the end of apartheid', *American Journal of Public Health*, 101(1), pp. 165–172.

Sylvia, S., Shi, Y., Xue, H., Tian, X., Wang, H., Liu, Q., Medina, A. and Rozelle, S. (2014) 'Survey using incognito standardized patients shows poor quality care in China's rural clinics', *Health Policy and Planning*, 30(3), pp. 322–333.

Takemura, Y., Ebisawa, K., Kakoi, H., Saitoh, H., Kure, H., Ishida, H. and Kure, M. (2005) 'Antibiotic selection patterns in acutely febrile new outpatients with or without immediate testing for C reactive protein and leucocyte count', *Journal of Clinical Pathology*, 58(7).

Tonin, M. and Vlassopoulos, M. (2010) 'Disentangling the sources of pro-socially motivated effort: A field experiment', *Journal of Public Economics*, 94(11), pp. 1086–1092.

Tonkin-Crine, S. K., Tan, P. S., van Hecke, O., Wang, K., Roberts, N. W., McCullough, A., Hansen, M. P., Butler, C. C. and Del Mar, C. B. (2017) 'Clinician-targeted interventions to influence antibiotic prescribing behaviour for acute respiratory infections in primary care: an overview of systematic reviews', *Cochrane Database of Systematic Reviews,* (9).

Vallejo-Torres, L. and Morris, S. (2018) 'Primary care supply and quality of care in England', *The European Journal of Health Economics*, 19(4), pp. 499–519.

van der Velden, A. W., Pijpers, E. J., Kuyvenhoven, M. M., Tonkin-Crine, S. K. G., Little, P. and Verheij, T. J. M. (2012) 'Effectiveness of physician-targeted interventions to improve antibiotic use for respiratory tract infections', *British Journal of General Practice*, 62(605).

Verbeke, G. and Lesaffre, E. (1997) 'The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data', *Computational Statistics & Data Analysis*, 23(4), pp. 541–556.

Volpp, K. G. M., Williams, S. V, Waldfogel, J., Silber, J. H., Schwartz, J. S. and Pauly, M. V (2003) 'Market reform in New Jersey and the effect on mortality from acute myocardial infarction', *Health Services Research*, 38(2), pp. 515–533.

Wadee, H., Gilson, L., Thiede, M., Okorafor, O. and McIntyre, D. (2003) 'Health Care Inequity in South Africa and the Public/Private Mix.' Geneva: UNRISD.

Wadge, H., Roy, R., Sripathy, A., Prime, M., Carter, A., Fontana, G., Marti, J. and Chalkidou, K. (2017) 'Evaluating the impact of private providers on health and health systems.' London: Imperial College.

Wadhera, R. K., Joynt Maddox, K. E., Yeh, R. W. and Bhatt, D. L. (2018) 'Public Reporting of Percutaneous Coronary Intervention Outcomes: Moving Beyond the Status Quo', *JAMA Cardiology*, 3(7), pp. 635–640.

Wagstaff, A. and Lindelow, M. (2008) ‘Can insurance increase financial risk?: The curious case of health insurance in China’, *Journal of Health Economics*, 27(4), pp. 990–1005.

Werner, R. M. and Asch, D. A. (2005) ‘The Unintended Consequences of Publicly Reporting Quality Information’, *JAMA*, 293(10), pp. 1239–1244.

Whaley, L. E., Businger, A. C., Dempsey, P. P. and Linder, J. A. (2013) ‘Visit complexity, diagnostic uncertainty, and antibiotic prescribing for acute cough in primary care: a retrospective study’, *BMC Family Practice*, 14(1).

WHO (2000) 'Appropriateness in health care services', *Report on a WHO workshop, Koblenz, Germany 23-25 March 2000*. Copenhagen: World Health Organisation

WHO (2001) 'WHO Global Strategy for Containment of Antimicrobial Resistance'. Geneva: World Health Organisation.

WHO (2015) 'Global action plan on antimicrobial resistance', *Ninth plenary meeting, World Health Assembly.* Geneva: World Health Organisation.

WHO (2018) 'WHO country cooperation strategy at a glance: South Africa'. Geneva: World Health Organisation.

WHO, World Bank and OECD (2018) 'Delivering quality health services: A global imperative for universal health coverage'. Geneva: World Health Organisation.

Wolinsky, A. (1993) ‘Competition in a Market for Informed Experts’ Services’, *The RAND Journal of Economics*, 24(3), pp. 380–398.

Wooldridge, J. M. (2003) ‘Cluster-Sample Methods in Applied Econometrics’, *American Economic Review*, 93(2), pp. 133–138.

Woolhandler, S. and Himmelstein, D. U. (2017) ‘The Relationship of Health Insurance and Mortality: Is Lack of Insurance Deadly?’, *Annals of Internal Medicine*, 167(6), pp. 424–431.

World Bank (2017) ‘Physicians (per 1,000 people)’, *WHO’s Global Health Workforce Statistics, supplemented by OECD country data.* Washington DC: The World Bank.

World Bank (2018) 'Overcoming Poverty and Inequality in South Africa:An Assessment of Drivers, Constraints and Opportunities'. Washington DC: The World Bank.

Xu, K., Soucat, A., Kutzin, J., Brindley, C. and Roubal, T. (2018) 'Public Spending on Health: A Closer Look at Global Trends'. Geneva: World Health Organisation.

Yip, W. C. (1998) ‘Physician response to Medicare fee reductions: changes in the volume of coronary artery bypass graft (CABG) surgeries in the Medicare and private sectors’, *Journal of Health Economics*, 17(6), pp. 675–699.

Zeckhauser, R. (1970) ‘Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives’, *Journal of Economic Theory*, 2(1), pp. 10–26.

Zhang, Y., Donohue, J. M., Lave, J. R., O’Donnell, G. and Newhouse, J. P. (2009) ‘The Effect of Medicare Part D on Drug and Medical Spending’, *New England Journal of Medicine*, 361(1), pp. 52–61.

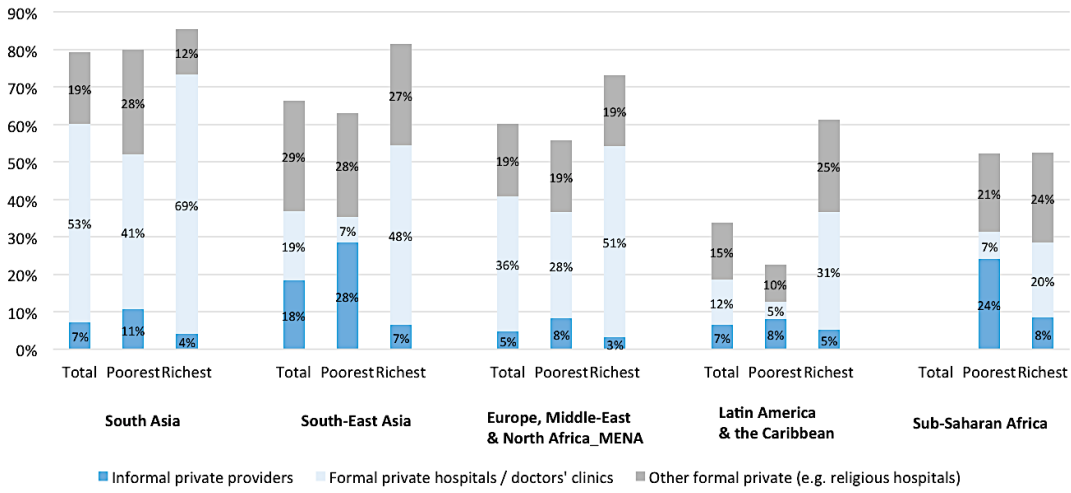
Zhou, R. A., Baicker, K., Taubman, S. and Finkelstein, A. N. (2017) ‘The Uninsured Do Not Use The Emergency Department More—They Use Other Care Less’, *Health Affairs*, 36(12), pp. 2115–2122.

Zuvekas, S. H. and Cohen, J. W. (2010) ‘Paying Physicians By Capitation: Is The Past Now Prologue?’, *Health Affairs*, 29(9), pp. 1661–1666.

Zweifel, P. and Manning, W. G. (2000) ‘Moral hazard and consumer incentives in health care’, in Pauly, M., McGuire, T. and Barros, P. *Handbook of Health Economics*, vol 1, pp. 409–459.

Appendices

Appendix A.1. Private provision of healthcare for diarrhoea and fever/cough in children under 5



Source. National Demographic and Health Surveys (DHS), between 2000 and 2008.
Notes. The figure summarises the source of care reported by all respondents with children under 5 that sought care for diarrhoea or fever/cough in the two weeks prior to the survey. The data is summarised by geographical region and the regional population wealth quintile of respondents.

Appendix A.2. Johannesburg: A microcosm of spatial inequality in South Africa

The city of Johannesburg – the largest metropolis, and financial centre of South Africa – is emblematic of the country's legacy of urban segregation, and the shortfalls in the redress policies that followed. Composed of 7 administrative regions (see Figure 2.2), the city remains effectively divided into the predominantly white, middle-class neighbourhoods in the north – around the former whites-only suburbs of Sandton and Randburg (in Regions B and E) – and the predominantly black, working-class neighbourhoods in the South, exemplified by Soweto (Region D).

Region A	Diepsloot, Midrand
Region B	Randburg, Northcliff
Region C	Roodeport
Region D	Greater Soweto
Region E	Sandton, Alexandra
Region F	Inner-city (CBD), Johannesburg South
Region G	Ennerdale, Orange Farm, Lenasia, Eldorado Park

Regions of the City of Johannesburg

This north-south divide along race and class lines is the combined result of two factors. First, there is the concentration of a declining (and northward shifting) manufacturing sector in the south, and a rapidly growing services sector in the north. Second, apartheid policies since 1950 restricted residential occupation in the northern neighbourhoods to whites only, and relocated black residents to the southern suburbs of Soweto, Eldorado Park and Lenasia.[326] By 2001, almost a decade after the end of legal spatial segregation, Crankshaw (2008) estimates that the white population still accounted for 73% of residents in the former whites-only northern neighbourhoods (despite comprising only 16% of the Johannesburg population). Moreover, all areas where the middle-class comprised more

[326] There are two exceptions to this north-south pattern of racial segregation in Johannesburg. One is the suburb of Alexandra in the northern region E, bordering Sandton, which was a blacks-only neighbourhood. It still remains a predominantly blacks-only neighbourhood, with black Africans comprising 99% of the area population according to the 2011 census. Another exception is Region A in the northern periphery, which was (and still is) predominantly resided by black Africans. The term "northern suburbs", as referenced in this thesis, refers to only those previously whites-only suburbs of Sandton and Randburg (Regions B and E, excluding Alexandra).

than two-thirds of the population were found in these predominantly white, northern neighbourhoods – concentrated around the areas of Sandton and Randburg.

Post-1994 government policies significantly expanded access to basic services, such as affordable housing, water, sanitation, electricity and public transport in deprived areas. However, spatial inequalities in access to economic opportunity - increasingly concentrated in the north – have perpetuated a "spatial poverty trap" among the working class neighbourhoods in the south (Pieterse and Owen, 2018). Despite being one of the richest municipalities in South Africa, around 43% of Johannesburg's 5 million residents live below the poverty line, and deprivation levels are highest in the city's south. High unemployment (over 23% in 2016, up from 6.8% in 1993), and a widening wage gap between low- and high-skilled workers, continues to exacerbate pre-existing inequalities. Unemployment is highest in the region of Soweto (35%), and black Africans comprise a disproportionately high share of low-wage and unemployed workers in the city (Pieterse and Owen, 2018). Health and human development outcomes are also aligned with these spatial patterns. Soweto had the highest density of HIV+ cases and AIDS-related deaths among all regions over the decade 2001-2011, followed by the other southern regions and Region A in the northern periphery. Conversely, Regions B, C and E (with the majority of all previously whites-only neighbourhoods) consistently had the highest rankings in the Human Development Index (HDI), which considers levels of literacy, education, life expectancy and income (HSRC, 2013).

References

Crankshaw, O. (2008) 'Race, Space and the Post-Fordist Spatial Order of Johannesburg', *Urban Studies*, 45(8), pp. 1692–1711.

HSRC (2013) 'The City of Johannesburg Economic Overview: 2013.' Cape Town: Human Sciences Research Council.

Pieterse, E. and Owen, K. (2018) 'Johannesburg: Confronting Spatial Inequality', *World Resources Report Case Study*. Washington DC: World Resources Institute.

Appendix B.1. Classifications of checklist items, diagnoses and treatments

	Clinical Case 1 Uncomplicated (viral) acute bronchitis in otherwise healthy young adult	Clinical Case 2 Complicated (bacterial) acute bronchitis in HIV+ young adult
A. Checklist items		
Essential history-taking	Duration of cough; temperature / fever; productive cough; coughing up blood; weight loss; night sweats; chest pain; breathing difficulty; smoking status; having TB before; HIV status; allergy to penicillin	Duration of cough; temperature / fever; productive cough; coughing up blood; weight loss; night sweats; chest pain; breathing difficulty; smoking status; having TB before; HIV status; allergy to penicillin; HIV history; any HIV treatment / ARV (antiretroviral) therapy; most recent viral load / CD4 count; any antibiotics (Bactrim) for pneumonia; any preventive drugs for TB
Recommended history-taking	Details about initial cold; having similar problem before; sore throat; ear ache; whistling noise with breaths; occupation / job status; contact with someone coughing / someone with TB; asthma (personal and family history); any regular medication; breathing fast; heart beating fast / palpitations; any allergies (in general)	Details about initial cold; having similar problem before; sore throat; ear ache; whistling noise with breaths; occupation / job status; contact with someone coughing / someone with TB; asthma (personal and family history); any regular medication; breathing fast; heart beating fast / palpitations; any allergies (in general)
Essential examinations	Take temperature; take blood pressure; take pulse rate; examine throat; listen to lungs; tap lungs (percuss)	Take temperature; take blood pressure; take pulse rate; examine throat; listen to lungs; tap lungs (percuss)
Recommended examinations	Palpate for lymph nodes (below ears); examine ears; listen to heart; check oxygen saturation	Palpate for lymph nodes (below ears); examine ears; listen to heart; check oxygen saturation
B. Diagnoses		
Correct / partially correct	Post-nasal drip; Coughing from bronchitis / bronchial irritation from coughing; Post-infectious cough; Acute bronchitis / bronchitis / inflammatory bronchitis; Tracheitis; Post-flu symptoms / after-effects of flu; Chest cold; Viral infection	Post-nasal drip; Coughing from bronchitis / bronchial irritation from coughing; Post-infectious cough; Acute bronchitis / bronchitis / inflammatory bronchitis; Tracheitis; Post-flu symptoms / after-effects of flu; Chest cold; Minor chest infection; Bacterial infection
Incorrect	Sino-bronchitis / chronic bronchitis; Allergies / hayfever; Rhinitis / laryngitis / pharyngitis / URTI; Throat inflammation / infection / irritation; Seasonal irritation; Lower respiratory infection; Minor chest infection; Chest inflammation / inflammation; Reactive / irritated airways; Sinusitis; Cough / persistent cough; Respiratory infection; Tonsilitis; Nasal irritation; Cold / flu; Asthma; TB; Bacterial infection	Sino-bronchitis / chronic bronchitis; Allergies / hayfever; Rhinitis / laryngitis / pharyngitis / URTI; Throat inflammation / infection / irritation; Seasonal irritation; Lower respiratory infection; Chest inflammation / inflammation; Reactive / irritated airways; Sinusitis, Cough / persistent cough; Respiratory infection; Tonsilitis; Nasal irritation; Cold / flu; Asthma; TB; Viral infection
C. Treatments		
Appropriate	Cough suppressants/expectorants; Analgesics; Throat preparations (e.g. lozenges)	Antibiotics (short course of Amoxicillin); Cough suppressants/expectorants; Analgesics; Throat preparations (e.g. lozenges)
Inappropriate / harmful	Antibiotics; Steroids; Nasal decongestants; Antihistamines; Bronchodilators; Probiotics; Vitamins; Other	Steroids; Nasal decongestants; Antihistamines; Bronchodilators; Probiotics; Vitamins; Other

Notes. For ease of comparison, listed items that are not shared across the two clinical cases are underlined. Any listed 'correct / partially correct' diagnosis for Clinical case 1 that was preceded by "bacterial" (e.g. "bacterial tracheitis") was classified as 'incorrect', whereas any preceded by "viral" (e.g. "viral bronchitis") was also considered 'correct / partially correct'. The classification of listed diagnoses apply to single diagnoses only. In a number of cases, multiple diagnoses were given, and protocols were established by the research team for classifying those. One example is where a 'Cold/Flu' diagnosis was given alongside a correct/partially-correct diagnosis (except just a 'viral infection'). Although 'Cold/Flu' on its own is incorrect (as classified above), the SPs presenting both clinical cases do report that a cold/flu preceded their current symptoms. As such, the combined diagnosis was also considered correct/partially-correct.

Appendix B.2. Q&A scripts for clinical cases

Clinical Case 1

POSSIBLE QUESTION ASKED BY THE DOCTOR/NURSE:	ANSWER GIVEN BY THE STANDARDISED PATIENT:
When did you have the flu/cold?	It started a week ago / last XXday.
Tell me more about the flu/cold?	It was like a normal cold. I had a blocked nose, a sore throat and my nose was running. I felt quite tired/I had no energy. And I was coughing. NOT: ~~shivering~~ ~~dizziness~~ ~~body pain~~
How long did the flu/cold last?	4-5 days
Tell me more about the cough?	It is really disturbing me. It is not going away.
How long have you had the cough?	About a week/ 7 days. Since the cold/flu started.
Did you take anything for the cold? / did you see a doctor?	I only took Panado/disprin and Medlemon / Stoney / Strepsils
Are you taking anything for the cough.	I've been taking cough syrup / Benylin/ Alcophyllex.
Is anything coming up when you cough?	Yes. A little
Is your cough dry?	Sometimes some white stuff comes up.
What you are coughing up – what does it look like?	White/clear stuff.
It is not green or yellow.	No
Is there any blood?	No
When are you coughing?	All the time. Both day and night. It really bothers me.
Do you have a fever?	No
Do/did you have any earache?	No
Is your throat sore now?	It is a bit irritated from the coughing.
Is the cough worse at night? Or worse in the morning?	Not really. But it is keeping keep me up at night.
Have you had this before?	I have had colds before and coughing. But this cough is worse than before.
Does anything make the coughing better / help the coughing?	I have been taking cough syrup. It doesn't help really.
Do you have any chest pain?	No
Are you in pain? / Do you have any pains?	Not really. My throat is sore from the coughing.
Is it painful when you breathe?	No
Do you feel short of breath/difficulty breathing?	No.
Do you have difficulty walking up the stairs or up a hill?	No.
Do you have any shortness of breath at night?	No
Have you had any wheezing/ whistling noise when you breathe?	Wheezing: What is that? No.
Is anyone else in your family/around you sick?	My flatmate / friend had the flu just before me. They are fine now.
Is there anyone around you that has been coughing for a long time / with a bad cough?	No

POSSIBLE QUESTION ASKED BY THE DOCTOR/NURSE:	ANSWER GIVEN BY THE STANDARDISED PATIENT:
Did you have any breathing problems / asthma as a child?	I don't think so. I've never heard that I did.
How is your appetite?	OK
Have you had any nausea or vomiting?	No
Have you had any diarrhoea?	No
Are you allergic to anything?	No, not that I know of.
Are you having night sweats?	No
Sweating a lot at night?	No
Are you losing weight? Have you lost weight recently?	No
Do you smoke? Have you ever smoked?	No
Do you drink?	Sometimes, with friends. OR never
Have you ever had TB?	No
Have you ever been tested for TB?	No
Have you ever had asthma?	No
Have you had any other lung problems before?	No
Are you HIV positive?	No
When did you last check for HIV?	About 2/4/6 months ago / at the beginning of the year
Why/Where/how were you last tested?	- I wanted to give blood - There was a campaign at work - I went with a friend - I broke up with my boyfriend when I found out he was cheating so I got tested
Do you have any other medical problem (Diabetes? Hypertension? Asthma? Pneumonia?)	No
Does anyone in your family have any medical problems (Diabetes? Hypertension? Asthma? Pneumonia?)	
Are you taking any other medication?	No. Just the cough syrup
Are you using oral contraceptives?	No
Are you pregnant?	No
When was your last menstrual period? / Have you missed your period? / Are your periods irregular?	2 weeks ago / No / No
What do you want me to do for you?	I am not sure. I just want to get better / I want this cough to go away. NOT: ~~prescription/treatment/drugs~~ ~~Sick note~~
Is there anything else wrong?	I am not sure what you mean.
Have you been travelling recently?/ Have you been out of the country recently?	No
Tell me about where you live?	Your current living situation. NOT: ~~Informal settlement, shack~~
Do you share a room with anyone?	No
Are you exposed to dust/fumes at work	No

Clinical Case 2

POSSIBLE QUESTION ASKED BY THE DOCTOR/NURSE:	ANSWER GIVEN BY THE STANDARDISED PATIENT:
When did you have the cold?	It started about 2 weeks ago.
Tell me more about the cold?	It was like a normal cold. I had a blocked nose, a sore throat and my nose was running. I felt quite tired/I had no energy. And I was coughing. NOT: ~~shivering~~ ~~dizziness~~ ~~body pain~~
How long did the cold last?	4-5 days
Tell me more about the cough?	It is really disturbing me. It is not going away.
How long have you had the cough?	About 2 weeks. Since the cold started.
Did you take anything for the cold? / did you see a doctor?	I only took Panado/disprin and Medlemon / Stoney / Strepsils
Are you taking anything for the cough.	I've been taking cough syrup / Benylin/ Alcophyllex.
Is anything coming up when you cough?	Yes. For the last 3 or 4 days. Some greenish stuff.
Is your cough dry?	No. Sometimes some greenish stuff comes up
What you are coughing up – what does it look like?	Green/Yellow stuff.
It is not green or yellow.	Yes it is.
Is there any blood?	No
When are you coughing?	All the time. Both day and night. It really bothers me.
Do you have a fever?	Yesterday in the afternoon I did feel a bit feverish. My friend at work felt my head and said I was hot. / I measured my temperature and it was 38 degrees.
Do/did you have any earache?	No
Is your throat sore now?	It is a bit irritated from the coughing.
Is the cough worse at night? Or worse in the morning?	Not really. But it is keeping keep me up at night.
Have you had this before?	I have had colds before and coughing. But this cough is worse than before.
Does anything make the coughing better / help the coughing?	I have been taking cough syrup. It doesn't help really.
Do you have any chest pain?	No
Are you in pain? / Do you have any pains?	Not really. My throat is sore from the coughing.
Is it painful when you breathe?	No
Do you feel short of breath/difficulty breathing?	No.
Do you have difficulty walking up the stairs or up a hill?	No.
Do you have any shortness of breath at night?	No
Have you had any wheezing/ whistling noise when you breathe?	Wheezing: What is that? No.
Is anyone else in your family/around you sick?	My flatmate / friend had the flu just before me. They are fine now.
Is there anyone around you that has been coughing for a long time / with a bad cough?	No

POSSIBLE QUESTION ASKED BY THE DOCTOR/NURSE:	ANSWER GIVEN BY THE STANDARDISED PATIENT:
Did you have any breathing problems / asthma as a child?	I don't think so. I've never heard that I did.
How is your appetite?	OK
Have you had any nausea or vomiting?	No
Have you had any diarrhoea?	No
Are you allergic to anything?	No, not that I know of.
Are you having night sweats?	No
Sweating a lot at night?	No
Are you losing weight? Have you lost weight recently?	No
Do you smoke? Have you ever smoked?	No
Do you drink?	Sometimes, with friends. OR never
Have you ever had TB?	No
Have you ever been tested for TB?	Yes.
When were you last tested for TB?	They have done the TB skin test and tested my sputum at the clinic. The last sputum was about 3 months ago. It was negative.
What was the result?	It was negative.
Have you ever had asthma?	No
Have you had any other lung problems before?	No
You say you are HIV positive?	Yes
How long have you known that you are HIV positive?	About 1 and a half years now.
How were you tested for HIV?	They used that finger prick test and they took blood from my arm.
How did you know that you were HIV positive? Were you sick when you discovered that you were HIV positive?	No. I was not sick / I completely well. I just had myself tested. - I wanted to give blood - There was a campaign at work - I went with a friend - I broke up with my boyfriend when I found out he was cheating so I got tested
Were you surprised to find that you were HIV positive.	Yes. It was a big shock. It took me some time to accept it.
How did you get the HIV?	I think it was my previous boyfriend/girlfriend. After I found out I had my suspicions and I went and asked them but s/he didn't really admit it.
Have you told your family about your status?	Yes.
Have you told your boyfriend/girlfriend about your status?	I don't have a boyfriend/girlfriend at the moment.
Have you had lots of boyfriends/girlfriends?	No. I was just unlucky with that one.
Are you on ARVs / taking treatment for the HIV?	Yes
Where are you getting the ARVs?	I go to XXX clinic, which is close to where I stay.
Why are you going to the public clinic for ARVs?	I don't have medical aid so have to go to the clinic for ARVs.
Why didn't you go to the clinic for this cough?	The queues are very long there. If you are sick you have to wait in the sick/acute queue which takes a long time. So it is better to go to private. / My next clinic visit is only in 2 weeks' time.

POSSIBLE QUESTION ASKED BY THE DOCTOR/NURSE:	ANSWER GIVEN BY THE STANDARDISED PATIENT:
Don't you have a regular private doctor? Why did you come to see me?	I have just got a new job nearby / This is close to where I work.
How often do you go to the clinic for your ARVs?	Every month to collect my drugs.
Which ARVs are you taking?	I am on the three-in-one tablet. Atripla /Tribuss. (Both pink oval tablets). I can't remember the 3 drug names.    
How many tablets are you on/taking?	One tablet a day / One tablet at night with supper.
When do you take your ARVs?	I take the tablet at night with supper.
When did you start the ARVs?	I was started after I was told I was positive. About 1 and a half year ago.
What was your CD4 count when you started the ARVs?	I don't remember exactly. I think it was about 450.
Have you changed your ARVs since you started?	No.
Did you bring the tablets with you?	No. I don't have them with me at work. I take them at home in the evening after supper.
Don't you have your clinic card with your HIV details?	Sorry. No I don't.
Do you take your ARVs properly/regularly?	Yes. I know that is important.
When was your last viral load test?	They only test it every 6 months. The last time was about 3 months ago.
What was your viral load?	It was not detected / not detectable
When was your last CD4 test?	About 3 months ago.
What was your CD4 count?	They told me it was more than 800.
Have you had any side effects from the ARVs?	I did have some nausea and dizziness when I first started. But it is better now.
Do you know about the side effects from the ARVs?	Yes. They told me. The serious one are problems with your liver…or your kidneys. They test my blood for that at the clinic.
Do you have any of these side-effects from the ARVs: nausea, diarrhoea, dizziness, insomnia, strange dreams, rash…..	Not at the moment. I did have some nausea and dizziness when I first started. But it is better now.
Is your boyfriend/girlfriend HIV positive?	No. I don't have a boyfriend/girlfriend at the moment.
Are you using condoms when you have sex?	I don't have a boyfriend/girlfriend at the moment.
Are you on Bactrim every month?	No
Are you taking any other drug other than the ARVs?	I am also taking the TB prevention drug.

POSSIBLE QUESTION ASKED BY THE DOCTOR/NURSE:	ANSWER GIVEN BY THE STANDARDISED PATIENT:
Are you on TB prevention drugs / IPT/Isoniazid preventive therapy?	Yes. They did my TB skin test and told me I had to take it for 3 years.
Are you taking Isoniazid?	Is that the drug to prevent TB? Yes I am also taking that.
Have you ever had enlarged/big lymph nodes?	No
Have you ever had thrush/meningitis?	What is that? No
Have you ever been in hospital for your HIV	No
Do you have any children?	No
Do you have any other medical problem (Diabetes? Hypertension? Asthma? Pneumonia?)	No
Does anyone in your family have any medical problems (Diabetes? Hypertension? Asthma? Pneumonia?)	
Are you taking any other medication?	The ARVs. And TB prevention tablet. And the cough syrup for the cough.
Are you using oral contraceptives?	No
Are you pregnant?	No
When was your last menstrual period? / Have you missed your period? / Are your periods irregular?	2 weeks ago / No / No
What do you want me to do for you?	I am not sure. I just want to get better / I want this cough to go away. NOT: ~~prescription/treatment/drugs~~ ~~Sick note~~
Is there anything else wrong?	I am not sure what you mean.
Have you been travelling recently?/ Have you been out of the country recently?	No
Tell me about where you live?	Your current living situation. NOT: ~~Informal settlement, shack~~
Do you share a room with anyone?	No
Are you exposed to dust/fumes at work	No

Appendix B.3. Role-specific Q&A scripts & insurance plan details

Patient Role A ("low-insured" patient)

POSSIBLE QUESTION ASKED BY THE DOCTOR (the doctor could ask you other questions)	ANSWER GIVEN BY THE STANDARDISED PATIENT:
Which medical aid / plan do you have?	I am on the Discovery Saver plan.
Are you paying for this consultation with your medical savings / through your medical aid?	Yes.
Are you paying / do you want to pay for this consultation in cash?	No, my medical aid will cover it. (NOTE: if the GP asks that you pay in cash AND CLAIM LATER, say 'Ok')
Do you have / do you know if you have enough medical savings to cover this consultation?	Yes, I checked. I have enough.
How much medical savings do you have left / how low are your savings?	It should be around **R1200**
How did you get your medical aid cover?	Through my work / my employer pays for it.
Have you claimed a lot recently / how did you run down your savings?	Just for one episode. I had some pain in my right knee recently, so I went to see a specialist about it. I had to have an X-ray and a lot of physio for that
How did the pain start / how did that happen?	I tripped and fell, and landed on my knee OR I was playing football, and it started hurting
When did it happen / when did the pain start?	A few weeks ago.
Is the knee pain better now?	Yes, the physio has helped a lot.
What did the knee specialist / doctor say? What was the injury / was it anything serious?	It's nothing serious. He said it was a routine injury, and I just need to rest the knee a bit.

Insurance plan details for Patient Role A:

GP Consultation Rate	R 430.40
Plan reimbursement rate	100%
Premiums per month	R 2773
Medical Savings per year	R 8316
How are GP visits paid for?	Through your Medical Savings
What happens when you run out of Medical Savings?	If you run out of Medical Savings for the year, the plan will cover 3 additional GP consultations (day-to-day extender benefit) if the GP is part of a specific Discovery network. After you've used up these 3 GP consultations as well, all other consultations in the same year will have to be paid for in cash. All other services, procedures or drugs prescribed during these consultations will also have to be paid in cash, if you run out of Medical Savings.
Co-payments	None, until the medical savings per annum are depleted

Patient Role B ("high-insured" patient)

POSSIBLE QUESTION ASKED BY THE DOCTOR (the doctor could ask you other questions)	ANSWER GIVEN BY THE STANDARDISED PATIENT:
Which medical aid / plan do you have?	I am on the Discovery Comprehensive plan.
Are you paying for this consultation with your medical savings / through your medical aid?	Yes.
Do you want to pay for this consultation in cash?	No, my medical aid will cover it. (NOTE: if the GP asks that you pay in cash AND CLAIM LATER, say 'Ok')
Do you have / do you know if you have enough medical savings to cover this consultation?	Yes, I checked. I have a lot.
How much medical savings do you have left?	I don't know exactly. But you should be able to check this on your system?
How did you get your medical aid cover?	Through my work / company.

Insurance plan details for Patient Role B:

GP Consultation Rate	R 430.40
Plan reimbursement rate	100%
Premiums per month	R 4882
Medical Savings per year	R 14640
How are GP visits paid for?	Through your Medical Savings
What happens when you run out of Medical Savings?	If you run out of Medical Savings for the year, the plan will cover unlimited additional GP consultations in the same year (day-to-day extender benefit) if the GP is part of a specific Discovery network. If you see a GP outside the network, or for any additional services, procedures and drugs prescribed during the consultations, you have to pay in cash once you finish your Medical Savings.
Co-payments	None, until your medical savings per year are depleted

Appendix B.4. Primary data collection tools

1. SP Debriefing Questionnaire

The tablet-based questionnaire, completed and submitted by SPs immediately after each GP consultation, was one of the two primary data collection tools in this study. The questionnaire was made up of 13 sections (groups of questions), as outlined below.

Questionnaire Section	Contents
1. **Introduction**	• GPS coordinates of the GP practice location. • Fieldworker's details (name, gender) • Patient role / clinical case to be played in consultation • GP name
2. **Arrival / Waiting Area**	• Time of arrival at facility • Questions on how and when fieldworkers' health insurance status was communicated to the receptionist • Number of other patients waiting to be seen by GP in waiting area (before and after consultation)
3. **Consultation Time**	• Consultation start and end times • GP time spent doing other things during consultation (e.g. talking on phone)
4. **Patient Role Reaction**	• GP's reactions (if any) to specific patient role played / SP opening statements (N.B. this section is only relevant to Part 1 of this study)
5. **Vitals Station**	• Any vitals taken prior to consultation (e.g. temperature, blood pressure)
6. **History-taking**	• Questions on which of a list of 25-30 history-taking items were completed (depending on clinical case)
7. **Examinations**	• Question on whether *any* examinations were completed; and if so, which of a list of 10-12 physical examinations were completed
8. **Tests**	• Questions on which of a list of 11 diagnostic tests were conducted / ordered during consultation
9. **Diagnosis**	• Question on whether any diagnosis was communicated by the GP; and if so, whether the diagnosis was volunteered or had to be asked for, and what was the specific diagnosis given
10. **Advice**	• Questions on which of a list of 10 recommendations / referrals were given by GP (e.g. advice on whether and under what circumstances to seek a follow-up consultation).
11. **Drugs**	• Any drugs dispensed, prescribed or recommended to buy OTC • Any advice given by GP with respect to taking drugs / potential side-effects • Instructions to fieldworkers to photograph dispensed drugs / prescription with tablet • Instructions to fieldworkers to place dispensed drugs / prescriptions into a sealable bag (labelled with correct fieldworker and GP name), for collection by research team
12. **Payment**	• Total cost of consultation • Question on whether dispensed drugs or diagnostic tests were charged for separately (if given) • Question on if consultation was paid for in cash or if GP would claim directly from insurance
13. **Satisfaction**	• Questions asking for a subjective assessment of the quality of the GP interaction • Fieldworkers' overall satisfaction with GP (on a scale of 1-10)

2. GP Interviews

The second primary data collection tool in this study was the face-to-face interview conducted with study participants. The interviews generally lasted 20-30 minutes, and were conducted at the GP practices between November 2018 and February 2019. A tablet-based interview tool was developed to guide the enumerators (including by displaying their script) and to capture the interview data. Enumerators from predominantly public health and nursing backgrounds were recruited and trained over the course of three days in November 2018 to carry out the interviews.[327] The timing of individual interviews was wholly dependent on GPs' availability. GPs were telephoned to schedule the interviews, and were asked for 30 minutes of their time. Enumerators followed a set script during these scheduling calls.[328]

The interviews consisted of five shorts parts, including a dictator game and an incentivised knowledge quiz, as detailed below.

a) Dictator Game

The first part of the interview involved a simple dictator game, to obtain a measure of GPs' individual altruism; specifically, their commitment to the wellbeing of patients. In this game, GPs were given R300 in cash in an envelope, and told that they could choose to keep all the money or donate some or all of it to a patient charity. The 'recipient' was chosen to be a patient charity, as the objective of the game was to elicit GPs' relative preferences for patient welfare in particular (rather than pro-social preferences in general), and it has been shown that the type of 'recipient' can influence allocator decisions (Eckel & Grossman 1996). A sealable, tamper-proof and opaque donation bag was provided inside the cash envelope, for GPs to deposit any cash they wished to donate. A list of four well-known patient charities, both local and national, were listed at the front of this bag for providers to choose from. Allowances were also made for providers to indicate another patient charity of their choice, if they wished.

The interviewer would begin the game by handing the cash envelope to the GP, and explaining all instructions following a set script. The script was carefully phrased to be as neutral as possible, as language can influence the game framing and subsequent outcomes (Aguiar et al. 2008). Any references to a "game" or "experiment" were avoided. GPs were told that the donation is entirely voluntary and anonymous; neither the interviewer nor the recipient patient charity would know if and how much they choose to donate. Experimenter and recipient blinding to participant donations is a common feature of dictator games, to avoid reputational or reciprocity concerns affecting donation choices (Hoffman et al. 1996; Hoffman et al. 1994). Recipient blinding may be particularly

[327] Some form of healthcare or health sciences experience was preferred in these enumerators, both to facilitate their comprehension and ease with some of the more technical parts of the interview, and to maintain credibility with the doctors.

[328] GPs were not told at this stage about the cash payoffs in the dictator game or knowledge quiz portions of the interview, to avoid these incentives influencing GPs' decisions to participate or undermining the purposes of the payoffs during the interview itself.

appropriate in experiments seeking to elicit physician preferences for patient welfare, given the asymmetries in information that tend to exist between physicians and patients in actual clinical settings, where some types of patient welfare-inducing physician effort may not be visible (or obvious) to the patient.

To reassure GPs of recipient blinding, they were told that donations to individual charities would be pooled before the donations were transferred to the charities, so that individual donation amounts were not identifiable. The interviewer also explained that they would turn around, while the GP decides how much to donate (if any) and seals the donation bag provided, to ensure interviewer blinding. GPs were further instructed to seal and hand back the donation bag even if they did not wish to donate anything, so that the interviewer cannot infer the decision to donate from the GP's retention of the bag.

Once the GP had finished sealing the donation bag, they were asked to inform the interviewer to turn back around, and hand the bag to them. At this stage, the interviewer would also ask the GP to sign a receipt confirming their receipt of the initial R300. The R300 endowment represented roughly 70% of the average consultation rate charged by these GPs for insured patients (approximately R434). While it was important to make the monetary payoff materially meaningful to the GPs, it may be argued that the absolute payoff amount is of little importance for measuring *relative* altruism (Carpenter et al. 2005); this study is ultimately interested in the distribution of altruism across the sample population, and how GPs' individual altruism *relative* to the sample population might explain within-sample variations in outcomes of interest.

b) Demographic and Practice Characteristics

The next part of the interview included a set of basic demographic and clinical practice questions. The demographic questions covered details on the GP's age, gender, ethnicity and nationality. Clinical practice questions covered the GP's years in professional practice; average daily patient load over the previous week; standard consultation rate for acute patients; contract status with specific medical insurance schemes; and drug dispensing status. Practice questions also asked whether the GP worked in a group practice, and if so, the number of other GP working there; and whether the GP only worked in the private sector, or also in the public sector.

c) Diagnosis & AMR Knowledge (Knowledge Quiz)

The decision was taken to incentivise most knowledge-based questions in this interview due to some (albeit limited) evidence that incentives can improve the quality of responses and reduce recall errors in surveys (Singer & Ye 2013), which are particular concerns when trying to measure the true knowledge of respondents. To minimise order and context effects in interview responses, whereby responses to certain questions may be influenced by preceding ones (Schwarz and Sudman, 1992), the ordering of questions within the knowledge quiz was randomised in each interview.

The first part of the quiz tested GPs' diagnostic knowledge on four patient case vignettes, presenting four respiratory illnesses with overlapping symptoms (uncomplicated (viral) acute bronchitis, complicated (bacterial) acute bronchitis, bacterial sinusitis, pneumonia). See Appendix B.5 (a) for a presentation of these vignettes. GPs were handed a typed copy of the vignettes to follow, while the interviewer read out each case. They were asked for the most likely diagnosis and pathogenic cause (virus, bacteria or neither) in each case. The order in which the four case vignettes were presented to each GP was randomised, to account for potential order effects in GP responses. The randomisation was programmed in the electronic questionnaire that was used by interviewers to complete the interviews, which generated a random ordering of the quiz questions for the interviewer to follow.

Before presenting the vignettes, the interviewer informed GPs that they would earn R20 for each correct answer to the eight questions in this section, and that they would receive their total cash reward at the end of the interview. To avoid any positive or negative encouragement effects on GPs' responses to remaining parts of the interview, it was decided to avoid revealing the quiz scores until the end.

The second part of the quiz covered five questions on GPs' understanding of AMR, including its definition, causes and risks. As with the diagnoses questions, GPs were told that each correct answer would be rewarded R20 at the end of the interview. Again to minimise order and context effects, these AMR questions were given at the end of the interview, to avoid biasing GP responses to previous questions on recommended treatment options (including antibiotics) – see details below.

d) Treatment Guidelines Knowledge

Following the diagnosis knowledge vignettes, GPs were presented with three further patient case scenarios where the diagnosis was revealed to them (see Appendix B.5 (b)). For each case, GPs were asked to indicate the most appropriate medications as recommended by the South African treatment guidelines. GPs could indicate multiple medications for each case. The purpose of this section was to assess GPs' knowledge of national treatment guidelines, accounting for patient characteristics and conditional on a correct diagnosis. The diagnoses given in the three cases were common cold, community-acquired pneumonia, and uncomplicated (viral) acute bronchitis.

As with the diagnosis vignettes, the ordering of the three case scenarios was randomised in each interview. However, it was decided not to incentivise these questions with cash rewards, as a 'correct' response is less clear in these cases. My main interest in these questions was to assess whether GPs made correct decisions as indicated in national guidelines for the prescription of antibiotics (and their type) in each case. It is more debatable whether other palliative medications can be considered appropriate or not, especially those that are not explicitly discouraged in the guidelines. Therefore, only the decision on antibiotic treatment was assessed in these questions.

e) Antibiotic Beliefs

GPs were informed that this part of the interview was not part of the knowledge quiz, but would ask for their opinions on the four patient cases presented in the diagnosis vignettes (see Appendix B.5(a)). For each case, GPs were presented with three questions, asking their opinions on the likelihood that the patient would recover faster *with* antibiotics (rather than without), the likelihood that other GPs *would* prescribe antibiotics to the same patient, and the likelihood that the patient would go elsewhere (visit another provider) next time they were ill if they were *not* given antibiotics for the case. For each question, the GP was asked to indicate a probability between 0% (meaning not likely at all) and 100% (indicating full certainty). The purpose of these questions was to gauge provider beliefs about the efficacy, peer norms, and patient expectations with regards to prescribing antibiotics in each patient case.

The interview concluded by giving GPs their overall score and total cash reward from the knowledge quiz. Each GP was also given a written summary of aggregate results from the SP visits in this part of the study, covering aggregate GP performance on recommended history-taking and examination completion, antibiotic prescribing, and costs of care.[329] Directions for contacting me and other members of the ESRC study team were also provided in the results summary sheet, in case of any clarifying questions or issues that GPs wished to flag.

References

Aguiar, F., Brañas-Garza, P. and Miller, L. M. (2008) 'Moral distance in dictator games', *Judgment and Decision Making*, 3(4), pp. 344–354.

Carpenter, J., Verhoogen, E. and Burks, S. (2005) 'The effect of stakes in distribution experiments', *Economics Letters*, 86(3), pp. 393–398.

Eckel, C. C. and Grossman, P. J. (1996) 'Altruism in Anonymous Dictator Games', *Games and Economic Behavior*. Academic Press, 16(2), pp. 181–191.

Hoffman, E., McCabe, K., Shachat, K. and Smith, V. (1994) 'Preferences, Property Rights, and Anonymity in Bargaining Games', *Games and Economic Behavior*, 7(3), pp. 346–380.

Hoffman, E., McCabe, K. and Smith, V. L. (1996) 'Social Distance and Other-Regarding Behavior in Dictator Games', *The American Economic Review*. American Economic Association, 86(3), pp. 653–660.

Schwarz, N. and Sudman (Ed.), S. (1992) *Context Effects in Social and Psychological Research*. Springer-Verlag New York.

Singer, E. and Ye, C. (2013) 'The Use and Effects of Incentives in Surveys', *The Annals of the American Academy of Political and Social Science*, 645, pp. 112–141.

[329] GPs that were part of the randomized experiment in Part 2 of this study (see Section 4.3.2) were not given this written feedback on aggregate study results at the end of their interviews. They were instead given individualized feedback or no feedback, depending on their experimental group assignment. The aggregate results were then shared with them at a later stage, after the experiment was concluded.

Appendix B.5 (a). Diagnosis knowledge – case vignettes

Patient Case	Description
A	A 25 year old man presents with three-day history of a blocked nose, sore throat and rhinorrhea. On examination he was found to have a temperature of 37.1ºC, a slight pharyngitis, a respiratory rate of 16 breaths per minute, and clear lungs.
B	An 18 year old woman presents with a 3-day history of fever, shortness of breath and a productive cough. On examination she is found to have a temperature of 38.2ºC, a respiratory rate of 26 breaths per minute, and crackles in both lungs.
C	A 25 year old man reports having a sore throat and rhinorrhoea for 10 days. However, over the last 5 days he has also developed a persistent cough throughout the day. The cough is productive of a white mucoid phlegm. On examination he was found to have a temperature of 37.1ºC, a respiratory rate of 17 breaths per minute, and clear lungs.
D	A 27 year old woman reports developing a cough after recovering from a cold. The cough has been persistent for the last 10 days and is productive of a yellow-green sputum. On examination she was found to have a temperature of 38.0ºC, a respiratory rate of 19 breaths per minute, and clear lungs.

Appendix B.5 (b). Treatment guidelines knowledge – case vignettes

Patient Case	Description
1	A 21 year old HIV-positive man suffering from a cold. His HIV is well-controlled and he is not allergic to penicillin.
2	A 38 year old woman with community-acquired pneumonia who is not allergic to penicillin.
3	A 36 year old HIV-negative man with a 10-day history of acute bronchitis, who is not allergic to penicillin.

Appendix B.6. Item Response Theory (IRT) assumptions

First applied in the context of psychometrics and educational testing, IRT is a model-based approach to developing and scoring tests (or checklists) for estimating latent individual traits (such as ability, or in this case, effort). Using maximum likelihood methods, IRT analysis calculates a weighted index score (the 'IRT score') that assigns greater weight to items that are more difficult (less likely to be completed) and that discriminate better among individuals in terms of the latent, unobserved trait of interest (Das and Hammer, 2005). IRT requires the following assumptions:

1. *Uni-dimensionality ('UD')*. This requires that the selected item set measures only one underlying trait – in this study's case, unobserved provider effort.
2. *No differential item functioning ('DIF')*. This requires that item completion (or non-completion) is only a function of underlying provider effort, and not any other provider- or consultation-level characteristic.
3. *Conditional independence ('CI')* of included items. That is, completion (or non-completion) of one item should not depend on the completion (and results) of other included items.

These assumptions are not straightforward to satisfy in the context of this study. On the no DIF assumption for instance, it is plausible that providers' performance in completing certain history-taking and examinations may depend on their diagnostic competence (knowledge). Conceptually, this issue is partially resolved by the theoretical modelling of knowledge as a pre-requisite for effort (alongside motivation), rather than as a direct and independent determinant of clinical practice (checklist completion). This notion is supported by evidence on a "know-do" gap in clinical performance in LMIC (Das and Gertler, 2007; Leonard and Masatu, 2010; Mohanan *et al.*, 2015), where knowledge does not always translate into appropriate practice – an intermediate variable (effort), perhaps driven by factors additional to knowledge (such as motivation), seems to be at play.

Typical history-taking and patient examination procedures also tend to follow a decision tree, where decisions to ask certain questions or complete certain examinations may be triggered by results (or patient responses) to previous completed items. This presents a potential violation of the third CI assumption. Whilst it is difficult to omit all such dependencies, this was a key consideration in determining which checklist items to retain in the IRT analysis, and some items were merged as a result. For example, the need to ask about a specific "penicillin allergy" may be negated by first asking about "any allergy". Similarly, if a provider asks the item about a "productive cough", he may expect the patient to also volunteer information about "coughing up blood"; if so, he may not follow up with the specific item on "coughing up blood". In these cases, I merged the two checklist items into one item (for example, "productive cough / coughing up blood") to be included in the IRT analysis, coded as 'completed' if the provider asked about one or both component items. Finally, the UD assumption is validated in the IRT analyses of by checking the eigenvalues from factor analysis of checklist items (Drasgow and Lissak, 1983).

Following Birnbaum (1968) and Das and Hammer (2005), I employ a three-parameter ('3pl') logistic function for modelling the probabilities that providers successfully complete individual checklist items, given their unobserved effort level. In addition to allowing for varying item characteristics (difficulty and discrimination levels), this allows for the possibility that certain checklist items can simply be guessed. Model fit is assessed through a chi-square test to check if model predictions for each item are significantly different from what is observed in the data.

References

Birnbaum, A. (1968) 'Some latent trait models and their use in inferring an examinee's ability.', *Statistical theories of mental test scores*. Addison-Wesley, pp. 397–472.

Das, J. and Gertler, P. J. (2007) 'Variations In Practice Quality In Five Low-Income Countries: A Conceptual Overview', *Health Affairs*. Health Affairs, 26(3), pp. w296–w309.

Das, J. and Hammer, J. (2005) 'Which doctor? Combining vignettes and item response to measure clinical competence', *Journal of Development Economics*, 78(2), pp. 348–383.

Drasgow, F. and Lissak, R. I. (1983) 'Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses.', *Journal of Applied Psychology*. US: American Psychological Association, 68(3), pp. 363–373.

Leonard, K. L. and Masatu, M. C. (2010) 'Using the Hawthorne effect to examine the gap between a doctor's best possible practice and actual performance', *Journal of Development Economics*, 93(2), pp. 226–234.

Mohanan, M., Vera-Hernández, M., Das, V., Giardili, S., Goldhaber-Fiebert, J. D., Rabin, T. L., Raj, S. S., Schwartz, J. I. and Seth, A. (2015) 'The Know-Do Gap in Quality of Health Care for Childhood Diarrhea and Pneumonia in Rural India', *JAMA Pediatrics*, 169(4), pp. 349–357.

Appendix B.7. Study information sheet (shared with ESRC study)

This study is investigating clinical decision-making by private and public primary care providers in Johannesburg. We thank you for your participation in this research.

The study has 4 components:

1. Fieldworker visits as patients

2-4 fieldworkers trained to act like patients will visit the clinic you work at over a period of 6-8 months.

You will not know who the fieldworker is because the study is blinded.

2. Follow-up telephone calls

To check whether you identified any of our fieldworker-patients.

3. Face-to-face interview

A 20-25 minute interview with you about primary health care and clinical decision-making.

The interview will include the opportunity to earn a small amount of money for your participation.

4. Private and confidential individual feedback

You will receive feedback on our findings from the fieldworker-patient visits to the clinic where you work.

Confidentiality

All information collected in this study is confidential. The results from this study may be published, but only aggregated results for groups of clinicians will be reported.

Your identity and your individual clinical performance will never be revealed or reported to anyone else.

Your participation

Your participation in this study is completely voluntary. There will be no negative consequences if you do not want to participate. You may discontinue your participation at any stage.

Potential benefits and risks

There are no personal risks to you from this study.

The face-to-face interview will include the opportunity to earn a small amount of money.

The information from this study will contribute to knowledge about the factors influencing clinical decision-making in primary care, in both the public and private sector.

Ethical aspects

Ethical approval for this study has been obtained from the University of the Witwatersrand Human Research Ethics Committee and the Ethics Committee of the London School of Economics and Political Science.

Ethical approval for this study has also been obtained from the District Research Committee, City of Johannesburg.

Questions

If you have any further questions about the survey, you can contact the researcher, Dr Duane Blaauw, at 082- 295-7377 or by email at duane.blaauw@wits.ac.za.

You can also contact the Wits University Human Research Ethics Committee, which is responsible for overseeing the ethical aspects of this study through Ms Zanele Ndlovu on (011) 717-1234.

Appendix B.8. Institutional ethics approval letters

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Houghton Street
London WC2A 2AE
United Kingdom

tel: +44 (0)20 7106 1202
email: rescon@lse.ac.uk

www.lse.ac.uk

Research Division

Arthika Sripathy
Department of Social Policy
A.Sripathy@lse.ac.uk

29th May 2018

Dear Arthika

Re: 'Physician altruism and health insurance: experimental evidence from South Africa' [REC ref# 000684]

I am writing with reference to the above research proposal. The Research Ethics Committee, having considered the documentation sent, is satisfied that the ethical issues raised by the proposed research have been properly taken into account and that adequate safeguards have been put in place. I am accordingly able on behalf of the Committee to confirm our approval of the application.

Please note that any significant changes to the research design must be reported to the Research Ethics Committee. Amendments to the research design that may affect participants and/or that may have ethical implications must be reviewed and approved by the Research Ethics Committee before commencement (or recommencement) of the project. The Research Ethics Committee may periodically conduct a selective audit of current research projects.

I would like to take this opportunity to wish you well with your research project.
If you have any further queries, please feel free to contact Lyn Grove, Research Division.

Yours sincerely,

Professor John Worrall
Chair of the Research Ethics Committee
cc. Lyn Grove, Research Division

R14/49 Dr Duane Blaauw et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M1611120

NAME: (Principal Investigator)	Dr Duane Blaauw et al
DEPARTMENT:	Centre for Health Policy School of Public Health Regions B and F, City of Johannesburg, Gauteng
PROJECT TITLE:	Determinants of Antibiotic Prescribing in Primary Care in South Africa: Studying Patient-Provider Interactions in the Private and Public Sectors
DATE CONSIDERED:	25/11/2016
DECISION:	Approved
CONDITIONS:	The Investigator has the responsibility to obtain permissions from the relevant research sites before any data may be collected
SUPERVISOR:	
APPROVED BY:	Professor P Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	06/02/2017

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report.** The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in November and will therefore be due in the month of November each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

06/02/2017
Date

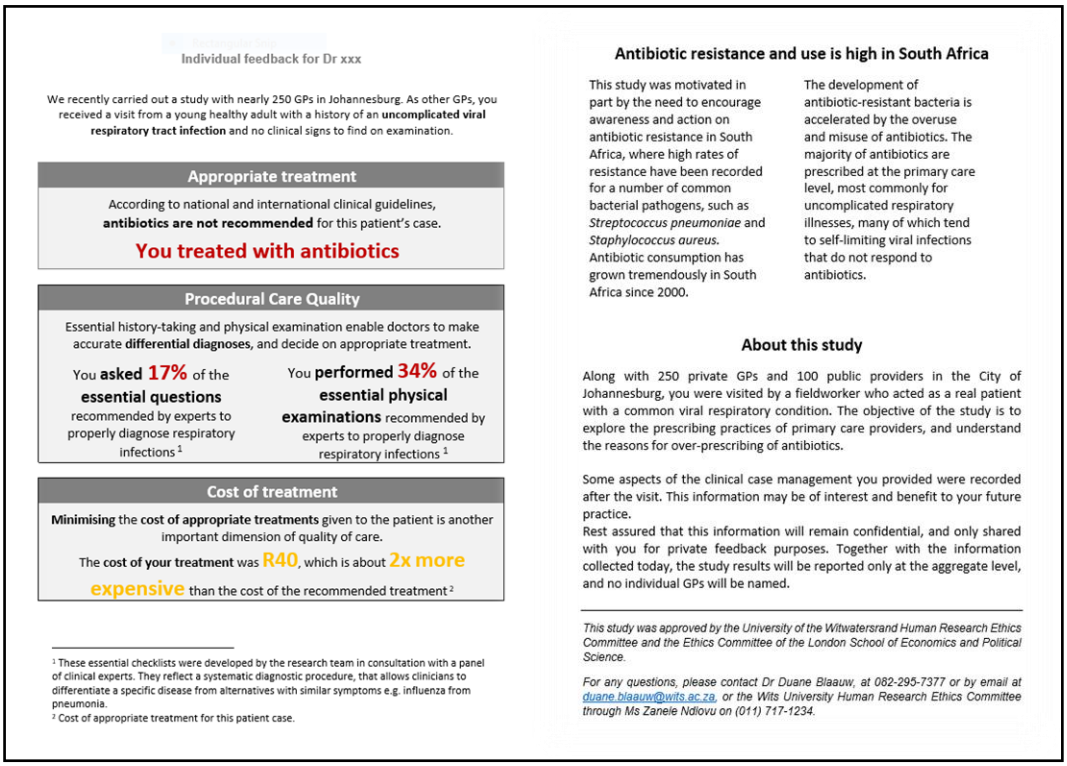
PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix B.9. Intervention delivery scripts for interviewers, by intervention group (study Part 2)

Treatment group	Control group
"As you may recall, Part 1 of this study [*the ESRC study*] involved sending standardised patients – that is, healthy actors trained to act as patients – to all participating providers. I can now offer you some private feedback on your individual results from the visits. This information is entirely confidential, and only shared with you for private feedback purposes. When we report the study results more broadly, we will only ever report the aggregate results - no individual providers will be named. The first page of the booklet inside summarises your results on a number of different aspects that we evaluated. Please read through the findings. I will try and answer any immediate questions that you might have. On the following page, there is also some general information on clinical guidelines for common conditions in primary care, which might be of interest."	"I will now give you some general information on clinical guidelines for common conditions in primary care, which might be of interest. Again, if you have any questions on the study or this interview, please feel free to contact the research team – their contact details are provided there as well."

Appendix B.10 (a). Template of feedback and educational leaflet for A&F treatment group

Pages 1 and 4 (front and back covers)



Individual feedback for Dr xxx

We recently carried out a study with nearly 250 GPs in Johannesburg. As other GPs, you received a visit from a young healthy adult with a history of an **uncomplicated viral respiratory tract infection** and no clinical signs to find on examination.

Appropriate treatment

According to national and international clinical guidelines, **antibiotics are not recommended** for this patient's case.

You treated with antibiotics

Procedural Care Quality

Essential history-taking and physical examination enable doctors to make accurate **differential diagnoses**, and decide on appropriate treatment.

You **asked** 17% of the **essential questions** recommended by experts to properly diagnose respiratory infections [1]

You **performed** 34% of the **essential physical examinations** recommended by experts to properly diagnose respiratory infections [1]

Cost of treatment

Minimising the **cost of appropriate treatments** given to the patient is another important dimension of quality of care.

The **cost of your treatment** was R40, which is about 2x more expensive than the cost of the recommended treatment [2]

[1] These essential checklists were developed by the research team in consultation with a panel of clinical experts. They reflect a systematic diagnostic procedure, that allows clinicians to differentiate a specific disease from alternatives with similar symptoms e.g. influenza from pneumonia.
[2] Cost of appropriate treatment for this patient case.

Antibiotic resistance and use is high in South Africa

This study was motivated in part by the need to encourage awareness and action on antibiotic resistance in South Africa, where high rates of resistance have been recorded for a number of common bacterial pathogens, such as *Streptococcus pneumoniae* and *Staphylococcus aureus*. Antibiotic consumption has grown tremendously in South Africa since 2000.

The development of antibiotic-resistant bacteria is accelerated by the overuse and misuse of antibiotics. The majority of antibiotics are prescribed at the primary care level, most commonly for uncomplicated respiratory illnesses, many of which tend to self-limiting viral infections that do not respond to antibiotics.

About this study

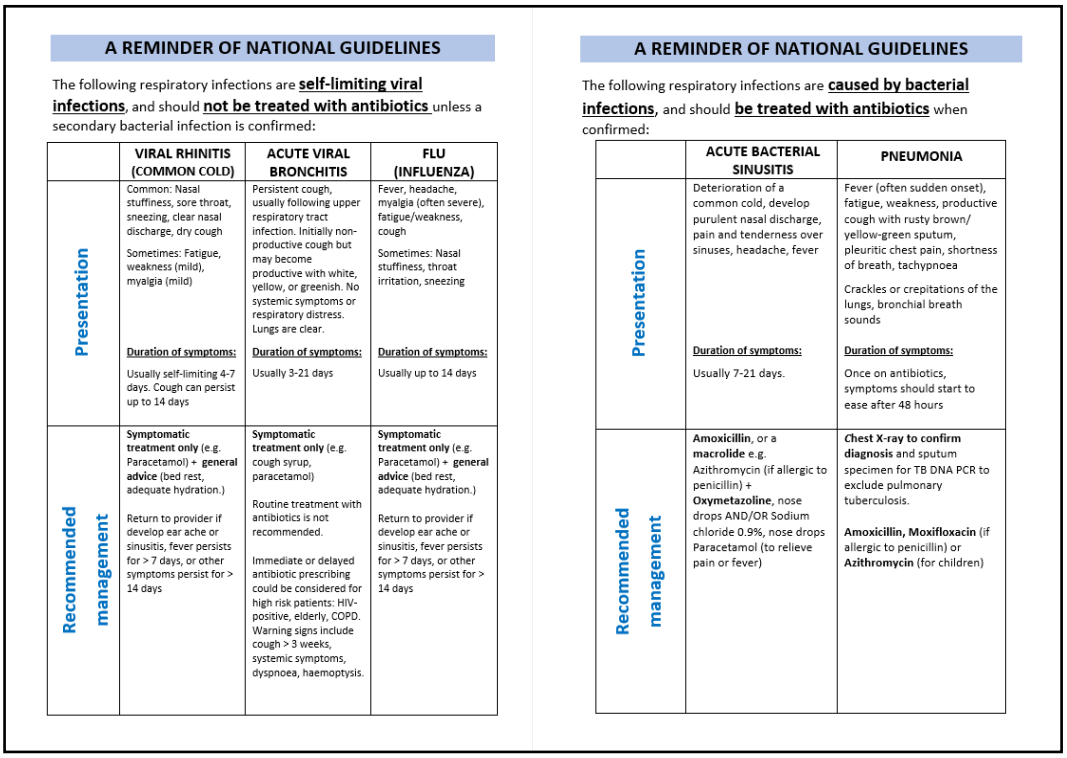
Along with 250 private GPs and 100 public providers in the City of Johannesburg, you were visited by a fieldworker who acted as a real patient with a common viral respiratory condition. The objective of the study is to explore the prescribing practices of primary care providers, and understand the reasons for over-prescribing of antibiotics.

Some aspects of the clinical case management you provided were recorded after the visit. This information may be of interest and benefit to your future practice.
Rest assured that this information will remain confidential, and only shared with you for private feedback purposes. Together with the information collected today, the study results will be reported only at the aggregate level, and no individual GPs will be named.

This study was approved by the University of the Witwatersrand Human Research Ethics Committee and the Ethics Committee of the London School of Economics and Political Science.

For any questions, please contact Dr Duane Blaauw, at 082-295-7377 or by email at duane.blaauw@wits.ac.za, or the Wits University Human Research Ethics Committee through Ms Zanele Ndlovu on (011) 717-1234.

Pages 2 and 3 (inside pages)



A REMINDER OF NATIONAL GUIDELINES

The following respiratory infections are **self-limiting viral infections**, and should **not be treated with antibiotics** unless a secondary bacterial infection is confirmed:

	VIRAL RHINITIS (COMMON COLD)	ACUTE VIRAL BRONCHITIS	FLU (INFLUENZA)
Presentation	Common: Nasal stuffiness, sore throat, sneezing, clear nasal discharge, dry cough Sometimes: Fatigue, weakness (mild), myalgia (mild) **Duration of symptoms:** Usually self-limiting 4-7 days. Cough can persist up to 14 days	Persistent cough, usually following upper respiratory tract infection. Initially non-productive cough but may become productive with white, yellow, or greenish. No systemic symptoms or respiratory distress. Lungs are clear. **Duration of symptoms:** Usually 3-21 days	Fever, headache, myalgia (often severe), fatigue/weakness, cough Sometimes: Nasal stuffiness, throat irritation, sneezing **Duration of symptoms:** Usually up to 14 days
Recommended management	**Symptomatic treatment only** (e.g. Paracetamol) + **general advice** (bed rest, adequate hydration.) Return to provider if develop ear ache or sinusitis, fever persists for > 7 days, or other symptoms persist for > 14 days	**Symptomatic treatment only** (e.g. cough syrup, paracetamol) Routine treatment with antibiotics is not recommended. Immediate or delayed antibiotic prescribing could be considered for high risk patients: HIV-positive, elderly, COPD. Warning signs include cough > 3 weeks, systemic symptoms, dyspnoea, haemoptysis.	**Symptomatic treatment only** (e.g. Paracetamol) + **general advice** (bed rest, adequate hydration.) Return to provider if develop ear ache or sinusitis, fever persists for > 7 days, or other symptoms persist for > 14 days

A REMINDER OF NATIONAL GUIDELINES

The following respiratory infections are **caused by bacterial infections**, and should **be treated with antibiotics** when confirmed:

	ACUTE BACTERIAL SINUSITIS	PNEUMONIA
Presentation	Deterioration of a common cold, develop purulent nasal discharge, pain and tenderness over sinuses, headache, fever **Duration of symptoms:** Usually 7-21 days.	Fever (often sudden onset), fatigue, weakness, productive cough with rusty brown/ yellow-green sputum, pleuritic chest pain, shortness of breath, tachypnoea Crackles or crepitations of the lungs, bronchial breath sounds **Duration of symptoms:** Once on antibiotics, symptoms should start to ease after 48 hours
Recommended management	**Amoxicillin**, or a **macrolide** e.g. Azithromycin (if allergic to penicillin) + **Oxymetazoline**, nose drops AND/OR Sodium chloride 0.9%, nose drops Paracetamol (to relieve pain or fever)	**Chest X-ray to confirm diagnosis** and sputum specimen for TB DNA PCR to exclude pulmonary tuberculosis. **Amoxicillin, Moxifloxacin** (if allergic to penicillin) or **Azithromycin** (for children)

Appendix B.10 (b). Template of educational leaflet for A&F control group

Pages 1 and 4 (front and back covers)

About this study

Along with 250 private GPs and 100 public providers in the City of Johannesburg, you have been visited by a fieldworker who acted as a real patient with a common viral respiratory condition. The objective of the study is to explore the prescribing practices of primary care providers, and understand the reasons for over-prescribing of antibiotics.

Some aspects of the clinical case management you provided were recorded after the visit. Once analysed, we will be happy to share this information with you. It may be of interest and benefit to your future practice. Rest assured that this information will remain confidential, and only shared for private feedback purposes with you.

Together with the information collected today, the overall results of the study will be reported only at the aggregate level, and none of the individual providers will be named.

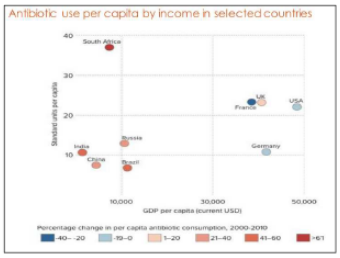
This study was approved by the University of the Witwatersrand Human Research Ethics Committee and the Ethics Committee of the London School of Economics and Political Science.

For any questions, please contact Dr Duane Blaauw, at 082-295-7377 or by email at duane.blaauw@wits.ac.za, or the Wits University Human Research Ethics Committee through Ms Zanele Ndlovu on (011) 717-1234.

Antibiotic resistance and use is high in South Africa

This study was motivated in part by the need to encourage awareness and action on antibiotic resistance in South Africa, where high rates of resistance have been recorded for a number of common bacterial pathogens, such as *Streptococcus pneumoniae* and *Staphylococcus aureus*. Antibiotic consumption has grown tremendously in South Africa since 2000.

The development of antibiotic-resistant bacteria is accelerated by the overuse and misuse of antibiotics. The majority of antibiotics are prescribed at the primary care level, most commonly for uncomplicated respiratory illnesses, many of which tend to self-limiting viral infections that do not respond to antibiotics.



Inside this leaflet, you will find some reminder about the **recommended management of respiratory infections**, according to the latest national and international clinical guidelines.

Pages 2 and 3 (inside pages)

A REMINDER OF NATIONAL GUIDELINES

The following respiratory infections are **self-limiting viral infections**, and should **not be treated with antibiotics** unless a secondary bacterial infection is confirmed:

	VIRAL RHINITIS (COMMON COLD)	ACUTE VIRAL BRONCHITIS	FLU (INFLUENZA)
Presentation	Common: Nasal stuffiness, sore throat, sneezing, clear nasal discharge, dry cough Sometimes: Fatigue, weakness (mild), myalgia (mild) **Duration of symptoms:** Usually self-limiting 4-7 days. Cough can persist up to 14 days	Persistent cough, usually following upper respiratory tract infection. Initially non-productive cough but may become productive with white, yellow, or greenish. No systemic symptoms or respiratory distress. Lungs are clear. **Duration of symptoms:** Usually 3-21 days	Fever, headache, myalgia (often severe), fatigue/weakness, cough Sometimes: Nasal stuffiness, throat irritation, sneezing **Duration of symptoms:** Usually up to 14 days
Recommended management	**Symptomatic treatment only** (e.g. Paracetamol) + **general advice** (bed rest, adequate hydration.) Return to provider if develop ear ache or sinusitis, fever persists for > 7 days, or other symptoms persist for > 14 days	**Symptomatic treatment only** (e.g. cough syrup, paracetamol) Routine treatment with antibiotics is not recommended. Immediate or delayed antibiotic prescribing could be considered for high risk patients: HIV-positive, elderly, COPD. Warning signs include cough > 3 weeks, systemic symptoms, dyspnoea, haemoptysis.	**Symptomatic treatment only** (e.g. Paracetamol) + **general advice** (bed rest, adequate hydration.) Return to provider if develop ear ache or sinusitis, fever persists for > 7 days, or other symptoms persist for > 14 days

A REMINDER OF NATIONAL GUIDELINES

The following respiratory infections are **caused by bacterial infections**, and should **be treated with antibiotics** when confirmed:

	ACUTE BACTERIAL SINUSITIS	PNEUMONIA
Presentation	Deterioration of a common cold, develop purulent nasal discharge, pain and tenderness over sinuses, headache, fever **Duration of symptoms:** Usually 7-21 days.	Fever (often sudden onset), fatigue, weakness, productive cough with rusty brown/ yellow-green sputum, pleuritic chest pain, shortness of breath, tachypnoea Crackles or crepitations of the lungs, bronchial breath sounds **Duration of symptoms:** Once on antibiotics, symptoms should start to ease after 48 hours
Recommended management	**Amoxicillin**, or a **macrolide** e.g. Azithromycin (if allergic to penicillin) + **Oxymetazoline**, nose drops AND/OR Sodium chloride 0.9%, nose drops Paracetamol (to relieve pain or fever)	**Chest X-ray to confirm diagnosis** and sputum specimen for TB DNA PCR to exclude pulmonary tuberculosis. **Amoxicillin, Moxifloxacin** (if allergic to penicillin) or **Azithromycin** (for children)

Appendix C.1. Determinants of non-participation in the GP interviews – provider characteristics and care quality outcomes

	Non-participation (1=Yes)
a) Provider characteristics	
Age	0.00
	(0.00)
Gender	0.05
	(0.06)
Is in a northern suburb	-0.09
	(0.06)
Number of GPs within 1 km radium (density)	0.01*
	(0.00)
b) Care quality outcomes	
Consultation length	-0.00
	(0.01)
Checklist completion (%)	-0.00**
	(0.00)
Essential exams (%)	-0.00***
	(0.00)
Essential history-taking (%)	-0.00
	(0.00)
IRT score	-0.06*
	(0.03)
Correct / partially correct diagnosis	-0.01
	(0.07)
Antibiotic prescribed	0.05
	(0.06)
Other inappropriate drugs prescribed	0.02
	(0.10)
Appropriate follow-up advice	-0.00
	(0.06)
Obs.	178

Notes: Column 2 shows estimates from univariate linear OLS regressions of provider attrition on a small set of provider characteristics (with available data on the full visits sample) and consultation outcomes of interest. Data on provider characteristics was taken from the *Medpages* database. The number of observations on "age" and "correct diagnosis" are marginally fewer than the stated 178 (at 172 and 177 respectively).
*** p<0.01, ** p<0.05, * p<0.1

Appendix C.2. IRT analysis of history-taking and examination checklist items (study Part 1)

Table A: Checklist items included in the IRT score

	Item type	Item description	Item complet. rate	Discrim. parameter	Standard error	Difficulty parameter	Standard error
1	History	Any regular medication	0.85	0.72	(0.27)	-2.62	(0.88)
2	History	Allergies (general) / penicillin allergy	0.93	1.29	(0.45)	-2.53	(0.63)
3	Examination	Listen to lungs	0.94	1.54	(0.53)	-2.42	(0.54)
4	Examination	Examine throat	0.94	1.83	(0.60)	-2.15	(0.41)
5	History	Cough duration	0.85	1.04	(0.31)	-1.98	(0.49)
6	History	Productive cough / coughing up blood	0.84	1.16	(0.33)	-1.79	(0.40)
7	History	Temperature / fever	0.83	1.17	(0.32)	-1.70	(0.37)
8	History	Occupation status	0.63	0.32	(0.19)	-1.69	(1.06)
9	Examination	Listen to heart	0.58	0.24	(0.18)	-1.46	(1.26)
10	History	Sore throat	0.77	1.07	(0.29)	-1.37	(0.32)
11	Examination	Examine ears	0.84	1.96	(0.53)	-1.35	(0.22)
12	History	Chest pain	0.68	0.73	(0.22)	-1.15	(0.37)
13	Examination	Measure blood pressure	0.65	0.77	(0.23)	-0.92	(0.32)
14	Examination	Palpate lymph nodes	0.68	1.19	(0.28)	-0.81	(0.21)
15	Examination	Take temperature	0.63	0.98	(0.25)	-0.66	(0.22)
16	History	Smoking status	0.57	0.59	(0.20)	-0.54	(0.32)
17	Examination	Measure pulse rate	0.57	1.38	(0.32)	-0.29	(0.15)
18	History	Asthma history	0.53	0.57	(0.20)	-0.25	(0.29)
19	History	Details about the initial cold	0.56	1.41	(0.31)	-0.24	(0.15)
20	History	Night sweats	0.31	1.04	(0.27)	0.94	(0.25)
21	History	Ear ache	0.25	1.58	(0.38)	1.01	(0.20)
22	History	Breathing difficulty	0.35	0.63	(0.21)	1.08	(0.41)
23	History	Weight loss	0.23	0.80	(0.26)	1.70	(0.50)
24	History	Exposure to anyone with TB	0.17	1.05	(0.33)	1.83	(0.46)
25	History	Similar problem before	0.12	0.67	(0.29)	3.16	(1.23)
26	Examination	Tap lungs (percuss)	0.07	0.77	(0.37)	3.60	(1.49)
27	Examination	Check oxygen saturation	0.13	0.41	(0.26)	4.69	(2.86)
28	History	HIV status	0.10	0.43	(0.32)	5.43	(3.86)
29	History	Whistling noise when breathing	0.19	0.26	(0.22)	5.70	(4.89)
30	History	TB history	0.15	0.16	(0.25)	10.64	(16.47)
Obs.		178					

Notes: The discrimination and difficulty parameters (and their standard errors) were estimated using a three parameter logistic model, which also allows for the possibility of item completion through guessing. The estimated 'pseudo-guessing' parameter was close to zero and insigificant, however, suggesting that the likelihood that doctors completed items on this checklist purely through guessing (and zero effort) is close to zero. Item completion rate is the share of all consultations (178) in which the specific item was completed. The difficulty parameter is a measure of the likelihood that a specific item is successfully completed in any given consultation, with higher parameter values indicating more difficult items. The discrimination parameter indicates how well an item can distinguish between high- and low-effort providers, with high-discrimination items having a higher correlation between their completion and provider effort.

Table B: Checklist items that were either excluded or merged in the IRT score

	Item type	Item description	Item completion rate
1	History	Productive/dry cough	0.83
2	History	Coughing up blood	0.07
3	History	Breathing fast	0.02
4	History	Heart beating fast / palpitations	0.02
5	History	Allergy to penicillin	0.41
6	History	Any allergies (in general)	0.87

Notes: Items 1 and 2 were merged into one item in the IRT analysis, as they are likely to overlap and potentially violate the conditional independence assumption necessary for the IRT score validity (see Section 5.3.1.1.1). Items 5 and 6 were also merged for the same reason. Items 3 and 4 were excluded as they were completed in only 2% of consultations, and their inclusion was problematic for the convergence of the maximum likelihood estimation procedure.

Columns 5 and 7 of Table A present estimates of the discrimination and difficulty parameters for each item from the IRT analysis, indicating their relative ability to discriminate between high- and low-effort physicians and their likelihood of completion in any given consultation, respectively. There appears to be very little correlation between an item's difficulty and its discriminating ability. For instance, the three most discriminating items are examination of the ears, examination of the throat and asking about earache. The first and last of these may be needed to rule out acute otitis media (AOM), while the second item is necessary for diagnosing an upper respiratory infection (rather than, or in addition to, a more systemic or lower respiratory condition). Yet, the most difficult items appear to be history-taking on a patient's TB history, HIV status, and the presence of any whistling noises when breathing (although the difficulty parameter point estimate for TB history is very imprecise). As expected, there is a high degree of correlation between the proportion of consultations in which a particular item was completed and its difficulty parameter.

Appendix C.3. Statistical approaches to handling cluster-correlated data

Four methods are generally proposed for handling cluster-correlated data, and the optimal approach is largely study-dependent (and to some extent, subject to disciplinary preferences (McNeish and Kelley, 2019)). The first approach is a paired-sample t-test, which is most appropriate when there are only two levels of treatment (as in the present case: high- or low-insured) and the primary objective is to test for differences in outcomes of interest across the two levels. This is equivalent to differencing the outcomes across treatment levels for each physician, and running a simple one-sample t-test of the differences. The limitation here is that the effects of additional (potentially confounding) covariates cannot be estimated or controlled for, unlike in a formal regression framework. Alternative approaches for handling clustering within linear regression frameworks include the marginal model (estimated through the generalised estimating equations (GEE) method (Liang and Zeger, 1986)), which can adjust OLS standard errors for clustering at the physician-level using a cluster-robust "sandwich" variance estimator. It requires no assumptions on the within-cluster correlational structure (Wooldridge, 2003), and can provide robust OLS parameter estimates when the number of clusters is sufficiently "large" (at the conservative end, more than 50 for balanced clusters (Cameron and Miller, 2015)). Another common approach to handling clustered data, particularly in the economics literature, is to include cluster-specific fixed effects in the linear regression model. This is equivalent to estimating the following:

$$Q_{ij} = \beta_0 + \beta_1 Insurance_{ij} + \delta_j + \varepsilon_{ij} \quad (1)$$

where δ_j are physician-specific indicators, that allow for unique intercepts and control for any general variation among physicians.[330] As the physician fixed effects control for all physician-level variation, it is not possible to estimate the effects of any covariates that do not vary within-physician. One advantage in this study context is that the fixed effects will automatically control for fieldworker-pair effects that can confound the estimated insurance effect (and therefore should be controlled for), as assigned fieldworker pairs do not vary within-physician in this experiment.

There are two common methods for obtaining the same cluster fixed effects estimate of β_1: OLS estimation of the mean-differenced version of model (2) ('within' estimation) or direct OLS estimation of (2) with the inclusion of physician-specific dummy variables ('least squares dummy variable (LSDV)' estimation). These two methods are generally considered equivalent; however, the first method is preferred in this study context for the following reason. The inclusion of cluster fixed effects may not absorb all within-cluster correlation in ε_{ij} (Cameron and Miller, 2015), and it is

[330] To allow that the insurance effect may also vary among physicians (i.e. to allow unique slopes, in addition to unique intercepts per physician), additional interaction terms of each physician-specific indicator and the insurance variable would need to be included. This may be very costly in terms of degrees of freedom, and will not be possible in some cases.

recommended to use the cluster-robust variance estimator even with the inclusion of cluster fixed effects (Arellano, 1987). In such cases where the cluster-robust variance estimator is used alongside cluster fixed effects, and the number of observations within clusters is very small (e.g. 2 in this case), the within estimator should be used instead of LSDV as the finite sample correction applied by STATA for correcting standard error biases in small samples is incorrect under LSDV (Cameron and Miller, 2010). [331]

An alternative approach, commonly employed in the behavioural sciences, is a linear mixed effects model (LMEM), where individual physicians are treated as random (rather than fixed) effects:

$$Q_{ij} = \beta_0 + \beta_1 Insurance_{ij} + u_{0j} + u_{1j} Insurance_{ij} + \varepsilon_{ij} \quad (2)$$

Here, $Insurance_{ij}$ is the only fixed effect, and physician-level variation is captured by allowing for random variation in the intercept and slope (insurance effect) of the estimated regression line by physician. By including a random intercept (by estimating the variance of the u_{0j}'s), we allow that physicians can vary in their average care quality choices. This is equivalent to estimating the fixed effects δ_j in (2). By further including a random slope (by estimating the variance of the u_{1j}'s), we also allow that physicians can vary in their care quality responses to patient insurance. This is equivalent to adding interaction terms between the δ_j's and $Insurance_{ij}$ in (2); although, as the LMEM only estimates the variances of the random effects (rather than the individual effects), we save several degrees of freedom. The LMEM nevertheless relies on stronger assumptions than both the marginal and fixed effects models, particularly in relation to the covariance structure of the random effects and distribution of the error terms. In general, it requires that all relevant random effects are included in the model (random intercepts and slopes), the random effects and error terms are normally distributed and their covariance structures properly specified, the random effects are independent of the error terms, and the distribution of error terms is homoscedastic (Laird and Ware, 1982). Several studies have investigated the effects of violations of these assumptions on the maximum likelihood inference of β_1 and its variance. It has been demonstrated that β_1 estimates are robust to non-normal random effect distributions (Verbeke and Lesaffre, 1997), non-normal or heteroscedastic error distributions (Jacqmin-Gadda *et al.*, 2007), and misspecification of the covariance structure (Liang and Zeger, 1986). However, the mixed effects model with random intercept and slope, as in (3), has been shown to be more robust to misspecification of the covariance structure than the mixed effects model with just a random intercept. It has also been shown that the cluster-robust variance estimator can correct for biased variance estimates of β_1 when the covariance structure is incorrectly specified (Liang and Zeger, 1986).

[331] LSDV estimation will inflate the cluster-robust variance estimate, particularly when the number of observations per cluster is very small (as in this study context).

The three linear regression approaches outlined above (marginal model with cluster-robust standard errors, cluster fixed effects, and LMEM) are preferred to response simplification (paired-sample t-test) given their ability to control for potential physician-level confounders of the insurance effect - most crucially, any fieldworker-pair effects.[332] The cluster fixed effects model is generally preferred to the LMEM due to its ease of specification, and the need for fewer assumptions for valid inference. Moreover, it automatically controls for all fixed physician-level confounders of the insurance effect without need for explicit specification of these confounders as model covariates.

References

Arellano, M. (1987) 'Practitioners' Corner: Computing Robust Standard Errors for Within-groups Estimators', *Oxford Bulletin of Economics and Statistics*, 49(4), pp. 431–434.

Cameron, A. C. and Miller, D. (2010) 'Robust Inference with Clustered Data', in Ullah, A. and Giles, D. (eds) *Handbook of Empirical Economics and Finance*. 1st edn. Chapman and Hall.

Cameron, A. C. and Miller, D. (2015) 'A Practitioner's Guide to Cluster-Robust Inference,' *Journal of Human Resources*, 50(2), pp. 317-372.

Jacqmin-Gadda, H., Sibillot, S., Proust, C., Molina, J.-M. and Thiébaut, R. (2007) 'Robustness of the linear mixed model to misspecified error distribution', *Computational Statistics & Data Analysis*, 51(10), pp. 5142–5154.

Laird, N. M. and Ware, J. H. (1982) 'Random-Effects Models for Longitudinal Data', *Biometrics*, 38(4), pp. 963–974.

Liang, K.-Y. and Zeger, S. L. (1986) 'Longitudinal data analysis using generalized linear models', *Biometrika*, 73(1), pp. 13–22.

McNeish, D. and Kelley, K. (2019) 'Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations.', *Psychological Methods*, 24(1), pp. 20–35.

Verbeke, G. and Lesaffre, E. (1997) 'The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data', *Computational Statistics & Data Analysis*, 23(4), pp. 541–556.

Wooldridge, J. M. (2003) 'Cluster-Sample Methods in Applied Econometrics', *American Economic Review*, 93(2), pp. 133–138.

[332] Fieldworker pair effects must be explicitly controlled for in the LMEM and marginal model through the inclusion of fieldworker-pair fixed effects. However, they will be automatically controlled for in the cluster fixed effects model in this study context.

Appendix C.4. Paired-sample t-test of differences in individual checklist item completion between low-insured and high-insured patient consultations

	(1) Low-insured patient		(2) High-insured patient		
	Mean	SD	Mean	SD	Difference
a. History-taking					
Cough duration +	0.82	(0.39)	0.88	(0.33)	0.06
Details about the initial cold	0.45	(0.50)	0.67	(0.47)	0.22***
Temperature / fever +	0.83	(0.38)	0.83	(0.38)	0.00
Productive cough +	0.84	(0.37)	0.82	(0.39)	-0.02
Coughing up blood +	0.07	(0.25)	0.08	(0.27)	0.01
Similar problem before	0.11	(0.32)	0.14	(0.34)	0.02
Weight loss +	0.21	(0.41)	0.25	(0.43)	0.03
Night sweats +	0.21	(0.41)	0.40	(0.49)	0.19***
Sore throat	0.80	(0.40)	0.74	(0.44)	-0.06
Chest pain +	0.71	(0.46)	0.65	(0.48)	-0.06
Ear ache	0.24	(0.43)	0.26	(0.44)	0.02
Breathing difficulty +	0.28	(0.45)	0.42	(0.50)	0.13**
Breathing fast	0.03	(0.18)	0.01	(0.11)	-0.02
Palpitations	0.02	(0.15)	0.01	(0.11)	-0.01
Whistling noise when breathing	0.21	(0.41)	0.17	(0.38)	-0.04
Smoking status +	0.56	(0.50)	0.58	(0.50)	0.02
Occupation status	0.64	(0.48)	0.62	(0.49)	-0.02
Exposure to anyone with TB	0.12	(0.33)	0.21	(0.41)	0.09*
TB history +	0.10	(0.30)	0.20	(0.40)	0.10**
Asthma history	0.57	(0.50)	0.49	(0.50)	-0.08
HIV status +	0.09	(0.29)	0.10	(0.30)	0.01
Penicillin allergy +	0.48	(0.50)	0.34	(0.48)	-0.15**
Penicillin allergy, conditional on antibiotic prescription (Obs. = 38)^	0.61	(0.50)	0.37	(0.49)	-0.24**
Any allergies (in general)	0.80	(0.40)	0.94	(0.23)	0.15***
Any regular medication	0.84	(0.37)	0.85	(0.36)	0.01
b. Physical examination					
Take temperature +	0.63	(0.49)	0.65	(0.48)	0.02
Measure blood pressure +	0.67	(0.49)	0.80	(0.46)	0.12**
Measure pulse rate +	0.54	(0.50)	0.73	(0.48)	0.19***
Examine throat +	0.97	(0.18)	0.91	(0.29)	-0.06**
Palpate lymph nodes	0.71	(0.46)	0.65	(0.48)	-0.06
Examine ears	0.86	(0.34)	0.82	(0.39)	-0.04
Listen to heart	0.76	(0.43)	0.40	(0.49)	-0.36***
Listen to lungs +	0.93	(0.25)	0.96	(0.21)	0.02
Tap lungs (percuss) +	0.08	(0.27)	0.07	(0.25)	-0.01
Check oxygen saturation	0.10	(0.30)	0.17	(0.38)	0.07
Observations	89		89		

+ indicates items that were considered *essential* by clinical experts, for enabling a differential diagnosis of the underlying condition.

^ This presents results from a paired-sample tttest of differences in the likelihood of a penicillin allergy check by insurance cover among the subsample of 38 physicians that prescribed an antibiotic to both high-insured and low-insured patients. Equivalent results were all found in an independent-sample ttest of differences among the subsample of 101 consultation observations where an antibiotic was prescribed. *** p<0.01, ** p<0.05, * p<0.1

Appendix C.5. Paired-sample t-test of differences in care quality outcomes between low-insured and high-insured patient consultations

	(1) Low-insured patient		(2) High-insured patient		
	Mean	SD	Mean	SD	Difference
a. Diagnostic effort					
Consultation length (mins.)	10.337	(4.22)	10.573	(6.52)	0.236
History-taking and examinations checklist completed (share)	0.481	(0.11)	0.497	(0.15)	0.016
Essential examinations completed (share)	0.646	(0.21)	0.697	(0.23)	0.051**
Essential history-taking completed (share)	0.434	(0.16)	0.463	(0.18)	0.281
IRT score	-0.111	(0.74)	0.111	(1.01)	0.221**
b. Diagnosis					
Gave a correct or partially-correct diagnosis	0.60	(0.49)	0.46	(0.50)	-0.136*
c. Treatment					
Prescribed an antibiotic	0.506	(0.50)	0.629	(0.49)	0.124**
Prescribed other inappropriate drugs	0.921	(0.27)	0.967	(0.18)	0.045
Gave appropriate follow-up advice	0.506	(0.50)	0.719	(0.45)	0.213***
Observations	89		89		

*** p<0.01, ** p<0.05, * p<0.1

Appendix C.6. Effect of insurance cover on care quality outcomes (mixed effects model specification)

	Diagnostic effort					Diagnosis & Treatment			
	(1) Consult. length	(2) Checklist completion	(3) Essential. exams	(4) Essential. history	(5) IRT score	(7) Correct /Partially - correct diagnosis	(9) Antibiotic	(10) Other inapprop. drugs	(11) Approp. follow-up advice
Panel A: provider-level random effects (RE)									
High-insured	0.24	0.02	0.05**	0.03	0.22**	-0.14**	0.12**	0.04*	0.21***
	(0.80)	(0.02)	(0.02)	(0.02)	(0.11)	(0.07)	(0.05)	(0.03)	(0.07)
Obs.	178	178	178	178	178	177	178	178	178
AIC	1103.19	-222.95	1595.98	1520.27	448.62	261.60	242.15	-23.03	249.81
Panel B: provider-level RE; fieldworker-pair fixed effects									
High-insured	0.25	0.01	0.05**	0.03	0.22**	-0.14**	0.12**	0.04	0.22***
	(0.80)	(0.02)	(0.02)	(0.02)	(0.11)	(0.07)	(0.05)	(0.03)	(0.07)
Obs.	178	178	178	178	178	177	178	178	178
AIC	1094.41	-213.65	1604.85	1498.29	455.06	271.32	253.16	-15.41	260.30
Mean of low-insured	10.34	0.48	0.65	0.43	-0.11	0.60	0.51	0.92	0.51

Notes: Standard errors are in parenthesis. In Panel A, estimates from a linear mixed effects model with provider-level random effects (intercept and slope with respect to insurance) are shown. In Panel D, results from the same mixed effects model with the addition of fieldworker-pair fixed effects are shown. The Akaike Information Criterion (AIC) is employed as a measure of model fit for maximum likelihood models. *** p<0.01, ** p<0.05, * p<0.1

Appendix C.7. Consultation-level predictors of select care quality outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	IRT score	Correct/ partially correct diagnosis	Correct/ partially correct diagnosis	Antibiotic prescribed	Antibiotic prescribed	Antibiotic prescribed
Consultation length	0.06***	0.01	0.01	-0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
IRT score			0.05		-0.13**	-0.14**
			(0.05)		(0.06)	(0.06)
Correct / partially correct diagnosis						0.04
						(0.08)
Obs.	178	177	177	178	178	177
R-squared	0.31	0.12	0.13	0.11	0.15	0.15
Sample mean	0.00	0.53	0.53	0.57	0.57	0.57

Notes: All models are linear OLS, with robust standard errors clustered at the provider-level shown in parenthesis. All model specifications include fieldworker fixed effects.
*** p<0.01, ** p<0.05, * p<0.1

Appendix C.8. Effect of insurance cover on care quality outcomes – robustness checks

	(1) SP detection	(2) Order effects	(3) Cash consult.	(4) Extreme values
Consultation length				
High-insured (effect)	0.14	0.25	0.39	-0.45
	(0.82)	(0.80)	(0.81)	(0.66)
Obs.	174	178	178	176
R-squared	0.00	0.01	0.04	0.01
Checklist completion (share)				
High-insured (effect)	0.01	0.02	0.02	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Obs.	174	178	178	174
R-squared	0.01	0.06	0.01	0.01
Essential exams (share)				
High-insured (effect)	0.05*	0.05**	0.05**	0.05**
	(0.02)	(0.02)	(0.02)	(0.03)
Obs.	174	178	178	174
R-squared	0.04	0.05	0.05	0.05
Essential history-taking (share)				
High-insured (effect)	0.03	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)
Obs.	174	178	178	178
R-squared	0.02	0.14	0.02	0.02
IRT score				
High-insured (effect)	0.21*	0.22**	0.23**	
	(0.11)	(0.11)	(0.11)	
Obs.	174	178	178	
R-squared	0.04	0.08	0.05	
Correct or partially-correct diagnosis				
High-insured (effect)	-0.14*	-0.13*	-0.13*	
	(0.07)	(0.07)	(0.07)	
Obs.	173	177	177	
R-squared	0.04	0.06	0.04	
Antibiotic prescribed				
High-insured (effect)	0.13**	0.12**	0.12**	
	(0.06)	(0.05)	(0.06)	
Obs.	174	178	178	
R-squared	0.06	0.09	0.06	
Other inappropriate drugs prescribed				
High-insured (effect)	0.05	0.04	0.05	
	(0.03)	(0.03)	(0.03)	
Obs.	174	178	178	
R-squared	0.03	0.03	0.03	
Appropriate follow-up advice				
High-insured (effect)	0.23***	0.21***	0.22***	
	(0.07)	(0.07)	(0.07)	
Obs.	174	178	178	
R-squared	0.11	0.11	0.10	

Notes: Standard errors are in parenthesis. All specifications are fixed effects (mean-differenced) models, estimated with linear OLS. Column (1) excludes providers that were assessed as having likely valid SP detections. Column (2) controls for the order of SP visits to each provider. Column (3) controls for consultations that were paid for in cash on-site. Lastly, column (4) excludes extreme values of continuous outcomes, crudely defined as those above or below 3 standard deviations of the sample mean. *** p<0.01, ** p<0.05, * p<0.1

Appendix C.9. Sub-sample analyses of insurance effect on diagnostic effort

	(1) Consultation length	(2) Checklist completion	(3) Essential exams	(4) Essential history	(5) IRT score
Panel A: Altruism					
High-insured	2.76	0.03	0.07	0.03	0.31*
	(2.07)	(0.03)	(0.05)	(0.04)	(0.18)
High-altruism	0.74	0.02	0.01	0.03	0.11
	(1.13)	(0.03)	(0.05)	(0.03)	(0.18)
High-insured*high-altruism	-3.16	-0.02	-0.03	-0.00	-0.09
	(2.22)	(0.04)	(0.06)	(0.05)	(0.24)
Obs.	150	150	150	150	150
R-squared	0.09	0.07	0.09	0.20	0.12
Panel B: GP density					
High-insured	1.11	0.03	0.07*	0.04	0.29**
	(1.32)	(0.02)	(0.03)	(0.04)	(0.14)
High-density	0.33	-0.02	-0.03	-0.01	-0.17
	(1.01)	(0.03)	(0.05)	(0.03)	(0.18)
High-insured*high-density	-1.67	-0.03	-0.04	-0.02	-0.16
	(1.66)	(0.03)	(0.05)	(0.05)	(0.22)
Obs.	178	178	178	178	178
R-squared	0.08	0.09	0.09	0.20	0.14
Panel C: Northern suburbs					
High-insured	-0.08	0.02	0.03	0.03	0.21
	(1.06)	(0.02)	(0.04)	(0.04)	(0.15)
Northern suburb	0.80	0.02	0.03	-0.00	0.16
	(1.00)	(0.03)	(0.05)	(0.03)	(0.18)
High-insured*northern suburb	0.82	0.00	0.03	-0.01	-0.01
	(1.72)	(0.03)	(0.05)	(0.05)	(0.22)
Obs.	178	178	178	178	178
R-squared	0.09	0.08	0.09	0.19	0.12
Mean of low-insured	10.34	0.48	61.24	43.45	-0.11

Notes: Standard errors are in parenthesis. All models are linear OLS with provider-level cluster-robust standard errors. All models include fieldworker-pair fixed effects. Panel A presents estimates of an interaction effects analyses between insurance and GP altruism, where "high altruism" indicates GPs who contributed the full amount of their dictator game endowment to a patient charity (this was the sample median contribution). Panels B and C present estimates of interaction effects between insurance and measures of GPs' competitive pressures. "High density" indicates GPs with a number of competing GPs located within a1km radium of their practice that is at or above the sample median (7).
*** p<0.01, ** p<0.05, * p<0.1

Appendix C.10. Sub-sample analyses of insurance effect on diagnosis & treatment outputs

	(1) Correct / partially-correct diagnosis	(2) Antibiotic	(3) Other inappropriate drugs	(4) Appropriate follow-up advice
Panel A: Altruism				
High-insured	-0.14	0.17**	-0.00	0.18*
	(0.13)	(0.07)	(0.05)	(0.11)
High-altruism	0.01	-0.26**	0.01	-0.26**
	(0.12)	(0.13)	(0.06)	(0.11)
High-insured*high-altruism	-0.07	-0.13	0.05	0.05
	(0.16)	(0.12)	(0.06)	(0.15)
Obs.	150	150	150	150
R-squared	0.10	0.15	0.11	0.15
Panel B: GP density				
High-insured	-0.20*	0.23***	-0.00	0.30***
	(0.10)	(0.07)	(0.03)	(0.10)
High-density	-0.12	0.11	-0.01	0.06
	(0.12)	(0.11)	(0.06)	(0.11)
High-insured*high-density	0.11	-0.21*	0.09	-0.16
	(0.15)	(0.11)	(0.06)	(0.14)
Obs.	177	178	178	178
R-squared	0.07	0.07	0.10	0.10
Panel C: Northern suburbs				
High-insured	-0.19**	0.14*	-0.00	0.20*
	(0.09)	(0.07)	(0.03)	(0.10)
Northern suburb	-0.01	-0.05	-0.05	-0.01
	(0.11)	(0.11)	(0.06)	(0.11)
High-insured*northern suburb	0.11	-0.04	0.10*	0.04
	(0.15)	(0.12)	(0.06)	(0.14)
Obs.	177	178	178	178
R-squared	0.07	0.07	0.10	0.09
Mean of low-insured	0.60	0.51	0.91	0.51

Notes: Standard errors are in parenthesis. All models are linear OLS with provider-level cluster-robust standard errors. All models include fieldworker-pair fixed effects. Panel A presents estimates of an interaction effects analyses between insurance and GP altruism, where "high altruism" indicates GPs who contributed the full amount of their dictator game endowment to a patient charity (this was the sample median contribution). Panels B and C present estimates of interaction effects between insurance and measures of GPs' competitive pressures. "High density" indicates GPs with a number of competing GPs located within a1km radium of their practice that is at or above the sample median (7).
*** p<0.01, ** p<0.05, * p<0.1

Appendix D.1. Summary of prescription items from insured SP visits (study Part 1)

	Total items prescribed	Average cost per prescribed item (ZAR)	No. of prescribed items per consultation						Difference in means
			All patients		Low-insured		High-insured		
			Mean	SD	Mean	SD	Mean	SD	
All drugs	590	79.52	3.31	(1.20)	3.17	(1.23)	3.46	(1.17)	0.29***
Branded (originator)	228	127.91	1.28	(1.01)	1.13	(0.92)	1.43	(1.10)	0.29**
Generic	361	55.23	2.02	(1.14)	2.03	(1.20)	2.02	(1.09)	-0.01
Appropriate drugs	133	54.80	0.75	(0.76)	0.69	(0.73)	0.81	(0.78)	0.12
Analgesics	28	48.43	0.16	(0.37)	0.11	(0.32)	0.20	(0.40)	0.09**
Cough syrups	96	55.23	0.54	(0.55)	0.55	(0.54)	0.53	(0.57)	-0.02
Throat preparations	9	71.89	0.05	(0.24)	0.02	(0.15)	0.08	(0.31)	0.06*
Inappropriate drugs	457	88.22	2.57	(1.06)	2.48	(1.06)	2.65	(1.06)	0.17*
Antibiotics	102	147.9	0.57	(0.51)	0.51	(0.50)	0.64	(0.51)	0.13**
Penicillins / Penicillin-clavulanates	51	122.03	0.28	(0.45)	0.24	(0.43)	0.33	(0.47)	0.09
Cephalosporins	6	224.33	0.03	(0.18)	0.03	(0.18)	0.03	(0.18)	0.00
Sulphonamides	1	120.83	0.01	(0.07)	0.00	(0.00)	0.01	(0.11)	0.01
Macrolides	38	174.28	0.21	(0.41)	0.21	(0.41)	0.21	(0.41)	0.00
Floruoquinolones	4	190.73	0.02	(0.15)	0.01	(0.11)	0.03	(0.18)	0.02
Tetracyclines	2	4.94	0.01	(0.11)	0.01	(0.11)	0.01	(0.11)	0.00
Nasal decongestants	91	98.61	0.51	(0.66)	0.51	(0.64)	0.52	(0.68)	0.01
Bronchodilators	56	79.16	0.31	(0.55)	0.31	(0.58)	0.31	(0.54)	0.00
Antihistamines	55	49.99	0.31	(0.46)	0.34	(0.48)	0.28	(0.45)	-0.06
Steroids	111	45.59	0.62	(0.53)	0.61	(0.54)	0.64	(0.53)	0.03
Probiotics	28	96.49	0.16	(0.38)	0.15	(0.36)	0.17	(0.41)	0.02
Other	14	113.64	0.07	(0.26)	0.07	(0.25)	0.09	(0.32)	0.02
Observations (no. consultations)	178								

Notes: Column (9) - the difference in means - presents results from a paired sample t-test of differences in the mean of each outcome between the two patient insurance types. *** p<0.01, ** p<0.05, * p<0.1

Appendix D.2. Consultation-level predictors of drug cost outcomes – all drugs

	(1) No. of drugs	(2) Average drug cost	(3) Total drug costs
Consultation length	0.01	0.26	0.05
	(0.02)	(0.71)	(2.41)
IRT score	-0.21	4.45	21.02
	(0.13)	(3.84)	(15.36)
Correct / partially correct diagnosis	0.02	-6.50	-26.04
	(0.21)	(5.90)	(25.76)
Antibiotic prescribed		18.63***	136.70***
		(7.00)	(28.69)
Obs.	177	177	177
R-squared	0.13	0.16	0.24
Sample mean	3.31	79.52	268.46

Notes: This analysis is with the sub-sample of 177 of a total 178 consultations where a diagnosis was given (which allows inclusion of the diagnosis variable in the multivariate regression analysis). All estimates are from multivariate regressions with linear OLS and provider-level cluster-robust standard errors (shown in parenthesis). All model specifications include fieldworker fixed effects. Costs are in South African Rand (ZAR).
*** p<0.01, ** p<0.05, * p<0.1

Appendix D.3. Consultation-level predictors of drug cost outcomes – unnecessary drugs

	(1) No. of inapprop. drugs	(2) Average inapprop. drug cost	(3) Total inapprop. drug costs	(4) Any branded drugs	(5) No. of branded drugs	(6) Average branded drug cost	(7) Total branded drug costs
Consultation length	0.02	0.36	0.52	-0.00	-0.02	0.71	-1.38
	(0.02)	(0.89)	(2.54)	(0.01)	(0.02)	(1.87)	(1.64)
IRT score	-0.05	2.33	29.10*	-0.04	-0.06	17.73**	12.45
	(0.13)	(5.53)	(15.20)	(0.05)	(0.11)	(8.28)	(15.17)
Correct / partially correct diagnosis	0.07	-6.94	-24.21	-0.07	-0.13	-0.66	-18.29
	(0.18)	(8.15)	(23.40)	(0.07)	(0.17)	(12.61)	(25.72)
Antibiotic prescribed		20.13**	138.34***	-0.01	0.22	42.63***	78.03***
		(9.86)	(27.19)	(0.07)	(0.19)	(12.25)	(26.10)
Obs.	177	176	177	177	177	136	177
R-squared	0.08	0.14	0.25	0.06	0.09	0.19	0.09
Sample mean	2.57	88.22	227.42	0.76	1.28	127.91	159

Notes: All estimates are from multivariate regressions with linear OLS and provider-level cluster-robust standard errors (shown in parenthesis). This analysis is with the sub-sample of 177 of a total 178 consultations where a diagnosis was given (which allows inclusion of the diagnosis variable in the multivariate regression analysis). All model specifications include fieldworker fixed effects. Costs are in South African Rand (ZAR).
*** p<0.01, ** p<0.05, * p<0.1

Appendix D.4. Consultation-level predictors of fee-for-service cost outcomes

	(1) Any FFS tests / procedures	(2) No. FFS tests / procedures	(3) Total consultation costs
Consultation length	0.02***	0.04***	-2.22
	(0.00)	(0.01)	(2.21)
IRT score	-0.03	-0.04	33.25***
	(0.02)	(0.03)	(12.55)
Correct / partially correct diagnosis	0.11***	0.16**	-25.37**
	(0.04)	(0.07)	(11.88)
Antibiotic prescribed	-0.02	-0.00	2.95
	(0.05)	(0.08)	(13.98)
Any FFS tests / procedures			198.48***
			(57.50)
Obs.	177	177	177
R-squared	0.23	0.28	0.30
Sample mean	0.08	0.12	469.58

Notes: This analysis is with the sub-sample of 155 consultations where a diagnosis was given (which allows inclusion of the diagnosis variable in the multivariate regression analysis). All estimates are from multivariate regressions with linear OLS and provider-level cluster-robust standard errors (shown in parenthesis). All model specifications include fieldworker fixed effects. Costs are in South African Rand (ZAR). FFS = fee-for-service. In column (3) the dependent variable "total consultation costs" excludes all drug costs (but includes the fee-for-service costs of any tests / procedures ordered).
*** p<0.01, ** p<0.05, * p<0.1

Appendix D.5. Determinants of non-participation in the GP interviews – care quantity and cost outcomes

	Attrition (1=Yes)
No. of drugs prescribed	-0.02
	(0.02)
Average prescribed drug cost	0.00*
	(0.00)
Total drug costs (total prescription cost)	0.00
	(0.00)
No. of inappropriate drugs prescribed	-0.01
	(0.03)
Average inappropriate drug cost	0.00
	(0.00)
Total inappropriate drug costs	0.00
	(0.00)
Likelihood of a branded drug prescription	0.05
	(0.06)
No. of branded drugs prescribed	0.01
	(0.03)
Average branded drug cost	0.00
	(0.00)
Total branded drug costs	0.00
	(0.00)
Any FFS tests / procedures	-0.03
	(0.10)
No. of FFS tests / procedures	-0.04
	(0.06)
Total consultation cost (excl. drugs)	-0.00
	(0.00)
Obs.	178

Notes: FFS = fee-for-service. Column 2 shows estimates from univariate linear OLS regressions of provider attrition on consultation outcomes of interest. All drug outcomes are on a per prescription basis.

*** p<0.01, ** p<0.05, * p<0.1

Appendix D.6. Sub-sample analyses of insurance effect on all drug expenditures

	(1) No. drugs	(2) Average drug cost	(3) Total drug cost
Panel A: Altruism			
High-insured	0.35*	14.66**	42.77
	(0.19)	(7.19)	(31.74)
High-altruism	-0.22	0.77	-39.82
	(0.34)	(9.34)	(45.91)
High-insured*high-altruism	-0.09	-6.12	10.93
	(0.23)	(9.49)	(38.81)
Obs.	150	150	150
R-squared	0.14	0.20	0.14
Panel B: GP density			
High-insured	0.27**	13.88**	48.89*
	(0.12)	(6.40)	(24.77)
High-density	0.24	17.33*	58.79
	(0.27)	(8.88)	(39.04)
High-insured*high-density	0.01	-10.10	-14.18
	(0.20)	(9.27)	(33.57)
Obs.	178	178	178
R-squared	0.16	0.18	0.13
Panel C: Northern suburbs			
High-insured	0.12	12.61**	38.18*
	(0.13)	(6.01)	(21.70)
Northern suburb	-0.16	16.99*	31.41
	(0.28)	(9.89)	(43.30)
High-insured*northern suburb	0.37*	-9.01	8.22
	(0.20)	(9.40)	(33.60)
Obs.	178	178	178
R-squared	0.16	0.18	0.12
Mean of low-insured	3.17	75.25	247.36

Notes: Standard errors are in parenthesis. All costs are in South African Rand (ZAR). All models are linear OLS with provider-level cluster-robust standard errors and fieldworker-pair fixed effects. Panel A presents estimates of interaction effects between insurance and GP altruism, where "high altruism" indicates GPs who contributed the full amount of their dictator game endowment to a patient charity (this was the sample median contribution). Panels B and C present estimates of interaction effects between insurance and measures of GPs' competitive pressures. "High density" indicates GPs with a number of competing GPs located within a 1km radium of their practice that is at or above the sample median (7). *** p<0.01, ** p<0.05, * p<0.1

Appendix D.7. Sub-sample analyses of insurance effect on unnecessary drug expenditures

	(1) No. of inapprop. drugs	(2) Average inapprop. drug cost	(3) Total inapprop. drug costs	(4) Any branded drugs	(5) No. of branded drugs	(6) Average branded drug cost	(7) Total branded drug costs
Panel A: Altruism							
High-insured	0.14	16.11*	30.26	0.18*	0.32	-26.85	3.56
	(0.18)	(8.75)	(32.77)	(0.11)	(0.19)	(20.38)	(29.84)
High-altruism	-0.20	8.02	-26.55	0.21*	-0.10	-11.34	-25.65
	(0.26)	(11.80)	(42.15)	(0.11)	(0.27)	(22.06)	(38.22)
High-insured*high-altruism	-0.05	-2.10	12.94	-0.18	-0.04	51.40**	55.03
	(0.22)	(11.74)	(39.35)	(0.13)	(0.26)	(25.40)	(39.45)
Obs.	150	150	150	150	150	113	150
R-squared	0.14	0.15	0.15	0.07	0.05	0.18	0.07
Panel B: GP density							
High-insured	0.16	15.90*	36.28	0.05	0.27*	1.55	28.93
	(0.15)	(8.45)	(25.21)	(0.08)	(0.16)	(17.31)	(28.64)
High-density	0.24	15.65	48.51	-0.02	0.12	3.23	16.40
	(0.23)	(11.21)	(36.93)	(0.09)	(0.21)	(18.78)	(32.09)
High-insured*high-density	-0.01	-10.60	-7.22	0.05	0.04	-6.03	1.98
	(0.20)	(11.28)	(34.10)	(0.11)	(0.23)	(22.72)	(36.41)
Obs.	178	177	178	178	178	136	178
R-squared	0.15	0.13	0.14	0.06	0.07	0.11	0.04
Panel C: Northern suburbs							
High-insured	-0.02	16.39**	26.83	0.06	0.41***	-11.38	40.59*
	(0.13)	(7.55)	(21.32)	(0.07)	(0.16)	(11.87)	(23.69)
Northern suburb	-0.18	22.15*	29.29	-0.03	0.21	-6.80	24.63
	(0.23)	(11.98)	(40.90)	(0.10)	(0.22)	(18.92)	(35.81)
High-insured*northern suburb	0.41**	-13.71	13.50	0.02	-0.28	22.09	-25.01
	(0.19)	(11.19)	(34.47)	(0.12)	(0.23)	(23.40)	(36.04)
Obs.	178	177	178	178	178	136	178
R-squared	0.14	0.14	0.13	0.06	0.07	0.11	0.04
Mean of low-insured	2.48	82.99	210.85	0.73	1.13	128.54	144.09

Notes: Standard errors are in parenthesis. All costs are in South African Rand (ZAR). All models are linear OLS with provider-level cluster-robust standard errors and fieldworker-pair fixed effects. Panel A presents estimates of interaction effects between insurance and GP altruism, where "high altruism" indicates GPs who contributed the full amount of their dictator game endowment to a patient charity (this was the sample median contribution). Panels B and C present estimates of interaction effects between insurance and measures of GPs' competitive pressures. "High density" indicates GPs with a number of competing GPs located within a 1km radium of their practice that is at or above the sample median (7). *** p<0.01, ** p<0.05, * p<0.1

Appendix D.8. Sub-sample analyses of insurance effect on fee-for-service expenditures

	(1)	(2)	(3)
	Any FFS tests / procedures	No. FFS tests / procedures	Total consultation cost (excl. drugs)
Panel A: Altruism			
High-insured	0.07	0.22	47.61**
	(0.05)	(0.14)	(21.58)
High-altruism	0.04	0.07	11.79
	(0.06)	(0.08)	(21.22)
High-insured*high-altruism	-0.01	-0.16	-20.19
	(0.08)	(0.15)	(34.19)
Obs.	150	150	150
R-squared	0.07	0.09	0.08
Panel B: GP density			
High-insured	0.02	0.11	31.87**
	(0.05)	(0.10)	(15.26)
High-density	-0.00	0.04	16.71
	(0.05)	(0.06)	(17.47)
High-insured*high-density	0.07	-0.02	-4.63
	(0.07)	(0.10)	(30.80)
Obs.	178	178	178
R-squared	0.07	0.08	0.07
Panel C: Northern suburbs			
High-insured	0.02	0.04	22.04*
	(0.05)	(0.06)	(11.75)
Northern suburb	0.02	0.07	38.84*
	(0.06)	(0.08)	(19.81)
High-insured*northern suburb	0.09	0.15	17.56
	(0.07)	(0.11)	(34.76)
Obs.	178	178	178
R-squared	0.08	0.11	0.11
Mean of low-insured	0.06	0.07	454.62

Notes: Standard errors are in parenthesis. All costs are in South African Rand (ZAR). All models are linear OLS with provider-level cluster-robust standard errors and fieldworker-pair fixed effects. Panel A presents estimates of interaction effects between insurance and GP altruism, where "high altruism" indicates GPs who contributed the full amount of their dictator game endowment to a patient charity (this was the sample median contribution). Panels B and C present estimates of interaction effects between insurance and measures of GPs' competitive pressures. "High density" indicates GPs with a number of competing GPs located within a 1km radium of their practice that is at or above the sample median (7). *** p<0.01, ** p<0.05, * p<0.1

Appendix D.9. Paired-sample t-test of differences in care quantity and cost outcomes between low-insured and high-insured patient consultations

	(1) Low-insured patient		(2) High-insured patient		Difference
	Mean	SD	Mean	SD	
d. All prescription items					
Number of drugs prescribed	3.17	(1.23)	3.46	(1.17)	0.29***
Average cost of prescribed drugs (ZAR)	75.25	(45.26)	83.79	(36.59)	8.54*
Total cost of drugs prescribed (ZAR)	247.36	(177.87)	289.57	(153.26)	42.21***
b. Inappropriate prescription items					
Number of inappropriate drugs	2.48	(1.06)	2.65	(1.06)	0.17*
Average cost of inappropriate drugs (ZAR)	82.99	(55.55)	93.07	(48.79)	10.09*
Total cost of inappropriate drugs (ZAR)	210.85	(171.20)	243.99	(151.50)	33.14**
c. Branded prescription items					
Likelihood of branded drug prescription	0.73	(0.45)	0.8	(0.40)	0.07
Number of branded drugs	1.14	(0.92)	1.43	(1.09)	0.29**
Average cost of branded drugs (ZAR)	130.59	(75.16)	130.57	(70.24)	-0.02
Total cost of branded drugs (ZAR)	144.09	(147.02)	173.91	(150.04)	29.81*
e. Fee-for-service items					
Additional fee-for-service tests / procedures billed	0.06	(0.23)	0.11	(0.32)	0.06*
Number of additional fee-for-service tests / procedures billed	0.07	(0.29)	0.17	(0.53)	0.10**
Total cost of consultation, excl. drugs (ZAR)	454.62	(78.37)	484.54	(138.71)	29.91**
Observations (no. consultations)	89		89		

*** p<0.01, ** p<0.05, * p<0.1

Appendix D.10. Effect of insurance cover on treatment intensity and care costs – all drugs (mixed effects model specifications)

	(1) No. drugs	(2) Average drug cost	(3) Total drug cost
Panel A: provider-level random effects (RE)			
High-insured	0.29***	8.54*	42.21***
	(0.10)	(4.49)	(16.08)
Obs.	178	178	178
AIC	516.17	1814.85	2296.91
Panel B: provider-level RE; fieldworker-pair fixed effects			
High-insured	0.28***	8.77*	41.59***
	(0.10)	(4.48)	(16.06)
Obs.	178	178	178
AIC	515.68	1812.81	2303.14
Mean of low-insured	3.17	75.25	247.36

Notes: Standard errors are in parenthesis. All costs are in South African Rand (ZAR). In Panel A, estimates from a linear mixed effects model with provider-level random effects (intercept and slope with respect to insurance) are shown. In Panel D, results from the same mixed effects model with the addition of fieldworker-pair fixed effects are shown. The Akaike Information Criterion (AIC) is employed as a measure of model fit for maximum likelihood models. All outcomes are per prescription (per consultation). *** p<0.01, ** p<0.05, * p<0.1

Appendix D.11. Effect of insurance cover on treatment intensity and care costs – unnecessary drugs (mixed effects model specifications)

	(1) No. inapprop .drugs	(2) Average inapprop drug cost	(3) Total inapprop drug cost	(4) Any branded drug	(5) No. branded drugs	(6) Average branded drug cost	(7) Total branded drug cost
Panel A: provider-level random effects (RE)							
High-insured	0.17*	10.24*	33.14**	0.07	0.29***	-0.83	29.81*
	(0.10)	(5.46)	(16.38)	(0.05)	(0.11)	(10.48)	(17.34)
Obs.	178	177	178	178	178	136	178
AIC	491.14	1883.82	2293.18	202.83	491.13	1553.32	2278.95
Panel B: provider-level RE; fieldworker-pair fixed effects							
High-insured	0.15	10.39*	32.36**	0.07	0.29***	-1.15	29.90*
	(0.09)	(5.45)	(16.27)	(0.05)	(0.11)	(10.40)	(17.35)
Obs.	178	177	178	178	178	136	178
AIC	488.60	1887.63	2296.59	213.69	503.23	1558.10	2292.84
Mean of low-insured	2.48	82.99	210.85	0.73	1.13	128.54	144.09

Notes: Standard errors are in parenthesis. All costs are in South African Rand (ZAR). In Panel A, estimates from a linear mixed effects model with provider-level random effects (intercept and slope with respect to insurance) are shown. In Panel D, results from the same mixed effects model with the addition of fieldworker-pair fixed effects are shown. The Akaike Information Criterion (AIC) is employed as a measure of model fit for maximum likelihood models. All outcomes are per prescription (per consultation). *** p<0.01, ** p<0.05, * p<0.1

Appendix D.12. Effect of insurance cover on treatment intensity and care costs – fee-for-service items (mixed effects model specifications)

	(1)	(2)	(3)
	Any FFS tests / procedures	No. FFS tests / procedures	Total consultation cost (excl. drugs)
Panel A: provider-level random effects (RE)			
High-insured	0.06*	0.10**	29.91**
	(0.03)	(0.05)	(14.77)
Obs.	178	178	178
AIC	34.96	162.12	2161
Panel B: provider-level RE; fieldworker-pair fixed effects			
High-insured	0.06*	0.10**	29.74**
	(0.03)	(0.05)	(14.78)
Obs.	178	178	178
AIC	46.45	174.08	2174.04
Mean of low-insured	0.06	0.07	454.62

Notes: Standard errors are in parenthesis. FFS = fee-for-service. All costs are in South African Rand (ZAR). In Panel A, estimates from a linear mixed effects model with provider-level random effects (intercept and slope with respect to insurance) are shown. In Panel D, results from the same mixed effects model with the addition of fieldworker-pair fixed effects are shown. The Akaike Information Criterion (AIC) is employed as a measure of model fit for maximum likelihood models. All outcomes are per prescription (per consultation). *** p<0.01, ** p<0.05, * p<0.1

Appendix D.13. Effect of insurance cover on treatment intensity and care costs – robustness checks

	(1) SP detection	(2) Order effects	(3) Cash consult.	(4) Extreme values
No. of drugs prescribed				
High-insured (effect)	0.29***	0.29***	0.29***	0.30***
	(0.10)	(0.10)	(0.10)	(0.10)
Obs.	174	178	178	176
R-squared	0.09	0.10	0.09	0.09
Average prescribed drug cost				
High-insured (effect)	7.31	8.53*	8.68*	10.48**
	(4.49)	(4.54)	(4.56)	(4.12)
Obs.	174	178	178	177
R-squared	0.03	0.04	0.04	0.07
Total drug cost (total prescription cost)				
High-insured (effect)	40.78**	42.10**	42.51**	53.21***
	(16.52)	(16.23)	(16.36)	(14.49)
Obs.	174	178	178	176
R-squared	0.07	0.08	0.07	0.14
No. of inappropriate drugs prescribed				
High-insured (effect)	0.16	0.17*	0.16	0.14
	(0.10)	(0.10)	(0.10)	(0.09)
Obs.	174	178	178	177
R-squared	0.03	0.05	0.03	0.02
Average inappropriate drug cost				
High-insured (effect)	8.88	10.00*	10.26*	11.36**
	(5.53)	(5.52)	(5.56)	(5.21)
Obs.	173	177	177	175
R-squared	0.03	0.04	0.04	0.05
Total inappropriate drug cost				
High-insured (effect)	31.51*	33.10**	33.53**	44.10***
	(16.82)	(16.57)	(16.66)	(14.85)
Obs.	174	178	178	176
R-squared	0.04	0.04	0.04	0.09
Likelihood of branded drug prescription				
High-insured (effect)	0.05	0.07	0.07	
	(0.05)	(0.06)	(0.06)	
Obs.	174	178	178	
R-squared	0.01	0.02	0.02	
No. of branded drugs prescribed				
High-insured (effect)	0.25**	0.29***	0.28**	0.29***
	(0.11)	(0.11)	(0.11)	(0.11)
Obs.	174	178	178	178
R-squared	0.06	0.10	0.09	0.07
Average branded drug cost				
High-insured (effect)	-0.02	-0.22	1.44	-6.74
	(11.34)	(11.46)	(11.45)	(9.30)
Obs.	134	136	136	135
R-squared	0.00	0.00	0.02	0.01

Appendix D.13. Effect of insurance cover on treatment intensity and care costs – robustness checks (cont.)

	(1) SP detection	(2) Order effects	(3) Cash consult.	(4) Extreme values
Total branded drug cost				
High-insured (effect)	27.02	29.66*	30.35*	32.96**
	(17.72)	(17.47)	(17.63)	(16.33)
Obs.	174	178	178	176
R-squared	0.03	0.04	0.03	0.05
Likelihood of any fee-for-service procedures				
High-insured (effect)	0.06*	0.06*	0.06*	
	(0.03)	(0.03)	(0.03)	
Obs.	174	178	178	
R-squared	0.03	0.03	0.03	
No. of fee-for-service procedures				
High-insured (effect)	0.10**	0.10**	0.10**	
	(0.05)	(0.05)	(0.05)	
Obs.	174	178	178	
R-squared	0.04	0.04	0.04	
Total consultation cost (excl. drugs)				
High-insured (effect)	30.60**	29.88**	22.03*	16.92**
	(15.19)	(14.94)	(12.74)	(6.67)
Obs.	174	178	178	174
R-squared	0.05	0.04	0.31	0.07

Notes: Standard errors are in parenthesis. All specifications are fixed effects (mean-differenced) models, estimated with linear OLS. Column (1) excludes providers that were assessed as having likely valid SP detections. Column (2) controls for the order of SP visits to each provider. Column (3) controls for consultations that were paid for in cash on-site. Lastly, column (4) excludes extreme values of continuous outcomes, crudely defined as those above or below 3 standard deviations of the sample mean. All costs are in South African Rand (ZAR).
*** p<0.01, ** p<0.05, * p<0.1

Appendix E.1. IRT analysis of clinical effort during baseline and end-line visits (study Part 2) – Clinical Case 1

Table A (a). Checklist items included in the IRT score at baseline

	Item type	Item description	Item complet. rate	Discrim. parameter	Standard error	Difficulty parameter	Standard error
1	History-taking	Cough duration	0.96	0.94	0.60	-3.79	1.96
2	History-taking	Productive cough / coughing up blood	0.83	0.96	0.38	-1.92	0.66
3	Examination	Examine throat	0.82	0.94	0.38	-1.86	0.64
4	Examination	Listen to lungs	0.88	1.60	0.60	-1.73	0.43
5	Examination	Listen to heart	0.64	0.49	0.26	-1.20	0.73
6	History-taking	Allergies (general) / penicillin allergy	0.71	1.21	0.41	-0.92	0.31
7	Examination	Measure blood pressure	0.61	0.56	0.27	-0.82	0.52
8	History-taking	Temperature / fever	0.67	1.06	0.35	-0.80	0.31
9	History-taking	Sore throat	0.65	0.94	0.33	-0.75	0.33
10	History-taking	Any regular medication	0.55	0.76	0.30	-0.27	0.31
11	History-taking	Occupation status	0.54	0.66	0.28	-0.23	0.34
12	Examination	Examine ears	0.52	0.94	0.33	-0.07	0.25
13	History-taking	Details about the initial cold	0.51	0.72	0.29	-0.03	0.31
14	Examination	Measure pulse rate	0.49	0.73	0.29	0.04	0.31
15	History-taking	Chest pain	0.47	1.15	0.37	0.12	0.22
16	Examination	Palpate lymph nodes	0.45	0.91	0.31	0.24	0.27
17	History-taking	Smoking status	0.44	1.21	0.37	0.24	0.22
18	History-taking	Asthma history	0.27	1.66	0.55	0.87	0.24
19	History-taking	Exposure to anyone with TB	0.24	1.58	0.51	1.03	0.27
20	Examination	Take temperature	0.40	0.36	0.25	1.12	0.94
21	History-taking	Weight loss	0.24	1.19	0.42	1.21	0.37
22	History-taking	Night sweats	0.34	0.57	0.29	1.22	0.66
23	History-taking	TB history	0.19	0.97	0.39	1.75	0.60
24	History-taking	Ear ache	0.13	1.37	0.53	1.80	0.50
25	History-taking	Breathing difficulty	0.16	0.76	0.40	2.41	1.10
26	History-taking	Similar problem before	0.10	1.05	0.49	2.46	0.91
27	History-taking	Whistling noise when breathing	0.08	0.70	0.48	3.77	2.29
28	Examination	Tap lungs (percuss)	0.11	0.57	0.40	3.87	2.53
29	History-taking	Breathing fast	0.05	0.42	0.56	7.13	9.10
30	History-taking	HIV status	0.14	0.09	0.33	19.29	68.31
31	Examination	Check oxygen saturation	0.06	0.12	0.49	23.19	95.30
Obs		80					

Notes: The discrimination and difficulty parameters (and their standard errors) were estimated using a three parameter logistic model, which also allows for the possibility of item completion through guessing. The estimated 'pseudo-guessing' parameter was close to zero and insigificant, however, suggesting that the likelihood that doctors completed items on this checklist purely through guessing (and zero effort) is close to zero. Item completion rate is the share of all consultations in which the specific item was completed. The difficulty parameter is a measure of the likelihood that a specific item is successfully completed in any given consultation, with higher parameter values indicating more difficult items. The discrimination parameter indicates how well an item can distinguish between high- and low-effort providers, with high-discrimination items having a higher correlation between their completion and provider effort.

Table A (b). Checklist items that were either excluded or merged in the IRT score at baseline

	Item type	Item description	Item completion rate
1	History-taking	Productive/dry cough	0.83
2	History-taking	Coughing up blood	0.09
3	History-taking	Heart beating fast / palpitations	0.02
4	History-taking	Allergy to penicillin	0.19
5	History-taking	Any allergies (in general)	0.66

Notes: Items 1 and 2 were merged into one item in the IRT analysis, as they are likely to overlap and potentially violate the conditional independence assumption necessary for the IRT score validity (see Appendix B.6). Items 4 and 5 were also merged for the same reason. Item 3 was excluded as it was not completed in any consultation.

Table B (a). Checklist items included in the IRT score at end-line

	Item type	Item description	Item complet. rate	Discrim. parameter	Standard error	Difficulty parameter	Standard error
1	Examination	Listen to lungs	0.94	0.93	0.55	-3.30	1.59
2	Examination	Listen to heart	0.64	0.25	0.27	-2.27	2.53
3	History-taking	Productive cough / coughing up blood	0.71	0.49	0.31	-1.96	1.24
4	History-taking	Cough duration	0.8	0.88	0.38	-1.81	0.70
5	Examination	Examine throat	0.9	2.54	1.17	-1.58	0.35
6	Examination	Measure blood pressure	0.74	0.85	0.36	-1.39	0.57
7	History-taking	Details about the initial cold	0.68	0.67	0.31	-1.19	0.60
8	History-taking	Temperature / fever	0.6	0.70	0.31	-0.64	0.43
9	History-taking	Allergies (general) / penicillin allergy	0.68	1.95	0.63	-0.58	0.22
10	Examination	Examine ears	0.64	1.34	0.48	-0.55	0.26
11	History-taking	Occupation status	0.56	0.81	0.33	-0.35	0.34
12	History-taking	Sore throat	0.53	0.46	0.28	-0.22	0.53
13	Examination	Measure pulse rate	0.54	1.04	0.39	-0.16	0.27
14	History-taking	Any regular medication	0.51	1.19	0.41	-0.04	0.24
15	History-taking	Smoking status	0.5	0.95	0.37	0.01	0.28
16	Examination	Take temperature	0.49	0.87	0.34	0.08	0.30
17	Examination	Palpate lymph nodes	0.48	1.43	0.49	0.11	0.22
18	History-taking	Asthma history	0.39	1.57	0.53	0.43	0.22
19	History-taking	Chest pain	0.44	0.52	0.29	0.51	0.52
20	History-taking	Night sweats	0.38	0.51	0.30	1.07	0.75
21	History-taking	Breathing difficulty	0.31	0.59	0.31	1.45	0.80
22	History-taking	Ear ache	0.13	1.73	0.80	1.62	0.46
23	History-taking	Exposure to anyone with TB	0.13	1.62	0.69	1.68	0.47
24	History-taking	Whistling noise when breathing	0.14	1.29	0.68	1.81	0.67
25	History-taking	Similar problem before	0.21	0.75	0.38	1.94	0.89
26	History-taking	Breathing fast	0.11	1.19	0.65	2.13	0.87
27	Examination	Check oxygen saturation	0.05	1.76	1.05	2.32	0.82
28	History-taking	Heart beating fast / palpitations	0.06	1.33	0.85	2.54	1.14
29	History-taking	Weight loss	0.26	0.32	0.32	3.26	3.13
30	Examination	Tap lungs (percuss)	0.14	0.58	0.41	3.37	2.20
31	History-taking	TB history	0.16	0.45	0.39	3.79	3.10
32	History-taking	HIV status	0.11	0.56	0.46	3.92	3.00
Obs		80					

Notes: The discrimination and difficulty parameters (and their standard errors) were estimated using a three parameter logistic model, which also allows for the possibility of item completion through guessing. The estimated 'pseudo-guessing' parameter was close to zero and insigificant, however, suggesting that the likelihood that doctors completed items on this checklist purely through guessing (and zero effort) is close to zero. Item completion rate is the share of all consultations in which the specific item was completed. The difficulty parameter is a measure of the likelihood that a specific item is successfully completed in any given consultation, with higher parameter values indicating more difficult items. The discrimination parameter indicates how well an item can distinguish between high- and low-effort providers, with high-discrimination items having a higher correlation between their completion and provider effort.

Table B (b). Checklist items that were either excluded or merged in the IRT score at end-line

	Item type	Item description	Item completion rate
1	History-taking	Productive/dry cough	0.66
2	History-taking	Coughing up blood	0.19
3	History-taking	Allergy to penicillin	0.22
4	History-taking	Any allergies (in general)	0.68

Notes: Items 1 and 2 were merged into one item in the IRT analysis, as they are likely to overlap and potentially violate the conditional independence assumption necessary for the IRT score validity (see Appendix B.6). Items 3 and 4 were also merged for the same reason.

Appendix E.2. Effects of private performance feedback – robustness checks

	(1)	(2)		(3)	(4)	(5)	(6)
	ITT	Manski-Horowitz bounds		Order effects	Hour-of-day effects	No. days since interven.	Additional covariates
		Lower (a)	Upper (b)				
Antibiotic prescription (1 = Yes)							
Treatment (1 = feedback given)	-0.14	-0.03	-0.32***	-0.23**	-0.24**	-0.23**	-0.21**
	(0.09)	(0.10)	(0.10)	(0.10)	(0.11)	(0.11)	(0.10)
Obs.	98	98	98	79	79	79	79
R-squared	0.52	0.43	0.46	0.56	0.55	0.55	0.73
Total drug costs, per consultation (ZAR)							
Treatment (1 = feedback given)	-33.82			-41.06	-41.44	-42.69	-25.32
	(25.88)			(25.68)	(25.51)	(25.92)	(29.31)
Obs.	98			79	79	79	79
R-squared	0.57			0.48	0.48	0.48	0.57
No. of drugs, per consultation							
Treatment (1 = feedback given)	-0.13			-0.41	-0.41	-0.43*	-0.40
	(0.27)			(0.25)	(0.25)	(0.26)	(0.28)
Obs.	98			79	79	79	79
R-squared	0.33			0.58	0.58	0.58	0.66
Average drug cost, per consultation (ZAR)							
Treatment (1 = feedback given)	-10.02			-8.52	-8.75	-8.54	-3.70
	(6.83)			(7.27)	(7.19)	(7.35)	(7.84)
Obs.	98			79	79	79	79
R-squared	0.48			0.56	0.57	0.56	0.68
Antibiotic prescription for HIV+ case (1 = Yes)							
Treatment (1 = feedback given)				0.05	0.04	0.03	0.05
				(0.12)	(0.12)	(0.12)	(0.13)
Obs.				79	79	79	79
R-squared				0.38	0.37	0.38	0.53
Steroid prescription (1 = Yes)							
Treatment (1 = feedback given)	-0.10			-0.16	-0.16	-0.17	-0.20*
	(0.09)			(0.11)	(0.11)	(0.11)	(0.11)
Obs.	98			79	79	79	79
R-squared	0.51			0.47	0.47	0.48	0.69
Other inappropriate drug prescription (1 = Yes)							
Treatment (1 = feedback given)	0.06			0.08	0.08	0.08	0.06
	(0.10)			(0.12)	(0.12)	(0.12)	(0.13)
Obs.	98			79	79	79	79
R-squared	0.34			0.28	0.32	0.28	0.48
Consultation time (mins.)							
Treatment (1 = feedback given)	1.06			1.78	1.84	1.96	1.60
	(1.40)			(1.71)	(1.71)	(1.72)	(1.90)
Obs.	98			80	80	80	80
R-squared	0.32			0.37	0.36	0.36	0.50

Notes: Standard errors are in parenthesis. All specifications are linear OLS with fieldworker fixed effects. Column (1) shows results from an ITT analysis with the initially randomized sample of 99 GPs, where missing outcomes are imputed with baseline values for the 19 GPs lost to non-participation or attrition. Column (2) estimates upper and lower bounds for the true effect (in the absence of any attrition) using the Manski-Horowitz method. The specifications in (1) and (2) do not control for all imbalanced covariates, due to the limited covariate data available for those 19 GPs that did not participate in the interviews. Column (3) controls for the order of SP visits to each provider (as each provider received two SP visits in this experiment, in a random order). Column (4) controls for the hour of day in which each SP visit occurred. Column (5) controls for the number of days between intervention delivery (in the GP interviews) and impact assessment (in the corresponding end-line SP visit). Column (6) controls for additional provider covariates – namely, all listed variables in Table 7.2, excluding the two variables on perceived patient expectations for antibiotics in the uncomplicated bronchitis case and perceived antibiotic prescribing norms for the complicated (HIV+) bronchitis case, which were highly correlated with some of the unbalanced covariates included in the main specification. *** p<0.01, ** p<0.05, * p<0.1

Appendix E.2. Effects of private performance feedback – robustness checks (cont.)

	(1)	(2)		(3)	(4)	(5)	(6)
	ITT	Manski-Horowitz bounds		Order effects	Hour-of-day effects	No. days since interven.	Additional covariates
		Lower (a)	Upper (b)				
Checklist completion (share)							
Treatment (1 =feedback given)	-0.04			-0.02	-0.02	-0.02	-0.02
	(0.03)			(0.03)	(0.03)	(0.03)	(0.04)
Obs.	98			80	80	80	80
R-squared	0.54			0.53	0.53	0.54	0.66
IRT score							
Treatment (1 = feedback given)	-0.29*			-0.18	-0.19	-0.20	-0.23
	(0.15)			(0.18)	(0.18)	(0.19)	(0.19)
Obs.	98			80	80	80	80
R-squared	0.60			0.59	0.59	0.59	0.73
Correct / partially-correct diagnosis							
Treatment (1 = feedback given)	-0.05			0.01	0.01	0.02	-0.13
	(0.11)			(0.14)	(0.13)	(0.14)	(0.14)
Obs.	93			75	75	75	75
R-squared	0.35			0.33	0.35	0.33	0.57

Notes: Standard errors are in parenthesis. All specifications are linear OLS with fieldworker fixed effects. Column (1) shows results from an ITT analysis with the initially randomized sample of 99 GPs, where missing outcomes are imputed with baseline values for the 19 GPs lost to non-participation or attrition. Column (2) estimates upper and lower bounds for the true effect (in the absence of any attrition) using the Manski-Horowitz method. The specifications in (1) and (2) do not control for all imbalanced covariates, due to the limited covariate data available for those 19 GPs that did not participate in the interviews. Column (3) controls for the order of SP visits to each provider (as each provider received two SP visits in this experiment, in a random order). Column (4) controls for the hour of day in which each SP visit occurred. Column (5) controls for the number of days between intervention delivery (in the GP interviews) and impact assessment (in the corresponding end-line SP visit). Column (6) controls for additional provider covariates – namely, all listed variables in Table 7.2, excluding the two variables on perceived patient expectations for antibiotics in the uncomplicated bronchitis case and perceived antibiotic prescribing norms for the complicated (HIV+) bronchitis case, which were highly correlated with some of the unbalanced covariates included in the main specification. *** p<0.01, ** p<0.05, * p<0.1