



Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Does Free Public Health Care Increase Utilization and Reduce Spending? Heterogeneity and Long-Term Effects

Peter Hangoma^{a,b}, Bjarne Robberstad^b, Arild Aakvik^b

^a University of Zambia, Zambia

^b University of Bergen, Norway

ARTICLE INFO

Article history:

Accepted 30 May 2017

Available online xxxxx

Key words:

user fees
utilization
medical spending
long term
heterogeneity
Zambia

SUMMARY

Zambia removed user fees in publicly supported–government and faith-based–health facilities in 54 out of 72 districts in 2006. This was extended to rural areas of previously unaffected districts in 2007. The natural experiment provided by the step-wise implementation of the removal policy and five waves of nationally representative household survey data enables us to study the impact of the removal policy on utilization and household health expenditure. We find that the policy increased overall use of health services in the short term and the effects were sustained in the long term. The increases were higher for individuals whose household heads were unemployed or had no or less education. The policy also led to a small shift in care seeking from private to publicly supported facilities, an effect driven primarily by individuals whose household heads were either formally employed or engaged in farming. The likelihood of incurring any spending reduced, although this weakened slightly in the long term. At the same time, there was an upward pressure on conditional health expenditure, i.e., expenditure was higher after removal of fees for those who incurred any spending. Hence, total (unconditional) household health expenditure was not significantly affected.

© 2017 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Health care was freely provided in many low- and middle-income countries (LMIC) before the mid 1980s. However, an increasing number of countries started introducing charges at the point of use, known as user fees (Akin, Birdsall, & De Ferranti, 1987; Yates, 2009), a wave of enthusiasm generated by an influential World Bank report (Akin *et al.*, 1987). Importantly, LMIC were facing severe health care financing challenges that affected the quality of health services. The two main arguments for user fees were that they would provide extra resources to the health system and act as a rationing device, preventing the frivolous use of health services. In practice, although user fee revenues accounted for 5–12% of total health system revenue at the central level, they accounted for 50–100% of non-salary operating costs at the facility level in most countries (Gilson, 1997). In Zambia, they constituted 27.6–80.7% of facility non-salary operating costs (Cheelo *et al.*, 2010). These resources were used to supplement salaries, finance community activities such as outreach, hire additional staff, and purchase supplies (Carasso, Lagarde, Cheelo, Chansa, & Palmer, 2012).

Apart from the health financing role, there has been debate on the demand effects of user fees, polarized by two sets of findings. The first set examines the short-term effect of introducing user fees (Asfaw, von Braun, & Klasen, 2004; Blas & Limbambala, 2001; Mwabu, Mwanza, & Liambila, 1995) or removing them (Masiye, Chitah, & McIntyre, 2010; Xu *et al.*, 2006) and finds that user fees deter access to care. But whether these changes in demand are due to frivolous use remains an open question. The second set of findings however shows that user fees do not necessarily reduce, but increase, the utilization of health care when accompanied by quality improvements (Akashi, Yamada, Huot, Kanal, & Sugimoto, 2004; Barber, Bonnet, & Bekedam, 2004; Litvack & Bodart, 1993; Mataria, Luchini, Daoud, & Moatti, 2007; Soucat *et al.*, 1997), which are made possible by the extra resources from user fees. Both sets of findings can be justified from a theoretical standpoint; the overall demand effect of changing the price of care, through user fees, depends on the responsiveness of demand to changes in price on one hand, and changes in quality on the other hand, and these pull in different directions.

However, no study has shown reductions in utilization when fees are removed to complement studies that find increases in

<http://dx.doi.org/10.1016/j.worlddev.2017.05.040>

0305-750X/© 2017 The Author(s). Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Please cite this article in press as: Hangoma, P., et al. Does Free Public Health Care Increase Utilization and Reduce Spending? Heterogeneity and Long-Term Effects. *World Development* (2017), <http://dx.doi.org/10.1016/j.worlddev.2017.05.040>

utilization when fees are introduced. This is despite indications that removal of user fees was accompanied by reduction in quality measures, e.g., drug availability, health worker motivation, etc., in most countries (Masiye et al., 2010; Meessen et al., 2011). A possible explanation is that although these quality measures reduced, they did not reduce enough to dilute the positive demand effects of user fee removal, chiefly because some commitments were made to compensate health facility revenue loss. However, these commitments were either delayed, not met, or if met, they were not sustained (Meessen et al., 2011), underscoring the importance of studying long-term policy effects. An alternative explanation is that individuals in these contexts, at least in the short-term, are more sensitive to price reductions than they are to reductions in quality. In markets where price is more important than quality, health services are generally of poor quality (Das, Hammer, & Leonard, 2008).

There are calls to move away from user fees toward health care financing systems based on pooling, such as social health insurance- or tax-based systems (World Health Organization, 2010), so as to achieve universal health coverage (UHC).¹ Yet others contend that LMIC have limited capacity to implement such systems because of a high informal sector share which makes it hard to collect tax or insurance contributions (Bitran, 2014; Wagstaff, 2010). Although policies exempting the poor from user fees have been historically unsuccessful, the Cambodian experience shows that a successful user fee policy can be implemented alongside a well-targeted equity fund (Bigdeli & Ir, 2010; Meessen, Damme, Tashobya, & Tibouti, 2007). The challenge again is that governance structures are weak in most LMIC (Leonard, Bloom, Hanson, O'Farrell, & Spicer, 2013) and for equity reasons, the solution boils down to removing user fees for all. However, there is doubt on whether removing user fees would enable the provision of care that improves health, even when demand rises, if no compensating supply side measures are taken to maintain or improve quality (Campbell, Oulton, McPake, & Buchan, 2011). Evidence shows that following user fee removals, individuals visiting public facilities had to rely on the private market for things such as drugs (Hadley, 2011; Nabyonga Orem, Mugisha, Kirunga, Macq, & Criel, 2011).

Under such supply side constraints, financial risk may remain high even with user fee removal. Additionally, a market for informal payments may be created. These payments could make up for the loss in incentives provided by user fees (Meessen et al., 2007), worsening financial risk (Barber et al., 2004). At the same time, individuals with higher ability to pay, wanting to get, now relatively scarce resources – e.g., drugs—would be willing to pay bribes or under the table payments. The widespread incidence of informal payments in low-income countries is well documented (Barber et al., 2004; Falkingham, 2004; Lindkvist, 2013). In Zambia, there is anecdotal evidence of the existence of informal payments (Hadley, 2011). Hence, whether the removal of user fees reduces medical spending is an empirical issue.

This study seeks to contribute to the literature examining the demand consequences of free health care, specifically, removal of user fees. Although the existing literature is rich and informative, it is plagued with a number of limitations (Dzakpasu, Powell-Jackson, & Campbell, 2014; Lagarde & Palmer, 2008). First, the widespread use of facility/administrative data possess severe limitations bordering on poor quality of this type of data in LMIC (Ashraf, Fink, & Weil, 2014; Lim, Stein, Charrow, & Murray, 2008; Sandefur & Glassman, 2015), and lack of detailed socioeconomic variables for examining heterogeneities (Masiye et al., 2010). Second most of the studies do not have access to policy changes that

permit the separation of the effect of user fees from other concurrent events, see for example Asfaw et al. (2004), Mwabu et al. (1995) and Xu et al. (2006). Some studies which have access to reasonable quasi-experiments face the challenge of having a few sample points—health facilities—threatening reliability as well as generalizability of findings. Although some studies have enlightened us on the effect of user fees in experimental situations (Cohen & Dupas, 2010; Kremer & Miguel, 2007; Powell-Jackson, Hanson, Whitty, & Ansah, 2014; Thornton, 2008), their validity may be limited when one has to imagine large and complex national scale interventions with system wide implications (Acemoglu, 2010). The dearth of evidence on the impact of complex user fee policies implemented at national scale motivated Ridde and Haddad (2009) to conclude that “African public health officials and decision makers are worried about the relationship between abolishing user fees and health care financing, and much remains to be done to provide them with the evidence they require.”

By combining several waves of huge nationally representative household survey data in Zambia for the period 1998–10 and the natural experiment provided by the step-wise implementation of the removal policy, we overcome some of the methodological and data challenges of the existing literature. Our identification strategy exploits the fact that in April 2006, the government of Zambia removed user fees in all publicly supported health facilities—government and faith-based (mission) facilities—at the primary level in 54 out of 72 districts classified as rural (MoH, 2007). Specifically, the removal policy stipulated that “All services for which clients were paying user/medical fees i.e consultation, treatment, admission, and diagnostic services shall be free” (MoH, 2007). The removal policy was extended to rural areas of the remaining 18 previously unaffected districts in June 2007. Thus, only urban areas of the 18 districts remained unaffected by both waves of the removal policy. We use difference-in-difference (DD) models and carry out a number of robustness checks to assess the validity of our identifying assumption.

This study contributes to the debate on free health care in general, and removal of user fees in particular, in at least four ways. First, our data enable us to examine the effect of the removal policy on overall use of health care. Previous studies in Zambia have only examined the effect of the first wave of removals on utilization of publicly supported health facilities (Lagarde, Barroy, & Palmer, 2012; Masiye et al., 2010; Onde, 2009), but have not been able to determine whether this could have been due to an increase in overall use of health care (uptake effect) or shift in care seeking from private facilities (switching effect), or indeed both. Switching may not improve health if quality of care in publicly supported and private facilities is the same, which appears to be the case (Basu, Andrews, Kishore, Panjabi, & Stuckler, 2012; Das et al., 2008; Powell-Jackson, Macleod, Benova, Lynch, & Campbell, 2015).² However, even if quality of care is the same but perverse incentives in private facilities unnecessarily increase the cost of care, then switching may improve social welfare by eliminating inefficiencies.

Second, access to rich household survey data permits us to examine heterogeneous effects of the removal policy by socioeconomic status. This is important in assessing whether removal of user fees elicited a higher utilization response from individuals from lower socioeconomic backgrounds as the policy intended.

Third, we provide evidence on the extent to which the removal policy affected household medical spending, an important starting point in discussing financial risk protection. Most of the studies have focused on utilization effects. This is an important gap

¹ UHC requires providing access to quality health care for all who need it, and ensuring that they are not impoverished as a result of accessing care.

² Here we refer to formal private-for-profit health facilities which constitute about 14% of health facilities in Zambia and are regulated by the Health Professions Council (HPC) (MoH, 2011).

because, as discussed earlier, user fee removal may not automatically translate into reduced financial risk protection. Descriptive evidence in Uganda, for example found no evidence of reduced household medical spending following removal of user fees (Nabyonga Orem et al., 2011; Xu et al., 2006). Even after the nationwide removal of user fees in Zambia, 10% of the population experienced catastrophic spending in 2013 (Masiye, Kaonga, & Kirigia, 2016). More credible evidence in the health insurance literature suggests that some form of free care, or subsidized care, while increasing utilization, may not reduce health spending (Ataguba & Goudge, 2012; Fink, Robyn, Sié, & Sauerborn, 2013; Liu & Zhao, 2014; Nguyen, 2012; Wagstaff, Lindelow, Jun, Ling, & Juncheng, 2009) and may actually increase financial risk (Wagstaff & Lindelow, 2008).

Fourth, we study the long-term effect of the removal policy. This shades light on the ability of the removal policy to sustain gains in utilization and medical spending since the demand effects of the removal policy may vary over time as quality of care changes. A few available studies find that gains in utilization are not sustained (Lagarde et al., 2012), while others find that they are (Nabyonga Orem, Mapunda, Musango, & Mugisha, 2013), but these studies are at best descriptive. Our study is the first to provide carefully isolated evidence of the long-term effects of a user fee policy implemented at national scale.

Results show that overall use of health services (uptake) increased, although there was a small shift in care seeking from private to publicly supported facilities (switching). The increase in overall use of health services was driven primarily by individuals from lower socioeconomic backgrounds. The proportion of individuals incurring any spending fell but there was an upward pressure on the amount of spending for those individuals still incurring any spending (conditional spending). As a result, total (unconditional) health spending was not significantly affected. The effects on utilization and medical spending observed in the short-term were maintained in the long-term.

The rest of this paper is organized as follows: Section 2 provides a short overview of the setting of this study and the user fee removal policy. Section 3 discusses data and identification. The empirical specification is given in Section 4, results in Section 5. We discuss our findings in Section 6 and conclude in Section 7.

2. Context and user fee removal

In 2015, Zambia's population was estimated at 15.5 million (CSO, 2016). With an urban share of 42%, it is one of the most urbanized countries in Africa. Zambia is classified as a lower middle-income country with GNP per capita of USD 1,810 (World Bank, 2015), but remains one of the most unequal countries in terms of income and social services (UNDP, 2014). In 2010, poverty levels were estimated at 60.5% with more than 83% of the labor force employed in the informal sector (CSO, 2012). The size and extent of informal economic activities possess a challenge to attempts by the government to expand the tax base. The disease burden is also high with malaria, tuberculosis, and HIV exerting a severe strain on the health system. The burden of non-communicable diseases is also on the increase (Institute for Health Metrics & Evaluation, Human Development Network, The World Bank, 2013).

Zambia is administratively divided into 10 provinces (increased from 9 in 2010), which are further divided into districts. The number of districts increased from 64 in 2010 to the current 106. Delivery of health care follows a 3-tier system, namely, primary, secondary, and tertiary level, and the referral system follows these levels. The primary level consists of health posts, health centers, and first-level hospitals. A few districts have first-level hospitals

but second-level and third-level hospitals are restricted to a few of the 10 provinces.

In terms of ownership, 80% of health facilities are owned by government, 6% by faith-based organizations (missions) and 14% are private-for-profit (MoH, 2011). The Health Professions Council (HPC) of Zambia regulates the registration and operations of all health facilities, including private ones. Outside these formal private-for-profit facilities, henceforth private facilities, the level of informality is not known, but appears to be very low.³ Government facilities have the widest coverage while private facilities are mainly concentrated in urban areas of all provinces. Faith-based health facilities are mostly located in rural areas and are normally classified as private-not-for-profit. We classify them as publicly supported, together with government facilities, because government provides funding and pays most of their health workers. In addition, government policies such as introduction or removal of user fees affect them equally (MoH, 2007).

Zambia has a mixed health care financing system with a heavy reliance on external financing. Between a third to half of total health expenditure (THE) is from external sources (Ministry of Health, 2009; World Health Organization, 2015). Zambia spends US\$86 per capita on health and 38% of this is from households (World Health Organization, 2015). For a long time, government spending on health has fallen far below the Abuja target of 15%. Health financing challenges have resulted in severe capacity constraints which have been driving most individuals utilizing publicly supported facilities to the private market for drugs and diagnostics. For example, a study in 2012 found that at the secondary and tertiary level, 18% of hospitals had no blood pressure measuring equipment, 22% had no X-ray machines, and 33% had no ultrasound machines (Institute for Health Metrics & Evaluation, 2014). The situation is even more severe at the primary level, the study found that about 40% of health centers, run by government or faith-based organizations, had no qualified staff, and most of these were rural.

From Independence in 1964, health care in Zambia was free at the point of use. As part of a number of structural reforms in the early 1990s however, user fees were introduced with exemptions for children below five years of age and adults above 65 years. User fee charges varied across the country and were agreed upon by health facility management and community representatives taking into account the local economy of the health facility catchment area.

In April 2006, a major health policy shift was announced. The government removed user fees at the primary level in all publicly supported health care facilities (government and private) in 54 districts designated as rural, leaving the rest of the districts (18) that were classified as urban unaffected. Moreover, individuals who went through the referral system continued to be exempt at higher levels of care. User fee removal was defined to include fees for registration, consultation, outpatient and inpatient care, X-ray, and laboratory tests (MoH, 2007). Fig. 1 shows the districts which were affected and unaffected by this first wave of removal of user fees. While user fees were removed for primary services only (clinics

³ An informative, though crude, check on the level of informality is to observe that the proportion of individuals who report using private facility in nationwide household surveys is far lower, e.g., 6% in 2010 (CSO, 2012), than the proportion of formal private health facilities, 14%, captured by HPC. If the proportion of people reporting using private facilities was close, equal, or more than the proportion of formally register private facilities, then there would be a high possibility of high informality. That said, the level of informality varies widely across countries even within the same region (Sheikh, Josyula, Zhang, Bigdeli, & Ahmed, 2017). This makes it hard to make judgments based on other regions, especially that the private-for-profit sector is far smaller in Zambia compared to other LMICs. More information about HPC and health facilities in Zambia can be accessed at <http://hpcz.org.zm/hpcz-publications/>.

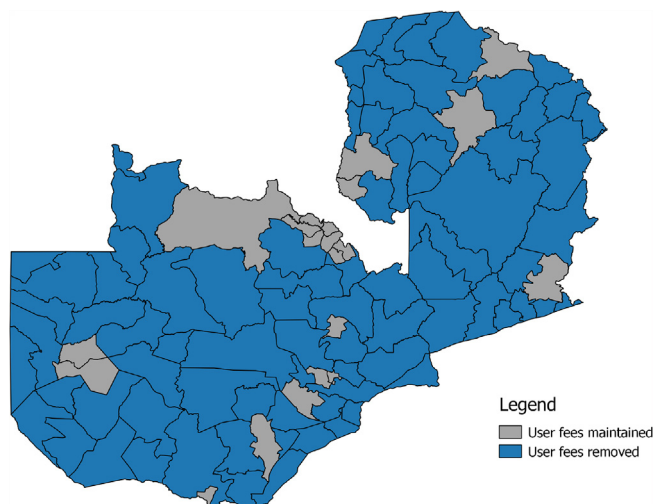


Figure 1. Zambia. Map shows districts where user fees were removed during the first wave of user fee removals. Three of the 18 district were each recently split into two. Hence the number of district where user fees were maintained on this map will count as 21 instead of 18.

and first-level hospitals) in publicly supported facilities, sweeping exceptions were given for districts to offer free care even at higher levels. For example, the policy specified that since many districts did not have Hospital Affiliated Health Centres at first-level hospitals, and many did not have these hospitals altogether, free services were to be provided at the secondary and tertiary hospitals in those districts (MoH, 2007).

The following year, in June 2007, the government acknowledged the shortcoming of removing user fees based on classifying the whole district as either rural or urban, and that there are areas within urban districts that were rural and vice versa. With this understanding, they extended the removal of user fees to rural areas of the 18 districts that were previously unaffected.

In particular, areas where user fees were removed depended on whether or not the district was located along the line of rail (railway). For those along the line of rail, user fees were removed in health facilities located more than 20 km away from the district administration. Similarly, user fees were removed in health facilities located more than 15 km away from the district administration for districts that were not located along the line of rail. Implicit in the removal policy is the idea that rural areas are located in the peripheries of the districts.

In 2012, user fees were finally removed throughout the country in all publicly supported health facilities.

3. Data and identification

(a) Data sources, sample, and outcome variables

Our data are from the 1998, 2002, 2004, 2006, and 2010 Living Conditions Monitoring Survey (LCMS). The LCMS is a large nationally representative repeated cross-sectional household survey conducted by the Central Statistical Office of Zambia (CSO). Although these surveys were also conducted before 1998, it was not possible to use these data because of changes in district boundaries making it impossible to isolate treated from control districts.

For medical spending, we do not include the 1998 data due to differences in components of medical spending that were collected in 1998 compared to subsequent surveys. While the 1998 survey asked about spending on a number of health services in the two weeks prior to the survey, the 2002 and subsequent surveys only

asked about total spending in the two weeks prior to the survey.⁴ It has been shown that more disaggregation and longer list of items, as is the case in the 1998 survey, yields higher levels of expenditures for similar households compared to less disaggregation (Beegle, Weerdt, Friedman, & Gibson, 2012; Jolliffe, 2001; Pradhan, 2009).

Our analysis focuses on individuals who reported any sickness or injury in the two weeks prior to the survey. We restrict our sample to individuals aged between 5 and 65. This is because individuals under 5 years and those over 65 were exempt from user fees even before the removal. The percentage (number) of individuals between the age of 5 and 65 reporting sickness or injury was 9% (7,018) in 1998, 12% (6,482) in 2002, 8% (8,182) in 2004, 7% (6,372) in 2006, and 11% (10,003) in 2010. Individuals who reported any sickness or injury were asked whether or not they consulted any health facility as a result of the sickness or injury. Those who reported consulting were then asked about the type of facility they consulted. Using this information, we define three measures of utilization. The first one, is an indicator variable equal to one if an individual reported consulting a publicly supported facility and zero if not. Individuals are said to have consulted a publicly supported facility if they visited a government health post, clinic, hospital, or a faith-based health facility for the illness or injury.⁵ This measures total response of utilization of publicly supported facilities and it may capture both the fact that there were more or less people using health services irrespective of provider (uptake effect) and that there was substitution between publicly supported and private facilities (switching effect). Thus, our second measure of utilization is also an indicator variable equal to one if an individual reported consulting a private health care facility and zero otherwise. This captures the switching effect, or more precisely, how the user fee policy affected the utilization of private facilities. The third measure is equal to one if an individual reported having made any consultation (publicly supported or private) and it measures the overall increase or reduction in health care use following the removal of user fees.

We point out that we only focus on utilization of formal private medical care. Individuals who reported consulting non-science-based care such as traditional or spiritual healers, constituting a very small share, e.g., 1% in 2010 (CSO, 2012), are classified as not having consulted any formal care. We do this also for individuals who reported self-medication, some of whom could have accessed drugs from informal drug stores. Drug stores and pharmacies are not classified as health facilities and are thus not part of the options for “type of facility visited” in the survey. However, such spending is included in total household medical expenditure.

Additionally, faith-based facilities, constituting about 6% of health facilities may fair better in quality than government facilities, a fact observed in many LMIC (Das et al., 2008). Thus, even if user fees were removed in both these publicly supported facilities, it would have been interesting to assess heterogeneous utilization responses in the two types of facilities. However, we are not able to conduct such an analysis. The challenge is that, due to the low proportion of mission facilities, estimates are unreliable. Our analysis thus reviews average effects in publicly supported facilities.

Medical expenditure in our data refers to the amount spent on consultation, medical examinations, drugs, and any form of self-medication in the past two weeks for individuals who reported being sick or injured regardless of where the spending took place, e.g., publicly supported, private health facility, drug store, etc. We define two measures of spending, one that looks at the proportion

⁴ This question existed in the 1998 survey but only referred to expenditure on self medication.

⁵ Faith based are also publicly supported because, just like, government health facilities, they rely on funding government grants. As such, they were also affected by the removal policy. See Section 2.

Table 1
Description of Treatment and Control groups

Group	Description	N
T1	Individuals residing in the 54 districts where user fees were removed in April 2006 at the primary level in all publicly supported health care facilities. User fees were maintained in the rest of the districts (18).	23,403
T2	Individuals residing in the rural areas of the 18 districts. They were affected by the second wave of removals in June 2007 ^a	4,140
Control	Individuals residing in urban areas of the 18 districts. These were not affected by the two waves of the removal policy in 2006 and 2007	10,514

^a The precise identification of these areas would require GPS coordinates which were unfortunately not collected in the surveys. We use the rural (urban) definition of the Central Statistical Office (CSO) to classify the areas in the 18 districts as treated (control). Our definition appears to distinguish treated and untreated areas in the 18 district very well. See [Appendix A.1](#) for details.

of individuals that incurred any spending (extensive margin) and the other measure which assesses the level of spending conditional on it being positive (intensive margin).

All outcomes are conditional on reporting any illness or injury in the two weeks prior to the survey.

(b) Identification and data description

We define the first wave of removal of user fees in early 2006 as Treatment 1 (T1) and the second wave of removals in June 2007 as Treatment 2 (T2). These two waves of removals partitioned the whole country such that only urban areas of the 18 districts remained unaffected, and these areas represent our control group. Individuals who reside in the areas where T1 was effected are defined as the T1 group. Using the T1 group and the control, we are able to identify the short-term (2006) and long-term (2010) effect of the first wave of removals, T1. Individuals residing in the rural areas of the 18 urban districts who were affected by the second wave of removals, T2, in 2007 are defined as the T2 group. This treatment group enables us to identify the effect, in 2010, of the second wave of removals. [Table 1](#) briefly describes our treatment and control groups.

Given multiple waves of data, we exploit geographical variation in the removal policy and use difference-in-difference models to estimate the short-term (2006) and long-term (2010) effects of providing free care in publicly supported health facilities. This is an intention-to-treat (ITT) effect of the removal policy. This is because, although the removal policy was meant for the primary level, it later allowed health facility officials, as discussed in [Section 2](#), to offer free services at the secondary and tertiary level in some areas. Thus, for those who reported visiting a hospital in our data, we are not able to distinguish whether they went to a paying one or not. But even where services were supposed to be free, compliance may not have been perfect. With deviation from policy and less than full compliance, ITT, though more conservative, is a more policy relevant effect.⁶

Our identifying assumption is that, in the absence of the removal policy, potential outcomes in areas where user fees were removed would have followed the same trend as non-treatment outcomes in the areas where user fees were not removed. A departure from trend is counted as the effect of the policy. This identifying assumption is fundamental in the difference-in-difference framework and it is called the *parallel trends* assumption. *Parallel trends* does not require that treated and control groups are the same, but that pre-treatment outcomes follow similar trends ([Godlonton & Okeke, 2016](#)). Formally, this assumption is assessed by checking that the differences in outcomes between the treated and control groups are the same at all time periods before treat-

ment (pre-treatment period). This implies that in a difference-in-difference model, if the pre-treatment interaction (treatment) effects are included, they should be zero, or statistically insignificant. We conduct placebo tests by including these pre-treatment treatment effects in our models. First, we visually assess this assumption.⁷

(i) Assessment of pre-treatment trends

The T1 and control group exhibited similar trends in utilization of publicly supported health facilities and private health facilities as well as both measures of medical spending throughout the pre-treatment period (1998–04) ([Fig. 2](#)).

For the T2 group and control, utilization rates of publicly supported facilities were the same throughout the pre-treatment period, as can be seen from the overlapping confidence intervals in the top left plot of [Fig. 3](#). The same can be said of the utilization of private facilities (bottom right of [Fig. 3](#)) and both measures of medical spending (left of [Fig. 3](#)).

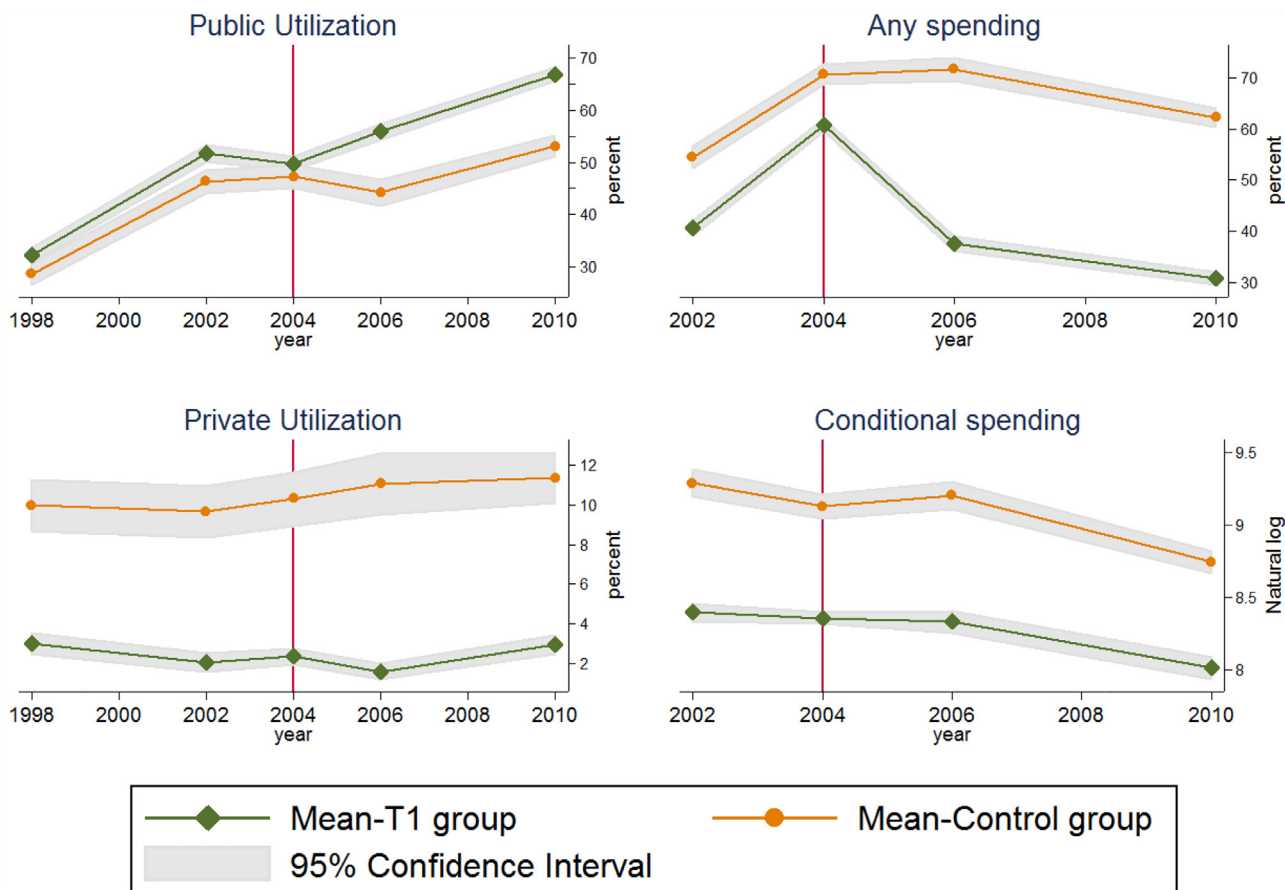
(ii) Baseline covariates and changes over time

The T1 and control group are generally similar in terms of household and demographic characteristics such as household size, age, sex, and marital status but differ in characteristics such as educational level, occupational status, and distance to a health facility ([Table 2](#), Column (1) and (2)). For instance, 12.2% of household heads had no education in the T1 group compared to 4% in the control group. Only 16.9% of individuals in the T1 group had their nearest health facility within 1 km compared to 45.6% in the control group. Although the T1 group had significantly fewer household heads who were unemployed, it also had a substantially lower proportion of household heads formally employed. Rather than the requirement that these characteristics (covariates) be the same in treated and control groups, what is important for our identification strategy is that changes in these covariates in the post-treatment period should be the same. Otherwise they could actually be the ones driving differences in utilization and medical spending that we could attribute to the removal policy. Columns (5) and (6) show that almost all characteristics did not change differently in the post-treatment period.

[Table 3](#) presents a comparison of characteristics in the T2 and control group. The picture is generally similar to that of the T1 and control group; although the T2 and control groups are broadly the same in terms of household and demographic characteristics such as household size, age, sex, and marital status, the T2 group is worse-off in terms of characteristics such as educational level, occupational status, and distance to a health facility ([Table 3](#), Column (1) and (2)). However, these characteristics generally change in the same way in the two groups over time.

⁶ [Hadley \(2011\)](#) documents cases where some health facilities charged different types of fees after the official fees were removed. [Masiye et al. \(2016\)](#) also found that some individuals reported having paid for care in primary publicly supported health centers even after the national-wide removal of user fees in 2013. ITT would only equal the average treatment on the treated (ATT) if there was full compliance and no policy deviations.

⁷ They are termed “pre-treatment treatment effects” because they try to measure the impact of the policy before it was actually implemented (treatment effect in the pre-treatment period). Pre-treatment treatment effects are not supposed to be significant if treated and control groups were being affected equally by other factors. If they are significant, then the true treatment effect is not identified.



Note: Medical spending only uses data for the period 2002--2010

Figure 2. Figure shows trends in average utilization and spending, and the associated 95% confidence intervals, for T1 group and control. The plots on the left show the percentage of individuals utilizing publicly supported health facilities (top) and private health facilities (bottom). On the right, we show the percentage of individuals incurring any spending (top) and conditional spending in natural logarithms (bottom).

4. Empirical specification

Our econometric model for evaluating the impact of the two waves of the removal of user fees, T1 and T2, is a difference-in-difference model of the form:

$$\begin{aligned}
 y_{it} = & \alpha_0 + \alpha_1 yr1998 + \alpha_2 yr2002 + \alpha_3 yr2006 + \alpha_4 yr2010 + \delta_1 T1_i + \delta_2 T2_i \\
 & + \delta_3 (T1_i \times yr2006) + \delta_4 (T1_i \times yr2010) + \delta_5 (T2_i \times yr2010) \\
 & + \delta_6 (T1_i \times yr1998) + \delta_7 (T1_i \times yr2002) \\
 & + \delta_8 (T2_i \times yr1998) + \delta_9 (T2_i \times yr2002) + \delta_{10} (T2_i \times yr2006) \\
 & + \beta' X_{it} + \gamma_1 \sum_r^{R-1} P_r + \gamma_2 \sum_r^{R-1} \sum_{t=1998,2002}^{2006,2010} (P_r \times yr_t) + \epsilon_{ijt}
 \end{aligned}
 \tag{1}$$

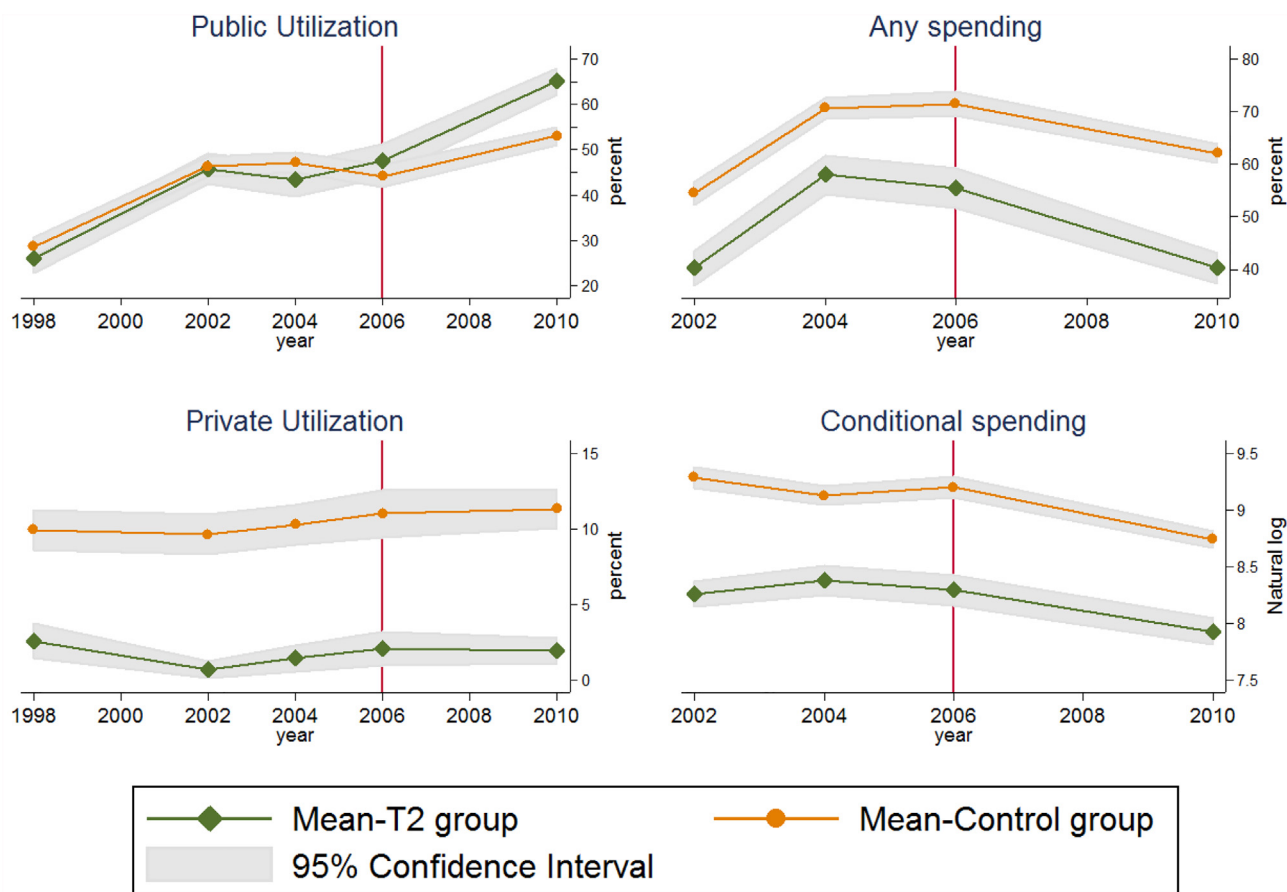
where y_{it} is the observed outcome for individual i in year t . The variables $yr1998$, $yr2002$, $yr2006$, and $yr2010$ are year dummies. They account for changes that may have affected utilization and medical spending in both treatment and control groups, typically national-level policies or economic shocks. The baseline year is 2004. $T1_i$ is a treatment indicator equal to 1 if individual i is in the T1 group, i.e., resides in any of the 54 treated districts. Similarly, $T2_i$ is 1 if they reside in the rural areas of the other 18 districts. $T1_i$ and $T2_i$ captures the influence of unobserved factors which are specific to these areas and have the potential to explain differences in utilization and medical spending between each treatment group and the control. X_{it} is a set of individual and household observable characteristics of i , e.g., education, employment status, household charac-

teristics, distance to health facility, etc. They account for any time varying individual and household characteristics that may have had differential impact on outcomes of treated and control groups. P_r is the r th province dummy, for $R = 9$, while $P_r \times yr_t$ are province by year dummies. They capture regional-level fixed effects and time-varying shocks that may have affected the treated and control groups differently. Supposing this is a linear model, the short-term difference-in-difference effect of T1 is given by the coefficient of the first interaction term δ_3 , while the coefficient on second interaction term, δ_4 , gives the long-term effect of T1. We measure the effect of the second wave of user fee removals, T2, by the coefficient, δ_5 . As a placebo test to formally assess the plausibility of the parallel trends assumption, a full set of pre-treatment interaction effects for both T1 and T2 are included. δ_6 and δ_7 should be zero if the T1 group and control followed similar pre-treatment trends. In the same vein, δ_8 , δ_9 , and δ_{10} should not be statistically different from zero if the T2 group and control followed similar pre-treatment trends.

To ensure that standard errors are not underestimated since the policy was implemented at the district level while the data are at the individual level, we cluster the standard errors at the district level (Bertrand, Duflo, & Mullainathan, 2004).

(a) Utilization

All three measures of utilization (public, private, and any utilization) are binary. For each of these outcomes, we fitted a Linear



Note: Medical spending only uses data for the period 2002--2010

Figure 3. Figure shows trends in average utilization and spending, and the associated 95% confidence intervals, for the T2 group and control. The plots on the left show the percentage of individuals utilizing publicly supported health facilities (top) and private health facilities (bottom). On the right, we show the percentage of individuals incurring any spending (top) and conditional spending in natural logarithms (bottom).

Probability Model (LPM) to Eqn. (1). Thus, the effect of the user fee removal policy are simply the coefficients, $\delta_3, \delta_4,$ and δ_5 . To save space, we only report these effects and the pre-treatment interaction coefficients $\delta_6, \delta_7, \delta_8, \delta_9,$ and δ_{10} .

To understand the distributional impact of the removal policy, we examine how utilization of various individuals across the socioeconomic distribution was affected. These socioeconomic variables should not have been affected by the removal policy. Thus, rather than focusing on income or household consumption, we look at educational level and occupation of the head of household. We categorize these variables and estimate a difference-in-difference model for each of these categories (sub-samples). Focusing on T1, we estimate the overall effect for each sub-sample, as opposed to short- and long-term effects as we did in Eqn. (1), in order to concentrate on examining heterogeneities.

Thus, for each of the sub-samples of these variables, we fit a difference-in-difference model of the form:

$$y_{it} = \theta_0 + \theta_1 T1_i + \theta_2 post_t + \theta_3 (T1_i \times post_t) + \theta_4 yr1998 + \theta_5 yr2002 + \beta' X_{it} + \gamma_1 \sum_r^{R-1} P_r + \gamma_2 \sum_r^{R-1} \sum_{t=1998,2002}^{2006,2010} (P_r \times yr_t) + \epsilon_{ijt} \quad (2)$$

where $post_t$ is 1 if an individual was observed in 2006 or 2010. Our parameter of interest, which we report for each sub-sample, is θ_3 . This analysis will give us an insight into which socioeconomic group began utilizing any health services after the removal (uptake effect) and which ones could have been switching from private to publicly

supported facilities where user fees were removed. Thus, we focused on two outcomes, any utilization and private utilization.

(b) Medical spending

As is common in household expenditure data, our medical spending data have many zeros and is particularly skewed. This is especially true because the removal policy eliminated medical spending for some individuals. Although OLS on log-transformed spending is simple and deals with the skewness problem, it yields biased estimates because it ignores the existence of zero expenditure. It is important to study the impact of the policy on the full distribution of medical spending. To achieve this, we examined the policy effect on two margins of the expenditure distribution, the extensive margin—whether or not an individual incurred any spending—and the intensive margin—the amount of spending conditional on it being positive. We then combined these two effects to yield a measure of the impact of the policy on the whole (unconditional) distribution of medical spending. This was achieved using a two-part model (TPM).

In the TPM, the extensive margin (first part) is modeled by fitting a probit model to Eqn. (1) and the intensive margin (second part) similarly fits Eqn. (1) using generalized linear models (GLM). The Box Cox test is used to select the link function while the family is chosen using the modified Park test. This leads us to a gamma family with a logarithmic link function.

Table 2
Baseline and changes in characteristics: T1 Group vs Control

Variable	Baseline (2004) Average		Change: 1998–2004	Change: 2002–2004	Change: 2004–2006	Change: 2004–2010
	T1 Group	Control	$\Delta T1 = \Delta Control$ (p-value)	$\Delta T1 = \Delta Control$ (p-value)	$\Delta T1 = \Delta Control$ (p-value)	$\Delta T1 = \Delta Control$ (p-value)
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.428	0.453	0.262	0.876	0.796	0.153
HH Male	0.784	0.750	0.045	0.348	0.424	0.779
H Size	6.165	6.072	0.786	0.670	0.882	0.161
Age	27.382	27.397	0.120	0.418	0.857	0.129
HH Married	0.768	0.659	0.013	0.146	0.100	0.217
HH College	0.048	0.121	0.866	0.725	0.978	0.588
HH Secondary	0.287	0.562	0.157	0.130	0.575	0.029
HH Primary	0.542	0.277	0.052	0.446	0.750	0.383
HH No Education	0.122	0.040	0.359	0.075	0.544	0.104
HH Employed	0.143	0.418	0.069	0.004	0.773	0.165
HH Self-Employed	0.090	0.306	0.054	0.055	0.394	0.450
HH Farming	0.707	0.071	0.831	0.000	0.143	0.006
HH Not Working	0.061	0.205	0.479	0.934	0.292	0.189
<i>Distance to Nearest Health Facility</i>						
≤1 km	0.169	0.456	0.007	0.381	0.616	0.839
1–10 Kms	0.612	0.529	0.029	0.159	0.462	0.172
11–20 Kms	0.146	0.000	0.980	0.241	0.182	0.003
≥21 Kms	0.073	0.015	0.326	0.605	0.807	0.046

Note: The table shows characteristics of individuals aged 5 to 65 who reported being sick or injured in the two weeks prior to each survey in the T1 group and control. The first and second columns report means of covariates at baseline (2004) for treated and control group, respectively. The rest of the columns report *p*-values that compare changes in the control and treated group from baseline to the relevant survey year. The *p*-values were computed using a simple difference-in-difference model of each characteristic (covariate). HH = Household Head and H = Household. All variables are dummies except H Size, and Age.

Table 3
Baseline and changes in characteristics: T2 Group vs Control

Variable	Baseline (2004) Average		Change: 1998–2004	Change: 2002–2004	Change: 2004–2006	Change: 2004–2010
	T2 Group	Control	$\Delta T2 = \Delta Control$ (p-value)	$\Delta T2 = \Delta Control$ (p-value)	$\Delta T2 = \Delta Control$ (p-value)	$\Delta T2 = \Delta Control$ (p-value)
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.399	0.453	0.427	0.357	0.061	0.023
HH Male	0.811	0.750	0.062	0.116	0.009	0.235
H Size	5.749	6.072	0.890	0.483	0.897	0.101
Age	25.299	27.397	0.229	0.595	0.859	0.313
HH Married	0.794	0.659	0.016	0.083	0.024	0.260
HH College	0.016	0.121	0.540	0.930	0.797	0.416
HH Secondary	0.636	0.277	0.648	0.413	0.981	0.621
HH Primary	0.253	0.562	0.888	0.144	0.534	0.358
HH No Education	0.095	0.040	0.794	0.730	0.439	0.274
HH Employed	0.131	0.418	0.087	0.114	0.900	0.893
HH Self Employed	0.123	0.306	0.344	0.919	0.168	0.819
HH Farming	0.684	0.071	0.365	0.073	0.520	0.866
HH Not Working	0.061	0.205	0.588	0.929	0.680	0.655
<i>Distance to Nearest Health Facility</i>						
≤1 km	0.178	0.456	0.063	0.788	0.020	0.691
1–10 Kms	0.644	0.529	0.029	0.159	0.111	0.342
11–20 Kms	0.131	0.000	0.274	0.513	0.846	0.253
≥21 Kms	0.047	0.015	0.711	0.501	0.704	0.183

Note: The table shows characteristics of individuals aged 5 to 65 who reported being sick or injured in the two weeks prior to each survey in the T2 group and control. The first and second columns report means of covariates at baseline (2004) for treated and control group, respectively. The rest of the columns report *p*-values that compare changes in the control and treated group from baseline to the relevant survey year. The *p*-values were computed using a simple difference-in-difference model of each characteristic (covariate). HH = Household Head and H = Household. All variables are dummies except H Size, and Age.

Because probit is a non-linear model, the interaction or marginal effect of the policy are not given by the interaction terms in Eqn. (1) (Ai & Norton, 2003). For the extensive margin (first part) thus, each interaction effect, e.g., for T1 in 2006, was computed as the following double difference:

$$\frac{\Delta^2 E}{\Delta yr2006 \Delta T1} = [\Phi\{yr2006 = 1, T1 = 1, X\} - \Phi\{yr2006 = 0, T1 = 1, X\}] - [\Phi\{yr2006 = 1, T1 = 0, X\} - \Phi\{yr2006 = 0, T1 = 0, X\}] \quad (3)$$

where Φ is the normal cumulative distribution function and the matrix X contains all covariates including region, other interaction terms, and year dummies.

For the second part (the GLM), the interaction effect was similarly computed as in Eqn. (3), although Φ is simply an exponent.

The overall effect on medical spending (unconditional medical spending) was estimated by combining the effect of the removal policy on each of two parts of the TPM. To see this, note that the unconditional expected value of medical spending is the joint expected value of medical spending, $E[y]$, which in turn is given

Table 4
Short- and long-term effects of user fee removal on overall, public, and private facility utilization

	Overall utilization (1)	Public utilization (2)	Private utilization (3)
<i>Short-term effect</i>			
T1 × yr2006	0.062(0.029)**	0.100(0.030)***	−0.034(0.017)*
<i>Long-term effects</i>			
T1 × yr2010	0.080(0.032)**	0.111(0.035)***	−0.032(0.014)**
T2 × yr2010	0.111(0.048)**	0.137(0.047)***	−0.027(0.016)
<i>Pre-treatment effects</i>			
T1 × yr1998	0.000(0.036)	0.004(0.033)	−0.004(0.022)
T2 × yr1998	0.018(0.064)	0.006(0.064)	0.011(0.023)
T1 × yr2002	−0.004(0.030)	0.009(0.032)	−0.013(0.016)
T2 × yr2002	−0.003(0.061)	0.014(0.059)	−0.011(0.016)
T2 × yr2006	0.078(0.065)	0.093(0.061)	−0.015(0.023)
Observations	31,887	31,887	31,887
T1 group baseline mean	0.521	0.497	0.024
T2 group baseline mean	0.447	0.433	0.014
Control group baseline mean	0.575	0.472	0.103

Note: Table shows the effect of the first and second waves of the removal policy, T1 and T2 respectively, on utilization of all facilities (Column 1), publicly supported facilities (Column 2) and private facilities (Column 3). Estimates are from DD linear probability models (LPM) corresponding to Eqn. (1) and are interpreted as percentage points. Standard errors clustered at the district level are reported in parentheses. The lower panel gives the baseline (2004) mean of each outcome for treated and control groups. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include distance to health facility, household size, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.

*Significant at 10%; **Significant at 5%; ***Significant at 1%.

by the product of the probability of incurring any spending—the probit part—and the average spending, given that spending is positive—the GLM part:

$$E[y] = \Phi(\cdot) \times E[y|y > 0] \quad (4)$$

The impact of the policy is given by how it changes this joint expected value. For example, the effect of T1 on overall medical spending in 2006 is given by taking the double difference (similar to Eqn. (3)) on Eqn. (4). Taking the double difference is a binary variable equivalent of differentiating the joint expectation twice.⁸

Standard errors are computed using a bootstrap procedure with 1000 replications that account for clustering at the district level.

5. Results

(a) Impact on individual utilization

(i) Overall impact

Results indicate that the removal of user fees in publicly supported health facilities increased overall utilization of health services (uptake effect), in addition to shifting use from private to publicly supported facilities (switching effect). However, the uptake effect was much greater, and stronger, than the switching effect (Table 4). In particular, The first wave of user fee removals, T1, increased overall utilization (uptake) of health services by 6.2 percentage points (pp) in the short term (Column (1)). The effect strengthened to 8 pp in the long term. However, utilization of publicly supported facilities increased more than the increase in overall utilization of health services in both the short- and long-term Column (2) shows that utilization of publicly supported facilities increased by 10 pp in the short-term and the effect was sustained at 11.1 pp in the long-term. The extra increase in utilization of publicly supported facilities was a result of switching from private- to publicly supported-facilities (Column(3)). Utilization of private facilities in the T1 group reduced by 3.4 pp in the short term with the effect sustained at 3.2 pp in the long term.

The second wave of removals (T2) increased the overall utilization of health services by 11.1 pp. Utilization of publicly supported

facilities increased by 13.7 pp of which 2.6 pp was due to switching from private facilities, although the switching effect was not significant. This insignificance is expected given the small proportion of individuals utilizing private facilities in T2 areas which renders standard errors to be imprecisely estimated.

All pre-treatment “treatment effects” for T1 are zero. The effect of the first wave of removals, T1, is thus identified. Similarly, all pre-treatment effects for T2 are not significantly different from zero. This suggests that utilization in the T2 group and the control also followed approximately parallel trends. This lends support to descriptive evidence of parallel trends on Fig. 2 in Section 3(a(i)).

(ii) Heterogeneous impact

Table 5 shows how individuals from different socioeconomic backgrounds responded to the removal of user fees. The policy increased overall utilization of health services (uptake) more for individuals from lower socioeconomic background than their better-off counterparts (Table 5). For individuals from higher socioeconomic background, the policy led to a shift in care seeking from private to publicly supported facilities (switching).

Beginning with education level, uptake was greater the lower the educational level of the head of household (Column (1)). Specifically, individuals whose household heads had no education increased their utilization most (10.6 pp). Uptake significantly increased by 6.2 pp for individuals whose household heads only had primary education and, although positive, the increase was not significant for individuals whose household heads had a secondary or college education. On the other hand, switching was generally driven by individuals coming from household heads with higher education (Column (2)).

For occupation status, uptake significantly increased by 11.2 pp for individuals whose household heads were not employed. Although positive, the overall use of health services did not increase significantly for individuals whose households heads were formally or self-employed. For individuals engaged in farming, the policy did not affect uptake. It however shifted their care seeking from private to publicly supported facilities. Individuals whose household heads were formally employed, despite not registering a statistically significant increase in uptake, also shifted care seeking from private to publicly supported facilities following the removal of user fees.

⁸ See Frondel and Vance (2013) for a discussion on interaction effects in non-linear two-part models.

Table 5
Heterogeneous effect of user fee removal on utilization

	Overall utilization (1)	Private utilization (2)
<i>Education Level of Household Head</i>		
No Education	0.106(0.063)*	0.013(0.021)
Primary	0.062(0.029)**	−0.022(0.009)**
Secondary	0.051(0.037)	−0.034(0.021)
College	0.033(0.049)	−0.031(0.034)
<i>Occupation of Household Head</i>		
Not Employed	0.112(0.050)**	−0.018(0.017)
Self Employed	0.024(0.041)	0.011(0.017)
Farming	0.003(0.042)	−0.077(0.021)***
Formally Employed	0.043(0.038)	−0.045(0.020)**

Note: Table shows the effect of the first wave of the user fee removal policy, T1, on utilization of all facilities (Column 1) and private facilities (Column 2) by educational level and occupational status of the head of household. Each reported coefficient is an estimate of θ_3 in a DD linear probability model (LPM) based on Eqn. (2). Thus, it is interpreted as a percentage point. Standard errors clustered at the district level are reported in parentheses. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include distance to health facility, household size, sex, age, and age squared of an individual, as well as household head characteristics such as marital status.

*Significant at 10%; **Significant at 5%; ***Significant at 1%.

(b) Impact on medical spending

The two waves of the removal policy significantly reduced the proportion of individuals incurring any medical spending but increased medical spending for those who continued to incur it, although this increase was not statistically significant (Table 6). As a consequence, both waves of the removal policy did not significantly impact overall medical spending.

The first wave of removals, T1, reduced the likelihood of incurring any spending in the short-term by 26.6 pp, and this effect reduced slightly to 19.3 pp in the long-term (Column 1, Table 6). These effects are highly significant at the 1% level. On the other hand, the second wave, T2, reduced the likelihood of incurring any spending by 8.7 pp. This effect is significant at the 10% level. On the other hand, estimates at the intensive margin (GLM part—Column 2) show that the amount of spending (conditional spending) did not change following the removal of user fees in the T1 and T2 groups. In fact, though not statistically significant, there was an upward pressure on the amount of spending. This upward pressure dampened the reduction in spending that resulted from reduced probability of incurring any spending. Thus, unconditional spending was left unchanged (Column 3).

All pre-treatment interaction effects for extensive (any spending) and intensive (conditional spending) part of medical spending are not statistically different from zero.

(c) Robustness

To assess the validity of our findings, we carried out a number of robustness checks (See the Appendix for details). First, a fully flexible difference-in-difference model proposed by Mora and Reggio (2012) and Mora (2015) was used to further test the common trends assumption. The assumption was met for all outcomes (Appendix A.2). Second, since our analysis is at the individual level while treatment varied at the regional level, we assessed the sensitivity of our findings to inclusion of an important regional-level variable, district funding.⁹ The inclusion of this key variable did not change our results (Appendix A.3). Third, the performance of difference-in-difference models can be assessed on how sensitive

⁹ Unfortunately, district funding was not available for 1998, 2002 and 2010. Hence, we could not include this variable in the main analysis.

the estimates are to the addition of covariates. Thus we estimated all models without covariates and with a limited set of covariates. Results were broadly similar (Appendix A.3). Fourth, we investigated whether the likelihood of reporting sickness or injury did not change differently between treated and control areas. If it did, there would be a possibility that the observed effects are actually driven by differential changes in sickness and not the policy change. Results show that changes in likelihood of reporting sickness and injury in treated and control areas were not different (Appendix A.4).

Fifth, we conducted tests to assess whether other factors (variables) were being affected which the user fee removal policy is not expected to affect. If these other variables were being affected, then there is a high chance that the effects we observe were actually not caused by the removal policy, but by other things. This falsification exercise ruled out such a possibility (Appendix A.5). Sixth, and lastly, we conducted randomization tests, which apart from ruling out rival explanations for the observed effects, also ensure that we do not falsely claim that results are significant when in fact not. This problem may arise because in difference-in-difference models with multiple time points, standard errors may be underestimated due to serial correlation (Bertrand et al., 2004). We did not find evidence that our standard errors could have been underestimated or that the effects were driven by rival explanations (Appendix A.6).

6. Discussion

Using data from Zambia, this paper shows that observed short-term increases in utilization of publicly supported facilities following the fall in price of care—due to removal of user fees—do not disappear in the long term. One possible explanation for these sustained effects could be that there were no significant supply-side constraints that could have lowered quality of care, and eventually health care use. An alternative hypothesis is that individuals care more about price than quality so that a fall in price accompanied by a fall in quality does not eventually reduce the initial increases in utilization. Two findings emerge from our study that may support this hypothesis. First, increases in overall use of care, or uptake—which accounted for most of the increase in use of publicly supported facilities—was driven by individuals from low socioeconomic backgrounds, a segment of the population that is likely to be more sensitive to price than quality (Ching, 1995; Gertler, Locay, & Sanderson, 1987). This effect was slightly larger in the long term. Second, and on the contrary, switching from private facilities—driven by relatively well-off individuals—was smaller, and actually insignificant for the T2 group, in the long term. Well-off individuals are likely to be more quality than price sensitive. These two findings suggests that utilization would be sustained in the long term even when the quality of care falls.

In fact, evidence suggests that quality of care may have fallen following user fee removal. For example, some publicly supported health facilities were not able to provide complete services such as drugs, medical examinations, etc., to match the increased demand so that individuals had to rely on the private market for these services (Hadley, 2011; Masiye et al., 2010; Ministry of Health, 2007; Onde, 2009).

This evidence on supply side constraints may also explain our finding that, on average, household medical spending did not reduce, even after the removal of user fees (recall that we look at spending regardless of utilizing any facility, and hence also capture spending in private drug shops). This finding is supported by studies showing that spending in both publicly supported facilities and private pharmacies/drug shops remained prevalent. For example, Hadley (2011) found that even after user fees were removed, there were additional payments for things such as having children

Table 6
Short- and long-term marginal effects of the removal of user fees on overall medical spending

	Any spending (Probit part) (1)	Conditional spending (GLM part) (2)	Unconditional spending (Probit + GLM) (3)
<i>Short-term effect</i>			
T1 × yr2006	−0.266(0.036)***	3.290(14.751)	−6.216(12.653)
<i>Long-term effects</i>			
T1 × yr2010	−0.193(0.047)***	17.055(12.546)	2.755(6.598)
T2 × yr2010	−0.087(0.050)*	6.775(10.981)	1.840(5.130)
<i>Pre-treatment effects</i>			
T1 × yr2002	0.005(0.040)	−12.250(24.527)	−3.773(11.093)
T2 × yr2002	0.003(0.052)	−19.567(22.259)	−6.797(10.372)
T2 × yr2006	−0.050(0.046)	0.706(15.928)	−2.249(10.444)
Observations	26,705	13,315	
T1 group baseline mean	0.607	17.177	
T2 group baseline mean	0.580	13.851	
Control group baseline mean	0.706	51.459	

Note: Table shows the effect of the first and second wave of the removal policy, T1 and T2 respectively, on medical spending for all individual who reported being sick or injured. Bootstrap standard errors (1000 replications) clustered at the district level are reported in parentheses. Estimates are from a two-part model where the first part (Column 1) models the probability of incurring any spending (extensive margin) using a probit model of Eqn. (1) with the 1998 year dummy dropped. The marginal effects of the probit model are computed according to Eqn. (3). The second part (Column 2) are effects of the policy on the intensive margin—amount of spending conditional on one having incurred any. The amounts are marginal effects computed according to Eqn. (3) from a GLM regression model (Gamma distribution and log-link) on Eqn. (1). Column (3) presents the effect of the policy on the whole conditional mean of medical spending (combination of the extensive and intensive margin). The lower panel gives the baseline (2004) mean of each outcome for treated and control groups. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include household size, distance to health facility, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.

*Significant at 10%; **Significant at 5%; ***Significant at 1%.

weighed at the facility, informal medical insurance arrangements, and book fees for maintaining medical records, among others. These additional charges are informal since all primary health services were to be free at the point of use. Hadley (2011) also found that even when drugs were available in health facilities, patients would be given prescriptions to buy drugs from private drug stores, most of which were owned by health facility workers.

Under such circumstances, it is not surprising that even if the proportion of people incurring any medical spending reduced, some people experienced an economically relevant, albeit statistically insignificant, increase in spending. It should be noted that even if the proportion of people incurring any spending (conditional spending) may have reduced, the absolute numbers may still increase due to the increase in utilization of health care. These people were, on average, now spending more than before the removal of user fees, and as mentioned earlier, this could have been for things like drugs in the private pharmacies and informal charges. User fee removal policies may thus induce such unintended effects especially if supply—side constraints are not looked at carefully. The increase in conditional spending appears to have diluted the reduction in spending that those who did not incur any spending experienced, so that average spending did not change significantly after removal of user fees.¹⁰ Evidence from Uganda also shows that removal of user fees did not reduce household medical spending (Nabyonga Orem et al., 2011; Xu et al., 2006). Studies in Zambia have documented catastrophic spending for individuals visiting publicly supported health facilities even after the nationwide removal of user fees (Masiye et al., 2016). As mentioned in Section 1, the health insurance literature shows that even when some segments of popu-

lation are provided with free care by fully subsidizing them, household medical spending may still not fall.

Coming back to findings on utilization, our results can be compared, and contrasted, with experimental evidence. A few of these studies find that removing user fees or providing health insurance does not increase overall use (uptake) of care but simply generate a switching effect (Levine, Polimeni, & Ramage, 2016; Powell-Jackson et al., 2015). Although we find evidence of switching in our natural experiment, our findings show that the increase in overall use of health service (uptake effect) was actually larger, and much stronger, than the switching effect. To the best of our knowledge, our study is the first to separate the magnitude of uptake and switching effects for a complex policy implemented at national scale. Mwabu et al. (1995) conducted a descriptive study of the effect of a temporary suspension of user fees in two districts of Kenya and found that utilization of publicly supported facilities increased while that of private facilities declined. Their analysis was however unable to tease out the magnitude of the increase in overall use of health services. Moreover, since it was a before and after study without a control group, it is hard to determine whether or not the observed changes in utilization patterns were due to the removal of user fees or other concurrent factors.

The welfare implication of our findings are not straight forward. On one side, welfare may have increased if the improved access to care by individuals of poor socioeconomic backgrounds reduced the duration as well as severity of their sickness. In this instance, the effect of such sickness on their labor productivity and earnings would have been eased. On the other side, if care was of poor quality, increased access may not have impacted positively on their health. In this case, the removal policy may have reduced welfare as resources were not being used efficiently.

(a) Limitations

The key limitation of our study is the non-availability of GPS information to exactly classify which households could have been

¹⁰ That total medical spending was unchanged is also consistent with the National Health Accounts (NHA) for Zambia which reviewed that the household share in Total Health Expenditure (THE) was 27% in 2006, increasing from 26.8 in 2005 (Ministry of Health, 2009). Actually, data show that household share in THE was, on average, higher in the period following the removal of user fees (World Health Organization, 2015).

treated in the second wave of the user fee removal policy, T2. Despite the fact that our classification performs extremely well, the long-term estimates are lower bound if there was any misclassification.

Secondly, our measure of medical spending does not fully characterize health expenditure because Living Conditions Monitoring Surveys only collects spending on consultation, purchase of drugs, and examinations. This leaves out other significant health care-related costs such as transportation which in some cases are more significant than payments at the point of care.

7. Conclusion

Consensus appears to have emerged that removing user fees is an effective strategy for improving Universal Health Coverage (UHC). A number of low- and middle-income countries have since removed user fees while others are considering to do so. The consensus draws heavily on studies that report dramatic increases in utilization. However, due to methodological and data limitations, the demand consequences, including changes in household spending, of such complex policy interventions still remain unclear.

To contribute in filling this gap in the literature, we exploit the step-wise implementation of removal of user fees in Zambia to create treatment and control groups, and use large nationally representative household surveys for the period 1998–10 to estimate effects in a difference-in-difference framework.

We find that the two waves of the user fee removal policy significantly increased overall rates of care seeking and led to a small shift in care seeking from private to publicly supported facilities in both the short and long terms. Overall rates of care seeking increased more for individuals from lower socioeconomic backgrounds. Despite the fact that the removal policy reduced the financial barrier to access by reducing the proportion of individual incurring any spending (extensive margin), and thus significantly increasing utilization, it exerted an upward pressure on spending for those who were incurring any spending (intensive margin/conditional spending). The intensive and extensive margin worked in opposite directions to leave total (unconditional) medical spending unchanged.

We conclude that free health care policies such as removing user fees may improve access to care but more attention should be given to supply-side factors, such as quality of care and adequacy of funding, to ensure that improved access translates into health improvements and reduced household expenditure.

Acknowledgment

We acknowledge the support given by the Fogarty International Center of the National Institutes of Health and PEPFAR under Award Number D43TW001035, the University of Bergen, the University of Zambia, and Grant No. 5R24TW008873 administered by the Fogarty International Center of the National Institutes of Health and funded

by OGAC and OAR. We would like to thank the Central Statistical Office (CSO) of Zambia for providing data used in this publication.

We also wish to thank two anonymous reviewers for their extremely helpful comments.

Appendix A

A.1. Assessing treatment classification for T2

The second wave of the removal policy, T2, extended the removal of user fees to the T2 group—rural areas of the remaining 18 districts. However, the rural/urban classification of areas within these 18 districts in our data is not exactly the same as defined in the removal policy. The classification in the data uses a more detailed criterion while the removal policy defined rural areas only on the basis of radius distance from the district administration centers of each of the 18 districts. Specifically, user fees were removed in all primary health facilities located outside a radius of 15 or 20 km from the districts centers, depending on localization relative to the line of rail. Ideally, we would need GPS coordinates for both households and facilities to precisely determine which household is actually located in a treated area. However, GPS coordinates were not collected in the surveys.

Since distances in the removal policy were determined in such a way that health facilities exempted from user fees would serve rural households, we use the rural classification from the Central Statistical Office (CSO) in the data to define a household as treated (belonging to the T2 group) if it is located in an enumeration area that is classified as rural. The classification of enumeration areas as either rural or urban by CSO is based on population size, economic activity (agricultural or not), and presence of basic modern facilities such as piped water, tarred roads, post office, and other services (CSO, 2012).

To assess how well this rural/urban definitions in the data (by CSO) performs in distinguishing treated from non-treated individuals, we check how the proportion of individuals incurring any medical spending changed between rural areas (supposedly treated-T2 group) and urban areas (supposedly controls). It is expected that after the removal of user fees in 2007, this measure falls significantly in areas that are supposedly treated, while it should be non-decreasing in areas deemed as controls. The data show that while the proportion of individuals incurring any spending in the supposedly treated areas fell by almost 20 percentage points (pp) between 2006 and 2010, it was almost unchanged in the supposed control (Table 7).

A.2. Fully flexible model

The model proposed by Mora and Reggio (2012) provides an alternative way of estimating the treatment effect of the removal

Table 7
Trends in medical spending in publicly supported health facilities

Variable	T2 Group				Control Group			
	2002	2004	2006	2010	2002	2004	2006	2010
Incurred any spending (%)	60.75	82.57	59.38	40.17	69.08	80.99	79.45	81.17
Conditional spending	8.35	12.55	12.85	7.17	42.68	43.10	45.53	30.68
Number of observations	873	701	673	1,032	1,969	1,972	1,586	2,391

Note: Table shows means medical spending for the T2 group and control. Note that, as opposed to computing spending for all individuals who reported being sick or injured as we did in the main analysis, medical spending here is conditional on utilizing publicly supported health facilities. Conditional Health Spending is in Zambian Kwacha (ZMW) at 2010 prices. The ZMW/USD exchange rate in 2010 was 5 Kwacha per USD.

Table 8
Short- and long-term effects of the removal of user fees: Fully flexible model (Mora, 2015)

PANEL A: Short- and long-term effect under different parallel assumptions							
	Overall utilization			Prob. of medical spending		Conditional spending	
	Parallel-1	Parallel-2	Parallel-3	Parallel-1	Parallel-2	Parallel-1	Parallel-2
T1 × yr2006	0.098*** (0.034)	0.108* (0.057)	0.090 (0.109)	−0.236*** (0.035)	−0.258*** (0.060)	0.060 (0.120)	−0.133 (0.569)
T1 × yr2010	0.102** (0.043)	0.119 (0.096)	0.066 (0.270)	−0.240*** (0.036)	−0.286*** (0.102)	0.019 (0.145)	−0.369 (0.388)
Observations	28,232	28,232	28,232	23,783	23,783	11,892	11,892
PANEL B: Test for Common Pre-treatment trends and equality of short- and long-term effects							
	Overall utilization		Prob. of medical spending		Conditional spending		
	F	p-value	F	p-value	F	p-value	
<i>Test for pre-treatment trends</i>							
H_0 : Common Trends	1.051	.5913	0.2805	0.5964	1.772	0.183	
<i>Test of equality of short- and long-term effects</i>							
Under Parallel-1							
H_0 : T1 × yr2006 = T1 × yr2010	0.018	0.892	0.024	0.877	0.103	0.748	
Under Parallel-2							
H_0 : T1 × yr2006 = T1 × yr2010	0.067	0.795	0.230	0.631	1.356	0.244	
Under Parallel-3							
H_0 : T1 × yr2006 = T1 × yr2010	0.021	0.885					

Note: The table shows OLS estimates in a fully flexible DD model for the T1 group and the control. The short- and long-term effect under parallel-1 assumes parallel paths and uses only 1 pre-treatment period (2004). The effects under parallel-2 uses the 2004 and 2002 pre-treatment data set and assumes parallel growths in outcomes. Under parallel-3, all pre-treatment data is used (2004, 2002 & 1998). Parallel-3 assumes quadratic growth in outcomes. The first test is a test of common trends which is equivalent to a test that the 3 (or 2 under medical spending) parallel path assumptions are equal. The last three tests checks whether the short-term and long-term effects are equal under each parallel assumption. Standard errors clustered at the district level are reported in brackets.

*Significant at 10%; **Significant at 5%; ***Significant at 1%.

policy while at the same time testing for the parallel trends assumption. As opposed to estimating one treatment effect of an intervention for each post-treatment data point (e.g., 2006 or 2010)—irrespective of the number of pre-treatment data points—the model proposed by Mora and Reggio (2012) allows one to estimate r treatment effects for each post-treatment data point, where r is the number of pre-treatment data points and each effect, β_r , is estimated under a different parallel assumption. Only when treatment effects under different parallel assumptions are equivalent is the parallel trends assumption met, and our assumption of common trends valid.

The advantage with the method proposed by Mora and Reggio (2012) is that even if the estimates from one post-treatment data point, e.g., for T1 × 2006, under different parallel assumptions are not equivalent, one is able to check how the effect varies, in both magnitude and significance across different parallel assumptions. Actually, Mora and Reggio (2012) show that the practice of allowing for flexibility in difference-in-difference models by including terms such as linear or quadratic trends imply different identifying assumptions from the ones the authors refer to, so that the reported effects may not be the true treatment effects.¹¹

We focus on the first wave of removals, T1, and estimate the fully flexible model on three outcomes, namely, overall (any) utilization, whether an individual incurred any spending or not, and the amount of spending conditional on spending. Given that we have three (two) pre-treatment periods for overall utilization (medical spending), we are able to estimate three (two) short-term and three (two) long-term effects of T1. For example, the computation of the three short-term and three long-term effect for overall utilization follows three stages. Firstly, the first set of short- and long-term effects is computed using the baseline data (2004) only. The identifying assumption in this case is the parallel paths assumption, which Mora and Reggio (2012) term parallel-1.

Then the second set of short-term and long-term effects is estimated using both the 2004 and 2002 pre-treatment data. The assumption here is parallel growths, called parallel-2 and it is equivalent to allowing for linear trends in a difference-in-difference model. The last set of effects are computed using all the pre-treatment datasets (1998, 2002, and 2004). This is called parallel-3 and its equivalent to allowing for quadratic trends in a difference-in-difference model.

We make use of the Stata package by Mora (2015) and focus on T1 which gives the short- and long-term effects. The models are estimated using ordinary least squares (OLS). Thus utilization and the first part of the two part model of expenditure applies a linear probability model (LPM) while the second part is OLS on log transformed expenditure.

Estimation results are presented in Table 8. The results are broadly consistent with our finding in the main analysis although the estimates and standard errors are slightly larger. Panel B shows test statistics and p -values for the parallel trends assumption test and the equivalence of short- and long-term effects. The common trends assumption was met for all outcomes, i.e. we do not reject H_0 of common trends at the 10% level. The tests also reveals that there was not statistically significant difference between short- and long-term effects.

A.3. Omitted variables and empty models

Although we controlled for individual covariates and included regional fixed effects as well as year by region effects, there may be concerns that other district-level time varying characteristics may bias our estimators. To gauge this possibility, we were able to collect data on district-level funding. Unfortunately, district funding was not available for 1998, 2002, and 2010. Hence, we could not include this variable in the main analysis. However, the two years, 2004 (before the first intervention) and 2006 (after) could be used to assess the sensitivity of estimates to possible omitted time varying district variables. District funding is perhaps one of the key variables that affects performance, quality of care,

¹¹ They applied their method to papers published in top economics journals and find that when flexible dynamics are applied, the significance of the results is affected in 6 of the 13 cases and only 3 in 9 cases was the common trends assumption satisfied.

Table 9
Short-term effect of user fee removal on overall, public, and private facility utilization

	Overall Utilization (1)	Public Utilization (2)	Private Utilization (3)
T1 × yr2006	0.079(0.033)**	0.120(0.032)***	-0.041(0.019)**
Observations	11,402	11,402	11,402
T1 group baseline mean	0.521	0.497	0.024
Control group baseline mean	0.575	0.472	0.103

Note: Table shows the effect of the first wave of the removal policy, T1, on utilization of all facilities (Column 1), publicly supported facilities (Column 2), and private facilities (Column 3) controlling for district-level funding. Funding data were only available for 2004 and 2006. Thus, estimates are only based on the 2004 and 2006 LCMS surveys. Reported coefficients are estimates from DD linear probability models (LPM) corresponding to Eqn. (1) and are interpreted as percentage points. Standard errors clustered at the district level are reported in parentheses. The lower panel give the baseline (2004) mean of each outcome for treated and control groups. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include district-level funding, distance to health facility, household size, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.

*Significant at 10%; **Significant at 5%; ***Significant at 1%.

Table 10
Short- and long-term marginal effects of the removal of user fees on overall medical spending

	Any spending (Probit part) (1)	Conditional spending (GLM part) (2)	Unconditional spending (Probit + GLM) (3)
T1 × yr2006	-0.265(0.037)***	2.946(11.662)	-5.355(9.991)
Observations	11,402	6,360	
T1 group baseline mean	0.607	17.177	
Control group baseline mean	0.706	51.459	

Table shows the effect of the first wave of the removal policy, T1, on medical spending for all individual who reported being sick or injured controlling for district-level funding. Funding data were only available for 2004 and 2006. Thus, estimates are only based on the 2004 and 2006 LCMS surveys. Bootstrap standard errors (1000 replications) clustered at the district level in parentheses. Estimates are from a two-part model where the first part (Column 1) models the probability of incurring any spending (extensive margin) using a probit model of Eqn. (1) with the 1998 year dummy dropped. The marginal effects of the probit model are computed according to Eqn. (3). The second part (Column 2) are effects of the policy on the intensive margin—amount of spending conditional on one having incurred any. The amounts are marginal effects computed according to Eqn. (3) from a GLM regression model (Gamma distribution and log-link) on Eqn. (1). Column (3) presents the effect of the policy on the whole conditional mean of medical spending (combination of the extensive and intensive margin). The lower panel gives the baseline (2004) mean of each outcome for treated and control groups. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include household size, distance to health facility, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.

*Significant at 10%; **Significant at 5%; ***Significant at 1%.

and availability of essential services and thus utilization as well as individual medical spending. The inclusion of this key district-level variable did not change our results (See Table 9 and 10 in the Appendix).

Still on the issue of control variables, the performance of difference-in-difference models can be assessed on how sensitive the estimates are to the addition of covariates. We estimated all models without control variables and with a limited set of control variables and results are presented (Table 11 and 12 in the appen-

Table 11
Short- and long-term effect of user fee removal on provider choice

	Overall Utilization (1)	Public Utilization (2)	Private Utilization (3)
<i>Short-term effect</i>			
T1 × yr2006	0.077(0.033)**	0.092(0.037)**	-0.015(0.021)
<i>Long-term effects</i>			
T1 × yr2010	0.10.8(0.042)**	0.113(0.040)***	-0.004(0.016)
T2 × yr2010	0.154(0.047)***	0.159(0.025)***	-0.005(0.015)
<i>Pre-treatment</i>			
T1 × yr1998	0.021(0.039)	0.012(0.031)	0.009(0.023)
T2 × yr1998	0.027(0.059)	0.012(0.056)	0.015(0.027)
T1 × yr2002	0.032(0.032)	0.029(0.035)	0.004(0.015)
T2 × yr2002	0.033(0.052)	0.034(0.050)	-0.001(0.014)
T2 × yr2006	0.072(0.062)	0.072(0.060)	-0.001(0.020)
Observations	36,319	36,319	36,319
T1 group baseline mean	0.521	0.497	0.024
T2 group baseline mean	0.447	0.433	0.014
Control group baseline mean	0.575	0.472	0.103

Note: Estimates from a difference-in-difference linear probability model (LPM) corresponding to Eqn. (1). Estimates are interpreted as percentage points. Standard errors clustered at the district level are reported in parentheses. All models are conditional on reporting sickness or injury in the two weeks prior to the survey. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include distance to health facility, household size, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.

*Significant at 10%; **Significant at 5%; ***Significant at 1%.

Table 12
Short- and long-term marginal effects of the removal of user fees on medical spendings

	Any spending (Probit part) (1)	Conditional spending (GLM part) (2)	Unconditional spending (Probit + GLM) (3)
<i>Short-term effect</i>			
T1 × yr2006	-0.223(0.032)***	5.080(12.340)	-1.824(6.487)
<i>Long-term effects</i>			
T1 × yr2010	-0.186(0.036)***	37.494(17.959)**	12.011(6.472)*
T2 × yr2010	-0.070(0.050)	10.905(17.075)	9.144(7.563)
<i>Pre-treatment effects</i>			
T1 × yr2002	-0.025(0.041)	4.747(13.625)	8.663(7.118)
T2 × yr2002	-0.001(0.047)	0.262(14.622)	8.676(7.822)
T2 × yr2006	-0.035(0.038)	6.705(14.844)	2.423(6.904)
Observations	31,039	15,175	
T1 group baseline mean	0.607	17.177	
T2 group baseline mean	0.580	13.851	
Control group baseline mean	0.706	51.459	

Note: Table shows marginal effects from a DD two part model and the combined effect of the two parts. All amounts are in Zambian Kwacha at 2010 prices. Bootstrap standard errors (1,000 replications) clustered at the district level are reported in parentheses. (1) presents probit model marginal effects at the extensive margin of medical spending (whether one incurred any spending or not). In (2), marginal effect at the intensive margin (those with positive spending) are presented. (3) presents the effect of the policy on the whole conditional mean of medical spending (combination of the extensive and intensive margin). All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include household size, distance to health facility, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.

*Significant at 10%; **Significant at 5%; ***Significant at 1%.

Table 13

Differences in the likelihood of reporting sickness between the treated (T1 & T2) and the control group

T1 × yr1998	T1 × yr2002	T1 × yr2006	T1 × yr2010	T2 × yr1998	T2 × yr2002	T2 × yr2006	T2 × yr2010
−0.015 (0.013)	−0.004 (0.012)	0.000 (0.008)	0.010 (0.010)	0.004 (0.012)	0.020 (0.013)	0.021 (0.013)	0.039* (0.023)
Observations							361,336

Note: Estimates from a difference-in-difference linear probability model (LPM) corresponding to Eqn. (1). Estimates are interpreted as percentage points. Standard errors clustered at the district level are reported in parentheses. The model includes a full set of year dummies, regional effects, and region by year interaction effects. Covariates include, household size, distance to health facility, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.

*Significant at 10%; **Significant at 5%; ***Significant at 1%.

dix). Estimates of the effect are broadly similar between models with and without covariates.

A.4. Differential selection into reporting sickness or injury

Since utilization is driven by the likelihood of reporting sickness or injury, and vice versa, there is a possibility of bias where the probability of reporting sickness/injury changes differently between treated and control areas. Table 13 shows that the proportion of individuals reporting sickness in treated and control areas did not change differently. However, the coefficient is significant at the 10% level for the T2 group in 2010. Of course since this particular regression is conducted on the whole dataset (361,336 observations), finding a significant effect at 10% may not necessarily imply bias. It may however also suggest that the removal of user fees led to a slight increase in reporting sickness or injury which is a form of moral hazard.

A.5. Falsification tests and compositional changes

There is still a possibility that other events affected treated and control groups differently and hence generated a false “treatment effect”. Although we see no other apparent reason or possible concurrent event that would significantly reduce the percentage of individuals incurring positive spending while at the same time drastically changing utilization, it is important to query this possibility. One way of doing this is to conduct falsification tests where we ask if the removal of user fees significantly affected covariates that are expected to be unresponsive to the removal policy. To achieve this, we fit a simple difference-in-difference models to all our covariates.¹²

In particular, very significant changes in key covariates such as employment status, distance to a health facility, education, especially in the post-treatment period, is evidence of other things happening other than removal of user fees. Tables 2 and 3, Columns 3–6, in Section 3(b(ii)) show that almost all key covariates were not affected differently in the T1^u group compared to the control.

A.6. Randomization tests

High correlation in observations over time (serial correlation) or within an areas (intra-cluster correlation) has a potential to underestimate standard errors. Since the user fee removal policy was implemented at the district level one needs to cluster the standard errors at the district level or aggregate the data in order to deal

¹² For example, to check for differential changes between T1 and control as reported in Table 2, we estimate difference-in-difference models of the form:

$$X_{citr} = \phi_0 + \phi_1 yr1998 + \phi_2 yr2002 + \phi_3 yr2006 + \phi_4 yr2010 + \phi_5 T1_d + \phi_6 (T1_d \times yr1998) + \phi_7 (T1_d \times yr2002) + \phi_8 (T1_d \times yr2006) + \phi_9 (T1_d \times yr2010) + \epsilon_{it}$$

where X_{citr} is covariate c for individual i residing in district j at time t and $T1_d = 1$ if district j was treated. We report p -values of the interaction effects ϕ_6 , ϕ_7 , ϕ_8 , and ϕ_9 in Table 2 and Table 3 for T2.

with the problem of within cluster correlation (Bertrand et al., 2004). We clustered standard errors at the district level.

Serial correlation is also a possible threat to the validity of our findings, and we have not addressed it yet. In a difference-in-difference strategy, outcomes are mostly positively correlated over time, and this is worsened by the fact that the treatment indicator itself is highly serially correlated; because it is kept on-equal to one-throughout the study period. This severely underestimates the standard errors of conventional DD models. Using randomly generated placebo interventions, Bertrand et al. (2004) found a false rejection rate, at the 5% level, of up to 44% even after standard errors are clustered at the group level or the data are aggregated to the group level. Since these were placebo interventions, significant effects (rejection rates) were expected in approximately 5% of the simulations.

To assess the extent of this problem, we conducted randomization experiments similar to those of Bertrand et al. (2004). This involves the 54 districts (T1 group) and the urban areas of the 18 districts (control group). If serial correlation is not a problem, the rejection rate in simulations where treatment is turned on in both 2006 and 2010 (call it the pooled model) should be close to the rejection rate where treatment effect is only turned on in either 2010 or 2006.¹³ We found a 6.1% rejection rate in simulation of the pooled model while the other model yielded a rejection rate of 7.1%, suggesting that serial correlation is not a problem in our setup.

References

- Acemoglu, D. (2010). Theory, general equilibrium, and political economy in development economics. *Journal of Economic Perspectives*, 24(3), 17–32.
- Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1), 123–129.
- Akashi, H., Yamada, T., Huot, E., Kanak, K., & Sugimoto, T. (2004). User fees at a public hospital in Cambodia: Effects on hospital performance and provider attitudes. *Social Science & Medicine*, 58(3), 553–564.
- Akin, J. S., Birdsall, N., & De Ferranti, D. M. (1987). *Financing health services in developing countries: An agenda for reform* (vol. 34). World Bank Publications.
- Asfaw, A., von Braun, J., & Klasen, S. (2004). How big is the crowding-out effect of user fees in the rural areas of Ethiopia? Implications for equity and resources mobilization. *World Development*, 32(12), 2065–2081.
- Ashraf, N., Fink, G., & Weil, D. N. (2014). Evaluating the effects of large scale health interventions in developing countries: The Zambian Malaria Initiative. In S. Edwards, S. Johnson, & D. N. Weil (Eds.). *African successes: Human capital* (Vol. 2). University of Chicago Press.
- Ataguba, J. E.-O., & Goudge, J. (2012). The impact of health insurance on health-care utilisation and out-of-pocket payments in South Africa. *Geneva Paper R I-Iss P*, 37(4), 633–654.
- Barber, S., Bonnet, F., & Bekedam, H. (2004). Formalizing under-the-table payments to control out-of-pocket hospital expenditures in Cambodia. *Health Policy and Planning*, 19(4), 199–208.
- Basu, S., Andrews, J., Kishore, S., Panjabi, R., & Stuckler, D. (2012). Comparative performance of private and public healthcare systems in low-and middle-income countries: A systematic review. *PLoS Medicine*, 9(6), e1001244.

¹³ In this experiment, for the pooled model, we randomly select 54 districts (regardless of treatment status) and designate them as “treatment” and let them keep the treatment status in 2010. We then record the treatment effect and standard error. We conduct this procedure 1,000 times each time taking note of the standard errors. In an alternative model, we randomly draw 54 districts and designate them as treated. In this case, however, treatment is turned on in either 2006 or 2010 but not both years.

- Beegle, K., Weerdt, J. D., Friedman, J., & Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of Development Economics*, 98(1), 3–18. Symposium on Measurement and Survey Design.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275.
- Bigdeli, M. & Ir, P. (2010). A role for user charges? thoughts from health financing reforms in Cambodia. Health systems financing—the path to universal coverage. The World Health Report.
- Bitran, R. (2014). Universal health coverage and the challenge of informal employment: Lessons from developing countries. Health, nutrition, and population (hnp) discussion paper, World Bank Group, Washington DC.
- Blas, E., & Limbambala, M. (2001). User-payment, decentralization and health service utilization in Zambia. *Health Policy and Planning*, 16(suppl 2), 19–28.
- Campbell, J., Oulton, J. A., McPake, B., & Buchan, J. (2011). Increasing access to 'free' health services: Are health workers not a missing link? *International Journal of Clinical Practice*, 65(1), 12–15.
- Carasso, B., Lagarde, M., Cheelo, C., Chansa, C., & Palmer, N. (2012). Health worker perspectives on user fee removal in Zambia. *Human Resources for Health*, 10(40).
- Cheelo, C., Chama, C., Pollen, G., Carasso, B., Palmer, N., Jonsson, D., ... Chansa, C. (2010). *Do user fee revenues matter: Assessing the influences of the removal of user fees on health financial resources in Zambia*. Department of Economics, The University of Zambia.
- Ching, P. (1995). User fees, demand for children's health care and access across income groups: The Philippine case. *Social Science & Medicine*, 41(1), 37–46.
- Cohen, J., & Dupas, P. (2010). Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *The Quarterly Journal of Economics*, 125(1), 1–45.
- CSO (2012). Living Conditions Monitoring Survey Report 2006 and 2010. Survey report, Central Statistical Office, Government of the Republic of Zambia, Lusaka, Zambia.
- CSO (2016). 2015 Living Conditions Monitoring Survey Report. Survey report, Central Statistical Office, Government of the Republic of Zambia, Lusaka, Zambia.
- Das, J., Hammer, J., & Leonard, K. (2008). The quality of medical advice in low-income countries. *Journal of Economic Perspectives*, 22(2), 93–114.
- Dzakpasu, S., Powell-Jackson, T., & Campbell, O. M. (2014). Impact of user fees on maternal health service utilization and related health outcomes: A systematic review. *Health Policy and Planning*, 29(2), 137–150.
- Falkingham, J. (2004). Poverty, out-of-pocket payments and access to health care: Evidence from Tajikistan. *Social Science & Medicine*, 58(2), 247–258.
- Fink, G., Robyn, P. J., Sié, A., & Sauerborn, R. (2013). Does health insurance improve health? Evidence from a randomized community-based insurance rollout in rural Burkina Faso. *Journal of Health Economics*, 32(6), 1043–1056.
- Frondel, M., & Vance, C. (2013). On interaction effects: The case of Heckit and two-part models. *Journal of Economics and Statistics*, 233(1), 22–38.
- Gertler, P., Locay, L., & Sanderson, W. (1987). Are user fees regressive? *Journal of Econometrics*, 36(1), 67–88.
- Gilson, L. (1997). The lessons of user fee experience in Africa. *Health Policy and Planning*, 12(3), 273.
- Godlonton, S., & Okeke, E. N. (2016). Does a ban on informal health providers save lives? Evidence from Malawi. *Journal of Development Economics*, 118, 112–132.
- Hadley, M. (2011). Does increase in utilisation rates alone indicate the success of a user fee removal policy? A qualitative case study from Zambia. *Health Policy*, 103(2–3), 244–254.
- Institute for Health Metrics and Evaluation (2014). Assessing Facility Capacity, Costs of Care, and Patient Perspectives. Policy Report, IHME, Seattle, WA.
- Institute for Health Metrics and Evaluation, Human Development Network, TheWorld Bank (2013). The Global Burden of Disease: Generating Evidence, Guiding Policy – Sub-Saharan Africa Regional Edition. Policy Report, IHME, Seattle, WA.
- Jolliffe, D. (2001). Measuring absolute and relative poverty: The sensitivity of estimated household consumption to survey design. *Journal of Economic and Social Measurement*, 27(1).
- Kremer, M., & Miguel, E. (2007). The illusion of sustainability. *The Quarterly Journal of Economics*, 122(3), 1007–1065.
- Lagarde, M., Barroy, H., & Palmer, N. (2012). Assessing the effects of removing user fees in Zambia and Niger. *Journal of Health Services Research & Policy*, 17(1), 30–36.
- Lagarde, M., & Palmer, N. (2008). The impact of user fees on health service utilization in low- and middle-income countries: How strong is the evidence? *Bulletin of the World Health Organization*, 86, 839–848C.
- Leonard, D. K., Bloom, G., Hanson, K., O'Farrell, J., & Spicer, N. (2013). Institutional solutions to the asymmetric information problem in health and development services for the poor. *World Development*, 48, 71–87.
- Levine, D., Polimeni, R., & Ramage, I. (2016). Insuring health or insuring wealth? An experimental evaluation of health insurance in rural Cambodia. *Journal of Development Economics*, 119, 1–15.
- Lim, S. S., Stein, D. B., Charrow, A., & Murray, C. J. (2008). Tracking progress towards universal childhood immunisation and the impact of global initiatives: A systematic analysis of three-dose diphtheria, tetanus, and pertussis immunisation coverage. *The Lancet*, 372(9655), 2031–2046.
- Lindkvist, I. (2013). Informal payments and health worker effort: A quantitative study from Tanzania. *Health Economics*, 22(10), 1250–1271.
- Litvack, J. I., & Bodart, C. (1993). User fees plus quality equals improved access to health care: Results of a field experiment in Cameroon. *Social Science & Medicine*, 37(3), 369–383.
- Liu, H., & Zhao, Z. (2014). Does health insurance matter? Evidence from China's urban resident basic medical insurance. *Journal of Comparative Economics*, 42(4), 1007–1020.
- Masiye, F., Chitah, B. M., & McIntyre, D. (2010). From targeted exemptions to user fee abolition in health care: Experience from rural Zambia. *Social Science & Medicine*, 71(4), 743–750.
- Masiye, F., Kaonga, O., & Kirigia, J. M. (2016). Does user fee removal policy provide financial protection from catastrophic health care payments? Evidence from Zambia. *PLoS ONE*, 11(1), 1–15.
- Mataria, A., Luchini, S., Daoud, Y., & Moatti, J.-P. (2007). Demand assessment and price-elasticity estimation of quality-improved primary health care in Palestine: A contribution from the contingent valuation method. *Health Economics*, 16(10), 1051–1068.
- Meessen, B., Damme, W. V., Tashobya, C. K., & Tibouti, A. (2007). Poverty and user fees for public health care in low-income countries: Lessons from Uganda and Cambodia. *The Lancet*, 368(9554), 2253–2257.
- Meessen, B., Hercot, D., Noirhomme, M., Ridde, V., Tibouti, A., Tashobya, C. K., & Gilson, L. (2011). Removing user fees in the health sector: A review of policy processes in six sub-Saharan African countries. *Health Policy and Planning*, 26(suppl 2), ii16–ii29.
- Ministry of Health (2007). Health Sector Joint Annual Review Report, 2006. Technical report, Ministry of Health, Government of the Republic of Zambia, Lusaka, Zambia.
- Ministry of Health (2009). Zambia National Health Accounts: 2003–2006. Final report, Ministry of Health, GRZ, Lusaka, Zambia.
- MoH (2007). Revised Guidelines on the Removal of User Fees in Government and Mission Health Facilities in Zambia. Technical report, Ministry of Health, Government of the Republic of Zambia, Lusaka, Zambia.
- MoH (2011). National Health Strategic Plan 2011–2015. Technical report, Ministry of Health, Government of the Republic of Zambia, Lusaka, Zambia.
- Mora, R. (2015). didq: A command for treatment-effect estimation under alternative assumptions. *Stata Journal*, 15(3), pp. 796–808(13).
- Mora, R. Reggion, I. (2012). Treatment effect identification using alternative parallel assumptions. Working Paper 12–33, Economic series (48), Departamento de Economía, Universidad Carlos III de Madrid.
- Mwabu, G., Mwanjia, J., & Liambila, W. (1995). User charges in government health facilities in Kenya: Effect on attendance and revenue. *Health Policy and Planning*, 10(2), 164–170.
- Nabyonga Orem, J., Mapunda, M., Musango, L., & Mugisha, F. (2013). *Long-term effects of the abolition of user fees in Uganda*. African health monitor paper. World Health Organization.
- Nabyonga Orem, J., Mugisha, F., Kirunga, C., Macq, J., & Criel, B. (2011). Abolition of user fees: The Uganda paradox. *Health Policy and Planning*, 26(suppl 2), ii41–ii51.
- Nguyen, C. V. (2012). The impact of voluntary health insurance on health care utilization and out-of-pocket payments: New evidence for Vietnam. *Health Economics*, 21(8), 946–966.
- Onde, R. (2009). Impact of abolishment of user fees in rural health centres: A case of Chongwe district. Master's thesis, School of Business, Copperbelt University, Kitwe, Zambia.
- Powell-Jackson, T., Hanson, K., Whitty, C. J., & Ansah, E. K. (2014). Who benefits from free healthcare? Evidence from a randomized experiment in Ghana. *Journal of Development Economics*, 107, 305–319.
- Powell-Jackson, T., Macleod, D., Benova, L., Lynch, C., & Campbell, O. M. R. (2015). The role of the private sector in the provision of antenatal care: A study of demographic and health surveys from 46 low- and middle-income countries. *Tropical Medicine & International Health*, 20(2), 230–239.
- Pradhan, M. (2009). Welfare analysis with a proxy consumption measure: Evidence from a repeated experiment in Indonesia. *Fiscal Studies*, 30(3–4), 391–417.
- Ridde, V., & Haddad, S. (2009). Abolishing user fees in Africa. *PLoS Med*, 6(1), e1000008.
- Sandefur, J., & Glassman, A. (2015). The political economy of bad data: Evidence from African survey and administrative statistics. *The Journal of Development Studies*, 51(2), 116–132.
- Sheikh, K., Josyula, L. K., Zhang, X., Bigdeli, M., & Ahmed, S. M. (2017). Governing the mixed health workforce: Learning from Asian experiences. *BMJ Global Health*, 2(2).
- Soucat, A., Gandaho, T., Levy-Bruhl, D., de Bethune, X., Alihonou, E., Ortiz, C., ... Knippenberg, R. (1997). Health seeking behaviour and household health expenditures in Benin and Guinea: The equity implications of the Bamako Initiative. *The International Journal of Health Planning and Management*, 12(S1), S137–S163.
- Thornton, R. L. (2008). The demand for, and impact of, learning hiv status. *American Economic Review*, 98(5), 1829–1863.
- UNDP (2014). Human Development Report 2014: Reducing Vulnerabilities and Building Resilience. Health, nutrition, and population (hnp) discussion paper, United Nations Development Program, New York, USA.
- Wagstaff, A. (2010). Social health insurance reexamined. *Health Economics*, 19(5), 503–517.
- Wagstaff, A., & Lindelow, M. (2008). Can insurance increase financial risk? The curious case of health insurance in China. *Journal of Health Economics*, 27(4), 990–1005.

- Wagstaff, A., Lindelow, M., Jun, G., Ling, X., & Juncheng, Q. (2009). Extending health insurance to the rural population: An impact evaluation of China's new cooperative medical scheme. *Journal of Health Economics*, 28(1), 1–19.
- World Bank (2015). World Development Indicators 2015. doi:10.1596/978-1-4648-0440-3. License: Creative Commons Attribution CC BY 3.0 IGO, World Bank, Washington DC.
- World Health Organization (2010). *The world health report-health systems financing: The path to universal coverage* Technical report. Geneva, Switzerland: World Health Organization.
- World Health Organization (2015). WHO Global Health Expenditure Database. Technical report. <http://apps.who.int/nha/database/Select/Indicators/en>. Accessed: 2015-06-28, type=National Health Accounts.
- Xu, K., Evans, D. B., Kadama, P., Nabyonga, J., Ogwal, P. O., Nabukhonzo, P., & Aguilar, A. M. (2006). Understanding the impact of eliminating user fees: Utilization and catastrophic health expenditures in Uganda. *Social Science & Medicine*, 62(4), 866–876.
- Yates, R. (2009). Universal health care and the removal of user fees. *The Lancet*, 373(9680), 2078–2081.