



Does free hospitalization insurance change health care consumption of the poor? Short-term evidence from Pakistan

Simona Helmsmüller¹ · Andreas Landmann²

Received: 4 April 2020 / Accepted: 28 September 2021 / Published online: 25 October 2021
© The Author(s) 2021

Abstract

We analyze short-term effects of free hospitalization insurance for the poorest quintile of the population in the province of Khyber Pakhtunkhwa, Pakistan. First, we exploit that eligibility is based on an exogenous poverty score threshold and apply a regression discontinuity design. Second, we exploit imperfect rollout and compare insured and uninsured households using propensity score matching. With both methods we fail to detect significant effects on the incidence of hospitalization. Whereas the program did not meaningfully increase the quantity of health care consumed, insured households more often choose private hospitals, indicating a shift towards higher perceived quality of care.

Keywords Health insurance · Universal health care · Program evaluation · Health care consumption · Pakistan

JEL classification O12 · I13 · I15 · O22

1 Introduction

In lower- and middle-income countries, economic inequity is linked to inequity in health. One of the chains by which these are bound together is through high out-of-pocket (OOP) expenditures for health. These affect poor households in two ways: First, they create financial distress, in particular in the case of expensive events, such as hospitalizations. Second, they create barriers to health care, contributing to

✉ Simona Helmsmüller
simonahel@gmail.com

Andreas Landmann
andreas.landmann@fau.de

¹ Bonn-Rhein-Sieg University of Applied Sciences, Sankt Augustin, Germany

² Present Address: Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany



a low health status and therefore potentially also lower ability to generate income. A straightforward approach to breaking this vicious cycle is to provide health insurance to the poor. Many recent reforms in lower- and middle-income countries around the world are thus establishing inclusive health insurance schemes, with the aim of not only reducing financial distress, but also to change health seeking behavior by reducing financial barriers.

In this paper, we explore whether fully subsidized insurance for hospitalization changes health service utilization of low-income households in Pakistan. In particular, we evaluate the Social Health Protection Initiative (SHPI) in the province of Khyber Pakhtunkhwa (KP), which grants fully subsidized health insurance to the poorest quintile of the population. By studying the patterns of inpatient care consumption, we not only investigate changes in the quantity of care consumed, but also study whether the composition of care changes. An especially relevant dimension here is the probability to seek care from private providers, which patients associate with higher quality in our study. To evaluate the effect of insurance coverage, we use two features of the program. First, we implement a regression discontinuity design, using the fact that eligibility for the program is based on a pre-defined, exogenous poverty score. Whereas this approach has a high internal validity, inference is valid only for households around the poverty cut-off score. Therefore, as complementary second approach, we exploit incomplete rollout and match insured to comparable, eligible but uninsured households using propensity score matching. These approaches allow us to calculate two separate effects: the intention-to-treat effect for households close to the cut-off and the average treatment effect on the treated for eligible households.

The results of both econometric approaches suggest that the SHPI did not have significant effects on the quantity of health care consumption, despite high levels of neglected health care. We find no increase in the propensity of using inpatient health care services, no increase in the share of individuals who visited a hospital more than once in the past year, and no decrease of neglected health care. However, we find evidence suggesting a change of provider choice from public to private facilities. This is consistent with a larger reduction of relative costs of private care vs public care in our data as well as with a small number of claims from public hospitals in administrative data, suggesting that public hospitals implemented the program less efficiently. Given the better resources and higher client satisfaction associated with private hospitals, we nevertheless interpret this as an important positive impact of the program. Should the demand shift from public to private providers not be in the interest of the government, however, additional programs to strengthen the capacity of public hospitals might be necessary.

Several studies have analyzed the effect of protecting low-income households through health insurance. Randomized control trials (RCT) on micro health insurance programs have shown some promising impacts in terms of financial protection (see Habib et al. 2016 for a recent review), access to medical services (e.g. Levine et al. 2016; Thornton et al. 2010), and social outcomes (e.g. Landmann and Froelich 2015; Froelich and Landmann 2018). In line with this, there is a move towards universal health coverage via a rapid expansion of state-funded health insurance arrangements across lower- and middle-income countries (Lagomarsino et al. 2012;



Reich et al. 2016). However, results from RCTs do not necessarily carry over to larger programs where limited absorptive capacity might hamper the effect (Mangham and Hanson 2010), and not every program design allows for plausible identification strategies. Whereas some quasi-experimental studies exist on health insurance reforms in countries such as India, China, and Indonesia (Wagstaff et al. 2009; Prinja et al. 2017; Vidyattama et al. 2014), evidence on the Pakistani program is scarce, even though it is a very relevant case for several reasons.

Pakistan is a lower middle income country with the sixth largest population in the world, where poverty and the risk of falling into poverty are still widespread. According to World Bank Indicators (2016), Government spending on health has been 0.82% of GDP until the SHPI became operational in 2016 and around 63% of health expenditure had to be paid out-of-pocket. Government spending is higher (1.36% of GDP) and out-of-pocket expenditure is slightly lower (56%) in the group of lower middle income countries on average, but India, which shares many challenges in the health sector, has similar numbers (0.94% and 63%).¹ This situation increases the need for inclusive insurance solutions, which many other lower and middle income countries have recently addressed through publicly funded health insurance schemes as well (Cotlear et al. 2015). While the fragmented nature of the health system with provincial responsibility for the health policies renders reforms more difficult, these might have particularly high effects. In addition, through the fully subsidized scheme with household enrollment based on a pre-existing poverty census, the program achieved remarkably high enrollment rates and mitigated the problem of adverse selection which challenges similar interventions in other countries (Banerjee et al. 2019; Asuming 2013).

At the same time, Pakistan features a dual health sector with both private and public providers operating in the same market. A similar situation exists in India, which has undergone large-scale reforms with far-reaching transformations in the health care market a few years earlier. In this context, large health financing reforms might shape the long-term character of the market and it is therefore worth studying how demand in each sector is affected by insurance. For Cambodia, Levine et al. (2016) find that insured households shift towards public hospitals, but in their case private hospitals were not empaneled by the program, which means that patients simply shift towards participating hospitals. This is also what Thornton et al. (2010) find in Nicaragua, where insured households were more likely to visit health care providers covered under the insurance. We contribute to this literature by studying a setting under which insurance coverage was in principle available at both public and private providers. Note that effective coverage might still have differed between the two sectors, as these face different incentives and dispose of different resources and governance structures for implementation. In fact, our finding of an increased usage of private care is consistent with a more efficient program implementation in private hospitals, suggesting the importance of including the private sector to increase absorptive capacity. Despite the relevance of the research question for public policy, there is virtually no evidence on the impact of large state-funded insurance schemes

¹ World Bank Indicators for 2016 are available at <http://data.worldbank.org>.



on public vs private health systems if insurance coverage is offered at both. By looking at demand-side effects, we thus contribute to closing this evidence gap in the context of a nascent health insurance system in Pakistan.

The rest of this paper is organized as follows. In Sect. 2 we provide the country context and program details. In Sect. 3 we present details of our dataset and summarize descriptive statistics. In Sect. 4 we explain our two main identification strategies and assess the plausibility of the underlying assumptions. Section 5 contains our main results on the usage of inpatient care and a brief analysis of heterogeneous effects. In Sect. 6 we discuss effect channels and challenges in implementation. The last section concludes.

2 Rationale of the intervention

2.1 Challenges in health care in Pakistan

Poor health is widespread in Pakistan. In its report from 2017, the World Health Organization (WHO) attests Pakistan to have the fifth highest burden of tuberculosis world-wide and the highest rate of malaria in the region, while being one of only three countries in the world where residual poliomyelitis (infantile paralysis) has not been eradicated. Hepatitis B and C, dengue and chikungunya show high prevalence, and leprosy and trachoma are still reported. Regarding non-communicable diseases, cancer, diabetes, respiratory and cardiovascular diseases are among the main causes of death. Maternal and child mortality are among the highest globally (WHO 2017).

With the abolition of the Federal Ministry of Health in 2011, health care management and regulation became the responsibility of the Provincial Governments. These maintain networks of multi-tiered health care providers, yet overall public spending on health care is very low. In consequence, the quality, in particular of the primary health care infrastructure, is limited, suffering from political interference and corruption, shortage of trained personnel, staff absenteeism, non-functioning facilities, and lack of medicines (ADB 2019; WHO 2013). Notably, the non-existence of public family physicians means that hospitals are often the first point of contact with the formal health care infrastructure. But even major district hospitals often lack specialized staff such as gynecologists, anesthetists or pediatricians (TRC 2012). Therefore, households often use private service providers (Government of Pakistan 2016), implying that most of the health expenditures must be borne by the patient (Nishtar et al. 2013; WHO 2017). Also in public hospitals, expenditures, such as for medications, are usually paid out-of-pocket.

Social security systems are not broadly spread and leave the large majority of the population uncovered (Nishtar et al. 2013).² Private health insurers, though existing,

² According to Nishtar et al. (2013), there are three vertical systems servicing 14.12% of the population: By the Armed Forces, by the Fauji Foundation for retired military servicemen, and by the Employees Social Security Institution for public servants. These are vertical, i.e., they have mutually exclusive service delivery infrastructures.



lack the depth of penetration, in particular into rural and poorer population groups, covering less than 3% of the population (Nishtar et al. 2013). While there are a number of micro health insurance schemes run by non-governmental organizations (NGOs), they have not achieved broad outreach. With one third of the population living on less than 1.5 USD per day and in the absence of affordable insurance, it is reasonable to assume that financial constraints lead to less than optimal health care among the poor population of Pakistan.

2.2 The Social Health Protection Initiative (SHPI)

Against this background, the Government of the Province of KP launched a large-scale program to improve access to health care, called the SHPI. With financial and technical assistance of the German KfW Development Bank, the program intends to reduce financial barriers to health care through the introduction of a subsidized health insurance. The program uses a pre-existing national poverty score, which had been assigned to all households in Pakistan based on a proxy means test (PMT) in 2010.³ All households below a pre-defined cut-off poverty score were selected to receive the insurance card at fully subsidized rates. The first phase of the program was officially launched in December 2015 in the four pilot districts Chitral, Kohat, Malakand, and Mardan. It covered households with poverty scores below 16.17, corresponding to the poorest 21% of households in this area (approx. 0.7 million people targeted). The program delivered the cards to beneficiaries via selected regional NGOs, who were in charge of forward campaigning (including but not limited to banners and call centers providing general information, radio announcements and posters to inform about dates of enrollment at village level) as well as the physical distribution of insurance cards at special card distribution centers (including permanent offices at district level and temporary offices at village level). Following the official enrollment dates, unenrolled eligible households should be contacted directly by the insurer via phone or in person (Oxford Policy Management 2016). In addition, the consulting company advising the program on behalf of the KfW Development Bank verified the distribution of cards via a limited number of spot checks. Six months after the official launch, the insurer reported an enrollment rate of 87.3% among the target population in the two pilot districts considered in our study (Oxford Policy Management 2017).⁴

During our study period, one insurance policy covered a household of seven members (assumed typical case: household head, spouse, four children and one elderly dependent). The benefit package addressed maternity-related care as well as non-maternity hospitalization, up to an annual limit of PKR 25,000 (238.25 USD)⁵

³ The Government revised the poverty score again only after our endline survey, so that during our study period and the five years prior to that the score remained unchanged for all households.

⁴ Whereas the program also foresaw voluntary, non-subsidized health insurance to the non-eligible population, no such product was on offer at the time of our study.

⁵ Exchange rate on December 31, 2015.



per person.⁶ This covered treatment for normal delivery and C-sections, as well as a pre-defined list of 497 medical procedures requiring hospitalization. Notably, the program did not cover outpatient care.⁷

The insured households could obtain these services at one of the empanelled hospitals, which include public and private health care providers.⁸ Prior to the distribution of insurance cards, the program identified and contacted potential hospitals for empanelment in the program, but was met with skepticism. Private providers were hesitant to join the network due to concerns regarding the reimbursement of costs, religious beliefs, or fear of stricter tax controls (Oxford Policy Management 2016, 2017). Public hospitals also showed little interest in the program until Government influence was used to encourage joining the program. Nevertheless, the program was able to empanel around one third of the candidate private hospitals, as well as the two main public hospitals in each district. During our survey period, however, some hospitals were de-paneled due to the use of unnecessary procedures or, in one case, a conflict of interest. Overall, during our study period, there were at least four public and seven private hospitals available for service provision at all times.⁹ Before the official launch, the program trained hospital staff and established service desks in each empaneled hospital for identification of beneficiaries, verification of eligible treatment and available balance, and claim management for cashless service provision. For further gatekeeping, a District Medical Officer employed by the insurer visited clients within 24 h after admission.

Fully subsidized premiums naturally lead to an adverse incentive structure for the insurance company: The Government transfers the insurance premiums for each enrolled household, hence creating a steady flow of income from the Government to the insurer. At the same time, the cost structure of the insurance company, which was also responsible for the distribution of insurance cards, is determined by actual usage. The insurance company would hence benefit from not informing insured individuals of the full benefit package. Therefore, a mandated awareness campaign accompanied each phase of card distribution, carried out by the implementing

⁶ We have administrative cost data only for a short period of time overlapping our study. Between January and July 2017, the median cost of treatment was 15,000 PKR in the two pilot districts considered here.

⁷ A second phase of the program, starting in January 2017, saw the gradual roll-out to the remaining districts and raised the poverty cut-off score to 26.75, thus covering approximately 51% of households in the district (approx. 14.4 million people targeted). The program also altered the benefits slightly, covering eight household members, raising the annual coverage limit and including tertiary care providers, but notably still restricting coverage to cases of inpatient care. Table A.1 in Online Appendix 1 provides an overview of the program features in both phases. Following the completion of our study, the Government initiated Phase 3, which extended the program to cover up to 69% of the population in the entire province of KP. Further extensions are planned with the aim of achieving universal health coverage.

⁸ Despite there being a number of NGOs active in the health sector in Pakistan, such as the Aga Khan Foundation, there are no NGO-run hospitals in our survey region.

⁹ Specifically, in Malakand the program started with three public hospitals, and three out of nine identified private hospitals. Later, one public and one private hospital were de-paneled, while one new private hospital joined. In Mardan, the program started with three public and five out of 14 identified private hospitals. Later, one public and two private hospitals were de-paneled, while one new private hospital joined (Oxford Policy Management 2016, Oxford Policy Management 2017).



insurance company as well as the NGOs. A further challenge was the identification of beneficiary households, which were selected based on the poverty census from 2010. This implies not only that the program does not necessarily target the currently poor, but also challenged the localization of households for enrollment given that addresses were partially outdated.¹⁰

The Government of the Province of KP is spearheading the program, supported by the KfW Development Bank with financial and technical cooperation. Considering the difficult political landscape of Pakistan, the Provincial Government had its own vested interest in the program which likely went beyond the distributional goals: At the time of our study, the Province of KP was governed by a different party than held power of the Federal Government. The Federal Government of Pakistan planned and slowly started rolling out a similar national social health insurance. While the Federal Government had not implemented the national scheme in the Province of KP at the time of our study and hence did not create competition in economic terms, it most certainly imposed political competition. The Provincial Government was hence politically motivated to make the SHPI widely known and clearly associated with their party. Nevertheless, limited awareness remained a concern, which we further address in Sect. 6.

2.3 Intended effects

The rationale behind the SHPI is that the insurance would lower the cost of hospitalization and that this would affect households along two dimensions. On the one hand, lower OOP expenditures should encourage an increased usage of health services and hence the quantity of health care consumed. Thus, the program would contribute to health improvement. On the other hand, lower OOP expenditures directly decrease the households' financial burden and reliance on more stressful coping strategies. Thus, the program would contribute to financial protection against health risks.¹¹ Whereas we acknowledge the importance of financial protection for the poor in its own right, we concentrate on the first aim in this study, i.e., improving health by increasing health care consumption.

3 Data and descriptive statistics

3.1 Survey data

We make use of household survey data collected specifically for the program evaluation. Four months prior to the start of the first program phase, we collected baseline

¹⁰ After our study, the Government initiated a new survey to update the poverty score, which might improve targeting in later phases.

¹¹ Additionally, the Government aimed at increasing quality and accountability of public hospitals by ensuring a client-based flow of funds through the program. We do not consider supply side effects in this study.



data (autumn 2015). We carried out the endline survey 12 to 15 months after the first program rollout (spring 2017). Prior to the design of this evaluation, the Provincial Government had selected four pilot districts for the first phase of the program, where the insurance was to be offered exclusively. We therefore collected data in these four districts as well as in four additional districts, initially intended as control districts. Political dynamics, however, led to an early extension of the program into control districts as well as differences in rollout across the four pilot districts. The data we use in this study therefore is from only two of the four pilot districts, where the initial rollout plan was largely followed and where our identification strategies are still valid (Malakand and Mardan).¹² We also use data of the control districts for some robustness checks.¹³ Online Appendix 2 summarizes the timeline of the SHPI roll-out and our surveys in the relevant districts.

Our sampling strategy is a multi-staged clustered approach. We randomly selected 24 union councils as survey clusters in the two pilot districts considered here. The poverty census of 2010 served as a sampling frame for the third and fourth stage: Stratified random sampling of 70 villages and then 1200 households in the two pilot districts. To increase power for our identification strategies, we additionally sampled 240 households below and 480 closely around the cut-off poverty score in the pilot districts (i.e., an additional 20% and 40% respectively in each survey cluster). Therefore, our baseline sample in the pilot districts consists of 1920 households of which 828 were eligible for the insurance. Figure 1 depicts the distribution of the poverty score (a) in the sampling frame, (b) in our random sample and (c) in our total sample, respectively, illustrating the degree of the two types of oversampling, i.e., below and around the cut-off score of 16.17.

Interviewing the same households in the baseline and endline study, we constructed a household panel dataset. We used computer-assisted personal interviews in both survey waves, allowing the collection of GPS coordinates, an efficient survey administration and, thus, a minimal level of attrition of under 2.5%. An additional 1.2% of the sample were dropped in the data cleaning process, leading to a panel dataset of 1842 households in the two pilot districts, of which 795 eligible households. We collected information on economic conditions, subjective well-being, the use of health care during childbirth, outpatient care, and neglected health care

¹² We also collected data in the two other pilot districts, namely Chitral and Kohat. In Kohat, however, our monitoring during the endline survey revealed several problems. Specifically, we find particularly high differences between official and self-reported enrollment in the urban areas. We also faced the highest attrition rate (7%) in this area. In addition, there were problems in the project implementation in this district with one hospital being suspected of fraud. We thus exclude the data from the whole district out of prudence. The district of Chitral, on the other hand, was hit by a severe flood just prior to the baseline. This negatively affected our data collection in terms of access to some areas. Also the empanellment of hospitals was much delayed, and the program became fully operational only after our endline survey, which led us to exclude this district as well. In Mardan, the second program phase started three months prior to the endline, which might create some first additional effects, but does not invalidate our empirical approach. We discuss implications for the regression discontinuity design in Sect. 4.1.

¹³ We selected the four control districts using an algorithm matching on publicly available socio-demographic indicators and health infrastructure. The second phase rolled out prematurely in two of these, but using a different cut-off score.



on household level. In light of the focus of the program on inpatient treatment, we recorded the history of inpatient care, including associated costs, and the subjective health status of each household member individually. This leads to a final panel sample size of 12,862 individuals, thereof 6007 eligible for insurance, when considering inpatient care. In the endline survey, we administered the same questionnaire, but added questions on the enrollment status and familiarity with the program.

3.2 Data quality and processing

Our local research partner pre-tested, translated, and implemented the questionnaires on tablet computers. To a large extent, items are based on a questionnaire which had been tested repeatedly and demonstrated high validity in previous projects. At the end of each survey day, supervisors uploaded the data from the tablets onto a server and we downloaded data in Germany for monitoring of interviewer performance and data quality. Daily quality control included automated consistency checks, spot checks, and follow-up phone calls. Comparing GPS coordinates of a household at baseline and endline guaranteed that indeed the same household was interviewed.

We winsorized quantitative variables which showed a large variation. The level of winsorizing depends on the initial variation of the specific variable and ranges from the 90th to the 99th percentile. We performed a principal component analysis of asset ownership to derive a variable for socio-economic standing (in the following denoted wealth index) and a principal component analysis of access to amenities such as toilets and drinking water to derive a variable for hygienic condition (in the following denoted hygiene index). For per capita household income, we account for economies of scale within the household and use the square root equivalent scale, i.e., we divide household income by the square root of household size. (An implication is that, e.g., a four-person-household has twice the monetary needs of a single person.)

We note that our survey might suffer from coverage error. This stems from the fact that the best available sampling frame, the poverty census, was collected in 2010 and is hence partly outdated. Moreover, in the absence of official addresses of most households, the identification of sampled households was a challenge and might have led to population subgroups being missing not-at-random. However, one should note that the SHPI used the same frame to determine program eligibility. While our results might not be fully representative, e.g., for young and newly formed or migrated households, they are internally consistent under the plausible assumption that all groups used for comparison in our identification strategies are likely to be similarly affected.

3.3 Baseline characteristics

Table 1 contains selected baseline characteristics of households and individuals in our panel samples, i.e., sampled households and their members in the two districts Malakand and Mardan with baseline as well as endline information. We separately present statistics on the full sample as well as on households eligible for insurance



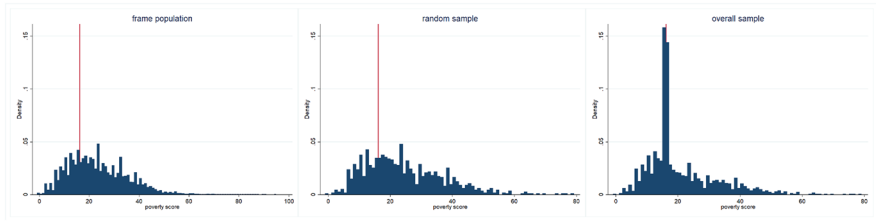


Fig. 1 Distribution of poverty score in sample UCs of Malakand and Mardan. This figure shows the distribution of poverty scores in the various samples. The overall sample (right panel) includes over-sampling below and additionally around the cut-off. Sample: BISP sampling frame ($N = 71,591$) and Household-level sample (panel, $N = 1842$). *Source* BISP survey (2010) and Baseline survey (2015). The poverty cut-off score (assigned in 2010) determining eligibility for the first phase of the SHPI is 16.17, indicated by the vertical red line in each panel. The figure illustrates the degree of two types of oversampling: below and around this cut-off. (Color figure online)

coverage, i.e., with a poverty score below 16.17. Note that the goal here is not to give a representative picture of the population but to describe the samples we are using for our analysis. These samples include oversampling below and around the cutoff, and therefore do not reflect average differences between eligible and non-eligible households in the population. We present statistics for the subsample of randomly selected households in Table A.2 in Online Appendix (differences to table below are marginal).

The average household in our full sample consists of 7.43 members and of 8.09 members in the subsample of eligible households. The members of eligible households are slightly younger (22 vs 23 years), more likely to be of school-aged (38% vs 33%), and a larger share has not completed primary school (63% vs 59%). Conversely, a smaller share of members has completed secondary school or higher (8% vs 12%). Consistently, the per capita household income among eligible households is around two thirds that of the full sample. There is a high gender disparity in education and work (not shown in table): Among male adults, 47.0% in our full sample have no formal education, and this percentage rises to 82.1% among female adults. Similarly, 67.2% of male adults have worked for pay in the year prior to the baseline survey, compared to only 3.3% of female adults. Overall, hygienic conditions are sub-optimal: Whereas 96% of households have electricity in their home, only 36% have a private flush toilet and only 12% have tap water supply in their residence. Travel time to the next hospital averages 44 minutes.¹⁴ Notably, awareness about insurance is virtually non-existing at baseline.

Regarding the use of health care services, 5% of individuals in the full sample reported an overnight stay in a hospital within the 12 months prior to baseline. To understand the socio-economic drivers of using inpatient services, we run three logit regressions including individual and household covariates with different proxies for poverty, namely the poverty score, the per capita household income, and the wealth index (results shown in Table 2). Older and female individuals

¹⁴ We did not ask to specify the medium of transport, so this likely differs across households.



Table 1 Baseline characteristics of full sample and subsample of eligible population (selected variables)

	Full sample			Eligible sample				
	Mean (1)	SD (2)	Min (3)	Max (4)	Mean (5)	SD (6)	Min (7)	Max (8)
Panel A: household-level variables								
Insurance status at endline	0.45	0.50	0.00	1.00	0.65	0.48	0.00	1.00
Poverty score	20.61	11.15	0.00	79.00	12.53	3.60	0.00	16.17
P.c. monthly income (sqr. root equiv.)	538.65	829.68	0.00	8888.89	388.80	611.17	0.00	8888.89
Wealth index	-0.25	1.92	-3.40	12.23	-0.56	1.56	-3.21	6.58
HH size	7.43	2.78	1.00	23.00	8.09	2.54	3.00	21.00
Electricity in HH	0.96	0.20	0.00	1.00	0.95	0.21	0.00	1.00
Tap water supply in residence	0.12	0.32	0.00	1.00	0.12	0.32	0.00	1.00
Private flush toilet	0.36	0.48	0.00	1.00	0.29	0.45	0.00	1.00
Reported dist. to next hosp. (minutes, win99)	43.62	26.35	0.00	150.00	44.41	26.90	0.00	150.00
Use of prof. assist. during childbirth	0.89	0.31	0.00	1.00	0.86	0.35	0.00	1.00
Case of neglected health care	0.14	0.35	0.00	1.00	0.16	0.36	0.00	1.00
Case of outpatient care	0.78	0.42	0.00	1.00	0.80	0.40	0.00	1.00
Citing health shock as a risk	0.94	0.23	0.00	1.00	0.95	0.22	0.00	1.00
Diff'ty finding money for health care ≥ 8 (scale 1/10)	0.48	0.50	0.00	1.00	0.49	0.50	0.00	1.00
Having heard of insurance	0.02	0.14	0.00	1.00	0.01	0.12	0.00	1.00
Observations	1842				795			
Panel B: member-level variables								
Age (win99)	23.17	18.17	1.00	90.00	22.25	17.45	1.00	90.00
School-aged (6 to 16)	0.33	0.47	0.00	1.00	0.38	0.49	0.00	1.00
Female	0.48	0.50	0.00	1.00	0.48	0.50	0.00	1.00
Prim. school not comp'd.	0.59	0.49	0.00	1.00	0.63	0.48	0.00	1.00
Comp'd sec. educ. or higher	0.12	0.32	0.00	1.00	0.08	0.28	0.00	1.00



Table 1 (continued)

	Full sample			Eligible sample				
	Mean (1)	SD (2)	Min (3)	Max (4)	Mean (5)	SD (6)	Min (7)	Max (8)
Worked for salary in previous month	0.19	0.39	0.00	1.00	0.18	0.39	0.00	1.00
Usage of inpatient care	0.05	0.22	0.00	1.00	0.04	0.21	0.00	1.00
Cost of last treatment (PKR, win99)	25,086	44,655	0	300,000	25,072	44,769	500	300,000
More than one admittance to hospital	0.26	0.44	0.00	1.00	0.26	0.44	0.00	1.00
Use of private hospital	0.31	0.46	0.00	1.00	0.30	0.46	0.00	1.00
Observations	12,862				6007			

This table shows the baseline characteristics of the full sample and the eligible subsample (households (members) with a poverty score below 16.17). Selected variables on the left, statistics on top

Sample: Households and their members in full or eligible sample (panel, varying N)

Source: Baseline survey (2015), insurance status from endline (2017)

Column (1) displays the mean for continuous/shares for binary variables in the full sample, Column (2) the standard deviation, Columns (3) the minimal and (4) the maximal value in the full sample. Columns (5) to (8) display the same statistics for the subsample of eligible households and their members

The suffix *win99* indicates that we winsorized the variable at the 99th percentile level. Monetary variables in PKR (100 PKR = 0.953 USD on December 31, 2015)

Note that the insurance status in the non-eligible sample is non-zero due to the roll-out of the second phase of the program in the district of Mardän shortly before our endline survey

Table A.2 in Online Appendix contains the statistics for the subsample of randomly selected households (excluding sampling). Differences are small and statistically insignificant



Table 2 Logit regression of hospitalization on individual and household characteristics, baseline

Admission to inpatient care	Logit 1			Logit 2			Logit 3		
	Coef.	SE	<i>p</i> value	Coef.	SE	<i>p</i> value	Coef.	SE	<i>p</i> value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Poverty score	0.004	0.005	0.497						
Per capita monthly HH income				− 0.009	0.006	0.134			
Wealth index							− 0.095	0.031	0.002
Female	0.229	0.085	0.007	0.228	0.086	0.008	0.222	0.084	0.008
Age	0.028	0.001	0.000	0.028	0.002	0.000	0.028	0.002	0.000
Household size	− 0.025	0.022	0.248	− 0.027	0.021	0.209	− 0.005	0.021	0.815
Hygiene index	0.031	0.040	0.449	0.009	0.040	0.822	− 0.028	0.044	0.532
Dist. to next hospital (min)	− 0.002	0.002	0.441	− 0.002	0.002	0.466	− 0.001	0.002	0.523
Const.	− 3.643	0.294	0.000	− 3.499	0.217	0.000	− 3.794	0.249	0.000

This table shows the coefficients of logit regressions of a dummy indicating admission to hospital on individual and household covariates. Covariates on the left, statistics on top

Sample: Member-level sample (panel, $N = 12,852$)

Source Baseline survey (2015)

Columns (1), (4), (7) display the coefficient estimates from the logit regressions, Columns (2), (5), (8) the standard errors and Columns (3), (6), (9) the *p* value of the two-sided tests that the coefficient is equal to zero, with one of three different proxies for poverty, respectively. Standard errors are adjusted for 24 clusters in union councils

Table A.3 in Online Appendix contains the results excluding childbirth related hospitalization. The general direction and significance of coefficients remains unchanged, but females are no longer significantly more likely to be admitted to hospital

are consistently more likely to consume inpatient care, where the gender effect is driven by childbirth related admissions (effect disappears when childbirth is excluded, see Table A.3 in Online Appendix). The results also suggest that poorer households consume significantly more inpatient care when using the wealth index as proxy for poverty. This is consistent with the fact that both wealth as well as health represent outcomes of long-term processes.

The conclusion that in our sample, the less wealthy are more likely to consume inpatient health care, does not necessarily imply that poor households are not restricted in their access to health care. Instead, the finding could be driven by higher health needs, as health and poverty are related by causality running in both directions (Wagstaff 2002). We therefore also check the relation of the wealth index with other important outcomes of interest, namely, a measure of subjective health status, using a private facility (conditional on being admitted), and neglected health care in Table 3. To do so, we repeat the regressions, controlling only for the evidently important covariates age and gender, but also including squared terms for a more flexible form. The wealth index and its square are correlated not only with admission to inpatient care (Column (1)), but also with the



Table 3 Usage patterns and health care needs

	Admission to inpatient care			Health status			Use of private vs public hospitals			Neglected health care		
	Coef.	SE	p value	Coef.	SE	p value	Coef.	SE	p value	Coef.	SE	p value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Wealth index	-0.094	0.032	0.003	0.004	0.017	0.813	0.296	0.079	0.000	-0.291	0.062	0.000
Wealth index sq.	0.005	0.006	0.479	0.003	0.002	0.082	-0.031	0.019	0.105	0.033	0.010	0.001
Female	0.223	0.084	0.008	-0.126	0.017	0.000	-0.447	0.167	0.007			
Age	0.022	0.006	0.001	0.005	0.002	0.008	-0.008	0.015	0.591			
Age sq.	0.000	0.000	0.294	-0.000	0.000	0.000	0.000	0.000	0.893			
Const.	-3.834	0.193	0.000	4.551	0.055	0.000	-0.225	0.304	0.458	-2.078	0.187	0.000
N	12,852			12,852			589			1849		

This table shows the coefficients of regressions on various outcomes (logit for binary outcomes, linear regression for health status variable). Covariates on the left, outcomes and statistics on top

Samples: Member-level and household-level samples (panel, varying N). Columns (7) to (9) conditional on reporting a case of inpatient care

Source: Baseline survey (2015)

Columns (1), (4), (7), (10) display the coefficient estimates from the regressions, Columns (2), (5), (8), (11) the standard errors and Columns (3), (6), (9), (12) the p values of testing that the coefficient is equal to zero, on different outcomes, respectively. Standard errors are adjusted for 24 clusters in union councils

We measured neglected health care on household level, hence we regress only on household-level covariates. Results are robust towards including the share of female, old, or young household members as covariates



subjective health status, which improves for individuals in wealthier households (Column 4). Also, wealthier households are more likely to visit a private hospital (Column 7), where care is frequently perceived to be of higher quality, and less likely to report an incident of neglected health care (Column 10).¹⁵ Our data therefore supports the hypothesis that poor households are indeed restricted in their access to health care, both in quantity and perceived quality.

4 Econometric approach

We use two identification strategies, which estimate different effects. First, we apply a sharp Regression Discontinuity Design (RDD) using the poverty score as running variable. This provides an estimate of an Intention to Treat (ITT) for observations around the cut-off. Second, we match insured and non-insured individuals and households on the propensity to receive insurance estimated from baseline values. This provides an estimate of an Average Treatment Effect on the Treated (ATT). Table 4 illustrates the different samples considered for the two estimators.

4.1 Regression discontinuity design (RDD)

We exploit the fact that there exists a pre-defined poverty cut-off score which exogenously determines program eligibility, creating an ideal set-up for an RDD approach. Figure 2 depicts the self-reported insurance status by poverty score using local polynomial smoothing in both considered districts. In Malakand, there is a large and significant drop in insurance enrollment at the cut-off. This drop is smaller in Mardan due to a pre-mature roll-out of the second phase, which led to enrollment of households with poverty scores between 16.17 and 26.75 in this district, but only three months prior to our endline survey. The figure displays the self-reported insurance status, hence also including enrollment under the second phase. Since our main outcome of interest, the usage of inpatient care, relates to a period of 12 months, it is more appropriate to consider households covered under the second phase as (largely) uninsured. We also estimated effects for the second phase in Mardan using the cut-off score of 26.75. We find no significant effects, possibly also due to the short time period of phase 2 implementation before our survey (max. 3 months) and the smaller data set (1232 households overall) in only one district. If anything, this should lead to a slight downward attenuation of the estimated effect.¹⁶ In our main

¹⁵ The result on the usage of private hospitals shown in the table is obtained by restricting the sample to individuals with a case of inpatient care. It also sustains, albeit less pronounced, when running the regression on the full sample, unconditional of a case of inpatient care.

¹⁶ For a quick back-of-the-envelope calculation, note that the effect in our model is reduced approximately by the average time the control group was covered (estimated as 2/12 months) times the share of the recently insured individuals above the cut-off (0.4 across both districts) over the share of insured below the cut-off (0.67), so by around 10%.



model specification we hence calculate intention-to-treat effects using a sharp RDD design with treatment determined by the poverty score only.

While the poverty score is assigned on household level, we measure key variables of interest on member level and use the member-level sample for estimation of effects on these. Potentially, all members of insured households were eligible to be insured, and since we focus on intention-to-treat effects here, we also use the poverty score as treatment indicator on member level. Figure B.2 in Online Appendix depicts the share of insured individuals in the member-level sample by poverty score for the two districts, essentially showing the same drop in the propensity to be insured.

We calculate local linear regression models to the left and right of the cut-off score using a triangular kernel, where the bandwidth is estimated to minimize the mean squared error as suggested in Calonico et al. (2017). We provide standard errors using heteroskedasticity-robust nearest neighbor variance estimators as provided by the Stata command `rdrobust` by the same authors.

The key assumption for the internal validity of the RDD approach is that the distributions of potential outcomes are smooth around the cut-off score. As this assumption is not directly testable, we in the following show different tests that underline the credibility of this main assumption.

4.1.1 No manipulation of poverty scores

An important assumption for the validity of RDD is that of no self-selection. In our setting, this implies that, while households might be able to manipulate the poverty score, they must be unable to precisely sort around the cut-off score (McCrary 2008). In general, self-selection is a threat if individuals are aware of the assignment rule, expect positive returns of participation in treatment and have sufficient time and resources to change their behavior to meet the assignment rule. To assess this risk in our setting, note that the poverty score was initially derived to determine eligibility to a nation-wide social program, the Benazir Income Support Program (BISP), and assigned to each household in 2010, long before the SHPI came to life. Since the BISP used the same cut-off score as the SHPI, this might have created an incentive for self-sorting into treatment. However, the poverty score is based on a proxy means test constructed from a scorecard with 23 variables. Households knew neither how the information was to be aggregated into a single poverty score nor was the cut-off score known in advance but set so as to cover the poorest 21% of the population in the program (Uddin et al. 2013).

More formally, we can test for discontinuities in the density of the poverty score at the cut-off. Cattaneo et al. (2020) provide a fully data driven approach to test for a discontinuity. When running this test for the random sample, we obtain insignificant results ($T = 0.844$, p value = 0.398), see Fig. B.4 in Online Appendix, suggesting that there was no manipulation of the poverty score. This is in line with the work of Ambler and De Brauw (2017) and Nawaz and Iqbal (2021), who also use



Table 4 Treatment and control groups in two estimators

	Treatment group	Control group
Regression discontinuity design (RDD)	Insured and uninsured HH/members from two pilot districts Poverty score $\in [16.17 - B, 16.17]$	Insured and uninsured HH/members from two pilot districts Poverty score $\in [16.17, 16.17 + B]$
Propensity score matching (PSM)	Insured HH/members from two pilot districts Poverty scores $\in [0, 16.17]$	Uninsured HH/members from two pilot districts Poverty scores $\in [0, 16.17]$

This table illustrates the different samples considered for the regression discontinuity design and propensity score matching estimators, respectively

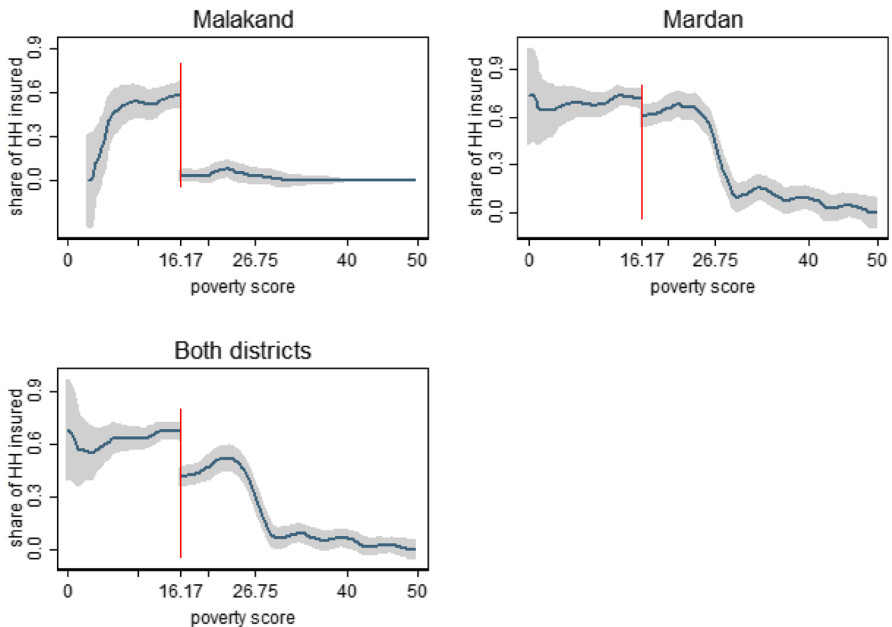


Fig. 2 Share of households insured (self-reported) by poverty score. This figure shows the average insurance rate, conditional on the poverty score. The solid blue lines and shaded grey areas are the predicted values and associated 95%-confidence intervals, respectively, based on local mean smoothing. The red vertical line indicates the cut-off score of 16.17. Sample: Household-level RDD sample (panel, $N_{\text{Malakand}} = 617$, $N_{\text{Mardan}} = 1232$). Source Endline survey (2017). (Color figure online)

an RDD with the same poverty score in Pakistan and run the same tests on their samples.¹⁷ The result is also supported by the report on the initial targeting survey of the national social program, which also finds no sharp break in the density of the

¹⁷ Ambler and De Brauw (2017) also use the undersmoothed estimators, which also brings insignificant results in our sample. We also run the Cattaneo et al. (2020) test on our member-level sample and obtain insignificant results. In addition, the density test proposed by McCrary (2008) also fails to detect a significant discontinuity.



poverty score and no significant jump at the threshold for baseline covariates and outcome variables (O'Leary et al. 2011).¹⁸

4.1.2 Confounding program

As mentioned above, the poverty score also determines eligibility to another social program, which includes an unconditional cash transfer. The program has been running since 2010 and was ongoing at the time of our survey. Most importantly, it uses the same cut-off score as the SHPI, which might lead to confounding effects. In fact, 84% of eligible households in our panel claim to have received transfers from the BISP program. However, the transfer was small (10% reporting 1000 PK and another 85% reporting 1500 PKR) and, most importantly, 80% of all households claim to have received these transfers already at baseline. We can therefore test whether the program had any effect on the outcomes considered in our study by estimating pseudo effects at baseline. Table 5 and Fig. 3 contain the results. We find no significant effects in any of the outcomes of interest, proving that the national social program does not confound our analysis. Table B.5 contains the pseudo effects at baseline for any subgroup considered in analysis of heterogeneous effects.

4.1.3 Continuity of covariates

Another test that is often run in similar studies is to check for continuity of covariates around the cut-off that could affect the outcome of interest. It is not a necessary condition for the continuity of potential outcomes, but it increases its plausibility. Regarding the choice of variables to assess, note that our running variable, the poverty score, is an unknown function (f) of some socio-demographic and economic covariates (X). In particular with discrete covariates, any specific value of $f(X)$ might only be attainable by a specific combination of X , such that similar but distinct values of $f(X)$ can only be attained by very different combinations of X . In this case, $E[x_i|f]$ becomes a non-smooth function, such that there might necessarily be discontinuities at the cut-off as well as many other points. As plausibility check, we therefore only test for discontinuities in variables which were not used to construct the poverty score. Table B.4 in Online Appendix shows that no significant discontinuity can be detected in any of these cases.

Regarding those covariates that were used to construct the poverty score, such as age, gender, and wealth, we believe the continuity of baseline outcomes (shown above) to be sufficient evidence for the continuity of potential outcomes around the cut-off, and hence the validity of the RDD approach. Nevertheless, we include the

¹⁸ The density test estimated in our overall estimation sample renders a test statistic of -3.209 with a p value of 0.0013 . The highly significant jump in the density is however due to our sampling strategy with a larger share of oversampling below than above the cut-off score.



covariates age, gender, and wealth in our regressions and find point estimates to be slightly larger in magnitude. Since inference remains however unaffected, we provide the more conservative estimates excluding covariates as main results and estimates controlling for covariates as robustness checks in Table C.13 in Online Appendix. Further pseudo-effect calculations on those covariates not used for calculating the poverty score give insignificant results, see Table B.10 in Online Appendix.

4.1.4 Further specification and falsification tests

To check robustness to tuning parameters, we estimate various specification tests: we cluster standard errors at UC-level, include covariates, apply a fuzzy design, use local constant regression models as well as polynomial order three, apply a uniform and an epanechnikov kernel, allow MSE-optimal bandwidth to differ below and above the cut-off, choose the bandwidth that optimizes the coverage error rate, and use the bandwidth estimators suggested by Imbens and Kalyanaraman (2012). Our results are robust regarding these different specifications, as shown in Tables C.11 and C.12 in Online Appendix.

We furthermore run an algorithm applying a series of pseudo cut-off scores in 0.1-steps from 14.69 to 17.69 on our four main outcome variables. Of the 116 thus calculated pseudo estimates, only three are significant at the 5% level, and none at the 1%.

Finally, we collapse our member-level sample by household and repeat our estimation regarding member-based outcomes on indicator variables on household level. This way we account for the fact that the poverty score was assigned on household, not individual level. Our results are robust to this aggregation, as shown in Table D.14 in Online Appendix.

4.2 Propensity score matching (PSM)

Whereas the regression discontinuity design has a high internal validity, its external validity is restricted to households and individuals around the cut-off. To also estimate average treatment effects on all treated, we supplement our analysis using a propensity score matching approach. This is possible because among eligible households, the program achieved self-reported enrollment rates of 65.2% of households. This is remarkably high,¹⁹ yet a sizable number of households targeted by the program did not report themselves insured in our survey, likely due to imperfections in program roll-out.

Two aspects on the definition of our treatment indicator are important to note. First, we make use of the self-reported insurance status, instead of the official status as per administrative data. We believe that households which are officially insured but not aware of this are more likely to behave as if they were uninsured and should

¹⁹ As comparison, Banerjee et al. (2019) report enrollment rates of 8% in the Indonesian national health insurance program of their study, which they managed to increase to 30% under a treatment arm with full program subsidization and assistance in the enrollment process.



Table 5 Pseudo-effects on inpatient care consumption at baseline

Outcome	RDD	
	β_{ITT}	SE
	(1)	(2)
Individual outcomes		
Usage of inpatient care	0.001	0.008
<i>N (left/right of cut-off)</i>	<i>3217/2557</i>	
Conditional on usage of inpatient care		
More than one admittance	0.006	0.080
<i>N (left/right of cut-off)</i>	<i>134/125</i>	
Usage of private vs public hospitals	0.021	0.079
<i>N (left/right of cut-off)</i>	<i>161/132</i>	
Household outcomes		
Neglected health care	– 0.011	0.030
<i>N (left/right of cut-off)</i>	<i>418/361</i>	

The bold values are the estimates, the italic values the number of observations. This is to differentiate between the important numbers (bold) and the supplementary numbers (italic)

This table shows results from RDD estimates on baseline variables, main outcome variables on the left, the different statistics on top

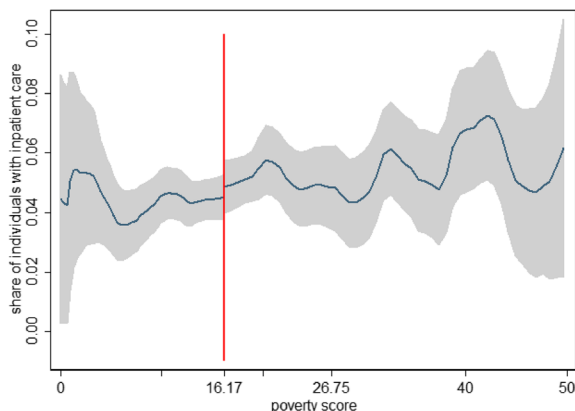
Sample: RDD sample (panel, varying N)

Source Baseline survey (2015)

Columns (1) and (2) show the coefficient and standard errors for the pseudo-intention-to-treat effect for households just below the cut-off score, estimated using a sharp regression discontinuity design. Estimated using local linear regression models with a triangular kernel and bandwidth estimated to minimize the mean squared error; S.E. as proposed in Calonico et al. (2014) (corresponding to our main model specification). Reported sample size refers to observations within selected bandwidth

The statistical significance is given as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$, with the null hypothesis being a zero effect size

Fig. 3 Inpatient care by poverty score at baseline. This figure shows the share of individuals with a case of inpatient care by poverty score, using local polynomial smoothing. Red vertical line indicates the cut-off score of 16.17. Sample: Member-level random sample (panel, $N = 12,863$). Source Baseline survey (2015). (Color figure online)



hence be part of the control group.²⁰ For ease of notation, we will use the term (*un-*) *insured* to refer to the self-reported insurance status from now on. Second, for estimating effects on individual level, we identify all members of an insured household as insured. Whereas in our survey, we ask for the insurance status of each household member individually, we find comparably high rates of false-positive and false-negative reporting. Initially, seven members of a household could enroll under the insurance, but this changed to eight members when Phase 2 was rolled out. In our complete sample, 28% of households with more than eight members claim that all are covered under the insurance, whereas only 78% of households with seven members or less report all members covered. The latter number increases to 91% if “do not know” answers are counted as enrolled. In the absence of administrative data for individuals we cannot further investigate the reasons for this. However, the fact that the “do not know” answer was chosen for 15.94% of the household members indicates that there might be substantial recall bias.

We exploit the fact that a third of the target households remain uninsured, and estimate the ATT using the following propensity score matching estimator:

$$\beta_{PSM} = \frac{1}{|I_1|} \sum_{i \in I_1} \left\{ Y_{1i} - \frac{\sum_{j \in I_0} Y_{0j} G\left(\frac{P_j - P_i}{B_n}\right)}{\sum_{k \in I_0} G\left(\frac{P_k - P_i}{B_n}\right)} \right\}, \tag{1}$$

where β_{PSM} is the statistic of interest, the average treatment effect on the treated. I_1 is the set of insured households within the region of common support, I_0 is the set of uninsured households, Y_{1i} is the outcome for an insured household, Y_{0j} for an uninsured household, $P_j = Pr_j(\textit{insured}|Z)$ is the propensity score, i.e., the probability that a household is insured conditional on a set of covariates Z , $G(\cdot)$ is the epanechnikov kernel, B_n the bandwidth. As was the case in the RDD, we also estimate effects on individual level, in which case I_1 is the set of individuals in insured households within the region of common support, I_0 is the set of individuals in uninsured households, Y_{1i} is the outcome for members in an insured household, Y_{0j} for a member in an uninsured household, $P_j = Pr_j(\textit{insured}|Z)$ is the propensity score, i.e. here, the probability that an individual is member of an insured household conditional on a set of covariates Z . Note that for each sample we consider (household, member, conditional, or subsamples thereof for the analysis of heterogeneous effects) we estimate the propensity scores anew and for this purpose re-run the algorithm that selects the set of covariates Z described below. That is, the set of covariates used for matching depends on the sample considered.

²⁰ Only 2.5% of ineligible households who report themselves insured in our survey are not insured according to administrative data. In contrast, 74.4% of eligible households who report themselves uninsured are registered as insured in administrative data. The numbers are in line with programs in other countries such as Philippines and Rwanda, where Lagomarsino et al. (2012) note that Government enrollment figures do not always match household survey data. In our case, three factors likely contribute to the deviance: (i) The household never received the card and the administrative data are fraudulent. (ii) The household was enrolled after our online survey. (iii) The household was enrolled, but the interviewed household member was not aware of it.



For the calculation of standard errors we account for the fact that propensity scores are estimated and that variables are clustered on the union council level by providing clustered bootstrapped standard errors (9999 repetitions).²¹ Note that we bootstrap the whole process of estimating propensity scores, imposing common support, matching observations, and estimating effects.

Two assumptions are key to this approach (Todd 2010): Conditional mean independence and common support, which we discuss in the following.

4.2.1 Common support

To ensure common support, we restrict our treatment sample to individuals or households with a propensity score above the 99th percentile score among the control group, as suggested in Caliendo and Kopeinig (2008). This eliminates 10.23% of households and 5.53% of household members in our respective treatment groups, for whom we have no suitable control observations. Furthermore, there are some gaps in the density of propensity scores in the control group. This is no concern though, as we have sufficient density to the left and right of these gaps for kernel matching. Nevertheless, we follow Smith and Todd (2005) and additionally drop 1% of our treatment observations at which the propensity score density of the control group is at its lowest. Figures B.7 and B.8 in Online Appendix demonstrate that common support is thus sufficiently ensured.²²

4.2.2 Conditional mean independence

Whether the conditional mean independence assumption is fulfilled is not directly testable, but hinges on the considered set Z for calculating propensity scores (Smith and Todd 2005). The lowest bias arises when Z includes all variables that simultaneously affect insurance status and considered outcomes. We see three factors that are important here: Non-random targeting (e.g. due to infrastructure or social status), non-random acceptance of the card (e.g. due to lack of education or trust in the government), and non-random awareness of having received a card (e.g. due to low valuation or knowledge about insurance). Any of these three systematic differences between insured and uninsured households is probably driven by a number of unobservable variables, such as geographic accessibility, quality of accessible health care, intensity of the awareness campaign, quality of education, or interviewer effects. Many of these are however likely to be geographically clustered. Indeed, enrollment rates in our sample of eligible households range from 40 to 80% in the 24 union councils of the two districts, as depicted in Fig. 4.²³ Correspondingly, we tested

²¹ Whereas Abadie and Imbens (2008) show that bootstrapping is invalid for nearest neighbor matching, they anticipate that the bootstrap is valid for kernel-based matching (which we use) due to its asymptotic linearity.

²² We concentrate on estimation of average treatment effects on the treated, hence we need not drop control observations for whom there is no match in the treatment group.

²³ Union councils are administrative units between the district and village level, which also served as survey clusters (second stage sampling unit).



whether the set of union council dummies contributes to explaining enrollment and find this to be the case (p value of an f test testing joint significance = 0.001). Therefore, we include the set of union council dummies in our set of covariates used to estimate propensity scores in our main model specification.

Other variables that might affect both, insurance status and outcomes, such as education, prior insurance knowledge or household size, are observable to us from the baseline survey. Tables B.6 and B.7 in Online Appendix show means of all collected baseline variables for insured and uninsured households. Along observable dimensions, the two groups differ significantly only in their willingness to take financial risks, with the uninsured households being more willing to bear risks. This is in line with the theory that risk-averse individuals have a higher incentive to seek insurance coverage. As Caliendo and Kopeinig (2008) note, omitting important variables can increase the bias in the estimates, which suggests including as many covariates as possible. However, over-parameterized models suffer from a lack of common support, which potentially increases the variance of the propensity score estimate. To balance the risk of bias and variance, we follow the algorithm described in Imbens and Rubin (2015, Chap. 13) for our main model specification. For this algorithm, we select a set of base variables, which we believe important for the selection model.²⁴ The algorithm searches for further baseline variables to be included linearly into the selection model in an iterative process, where in each step the variable yielding the largest likelihood ratio statistic is included in the model, until all these statistics are smaller than one. The iterative process leads us to include an additional ten variables on household and four variables on member level. Finally, we select quadratic and interaction terms of all selected variables to be included into the selection model using the same iterative process as before.²⁵ This time, we stop when all likelihood ratio statistics are smaller than 2.7, as in Imbens and Rubin (2015). This leads us to include another 20 interaction terms on household and 23 terms on member level. In conclusion, in our main model specification, we match households on 36 and individuals on 40 linear and interaction baseline variables in addition to union council dummies. We repeat the same variable selection algorithm for any subsample analysis.

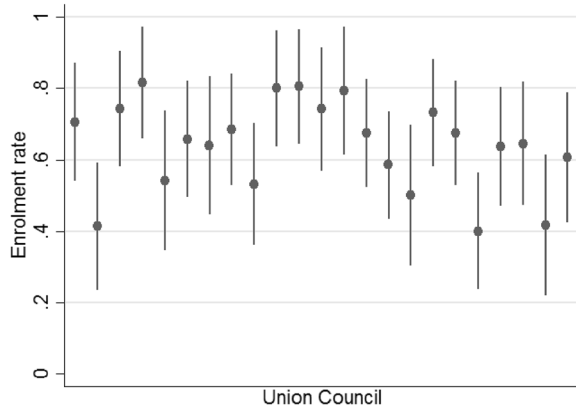
Tables B.6 and B.7 in Online Appendix demonstrate the achieved balancing on union councils and baseline variables. Figures B.5 and B.6 in Online Appendix show the distribution of the poverty score among insured and matched uninsured samples, underlining the credibility of the conditional mean independence assumption.

²⁴ In addition to the union council dummies, we chose the poverty score, the average monthly household income (winsorized), the household size, household-level usage of inpatient care (extensive margin and number of household members treated), as well as the minimum of the reported health status over all household members for the household-level sample. On member level, the set additionally includes age, gender, subjective health status, admittance to hospital (whether admitted at all and whether admitted more than once), and dummy variables for whether the member completed primary school and whether the member completed senior or higher education.

²⁵ We refrain from including interactions with the union council dummies, as this would lead to overfitting, violating the common support assumption.



Fig. 4 Enrollment rates in main sample of eligible households across 24 union councils. This figure shows the enrollment rates per union council in the two pilot districts Malakand and Mardan. Sample: Household-level PSM sample (panel, $N = 795$). Source Endline survey (2017). Union councils are geo-administrative units two levels below the district which served as sampling clusters (second stage sampling unit)



4.2.3 Additional model specifications

To test the sensitivity of our propensity score estimates regarding the set of covariates Z , we additionally estimate propensity scores using (a) variables selected by the Imbens algorithm without pre-selecting any variables for inclusion, (b) only the pre-selected variables and UC-dummies, (c) the variables selected by the Imbens algorithm but without UC-dummies, and d) linear and interaction variables selected using lasso methods with crossvalidation. Tables B.8 and B.9 in Online Appendix show the number of variables selected in each of these specifications and the log likelihood function using probit estimation. In both datasets, the latter is maximized using the set of covariates from our main model specification. The tables furthermore show the correlation coefficients of the log odds ratio of propensity scores calculated using our main model and the four other specifications. The correlation with other specifications is lowest for the set selected using lasso methods and the set of pre-selected variables and UC-dummies. We therefore also estimate effects using these model specifications. Results remain however unchanged, as illustrated in Table C.13 in Online Appendix.

4.2.4 Further robustness and falsification tests

To check robustness, we cluster standard errors on household-level and we restrict estimation to the random sample. Our results are robust regarding these different specifications, see Table C.13 in Online Appendix. Furthermore, we apply augmented inverse-probability weighted regression estimation. For this doubly robust method, we specify an outcome as well as a treatment model. Whereas the latter corresponds to the model we use for propensity score estimation (i.e. covariates selected using the Imbens algorithm for each subset), the outcome model is inspired by Table 3 in that we include covariates that were correlated



with the outcomes at baseline.²⁶ Our results are robust to this alternative estimation method.

5 Results

5.1 Main results

Our main outcome variables concern the use of inpatient care. In our endline survey, we asked for each household member separately whether that member experienced a case of inpatient care in the past 12 months (admittance to hospital). If answering affirmatively, we also asked how often the individual was admitted to hospital within that timeframe, and what type of hospital she visited (private vs public). Furthermore, on household level we asked whether any household member faced an accident or illness where inpatient care was considered but not sought within the past 12 months (neglected health care). For all these four key outcomes, we estimate the effects of providing insurance coverage using both, the Regression Discontinuity Design (RDD) and as supplementary estimates Propensity Score Matching (PSM).

Table 6 contains the results from the RDD and the PSM estimations. For example, in the first line the outcome considered is whether an individual has sought inpatient care in the past 12 months. The mean among the matched control group in the PSM sample is 0.059 and we estimate a negative and insignificant coefficient of -0.002 in the PSM estimation, with a standard error of 0.011. The sample consists of 2526 uninsured and 3638 insured household members. Our RDD estimation also yields a coefficient of -0.002 with a standard error of 0.007, where we rely on 3118 observations below and 2526 observations above the cut-off. Note that the reported sample size refers to the area of common support (PSM sample) and the observations within the selected bandwidth around the cut-off score (RDD sample) respectively, implying that these numbers change across regression specifications.

Despite the large number of observations at our disposition, we find no significant effects of the program on the usage of inpatient care, neither locally around the cut-off (RDD) nor averaged across all treated (PSM). Even when accounting for clustering effects, standard errors are limited to one percentage point, such that we would have detected effect sizes of less than two percentage points as significant (one third of the control mean). There are also no effects when analyzing the corresponding outcomes on household level, i.e., considering as outcomes either the propensity that any member in a given household is admitted to hospital or the number of household members admitted, see Table D.14 in Online Appendix. In other words, we can exclude short-term transformative changes in seeking hospitalization in our sample.

²⁶ That is, we regress the usage of inpatient care on wealth index, gender and age; the variables more than one admittance and usage of private hospitals on wealth and gender, and the variable neglected health care on wealth and squared wealth index.



Among individuals who reported a case of inpatient care, we also look at the share of individuals with more than one stay at a hospital and also find no effect here. As we did not find an effect of the program on the probability of using any inpatient care before, we believe in the validity of this result, even though the sample restriction to those with inpatient care might in principle be endogenous. Furthermore, note that our sample size is much smaller here.²⁷

We also do not observe a decrease in the share of households with neglected health care. Again, precision of the coefficients is limited, but both the RDD as well as the PSM point estimates are very close to zero. Note that 90% of households reporting a case of neglected health care stated that this was because they could not afford the cost of treatment in a hospital, suggesting that a functioning insurance scheme could have had an impact on this variable.

Suggestive evidence in line with these null effects also comes from households with childbirth. Given that the insurance explicitly covers maternity care, we would expect a particularly strong increase in the usage of professional assistance during childbirth in these households. Unfortunately, there are too few childbirths in our sample to run a proper matching procedure, but a simple comparison of beneficiary groups in the RDD and PSM samples does not reveal any significant differences.²⁸

While the quantity of inpatient care consumed seems to remain largely unchanged, usage patterns may nevertheless have changed. Specifically, we find a significant increase in the usage of private vs public hospitals in the RDD estimation as illustrated in Fig. 5. This result is robust against different bandwidth specifications, as illustrated in Fig. 6, and also holds when analyzing the result on household level, see Table D.14 in Online Appendix.

In addition, the effect of 6.8 percentage points calculated in the PSM estimation is, albeit insignificant, sizeable and in the expected direction, increasing the share of individuals visiting a private instead of a public hospital by 18.06%. This result is in line with administrative data: In their progress report for January to June 2017, the consultancy supporting the program on behalf of the KfW Development Bank notes that 95.59% of admissions in the two districts were registered in private hospitals (Oxford Policy Management 2017).²⁹ We draw further descriptive evidence from a separate section of the questionnaire, where we asked households whether they have used the card, at what type of hospital, and whether this was the first time they visited that facility. Among respondents who used their card at a private facility,

²⁷ To avoid overfitting, we therefore repeat the calculation of propensity scores for this subsample and include only linear terms and no interaction terms in the estimation model.

²⁸ We have 113 cases of childbirth within a 2-points interval around the cut-off score. Yet, the RDD estimation also does not find a significant effect on the usage or professional assistance at childbirth. In our PSM sample, we observe only 35 uninsured and 80 insured households with childbirth, and this sample size is not sufficient to ensure common support for PSM estimation. Regressing the use of professional assistance during childbirth on the insurance status among the PSM sample yields no significant result.

²⁹ At the time of our study, the program had empanelled seven private and four public hospitals in the two districts, and for each private facility there is one public facility in immediate proximity (Oxford Policy Management 2017).



Table 6 Effects on inpatient care consumption

Outcome	Control	PSM		RDD	
	Mean	β_{mATT}	SE	β_{ITT}	SE
	(1)	(2)	(3)	(4)	(5)
Individual outcomes					
Usage of inpatient care	0.059	– 0.002	0.011	– 0.002	0.007
<i>N</i> (uninsured/insured, left/right of cut-off)		<i>2526/3638</i>		<i>4336/3227</i>	
Conditional on usage of inpatient care					
More than one admittance	0.195	0.022	0.077	0.062	0.054
<i>N</i> (uninsured/insured, left/right of cut-off)		<i>107/212</i>		<i>255/174</i>	
Usage of private vs public hospitals	0.383	0.068	0.075	0.237***	0.073
<i>N</i> (uninsured/insured, left/right of cut-off)		<i>101/202</i>		<i>186/131</i>	
Household outcomes					
Neglected health care	0.062	0.007	0.026	0.001	0.022
<i>N</i> (uninsured/insured, left/right of cut-off)		<i>277/461</i>		<i>581/466</i>	

The bold values are the estimates, the italic values the number of observations. This is to differentiate between the important numbers (bold) and the supplementary numbers (italic)

This table shows our main results, the effect of free hospitalization insurance on inpatient care consumption. Outcome variables on the left, different econometric models and statistics on top

Samples: Member-level and household-level PSM and RDD samples (panel, varying *N*)

Source Endline survey (2017)

Column (1) displays the mean for the matched controls of uninsured, but eligible households (poverty score below 16.17). Columns (2) and (3) show the coefficient and standard error for the average treatment effect on the treated, estimated using propensity score kernel matching. Columns (4) and (5) show the coefficient and standard errors for the intention-to-treat effect for households just below the cut-off score of 16.17, estimated using a sharp regression discontinuity design

Note on PSM SE are derived by bootstrapping the whole process of estimation of propensity scores, restricting the sample to common support, matching, and ATT estimation. Unit of clustering is the union council. Number of bootstraps: 9999. Reported sample size refers to area of common support (overall sample size: 795 households with 6007 members). Table C.13 in Online Appendix contains results of various robustness checks. Inference remains unchanged across various alternative specifications

Note on RDD Estimated using local linear regression models with a triangular kernel and bandwidth estimated to minimize the mean squared error; SE as proposed in Calonico et al. (2014). Reported sample size refers to observations within selected bandwidth (overall sample size: 1842 households with 12,862 members). Table C.11 in Online Appendix contains results of various robustness checks. Inference remains unchanged across various alternative specifications

The statistical significance is given as follows: * $p < 0.1$; ** $p < 0.05$, and *** $p < 0.01$, with the null hypothesis of the two-sided test being a zero effect size

79.45% visited this facility for the first time, compared to only 36.67% among public facility card users.

The shift from public to private hospitals constitutes an improvement of health for the beneficiaries if and only if private hospitals provide better quality of care. However, whereas public hospitals are hardly monitored, private hospitals do not



Fig. 5 RDD plot for use of private vs public hospitals. This figure illustrates the result for the impact of the insurance the insurance on the use of private vs public hospitals. Small confidence intervals in the bins just above and below the cut-off score reflect our sampling strategy. Sample: Member-level RDD sample with case of inpatient care (panel, $N_{total} = 742$). Source Baseline survey (2015)

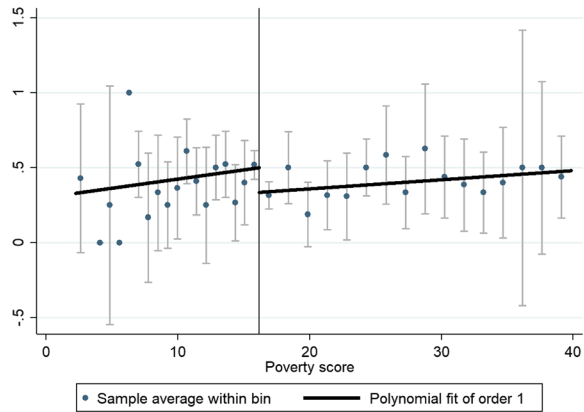
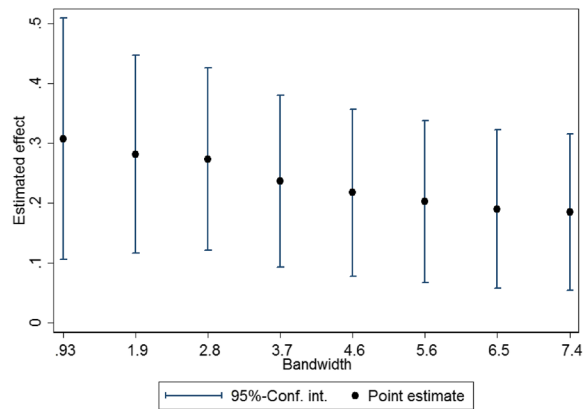


Fig. 6 RDD estimates for use of private vs public hospitals, different bandwidths. This figure shows the point estimates and confidence intervals for the impact of the insurance on the use of private vs public hospitals for different bandwidths and accounting for clustering on household level. Main specification uses bandwidth of 3.71. Sample: Member-level RDD sample with case of inpatient care (panel, varying sample size depending on bandwidth). Source Baseline survey (2015)



even register, rendering it notoriously difficult to measure quality of care.³⁰ Though private hospitals seem to perform better regarding governance and resources, it is unclear whether this transforms into better health outcomes, as private hospitals may overtreat common diseases while referring difficult cases to public tertiary hospitals.³¹ Nevertheless, suggestive evidence comes from our baseline survey, where we asked respondents to rate the health status of household members with a case of inpatient care at the worst time of their illness, before, and after hospitalization on

³⁰ A recent assessment of hospitals in the province of KP led by the Asian Development Bank paints a rather daunting picture of health care quality, listing among other challenges political interference and corrupt practices, serious lack of space, workforce, and drug supplies, as well as issues related to infection control (ADB 2019). The review comprised 37 hospitals, including two private ones, and while this is hardly a representative review of the private sector, the described governance challenges related to nepotism and corruption are likely to be dominant in the public sector.

³¹ We find no evidence for this in our data. Whereas the diseases treated in public and private hospitals are similar, the number of cases per diseases and type of hospital might be too low to detect significant differences.



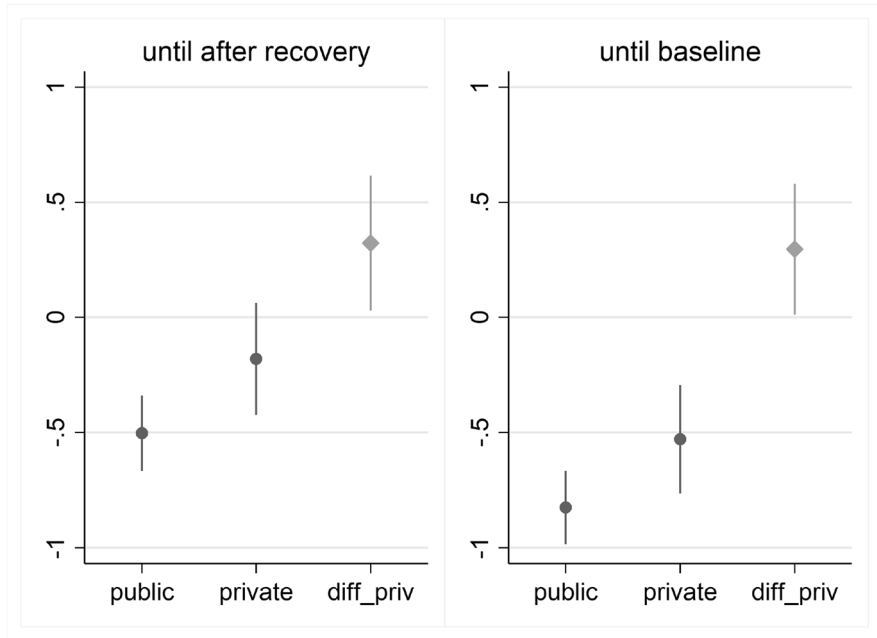


Fig. 7 Average drop in health by provider choice. This figure shows the difference between the health status before the illness that led to hospitalization and after hospitalization (left panel), respectively at baseline (right panel). On average, the health status worsened after hospitalization, but private hospitals restored health to a higher level than public hospitals. Sample: Conditional sample (baseline, $N = 508$). *Source* Baseline survey (2015)

a scale of 1 to 5. Regressing the health status after hospitalization on whether a private hospital was chosen yields a significant and positive coefficient even when controlling for health status before hospitalization. This relation is illustrated in Fig. 7.

Moreover, 65.82% of respondents in our PSM sample rather agreed than disagreed with the statement that private facilities provide better quality of service than public facilities in the endline survey.³² Most importantly, as we have laid out in Sect. 3, wealthier clients are significantly more likely to visit private hospitals. Specifically, 17.93% of individuals with a case of inpatient care in the lowest wealth quintile visited a private hospital at baseline, whereas that share increases to 44.87% in the highest wealth quintile. We can reasonably assume that individuals would not be willing to pay higher prices in private hospitals if these were not at least perceived to provide better care. Therefore, we associate the observed behavior change in provider choice caused by the insurance with an increase in subjective quality of care.

³² Shabbir and Malik (2016) provide further circumstantial evidence by finding patients of private hospitals in Islamabad to be more satisfied than patients of public hospitals.



5.2 Heterogeneous effects

Average treatment effects might mask heterogeneity regarding demographic or socio-economic characteristics. We therefore repeat the estimation of treatment effects on our main outcome variable, the propensity to use any inpatient care, for selected subsamples with particularly high health financing needs.³³ We look at female household members, at adults above the age of 16, at members with self-rated health status below median at baseline (i.e., below perfect health), and at households with below-median wealth.³⁴ Table 7 contains the results of the subsample analysis. We find no significant effects for any of the four subgroups. Note that the control mean in the overall PSM sample was 0.059, underlining that the subgroups considered here are the high-risk groups.

6 Discussion

In this section, we discuss our finding of shifts towards private care without an increase in overall hospitalization. Let us first emphasize that the RDD and PSM approaches meaningfully complement each other, because they allow us to look at effects on two different populations (intention-to-treat effect at the cutoff vs average treatment effect on the treated), and thereby provide a more complete picture.³⁵ Also, they complement each other in overcoming relative weaknesses. For example, the RDD estimate is based on an exogenous eligibility cutoff, which might be a less noisy measure of ‘effective’ coverage than self-reported insurance status. However, in a context of imperfect rollout and awareness, it is also valuable to have the PSM estimates, which are based on self-reported enrollment, to confirm results.

One important aspect in the interpretation of the findings is that the endline survey took place 12 to 15 months after the distribution of insurance cards. This might be too short for results to materialize, for example, because households might need longer to change behavior, or because hospitals might need longer to set up the required procedures. Whereas we agree that the program likely needed more time to reach its full potential, we do not believe that inertness to change behavior is the main reason for this in this setting.³⁶ In line with this, absolute claim numbers in our

³³ Note that for the other outcomes analyzed before, subsample analyses suffer from the limited number of observations.

³⁴ For these subsamples, we calculate pseudo RDD-effects at baseline to check the plausibility of the RDD assumptions. There are no significant effects, as illustrated in Table B.5. For the PSM estimation, we calculate propensity scores based on covariates Z_j , which are selected anew for each subsample from the set of all baseline covariates following the Imbens algorithm.

³⁵ Note that we do not focus on the validity of the assumptions underlying our empirical approach here. Those are discussed in Sect. 4, where we present supporting evidence as far as possible. Specifically, we test the assumptions wherever our data allows, conduct a range of plausibility and sensitivity tests (see Tables C.11 and C.13 in Online Appendix), and run placebo analyses using baseline variables, alternative cut-offs, and control districts.

³⁶ At baseline, 72% of households with a case of outpatient care four weeks prior to our survey received this care at a (public or private) hospitals. Households are hence familiar with hospital visits and the



Table 7 Heterogeneous effects on usage of inpatient care

Subgroup	Control	PSM		RDD	
	Mean	β_{rmATT}	SE	β_{ITT}	SE
	(1)	(2)	(3)	(4)	(5)
Female household members <i>N (uninsured/insured, left/right of cut-off)</i>	0.064	0.005 <i>1017/1740</i>	0.013	0.004 <i>1633/1253</i>	0.012
Adults above 16 years <i>N (uninsured/insured, left/right of cut-off)</i>	0.097	– 0.020 <i>1051/1823</i>	0.022	– 0.021 <i>1584/1266</i>	0.014
Baseline health status below median (< 5) <i>N (uninsured/insured, left/right of cut-off)</i>	0.093	– 0.002 <i>825/1281</i>	0.017	– 0.000 <i>1419/1108</i>	0.015
Wealth index below median (< –0.60) <i>N (uninsured/insured, left/right of cut-off)</i>	0.079	– 0.015 <i>1065/1772</i>	0.023	– 0.002 <i>1574/1362</i>	0.114

The bold values are the estimates, the italic values the number of observations. This is to differentiate between the important numbers (bold) and the supplementary numbers (italic)

This table shows heterogeneous effects of hospitalization insurance on the usage of inpatient care. The different subgroups considered are indicated on the left, econometric models and statistics on top

Samples: Subgroups of the PSM and RDD samples (panel, varying N)

Source Endline survey (2017)

Column (1) displays the mean for the matched controls of uninsured, but eligible households (poverty score below 16.17). Columns (2) and (3) show the coefficient and standard error for the average treatment effect on the treated, estimated using propensity score kernel matching. Columns (4) and (5) show the coefficient and standard errors for the intention-to-treat effect for households just below the cut-off poverty score of 16.17 estimated using a sharp regression discontinuity design

Note on PSM SE are derived by bootstrapping the whole process of estimation of propensity scores, restricting the sample to common support, matching and ATT estimation. Unit of clustering is the union council. Number of bootstraps: 9999. Reported sample size refers to area of common support @publishing services Please add line brakes suggest by the authors

Note on RDD Estimated using local linear regression models with a triangular kernel and bandwidth estimated to minimize the mean squared error; SE as proposed in Calonico et al. (2014). Reported sample size refers to observations within selected bandwidth. The statistical significance is given as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$, with the null hypothesis being a zero effect size

two study districts reach relatively stable levels within the first two to three months of insurance introduction and only increase after the second phase of the program is introduced (around the timing of the endline survey). We illustrate this fact in Fig. 8, where we plot the number of claims in the two districts respectively as per administrative data of the program. Note that Phase 2 started at different points of time in the two districts and saw a notable increase of enrollment from 21% of the population to 51%.

The overwhelming majority of these claims come from private providers, in line with our finding of an increased propensity to visit private rather than public

Footnote 36 (continued)

required behavioral change is smaller than it might be in settings with a functioning primary health care system.



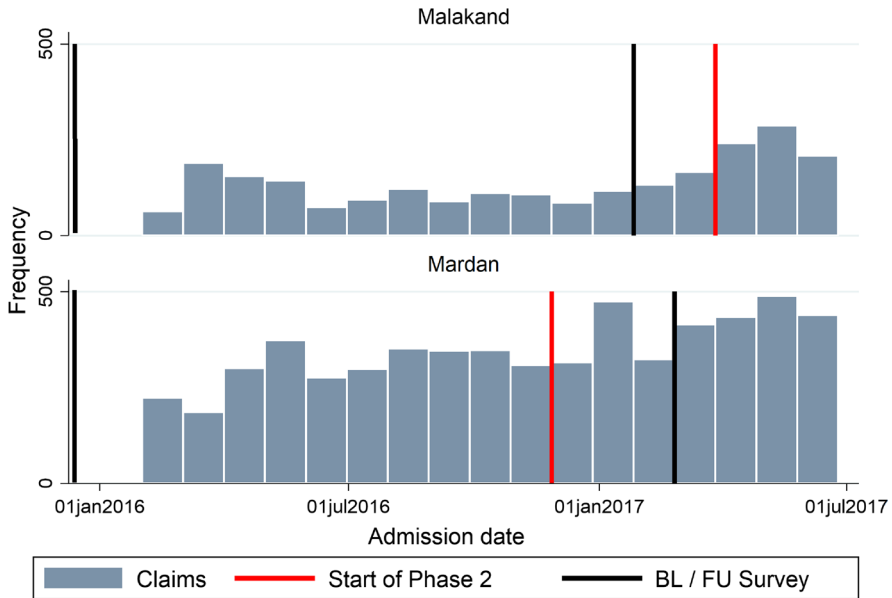


Fig. 8 Absolute number of claims over time (admin data). This figure shows the absolute number of claims per district over the first one and a half years of the program. *Source* Administrative data of the SHPI program. Red lines indicate the start of Phase 2 in the respective districts (December 2016 in Mardan, April 2017 in Malakand) and black lines the timing of the baseline and follow-up surveys. (Color figure online)

hospitals. A plausible explanation might be that public hospitals were slow to implement required procedures and hence effective coverage was only provided in private hospitals. In fact, private hospitals might have faced greater incentives to participate in the program as the insurance allowed them to attract clients who could previously not afford their services, while at the same time possessing a more flexible governance structure. We therefore cannot reject the possibility of supply side constraints in the public sector driving the observed change in provider choice. This is a relevant consideration as it illustrates the importance of empaneling private health care providers in large-scale programs with potential public capacity limitations. We have to keep in mind, however, that despite the apparent capacity to absorb patients in private facilities, we do not observe an increase in overall inpatient care consumption. This prompts the question whether there are other bottlenecks in the new health insurance scheme.

One possibility is that a lack of information restricts beneficiaries from effectively using the insurance. Given that the government pays full premiums to the insurer based on insurance cards distributed, and that the costs faced by the insurer are driven by actual usage, the insurer has little incentive to provide comprehensive information to facilitate utilization. Information provided by the regional NGOs might also be incomplete in this principal-agent setting. Our endline survey contains knowledge questions about the insurance program, in particular which treatments are covered (inpatient and/or outpatient) and which hospitals would accept the card (public and/



Table 8 Evidence on program limitations

Outcome	Control	PSM		RDD	
	Mean	β_{rmATT}	SE	β_{ITT}	SE
	(1)	(2)	(3)	(4)	(5)
Panel A: effects on usage of inpatient care in different subsamples (member-level)					
HH with good knowledge on program details	0.069	– 0.018	0.015	– 0.001	0.008
<i>N</i> (uninsured/insured, left/right of cut-off)		<i>2007/2313</i>		<i>2863/2755</i>	
HH living within 10 km distance to next empaneled hos.	0.066	– 0.014	0.022	– 0.020	0.011
<i>N</i> (uninsured/insured, left/right of cut-off)		<i>1061/1881</i>		<i>1861/1431</i>	
HH living within 40 min travel distance to next hospital	0.049	– 0.005	0.012	0.007	0.010
<i>N</i> (uninsured/insured, left/right of cut-off)		<i>1061/1419</i>		<i>1713/1326</i>	
Panel B: effects on financial outcomes conditional on inpatient usage (member-level)					
Log out-of-pocket expenditures (PKR, win99)	8.947	– 0.256	0.310	– 0.266	0.256
<i>N</i> (uninsured/insured, left/right of cut-off)		<i>107/212</i>		<i>178/136</i>	
Log of cost for diagnosis and treatment (PKR, win99)	4.833	– 0.261	0.562	– 0.601	0.567
<i>N</i> (uninsured/insured, left/right of cut-off)		<i>107/212</i>		<i>207/151</i>	
Log of cost for medicines (PKR, win99)	8.238	– 0.316	0.398	– 0.030	0.344
<i>N</i> (uninsured/insured, left/right of cut-off)		<i>107/212</i>		<i>204/147</i>	
Sleepless night due to hospital costs (PKR, win99)	0.486	0.083	0.095	0.028	0.066
<i>N</i> (uninsured/insured, left/right of cut-off)		<i>107/212</i>		<i>235/170</i>	

The bold values are the estimates, the italic values the number of observations. This is to differentiate between the important numbers (bold) and the supplementary numbers (italic)

This table shows evidence for the discussion on program limitations. Different subgroups (Panel A) or outcomes variables (Panel B) on the left, statistics on top

Samples: (Subsamples of) PSM sample and RDD sample (panel, varying *N*)

Source Endline survey (2017)

Column (1) displays the mean for the matched controls of uninsured, but eligible households (poverty score below 16.17). Columns (2) and (3) show the coefficient and standard error for the average treatment effect on the treated, estimated using propensity score kernel matching. Columns (4) and (5) show the coefficient and standard errors for the intention-to-treat effect for households just below the cut-off score of 16.17, estimated using a sharp regression discontinuity design

Note on PSM S.E. are derived by bootstrapping the whole process of estimation of propensity scores, restricting the sample to common support, matching and ATT estimation. Unit of clustering is the union council. Number of bootstraps: 9999. Reported sample size refers to area of common support

Note on RDD Estimated using local linear regression models with a triangular kernel and bandwidth estimated to minimize the mean squared error; SE as proposed in Calonico et al. (2014). Reported sample size refers to observations within selected bandwidth

The statistical significance is given as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$, with the null hypothesis of the two-sided test being a zero effect size

The suffix *win99* indicates that we winsorized the variable at the 99th percentile level

Table D.15 in the Online Appendix shows results for non-logarithmic financial outcomes



or private). To test whether information is indeed an important factor, we restrict our sample to those households who answered both these questions correctly and repeat the estimation. We display results in Table 8 (first line of Panel A).³⁷ The PSM estimate is negative and insignificant, while the RDD estimate is insignificant as well, and very close to zero. These results do not suggest that the program led to higher utilization among those with better knowledge of insurance details.

Another barrier might be that the program restricts the choice of care providers to specific, empaneled hospitals. At the time of our endline survey, these included two public and three private hospitals in the district of Malakand, and two public and four private hospitals in the district of Mardan. All hospitals are in city centers, hence accessibility remains an issue in rural areas. We measure the distance to these hospitals using GPS data and find a median of 9.6 km. Note that this is the geographic distance calculated from GPS coordinates and likely only proxies accessibility. We also asked respondents about their travel time to the next hospital, including non-empaneled ones, and report a median of 40 minutes. To analyze to which extent distance to hospitals restricted the program's impact, we repeat the estimation of effects for households which live within a below-median distance to a hospital, i.e., within 10 km to an empaneled hospital or within 40 minutes away from any hospital. We display results in Table 8 (second and third line of Panel A). Again, we find no evidence of a program impact on overall inpatient service utilization in these subsamples.

An increase in the consumption of inpatient care, however, is only plausible if two conditions are fulfilled. First, the insurance should achieve financial protection, i.e., it should decrease costs of seeking inpatient care. The second condition is that costs of treatment should actually influence inpatient service utilization. In the endline survey, we asked respondents about the cost of treatment born out-of-pocket, which we use to assess the first condition. We estimate the effect of the program on total expenditures and on the individual cost positions for diagnosis and treatment, and medication.³⁸ We present results in Table 8, Panel B.³⁹ Note that we have only a very limited sample size, as we only consider individuals who reported a case of inpatient care, while at the same time here considering a variable with high variation. Therefore, coefficients are not significant, even though we estimate negative and sizable effects (suggesting a cost decrease of around 30%). Additionally, we asked whether those with a hospitalization case experienced sleepless nights due

³⁷ As before in the analysis of heterogeneous effects, we calculate pseudo RDD-effects at baseline for the subsamples considered here to check the plausibility of the RDD assumptions. There are no significant effects, as illustrated in Table B.5 in Online Appendix. For the PSM estimation, we calculate propensity scores based on covariates Z_j , which are again selected anew for each subsample from the set of all baseline covariates following the same procedure as for the complete PSM sample, described above.

³⁸ We asked separate questions for total costs and individual cost positions to test for possible side payments demanded by the hospital staff. We find no evidence for this. Note that we did not measure opportunity costs such as forgone wages, except for transportation, meals, and accommodation for accompanying family members, which we find to be a negligible portion of total expenditures.

³⁹ Due to the highly skewed distribution of the quantitative cost variables, we use log values as outcomes.



to the related costs. In this case, the coefficient is positive, though insignificant. An explanation for the counterintuitive result on this subjective measure might be an attention bias, given that we previously had asked only insured households about their insurance status and understanding of insurance principles. In summary, our data are inconclusive when it comes to financial protection achieved by the program. It is consistent with a possible decrease in hospitalization costs, though.

Even if the program was successful in decreasing the financial burden of inpatient care, it does not necessarily lead to more utilization. In Sect. 3, we showed that using inpatient care in general does not increase with financial wealth, suggesting a low risk of moral hazard for an insurer. With higher wealth, however, we observe an increase in private care, which patients often associated with higher quality. In other words, individuals with urgent health problems might visit a hospital irrespective of their wealth. This is consistent with evidence that hospitalization (in contrast to outpatient care visits) is not very sensitive to cost sharing by an insurer (e.g. Finkelstein 2007). Poorer individuals, however, seem to seek care at cheaper public facilities. We compare average costs of public and private hospitalization in our data. Indeed, we see that private care is much more expensive than public care at baseline in our sample of interest (39,000 vs. 23,000 PKR).⁴⁰ Interestingly, it descriptively seems like this difference shrinks after the program rollout only for the insured, driven by a strong decrease in private care costs.⁴¹ Given these observations, it may not be too surprising that instead of an overall increase in hospitalization, we measure a shift towards private facilities as a result of the program. For a dual health system with public and private providers operating in the same market, this is a highly relevant result, as the insurance program might also shape the composition of the market in the long term.

7 Conclusion

Providing free health insurance to a large number of poor households is an intuitive approach to increase health care consumption. The rationale is that high OOP expenditures not only pose a financial risk, but also restrict poor households' access to inpatient care. We analyze the effect of the SHPI, which provided free hospitalization insurance to the poorest 21% of the population in the Pakistani province of KP, on health care seeking behavior. To this end, we apply a regression discontinuity approach, comparing households just above and just below the exogenous poverty

⁴⁰ The diseases treated in public and private facilities at baseline are largely the same in our data, but sample size per disease is too low to comment on statistical significance. A notable exception is that whereas only 5.22% of patients were treated for appendicitis in public facilities, the rate is 19.02% among patients in private facilities.

⁴¹ After the insurance rollout average private care costs are 20,000 PKR and public care costs 13,000 PKR for the insured, while the difference is much larger for the non-insured (30,000 vs. 13,000 PKR). So in particular the relative decrease in costs for private care seems to be larger for the insured at endline. The difference in cost for private care between insured and non-insured was in fact the reverse at baseline (41,000 PKR for the insured vs 35,000 PKR for the non-insured).



cut-off score, and a propensity score matching approach, comparing insured and uninsured but eligible households. While the former has a higher internal validity, it provides estimates of intention-to-treat effects for households around the cut-off poverty score only. In contrast, propensity score matching relies on more restricting assumptions, but provides average treatment effects on the treated. In this sense the two identification strategies complement each other.

In our study, we find that insured households do not increase the quantity of inpatient care consumed and have the same propensity to neglect their health care as uninsured households. Large-scale multi-stakeholder programs like the SHPI naturally face many challenges in implementation, including limited awareness and insufficient empanelment of hospitals. Yet, we find no support in our data that these factors seriously impaired the program's impact. Also, we measure impact only one year after program introduction. Whereas we concur that the program might develop larger impact over a longer period of time, administrative data confirms that the program was largely operational within the considered time period, in particular in private hospitals. To check whether heterogeneity masks effects for some subgroups, we repeat estimations separately for several high-risk groups, but also fail to detect significant increases in inpatient care.

Importantly, we do however observe a sizable increase in the propensity of visiting a private instead of a public hospital. This result is in line not only with administrative data, but also with a larger decrease of reported care costs for insured individuals in private compared to public hospitals. Since patients in Pakistan often consider private hospitals to provide higher quality of care, this is an important and policy-relevant effect of the program, which might thus contribute to a more equitable access to high-quality care. Whether strengthening the private sector to overcome possible supply side constraints in the public sector leads to desirable outcomes in the long run is an open question, though. Given that there are a number of countries with mixed health systems moving towards universal insurance coverage, including India and Indonesia, further research on the long-term effects on public vs private market sectors seems promising.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1057/s10713-021-00069-0>.

Acknowledgements The survey and administrative data were produced and procured for the evaluation of the SHPI program, which was financed by the KfW Development Bank as part of a collaboration with its evaluation unit. We are grateful for the latter's continued support and insights. We also would like to thank Markus Frölich, Manuela De Allegri, Remi Maier-Rigaud, Torben Fischer, Christoph Strupat, and seminar participants at C4ED (internal seminar) and St John's University (GRIR conference on Risk Considerations and Insurance in Developing Countries) for valuable comments and feedback on earlier versions of this work. Katja Hofmann, Fawad Ejaz, and Jan Berkes provided excellent research assistance. The authors take sole responsibility for any remaining errors

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.



Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abadie, A., and G. Imbens. 2008. On the failure of the bootstrap for matching estimators. *Econometrica* 76 (6): 1537–1557.
- ADB. 2019, October. *Khyber Pakhtunkhwa Health Sector Review - Hospital Care*. Technical report, Asian Development Bank.
- Ambler, K., and A. De Brauw. 2017. *The impacts of cash transfers on women's learning from Pakistan's BISP program*. Washington, DC: World Bank.
- Asuming, P.O. 2013. *Getting the poor to enroll in health insurance, and its effects on their health: Evidence from a field experiment in Ghana*. Job Market Paper. Columbia University.
- Banerjee, A., A. Finkelstein, R. Hanna, B.A. Olken, A. Ornaghi, and S. Sumarto. 2019. *The challenges of universal health insurance in developing countries: Evidence from a large-scale randomized experiment in Indonesia*. Technical report. National Bureau of Economic Research.
- Caliendo, M., and S. Kopeinig. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 22 (1): 31–72.
- Calonico, S., M.D. Cattaneo, M.H. Farrell, and R. Titiunik. 2017. rdrobust: Software for regression-discontinuity designs. *The Stata Journal* 17 (2): 372–404.
- Calonico, S., M.D. Cattaneo, and R. Titiunik. 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82 (6): 2295–2326.
- Cattaneo, M.D., M. Jansson, and X. Ma. 2020. Simple local polynomial density estimators. *Journal of the American Statistical Association* 115 (531): 1449–1455.
- Cotlear, D., S. Nagpal, O. Smith, A. Tandon, and R. Cortez. 2015. *Going universal: How 24 developing countries are implementing universal health coverage from the bottom up*. Washington, DC: World Bank.
- Finkelstein, A. 2007. The aggregate effects of health insurance: Evidence from the introduction of medicare. *The Quarterly Journal of Economics* 122 (1): 1–37.
- Froelich, M., and A. Landmann. 2018. Effects of insurance on child labour: Ex-ante and ex-post behavioural changes. *The Journal of Development Studies* 54 (6): 1002–1018.
- Government of Pakistan. 2016. *Pakistan Social and Living Standards Measurement Survey 2014/2015*. Technical report. Statistics Division, Pakistan Bureau of Statistics.
- Habib, S.S., S. Perveen, and H.M.A. Khuwaja. 2016. The role of micro health insurance in providing financial risk protection in developing countries a systematic review. *BMC Public Health* 16 (1): 281.
- Imbens, G.W., and D.B. Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge: Cambridge University Press.
- Khan, S.A. 2016. Comparison of Federal and Khyber Pakhtunkhwa health insurance schemes, qualitative analysis of their proximity to universal health coverage. Master's thesis. Khyber Medical University.



- Lagomarsino, G., A. Garabrant, A. Adyas, R. Muga, and N. Otoo. 2012. Moving towards universal health coverage: health insurance reforms in nine developing countries in africa and asia. *The Lancet* 380 (9845): 933–943.
- Landmann, A., and M. Frölich. 2015. Can health-insurance help prevent child labor? An impact evaluation from Pakistan. *Journal of Health Economics* 39: 51–59.
- Levine, D., R. Polimeni, and I. Ramage. 2016. Insuring health or insuring wealth? an experimental evaluation of health insurance in rural cambodia. *Journal of Development Economics* 119: 1–15.
- Mangham, L.J., and K. Hanson. 2010. Scaling up in international health: what are the key issues? *Health Policy and Planning* 25 (2): 85–96.
- McCrary, J. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142: 698–714.
- Nawaz, S., and N. Iqbal. 2021. How cash transfers program affects environmental poverty among ultra-poor? insights from the bisp in pakistan. *Energy Policy* 148: 111978.
- Nishtar, S., T. Boerma, S. Amjad, A.Y. Alam, F. Khalid, and ul Haq, I., and Mirza, Y. A. 2013. Pakistan's health system: Performance and prospects after the 18th Constitutional Amendment. *The Lancet* 381 (9884): 2193–2206.
- O'Leary, S., Cheema, I., Hunt, S., Carraro, L., and Pellerano, L. 2011. *Benazir Income Support Programme impact evaluation. Baseline survey report*. Oxford: Oxford Policy Management.
- Oxford Policy Management. 2016. April. *Social Health Protection Initiative, Government of Khyber Pakhtunkwa—inception report*. Technical report. Oxford Policy Management.
- Oxford Policy Management. 2017. September. *Social Health Protection Health Insurance Programme - consulting services for programme implementation, Pakistan - progress report (January to June 2017)*. Technical report. Oxford: Oxford Policy Management.
- Prinja, S., A.S. Chauhan, A. Karan, G. Kaur, and R. Kumar. 2017. Impact of publicly financed health insurance schemes on healthcare utilization and financial risk protection in India: A systematic review. *PLoS ONE* 12 (2):e0170996.
- Reich, M.R., J. Harris, N. Ikegami, A. Maeda, C. Cashin, E.C. Araujo, K. Takemi, and T.G. Evans. 2016. Moving towards universal health coverage: Lessons from 11 country studies. *The Lancet* 387 (10020): 811–816.
- Shabbir, A., and S.A. Malik. 2016. Measuring patients' healthcare service quality perceptions, satisfaction, and loyalty in public and private sector hospitals in Pakistan. *International Journal of Quality and Reliability Management* 33 (5): 538–557.
- Smith, J.A., and P.E. Todd. 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics* 125 (1–2): 305–353.
- Thornton, R.L., L.E. Hatt, E.M. Field, M. Islam, F. Solis Diaz, and M.A. Gonzalez. 2010. Social security health insurance for the informal sector in nicaragua: A randomized evaluation. *Health Economics* 19 (S1): 181–206.
- Todd, P.E. 2010. Matching estimators. In *Microeconometrics*, 108–121. London: Palgrave Macmillan.
- TRC. 2012, June. *Health facility assessment, Khyber Pakhtunkhwa provincial report*. Tech. Rep.
- Uddin, B.R., S. Rafi, H. Uddin, and A. Khurshid. 2013. *Targeting survey spot check* (Tech. Rep.). Benazir Income Support Programme (Innovative Development Strategies).
- Vidyattama, Y., R. Miranti, and B.P. Resosudarmo. 2014. The role of health insurance membership in health service utilisation in Indonesia. *Bulletin of Indonesian Economic Studies* 50 (3): 393–413.
- Wagstaff, A. 2002. Poverty and health sector inequalities. *Bulletin of the World Health Organization* 80: 97–105.
- Wagstaff, A., M. Lindelow, G. Jun, X. Ling, and Q. Juncheng. 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.
- WHO. 2013. *Pakistan: Health system profile*. Technical report. World Health Organization Regional Office for the Eastern Mediterranean.
- WHO. 2017. *Annual report 2017*. Technical report. World Health Organization. www.healthkp.gov.pk/wp-content/uploads/2018/03/annual-report2017.pdf.

