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

Simona Helmsmüller [*] Andreas Landmann [†]

January 31, 2020

Working Paper

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 imperfect rollout and compare insured and uninsured households using propensity score matching. Second, we exploit that eligibility is based on an exogenous poverty score threshold and apply a regression discontinuity design. 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, microinsurance, program evaluation, health care consumption, Pakistan
JEL Codes: O12, I13, I15, O22

[*] Corresponding Author. University of Mannheim, Department of Economics, L7, 3-5, 68131 Mannheim, Germany, simona.helmsmueller@gess.uni-mannheim.de

[†] University of Göttingen

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 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. A particularly 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 exploit incomplete rollout and match insured to comparable, eligible but uninsured households using propensity score matching. Second, we implement a regression discontinuity design, using the fact that eligibility for the program is based on an exogenous poverty score.

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 (which is consistent with a larger reduction of relative costs of private care versus public care which we observe in our data). Given the better resources and higher client satisfaction associated with private hospitals, we 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 strength the capacity of public hospitals might be necessary.

Several studies have analyzed the effect of protecting low-income households through health insurance and these have shown some promising impacts in terms of financial protection (Habib, Perveen, and Khuwaja 2016), access to medical services (e.g. Juetting 2004; Wagstaff, Lindelow, Jun, Ling, and Juncheng 2009), 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, including India, China, Indonesia, and most recently also Pakistan. Whereas some studies exist on the reforms in India, China, and Indonesia (Banerjee et al. 2019, Wagstaff et al. 2009, Prinja, Chauhan, Karan, Kaur, and Kumar 2017, Vidyattama, Miranti, and Resosudarmo 2014), evidence on the Pakistani case 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. Government spending on health has been well below one percent of GDP until recently, and around 90 percent of private expenditure had to be paid out-of-pocket in 2016 (World Bank Indicators 2016).[1] This situation increases the need for inclusive insurance solutions. 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 (Asuming 2013, Banerjee et al. 2019).

At the same time, Pakistan features a dual health sector with both private and public providers operating in the same market. 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. 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. Despite the importance of the research question for public policy, there is virtually no evidence on the impact of state-funded insurance schemes on public versus private health systems. 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 Section 2 we provide the country context and program details. In Section 3 we present details of our dataset and summarize descriptive statistics. In Section 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 Section 6 we discuss effect channels and challenges in implementation. The last section concludes.

[1] World Bank Indicators until 2016 are available at http://data.worldbank.org/country/pakistan.

2 Rationale of the intervention

2-1 Challenges in health care in Pakistan

Poor health is wide-spread 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 has become 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 a political interference and corruption, shortage of trained personnel, staff absenteeism, non-functioning facilities, and lack of medicines (ADB 2019, WHO 2013). 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, 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 of 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

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

the German Development Bank (KfW), 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 based on a proxy means test (PMT) in 2010. 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 KfW verified the distribution of cards via a limited number of spot checks. By end of June 2016, the insurer reported 87.3% of the target population as enrolled (in the two pilot districts considered in our study) (Oxford Policy Management 2017).[3]

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)[4] per person.[5] This covered treatment for normal delivery and C-sections, as well as a pre-defined list of 497 medical procedures requiring hospitalization. Notably, outpatient care is not covered by the program.[6]

The insured households could obtain these services at one of the empanelled hospitals, which include public and private health care providers. Prior to the distribution of insurance cards, the program identified and informed 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, because of religious beliefs, or out of fear of stricter tax controls (Oxford Policy Management 2016, Oxford Policy Management 2017). Public hospitals also showed little interest in the program until Government influence was used to encourage

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

[4]Exchange rate on December 31, 2015.

[5]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.

[6]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 Appendix A.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.

joining the program. Nevertheless, the program was able to empanel around one third of the candidate private hospitals, as well as 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.[7] Prior to card distribution, 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 hours 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 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.[8]

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 caused 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 Section 6.

[7] 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).

[8] After our study, the Government initiated a new survey to update the poverty score, which will lead to improved targeting in later phases.

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.[9] 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 data (autumn 2015). We carried out the endline survey twelve to fifteen 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 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 these pilot districts, where the initial rollout plan was largely followed and where our identification strategies are still valid (Malakand and Mardan).[10] For robustness checks, we also use data of two control districts, namely Lower Dir and Swat.[11] Appendix A.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

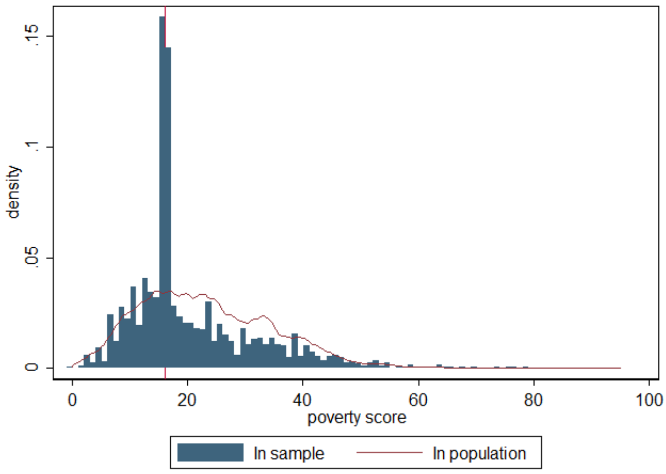
[9]Additionally, the Government aimed at increasing quality and accountability of public hospitals by ensuring a client-based flow of funds through the program. We do not consider supply-side effects in this study.

[10]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 Section 4-2.

[11]We had selected four control districts using an algorithm matching on publicly available socio-demographic indicators and health infrastructure. However, the pre-mature roll-out of the second phase in two of those districts renders them unusable.

2010 served as a sampling frame for the third and fourth stage: Stratified random sampling of 70 villages and then 1,200 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. This means that in each survey cluster we additionally sampled 20 percent more households at random below the cutoff and 40 percent more households closely around the cutoff. Therefore, our baseline sample in the pilot districts consists of 1,920 households of which 828 were eligible for the insurance. Figure 1 depicts the distribution of the poverty score in the population and in our sample respectively, in the selected union councils of the two pilot districts, illustrating the degree of oversampling below and around the cut-off score of 16.17.

Figure 1: Distribution of poverty score in sample and population in selected UCs of Malakand and Mardan



► *Note*: This figure shows the distribution of the poverty score in our sample (blue bars) and in the sampling frame (red line) in selected union councils (UCs) of Malakand and Mardan.
► Sample: Household-level sample (panel, N=1,842) and BISP sampling frame (N=71,591).
► Source: Baseline survey (2015) and BISP survey (2010).
► The poverty cut-off score (assigned in 2010) determining eligibility for the first phase of the SHPI is 16.17. The figure illustrates the degree of oversampling below and around this cut-off.

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 1,842 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 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 6,007 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. We provide further information on the survey methodology, including questionnaires and sampling strategy, in the Online Appendix.

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. 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 where the variation was identified as large. 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 should still be internally consistent as 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 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 the Appendix (differences to table below are marginal).

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

	Full sample				Eligible sample			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(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	8,888.89	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
Tab 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
Dif'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	1,842				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
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				6,007			

- *Note*: 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 Mardan shortly before our endline survey.
- Table A.2 in the Appendix contains the statistics for the subsample of randomly selected households (without oversampling). Differences are small and statistically insignificant.

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 versus 23 years), more likely to be of school-aged (38% versus 33%), and a larger share has not completed primary school (63% versus 59%). Conversely, a smaller share of members has completed secondary school or higher (8% versus 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.

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

Admission to inpatient care	Logit 1 Coef. (1)	Std. Error (2)	P-value (3)	Logit 2 Coef. (4)	Std. Error (5)	P-value (6)	Logit 3 Coef. (7)	Std. Error (8)	P-value (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

- *Note*: This table shows the coefficients of a logit regression 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 regression, Columns (2), (5), (8) the standard error and Columns (3), (6), (9) the p-value of the two-sided test 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 the Appendix contains the results excluding childbirth related hospitalization. The general direction and significance of coefficients remains unchanged, but the magnitude of the correlation between wealth index and hospitalization is larger.

Similarly, 67.2% of male adults have done any work 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.[12] 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 hospital within the twelve months prior to baseline. To understand the socioeconomic drivers of using inpatient services, we run three logit regressions including individual and household covariates with different proxies for poverty (results shown in Table 2). Older and female individuals 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 the Appendix A.4). 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 (if admitted somewhere), 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 including squared terms for a more flexible form. The wealth

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

index and its square are strongly correlated not only with admission to inpatient care (Column (1)), but also with the subjective health status, which improves for individuals in wealthier households (Column 4). Also, wealthier households are more likely to visit a private hospital, where care is frequently perceived to be of higher quality, (Column 7) and less likely to report an incident of neglected health care (Column 10).[13] Our data therefore supports the hypothesis that poor households are indeed restricted in their access to health care, both in quantity and perceived quality.

Table 3: Usage patterns and health care needs

	Admission to inpatient care			Health status			Use of private versus public hospitals			Neglected health care		
	Coef.	Std. Error	P-value	Coef.	Std. Error	P-value	Coef.	Std. Error	P-value	Coef.	Std. Error	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			1,849		

- *Note*: 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 error and Columns (3), (6), (9), (12) the p-value of testing that the coefficient is equal to zero, on different outcomes respectively. Standard error 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.

[13] 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.

4 Econometric approach

We use two identification strategies, which estimate different effects. First, 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). Second, we apply a sharp Regression Discontinuity Design (RDD) using the poverty score as running variable. This provides an estimate of a Local Average Treatment Effect (LATE).[14] Table 4 illustrates the different samples considered for the two estimators.

Table 4: Treatment and control groups in two estimators

	Treatment group	Control group
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 $\in [0, 16.17]$
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 $\in [16.17, 16.17+B]$

▸ *Note*: This table illustrates the different samples considered for the propensity score matching and regression discontinuity design estimators respectively.

4-1 Propensity score matching (PSM)

Among eligible households in our PSM sample, the program achieved self-reported enrollment rates of 65.2% of households. This is remarkably high[15], yet a sizable number of households which were targeted by the program did not report themselves insured in our survey, likely due to imperfections in program roll-out. It is important to note that 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 hence be part of the control group.[16] For ease of notation, we will use the term *(un-)insured* to refer to the self-reported insurance status from now on.

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

[14] In experimental designs such as randomized control trials, often the ATT is an intention-to-treat effect using program eligibility as treatment indicator. The LATE is then calculated using an instrumental variable approach where actual or self-reported participation is instrumented via eligibility to treatment. Note that in this quasi-experimental study we explicitly diverge from this convention.

[15] 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 premium subsidization and assistance in the enrolment process.

[16] 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. Three factors likely contribute to the deviance: (i) The household never received the card and the administrative data is fraudulent. (ii) The household was enrolled after our endline survey. (iii) The household was enrolled, but the interviewed household member was not aware of it.

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

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, $P_j = Pr_j(insured|Z)$ is the propensity score, i.e., the probability of being insured conditional on a set of covariates Z, $G()$ is the epanechnikov kernel, B_n the bandwidth.

Two assumptions are key to this approach (Todd 2010): Conditional mean independence and common support. 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.[17] Many of these variables, such as education, prior insurance knowledge or household size, are observable to us from the baseline survey and we systematically include them in our matching procedure. Tables B.4 and B.5 in the Appendix show means of all collected baseline variables for insured and uninsured households. 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. In addition, there are 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 which are 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 Figure 2.[18] Correspondingly, we tested 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). We therefore estimate propensity scores using a probit model with a vector of union-council dummies and selected baseline variables, and include linear as well as quadratic terms.[19] Tables B.4 and B.5 in the Appendix demonstrate the achieved balancing on union councils and baseline variables. Figures B.2 and B.3 show the distribution of the poverty score among insured and matched uninsured samples, underlining the credibility of the conditional mean independence assumption.

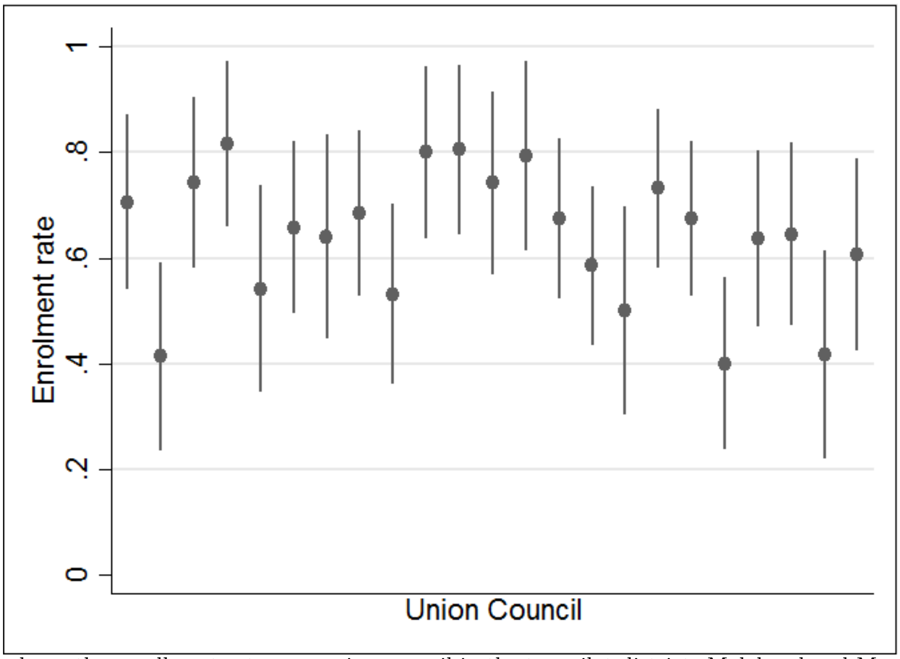
To ensure common support, we restrict our treatment sample to individuals or households with a propensity score above the 99th percentile score among the control group, as suggested

[17] 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).

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

[19] The selection of baseline covariates to be included in the propensity score estimation is an important step in our econometric approach. 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 procedure described in Imbens and Rubin (2015) (chapter 13) and summarized in Appendix B.1.

Figure 2: Enrolment rates in main sample of eligible households across 24 union councils



► *Note*: 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 level which served as sampling clusters (second stage sampling unit).

in Caliendo and Kopeinig (2008). This eliminates approximately 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.4 and B.5 in the Appendix demonstrate that common support is thus sufficiently ensured.[20]

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 (9,999 repetitions).[21] Note that we bootstrap the whole process of estimating propensity scores, imposing common support, matching observations, and estimating effects. We use clustering at the union council level, but as robustness check also clustered at household-level for member-level variables. The differences are marginal and do not affect inference.

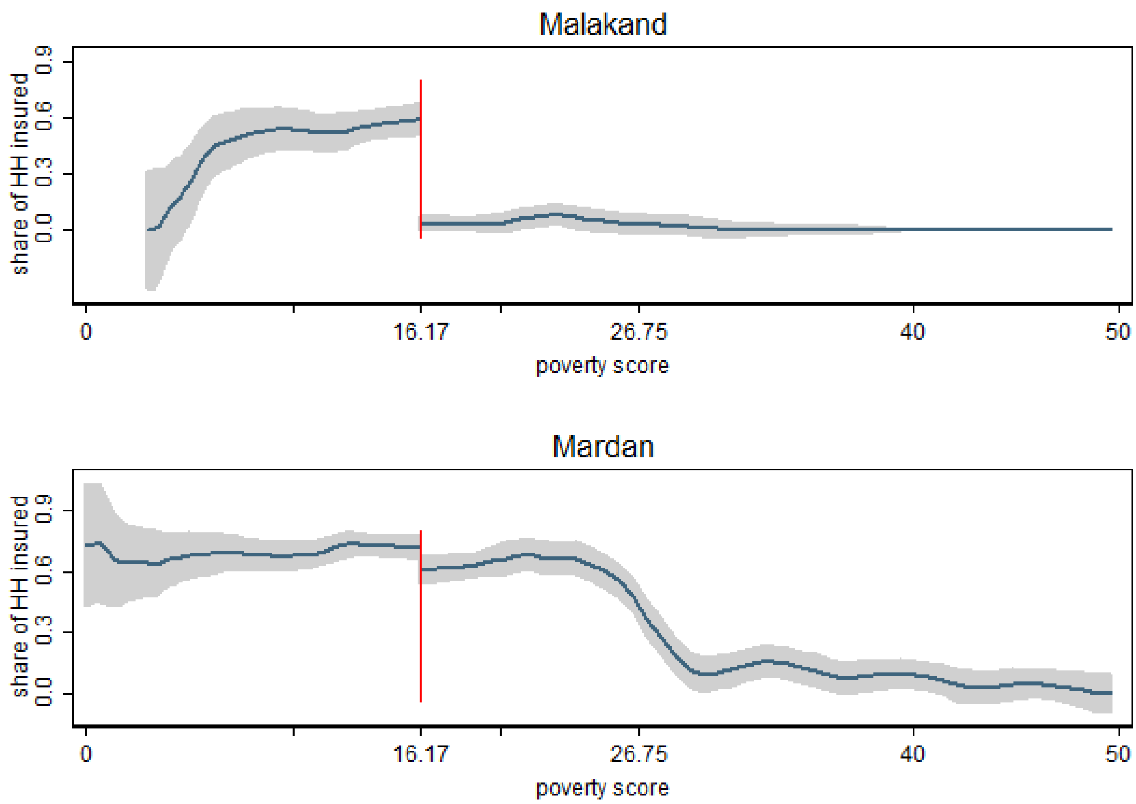
[20]We concentrate on estimation of average treatment effects on the treated, hence we need not drop control observations for whom there is no match in the treatment group.

[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.

4-2 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 3 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 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, was measured covering a period of twelve months, it is more appropriate to consider households covered under the second phase as (largely) uninsured. If anything, this should lead to a slight attenuation of the estimated affect.[22] Also note that we here calculate intention-to-treat effects using a sharp RDD design with treatment determined by the poverty score only.

Figure 3: Share of households insured (self-reported) by poverty score



► *Note:* 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, $N_{Malakand} = 617$, $N_{Mardan} = 1,232$).
► Source: Endline survey (2017).

[22]For a quick back-of-the-envelop calculation, note that the effect will be reduced approximately by the average time the control group was covered (estimated as 2/12 months) times the share of 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%.

We calculate local linear regression models to the left and right of the cut-off score, where the bandwidth is estimated to minimize the mean squared error as suggested in Imbens and Kalyanaram (2009). As there is the usual trade-off between bias and efficiency, we cross-checked results for a selection of alternative bandwidths as well, but did not find this to affect inference. For household-level variables, we calculate unclustered standard errors as we compare households within the unit of clustering (union council). For member-level variables, we cluster standard errors at the household-level to account for within-household correlation.

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. 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. However, the density test proposed by McCrary (2008) fails to detect a significant discontinuity, as illustrated in Figure B.7 in the Appendix, which suggests that there was no systematic manipulation of the assignment variable.

Another important assumption for the internal validity of the RDD approach is that the distributions of outcome variables, conditional on treatment status, are smooth around the cut-off score. This might be a problematic point, as the poverty score was initially derived to determine eligibility to a nation-wide social program, which includes among other benefits an unconditional cash transfer. 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 1,000 PK and another 85% reporting 1,500 PKR) and, most importantly, 80% of households claim to have received these transfers already at baseline. While the smoothness assumption is not directly testable, we make use of the panel structure and calculate pseudo-effects on the outcome variables using baseline data. We find no significant effects in any of the outcomes of interest, as illustrated in Appendix B.6. This suggests that the national social program did not create a discontinuity at the cutoff in the outcomes of interest of our study, and hence does not confound our analysis.

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 twelve 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 versus 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 twelve months (neglected health care). For all these four key outcomes, we estimate the effects of providing insurance coverage using both Propensity Score Matching (PSM) and a Regression Discontinuity Design (RDD).

Table 5 contains the results from both the PSM and the RDD estimations. For example, in the first line the outcome considered is whether an individual has sought inpatient care in the past twelve 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 2,111 uninsured and 3,638 insured household members. Our RDD estimation yields a similar coefficient of -0.003 with a standard error of 0.011, where we rely on 2,988 observations below and 2,466 observations above 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.

Table 5: Effects on inpatient care consumption

Outcome	Control Mean (1)	PSM β_{ATT} (2)	PSM S.E. (3)	RDD β_{LATE} (4)	RDD S.E. (5)
Individual outcomes					
Usage of inpatient care	0.059	**-0.002**	0.011	**-0.003**	0.011
N (uninsured/ insured, left/ right of cut-off)		*2,111/ 3,638*		*2,988/ 2,466*	
conditional on usage of inpatient care					
More than one admittance	0.195	**0.022**	0.077	**0.087**	0.087
N (uninsured/ insured, left/ right of cut-off)		*107/ 212*		*208/ 152*	
Usage of private versus public hospitals	0.383	**0.068**	0.075	**0.279*** **	0.088
N (uninsured/ insured, left/ right of cut-off)		*101/ 202*		*145/ 118*	
Household outcomes					
Neglected health care	0.062	**0.007**	0.026	**0.006**	0.023
N (uninsured/ insured, left/ right of cut-off)		*461/ 277*		*518/ 425*	

- *Note*: 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 local average treatment 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: 9,999. Reported sample size refers to area of common support (overall sample size: 795 households with 6,007 members).
- *Note on RDD:* Estimated using local linear regression models on both sides of the cutoff score, and reported analytic S.E. are based on the regressions. Optimal bandwidth from Imbens and Kalyanaram (2009) minimizing the mean squared error. Reported sample size refers to observations within selected bandwidth (overall sample size: 1,842 households with 12,862 members).
- The statistical significance is given as follows: * indicates $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$, with the null hypothesis of the two-sided test being a zero effect size.

Despite the large number of observations at our disposition, we find no significant effects of the program on the usage of inpatient care, neither averaged across all treated (PSM) nor locally around the cut-off (RDD). 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). In other words, we can exclude short-term transformative changes in seeking hospitalization in our sample.

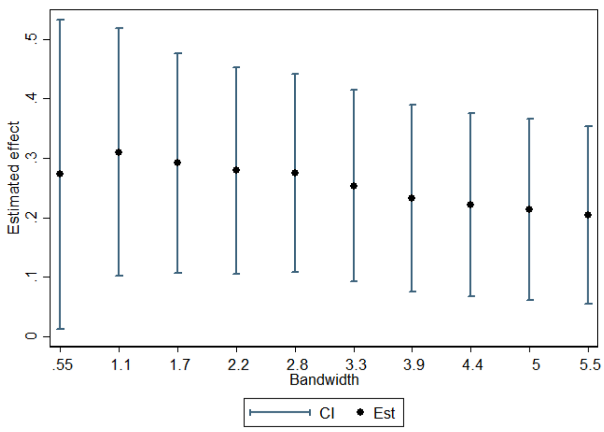
Among individuals who reported a case of inpatient care, we also look at the share of individuals with more than one stay at a hospital and also find no effect here. As we did not find an effect of the program on the probability of using any inpatient care before, we believe in the validity of this result, 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.[23]

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 PSM as well as the RDD point estimates are very close to zero. Note that 90 percent 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 childbirths. 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 PSM and RDD samples does not reveal any significant differences.[24]

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 versus public hospitals in the RDD estimation. This result is robust against different bandwidth specifications, as illustrated in Figure 4.

Figure 4: RDD ESTIMATES FOR USE OF PRIVATE VERSUS PUBLIC HOSPITALS, DIFFERENT BANDWIDTHS



► *Note*: This figure shows the point estimates and confidence intervals for the impact of the insurance on the use of private versus public hospitals for different bandwidths and accounting for clustering on household level. Main specification uses bandwidth of 2.20.
► Sample: Member-level RDD sample with case of inpatient care (panel, N=690).
► Source: Endline survey (2017).

[23]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.

[24]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. 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 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 KfW notes that 95.59% of admissions in the two districts were registered in private hospitals (Oxford Policy Management 2017).[25] 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, 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 would constitute an improvement of health for the beneficiaries, if private hospitals provide better quality of care. However, whereas public hospitals are hardly monitored, private hospitals do not even register, rendering it notoriously difficult to measure quality of care.[26] 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. Therefore, we restrict ourselves to subjectively perceived quality of care. Evidence that patients associate higher quality of care with private hospitals comes from our endline survey, where 65.82% of respondents in our PSM sample in the endline survey rather agreed than disagreed with the statement that private facilities provide better quality of service than public facilities.[27] Most importantly however, as we have laid out in Section 3-3, 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 reasonably assume that individuals would not be willing to pay higher prices in private hospitals if these were not 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.

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

[25] 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).

[26] 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.

[27] 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.

high health financing needs.[28] 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.[29] Table 6 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.

Table 6: Hetereogeneous effects on usage of inpatient care

Subgroup	Control Mean (1)	PSM β_{ATT} (2)	PSM S.E. (3)	RDD β_{LATE} (4)	RDD S.E. (5)
Female household members	0.064	**0.005**	0.013	**0.006**	0.015
N (uninsured/ insured, left/ right of cut-off)		*1,017 / 1,740*		*1,253 / 1,633*	
Adults above 16 years	0.097	**-0.020**	0.022	**-0.025**	0.019
N (uninsured/ insured, left/ right of cut-off)		*1,051 / 1,823*		*1,266 / 1,584*	
Baseline health status below median (< 5)	0.093	**-0.002**	0.017	**-0.007**	0.017
N (uninsured/ insured, left/ right of cut-off)		*825 / 1,281*		*1,108 / 1,419*	
Wealth index below median ($< -.60$)	0.079	**-0.015**	0.023	**-.002**	0.138
N (uninsured/ insured, left/ right of cut-off)		*1,065 / 1,772*		*1,362/ 1,574*	

- *Note*: 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 (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 local average treatment effect for households just below the cut-off poverty 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: 9,999. Reported sample size refers to area of common support.
- *Note on RDD:* Estimated using local linear regression models on both sides of the cutoff and reported analytic S.E. are based on the regressions. Optimal bandwidth from Imbens and Kalyanaram (2009) designed to minimize the mean squared error. Reported sample size refers to observations within selected bandwidth.
- The statistical significance is given as follows: * indicates p < 0.1, ** p < 0.05, and *** p < 0.01, with the null hypothesis being a zero effect size.

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 PSM and RDD approaches meaningfully complement each other, because they allow us to look at effects on two different populations (average treatment effect on the treated versus local intention-to-treat effect at the cutoff), and thereby provide a more complete picture.[30] Also, they complement each other in overcoming relative weaknesses. For example, PSM relies on comparing households based on self-reported enrollment. We argued that self-reported might be more relevant than official coverage in a

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

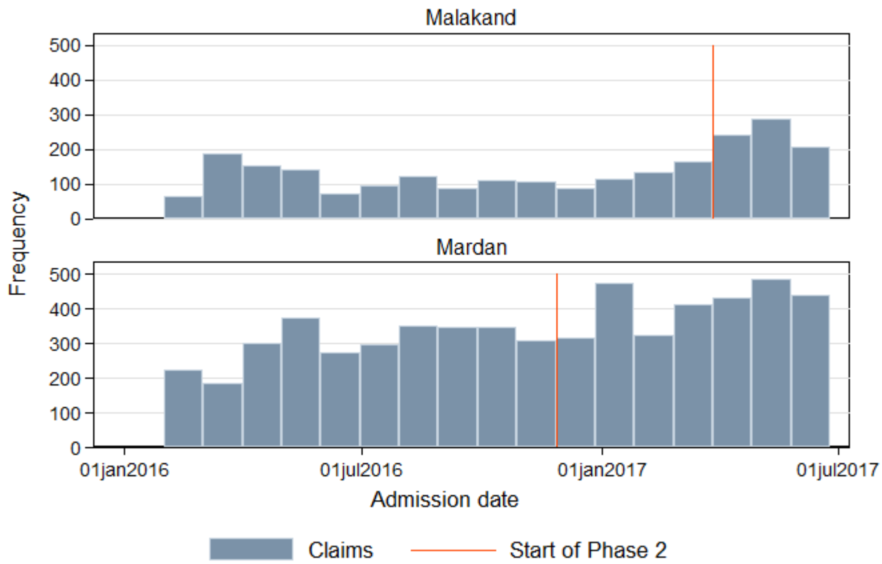
[29] For these subsamples, we calculate propensity scores based on covariates Z_j, which are selected anew for each subsample from the set of all baseline covariates following the same procedure as for the complete PSM sample, described in Appendix B.1.

[30] Note that we do not focus on the validity of the assumptions underlying our empirical approach here. Those are discussed in Section 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, and run placebo analyses using baseline and control district observations.)

context of imperfect rollout and awareness. Nevertheless, both self-reported as well as official coverage might be noisy measures of 'effective' coverage; and it is therefore valuable to have the RDD estimate, which is based on an exogenous eligibility cutoff, to confirm results.

One important aspect in the interpretation of the findings is that the endline survey took place twelve to fifteen 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. In line with this, absolute claim numbers in our 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 Figure 5, 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%.

Figure 5: Absolute number of claims over time (admin data)



► *Note*: 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).

There are likely more persistent reasons why the program did not increase overall inpatient care consumption. 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/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 7 (first line of Panel A). 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 insurance leads to more 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.[31] 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 empanelled hospital or within 40 minutes away from any hospital. We display results in Table 7 (second and third line of Panel A). Again, we find no evidence for 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 used 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.[32] We present results in Table 7, Panel B.[33] 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 percent). Additionally, we asked whether those with a hospitalization case experienced sleepless nights due 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 is

[31] Originally, one more hospital was empanelled in Mardan but dropped due to conflict of interest of the owner. Contracts with two of the empaneled private hospitals in Mardan were suspended due to using unnecessary procedures in 2017.

[32] 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 found to be a negligible portion of total expenditures.

[33] Due to the highly screwed distribution of the quantitative cost variables, we use log values as outcomes.

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 Section 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).[34] Interestingly, it descriptively looks like this difference shrinks after the program rollout only for the insured, driven by a strong decrease in private care costs.[35] 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.

[34]The diseases treated in public and private facilities at baseline were 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.

[35]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 seem 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 noninsured).

Table 7: Evidence on program limitations

Outcome	Control Mean (1)	PSM β_{ATT} (2)	PSM S.E. (3)	RDD β_{LATE} (4)	RDD S.E. (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**	0.011
N (uninsured/ insured)		*2,007/ 2,313*		*2,337/ 2,375*	
HH living within 10 km distance to next empaneled hos.	0.066	**-0.014**	0.022	**-0.022**	0.013
N (uninsured/ insured)		*1,061/ 1,881*		*1,220/ 1,451*	
HH living within 40 min travel distance to next hospital	0.049	**-0.005**	0.012	**0.003**	0.011
N (uninsured/ insured)		*1,061/ 1,419*		*1,196/ 1,465*	
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	**-0.240**	0.260
N (uninsured/ insured)		*107/ 212*		*127/ 142*	
Log of cost for diagnosis and treatment (PKR, win99)	4.833	**-0.261**	0.562	**-0.498**	0.661
N (uninsured/ insured)		*107/ 212*		*134/ 168*	
Log of cost for medicines (PKR, win99)	8.238	**-0.316**	0.398	**-0.032**	0.364
N (uninsured/ insured)		*107/ 212*		*131/ 155*	
Sleepless night due to hospital costs (PKR, win99)	0.486	**0.083**	0.095	**0.021**	0.099
N (uninsured/ insured)		*107/ 212*		*147/ 204*	

- *Note*: 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 local average treatment 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: 9,999. Reported sample size refers to area of common support.
- *Note on RDD:* Estimated using local linear regression models on both sides of the cutoff and reported analytic S.E. are based on the regressions. Optimal bandwidth from Imbens and Kalyanaram (2009) designed to minimize the mean squared error. Reported sample size refers to observations within selected bandwidth.
- The statistical significance is given as follows: * indicates $p < 0.1$, ** $p < 0.05$, 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.

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 propensity score matching approach, comparing insured and uninsured but eligible households, and a regression discontinuity approach, comparing households just above and just below the exogenous poverty cut-off score. While the latter has a high 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 assumption, 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. 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. 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 versus private market sectors seems promising.

Acknowledgments

The survey and administrative data were produced and procured for the evaluation of the SHPI program, a work which is largely independent from this paper. That evaluation was a cooperation of the authors with the KfW Development Bank 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, and Christoph Strupat for valuable comments and feedback on earlier versions of this work. The authors take sole responsibility for any remaining errors.

Katja Hofmann, Fawad Ejaz, and Jan Berkes provided excellent research assistance.

Bibliography

Abadie, A., & Imbens, G. W. (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* (Tech. Rep.). Asian Development 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., Finkelstein, A., Hanna, R., Olken, B. A., Ornaghi, A., & Sumarto, S. (2019). *The challenges of universal health insurance in developing countries: Evidence from a large-scale randomized experiment in Indonesia* (Tech. Rep.). National Bureau of Economic Research.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72.

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., & Landmann, A. (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, March). *Pakistan Social and Living Standards Measurement Survey 2014/ 2015* (Tech. Rep.). Statistics Devision, Pakistan Bureau of Statistics.

Habib, S. S., Perveen, S., & Khuwaja, H. M. A. (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., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Juetting, J. P. (2004). Do community-based health insurance schemes improve poor peoples access to health care? Evidence from rural Senegal. *World Development*, *32*(2), 273–288.

Khan, S. A. (2016). *Comparison of Federal and Khyber Pakhtunkhwa health insurance schemes, qualitative analysis of their proximity to universal health coverage* (Unpublished master's thesis). Khyber Medical University.

Landmann, A., & Froelich, M. (2015). Can health-insurance help prevent child labor? An impact evaluation from Pakistan. *Journal of Health Economics*, *39*, 51–59.

McCrary, J. (2008). Manipulation of the running variable in the regression disconinuity design: A density test. *Journal of Econometrics*, *142*, 698-714.

Nishtar, S., Boerma, T., Amjad, S., Alam, A. Y., Khalid, F., ul Haq, I., & Mirza, Y. A. (2013). Pakistan's health system: Performance and prospects after the 18th Constitutional Amendment. *The Lancet*, *381*(9884), 2193–2206.

OLeary, S., Cheema, I., Hunt, S., Carraro, L., & Pellerano, L. (2011). *Benazir Income Support Programme impact evaluation: Baseline survey report.* Oxford Policy Management Oxford.

Oxford Policy Management. (2016, April). *Social Health Protection Initiative, Government of Khyber Pakhtunkwa - inception report* (Tech. Rep.). Oxford Policy Management.

Oxford Policy Management. (2017, September). *Social Health Protection Health Insurace Programme - consulting services for programme implementation, Pakistan - progress report (january to june 2017)* (Tech. Rep.). Oxford Policy Management.

Prinja, S., Chauhan, A. S., Karan, A., Kaur, G., & Kumar, R. (2017). Impact of publicly financed health insurance schemes on healthcare utilization and financial risk protection in India: A systematic review. *PLoS ONE*, *12*(2). doi: 10.1371/journal.pone.0170996

Shabbir, A., & Malik, S. A. (2016). Measuring patients' healthcare service quality perceptions, satsifaction, and loyalty in public and private sector hospitals in pakistan. *International Journal of Qualtiy and Relability Management*, *33*(5), 538-557.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, *125*(1-2), 305–353.

Todd, P. E. (2010). Matching estimators. In *Microeconometrics* (pp. 108–121).

TRC. (2012, June). *Health facility assessment, Khyber Pakhtunkhwa provincial report* (Tech. Rep.).

Uddin, B. R., Rafi, S., Uddin, H., & Khurshid, A. (2013). *Targeting survey spot check* (Tech. Rep.). Benazir Income Support Programme. (Innovative Development Strategies)

Vidyattama, Y., Miranti, R., & Resosudarmo, B. P. (2014). The role of health insurance membership in health service utilisation in Indonesia. *Bulletin of Indonesian Economic Studies*, *50*(3), 393–413. doi: 10.1080/00074918.2014.980380

Wagstaff, A. (2002). Poverty and health sector inequalities. *Bulletin of the World Health Organization*, *80*, 97–105.

Wagstaff, A., Lindelow, M., Jun, G., Ling, X., & Juncheng, Q. (2009). Extending health insurance to the rural population: An impact evaluation of China's new cooperative medical scheme. *Journal of Health Economics*, *28*(1), 1–19.

WHO. (2013). *Pakistan: Health system profile* (Tech. Rep.). World Health Organization

Regional Office for the Eastern Mediterranean.

WHO. (2017). *Annual report 2017* (Tech. Rep.). World Health Organization. (http://www.healthkp.gov.pk/wp-content/uploads/2018/03/annualreport2017.pdf)

Appendices

A Additional information on the program and survey

A.1 Features of the two programme phases

The following table is taken from Khan (2016).

Table A.1: Features of the two programme phases

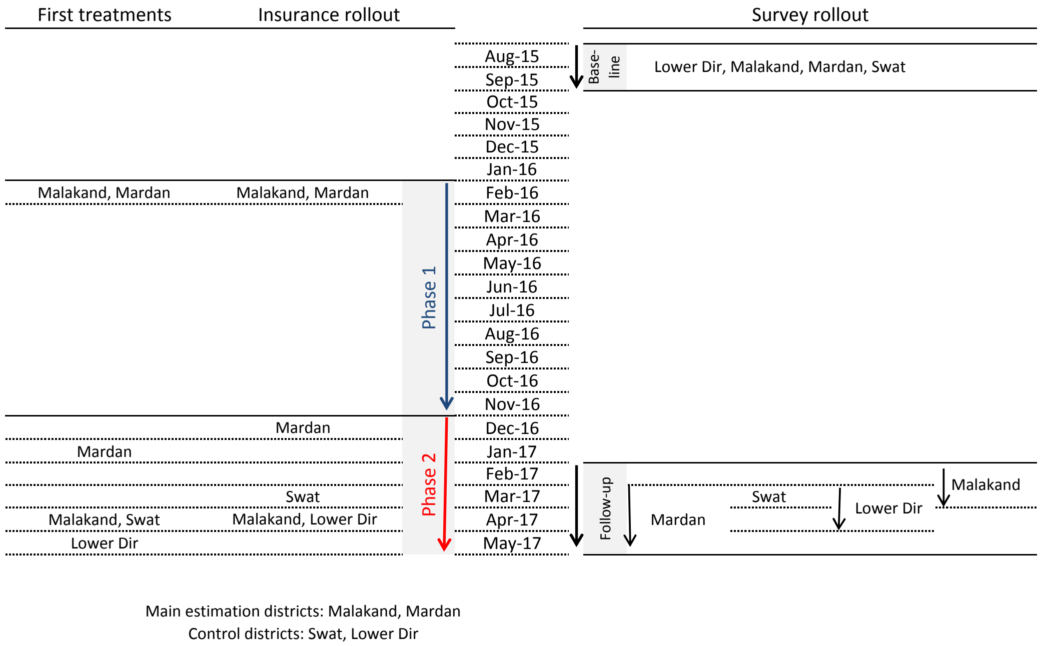
Salient features	**SHPI Phase-1**	**SHPI Phase-2**
Area	**Four (4) districts**	**Entire province, i.e. twenty-six (26) districts**
Total funding	PKR 1.4 billion	PKR 5.4 Billion
Source of funding	Kreditanstalt für Wiederaufbau (KfW) + Government of KP	Government of KP
Funding (PKR)	Total cost: PKR 1.4 billion, KfW share; PKR 1,233 million (88%) KP share; PKR 1,66 million (12%)	Total cost: PKR 5.4 billion, all through Government's general revenue
Premium	PKR 1,661/- household	PKR 1,549/- household
Project launch	**15th December 2015**	**31st August 2016**
Duration of project	ADP scheme for 5 years	ADP schemes for 2 years
Who is covered		
Percentage of population	**21% poorest population of target districts**	**51% poorest population of entire KP**
Enrolment criteria	**Families with poverty score of 16.17 or less**	**Families with poverty score of 26.75 or less**[36]
Family size	**7 persons per household**	**8 persons per household**
Total population covered	0.1 Million households, comprised of 7 members, hence 0.7 million people are covered	1.8 million households, comprised of 8 members, hence 14.4 million people are covered
What is covered		
Type of services	**Inpatient services only**	**Predominantly inpatient services**
Outpatient cover	Maternity services only	Maternity services and cancer care
Secondary diseases	Almost all needing admission	All needing admission
Tertiary conditions	None	Yes, limited tertiary cover
What amount of expenditure is covered		
Mode of payment	Beneficiary-provider interaction is cashless. The insurer pays the providers.	Beneficiary-provider interaction is cashless. The insurer pays the providers.
Premium (Government paid)	PKR 1,661/- household	PKR 1,549/- household, it includes PKR 50 for stop loss coverage
Upper limit (secondary care)	**PKR 25,000/- per person per year, PKR 175,000/- per household per year**	**PKR 30,000/- per person per year, PKR 240,000/- per household per year**

[36] Officially, the cut-off score is 24.51. However, the administrative data reveals that households up to 26.75 were enrolled in our surveyed districts.

Upper limit (tertiary care)	None	PKR 3,000/- per household per year
Wage replacement	None	PKR 250/- per day for 3 days
Tertiary transportation	None	PKR 2,000/- after discharge from a tertiary care
Maternity transportation	None	PKR 1,000/- post-delivery at hospital
Burial allowance	None	PKR 10,000/- at death of insured household member, during a hospital admission
OPD voucher	None	One OPD visit after discharge from hospital

A.2 Timeline of program and survey

Figure A.1: Timeline of program implementation and survey



► *Note*: This figure illustrates the months of baseline and endline survey, enrollment and first reported card usage in hospitals for each of the four districts considered here.
► We estimate effects for households in the districts Malakand and Mardan. During the time of our study, Malakand saw roll-out of Phase 1 only, whereas Phase 2 was initiated three months prior to the endline survey in Mardan.
► For robustness checks, we use the districts Lower Dir and Swat. In Lower Dir, enrollment for Phase 2 started after the endline survey. In Swat, enrollment for Phase 2 was taking place in parallel to our follow-up survey, but hospitals had not started offering treatment yet.

A.3 Baseline characteristics for random sub-sample

Table A.2: Baseline characteristics of full sample and subsample of eligible population (selected variables)

	Full sample				Eligible sample			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Household-level variables								
Insurance status at endline	0.38	0.49	0.00	1.00	0.66	0.48	0.00	1.00
Poverty score	24.31	12.56	0.00	79.00	11.17	3.41	0.00	16.17
P.c. monthly income (sqr. root equiv.)	607.71	901.10	0.00	7500.00	373.72	535.92	0.00	4375.00
Wealth index	-0.08	2.02	-3.40	12.23	-0.64	1.54	-3.21	5.79
HH size	7.12	2.74	1.00	23.00	8.02	2.46	3.00	21.00
Electricity in HH	0.97	0.18	0.00	1.00	0.94	0.23	0.00	1.00
Tab water supply in residence	0.11	0.32	0.00	1.00	0.11	0.31	0.00	1.00
Private flush toilet	0.38	0.49	0.00	1.00	0.26	0.44	0.00	1.00
Reported dist. to next hosp. (minutes, win99)	43.76	26.21	0.00	150.00	45.66	27.51	0.00	150.00
Use of prof. assist. during childbirth	0.90	0.31	0.00	1.00	0.86	0.35	0.00	1.00
Case of neglected health care	0.15	0.35	0.00	1.00	0.18	0.38	0.00	1.00
Case of outpatient care	0.78	0.42	0.00	1.00	0.79	0.41	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
Dif'ty finding money for health care ¿= 8 (scale 1/10)	0.47	0.50	0.00	1.00	0.51	0.50	0.00	1.00
Having heard of insurance	0.02	0.15	0.00	1.00	0.01	0.12	0.00	1.00
Observations	1,146				566			
Panel B: Member-level variables								
Age (win99)	23.97	18.70	1.00	90.00	21.99	17.41	1.00	90.00
School-aged (6 to 16)	0.30	0.46	0.00	1.00	0.40	0.49	0.00	1.00
Female	0.48	0.50	0.00	1.00	0.49	0.50	0.00	1.00
Prim. school not comp'd.	0.57	0.50	0.00	1.00	0.66	0.47	0.00	1.00
Comp'd sec. educ. or higher	0.13	0.34	0.00	1.00	0.07	0.26	0.00	1.00
Worked for salary in previous month	0.20	0.40	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	0.20	0.00	1.00
Cost of last treatment (PKR, win99)	24,613	42,664	0	300,000	25,587	47,216	500	300,000
More than one admittance to hospital	0.23	0.42	0.00	1.00	0.24	0.43	0.00	1.00
Use of private hospital	0.28	0.45	0.00	1.00	0.27	0.45	0.00	1.00
Observations	7,687				4,238			

- *Note*: 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.
- Samples: Households and their members in full or eligible sample (varying N). To obtain representative statistics, the full sample contains only randomly selected households, i.e., excluding oversampling below and around cut-off. Eligible sample contains randomly selected households below cut-off, i.e., excluding oversampling around cut-off. (60% of eligible sample are part of full sample, 40% are oversampled randomly below cut-off.)
- 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).

A.4 Non-childbirth hospitalization

Table A.3: Logit regression of non-childbirth related hospitalization on individual and household characteristics, baseline

Admission to inpatient care	Logit 1 Coef. (1)	Logit 1 Std. Error (2)	Logit 1 P-value (3)	Logit 2 Coef. (4)	Logit 2 Std. Error (5)	Logit 2 P-value (6)	Logit 3 Coef. (7)	Logit 3 Std. Error (8)	Logit 3 P-value (9)
Poverty score	-0.001	0.006	0.862						
pcear				-0.008	0.006	0.159			
Wealth index							-0.113	0.028	0.000
Female	0.071	0.087	0.416	0.069	0.087	0.429	0.062	0.086	0.469
Age	0.029	0.002	0.000	0.029	0.002	0.000	0.030	0.002	0.000
Household size	-0.032	0.022	0.149	-0.032	0.021	0.135	-0.006	0.022	0.785
Hygiene index	0.017	0.043	0.697	0.004	0.040	0.912	-0.042	0.045	0.344
Dist. to next hospital (min.)	-0.002	0.003	0.425	-0.002	0.003	0.440	-0.002	0.003	0.504
Const.	-3.523	0.308	0.000	-3.487	0.225	0.000	-3.820	0.261	0.000

- *Note*: This table shows the coefficients of a logit regression of a dummy indicating non-childbirth related 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 regression, Columns (2), (5), (8) the standard error and Columns (3), (6), (9) the p-value of testing that the coefficient is equal to zero, with one of three different proxies for poverty respectively. Standard error are adjusted for 24 clusters in union councils.

B Additional information on econometric approach

B.1 Estimation of propensity scores

We here describe the estimation of propensity scores for our PSM sample, based on the methodology suggested in Imbens and Rubin (2015).

First, we select a set of base variables, which we believe important for the selection model, guided by economic theory and program knowledge. As described in the main text, we believe the union council to be of fundamental importance as it captures information on otherwise unobservable characteristics. Furthermore, we linearly include 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. We denote this set of variables the base variables K_B. For estimation of propensity scores on household member level, the set furthermore 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.

Second, we search for further variables to be included linearly into the selection model in an iterative process. To start this process, we estimate logit models of assignment to treatment on the base variables in addition to one more variable of a choice set. Subsequently, we calculate likelihood ratio test statistics and test the null hypothesis that the additional variable has a zero coefficient. Of all variables in the choice set, we include the variable with the largest likelihood ratio statistic in the base model. Iteratively, we repeat this step, each time testing another variable and including the one with the largest likelihood ratio statistic. We stop when all these statistics are smaller than one. Our choice set includes all the baseline variables which are not already in K_B (naturally, household level variables are used for household matching and member level variables for member matching). The iterative process leads us to include twelve variables on household and four variables on member level. We denote the union of K_B and selected additional variables as K_L, which will enter the selection model linearly.

Third, we select higher order terms to be included into the selection model. In line with Imbens and Rubin 2015, we restrict ourselves to quadratic and interaction terms of the variables in the set K_L and refrain from including higher order terms. We create interaction terms for all but the union-council-dummy variables and run the same iterative process as before. We refrain from including interactions with the union-council-dummies, because the thus achieved over-fitting violates the common support assumption. 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 18 terms on member level.

B.2 Balancing of baseline variables in unmatched and matched sample

On average, insured households are slightly smaller and composed of slightly better educated members. They are also a little less poor, live in relatively larger villages, with a smaller distance to the next empanelled hospital as measured by GPS data. None of these differences are large in magnitude or highly significant, and only few are marginally significant. Notably, prior insurance knowledge is virtually non-existing in any of the two sub-population. Even more so, indicators generated at household member level, for which we can draw on a much larger sample size, are very well balanced in all variables, e.g., usage of inpatient care and cost thereof. We also compared the prevalence of different causes for inpatient care among household members (results not shown in the tables). Abdominal pain and diarrhea was cited more often among insured households (6.5% versus 2% for each disease, p-value 0.09), but there are no relevant differences in the incidence rates of the most common problems such as appendicitis (8-10%), heart attack (6-10%), diabetes (5-9%) or malaria (6%).

Table B.4: Balancing before and after matching - Household level

Variable	Unmatched Matched	Mean Insured (1)	Uninsured (2)	%bias (3)	% reduct. |bias| (4)	t-test t (5)	p> $\|t\|$ (6)
UC1	U	0.0344	0.0361	-0.900	.	-0.120	0.904
	M	0.0347	0.0444	-5.200	-469.5	-0.750	0.453
		.	.	.	.	.	.
UC2	U	0.0258	0.0614	-17.50	.	-2.420	0.0160
	M	0.0260	0.0259	0	99.80	0.0100	0.993
		.	.	.	.	.	.
UC3	U	0.0516	0.0325	9.500	.	1.220	0.222
	M	0.0521	0.0447	3.700	61.30	0.520	0.601
		.	.	.	.	.	.
UC4	U	0.0624	0.0253	18.20	.	2.280	0.0230
	M	0.0607	0.0507	4.900	73	0.660	0.508
		.	.	.	.	.	.
UC5	U	0.0280	0.0397	-6.500	.	-0.870	0.382
	M	0.0282	0.0244	2.100	67.20	0.370	0.715
		.	.	.	.	.	.
UC6	U	0.0473	0.0433	1.900	.	0.250	0.802
	M	0.0477	0.0516	-1.900	2.800	-0.270	0.787
		.	.	.	.	.	.
UC7	U	0.0344	0.0325	1.100	.	0.140	0.889
	M	0.0347	0.0340	0.400	64.20	0.0600	0.954
		.	.	.	.	.	.
UC8	U	0.0559	0.0433	5.800	.	0.750	0.452
	M	0.0564	0.0816	-11.60	-99.80	-1.510	0.132
		.	.	.	.	.	.
UC9	U	0.0366	0.0541	-8.400	.	-1.140	0.254
	M	0.0369	0.0428	-2.900	66	-0.460	0.643
		.	.	.	.	.	.
UC10	U	0.0516	0.0253	13.70	.	1.740	0.0830
	M	0.0521	0.0458	3.300	76.10	0.440	0.658
		.	.	.	.	.	.
UC11	U	0.0538	0.0253	14.60	.	1.850	0.0650
	M	0.0521	0.0489	1.600	88.90	0.220	0.827
		.	.	.	.	.	.
UC12	U	0.0409	0.0289	6.500	.	0.840	0.400
	M	0.0391	0.0268	6.700	-2.400	1.040	0.297
		.	.	.	.	.	.
UC13	U	0.0323	0.0217	6.500	.	0.840	0.400
	M	0.0325	0.0300	1.500	76.40	0.220	0.827

continued

Table B.4: Balancing before and after matching - Household level

Variable	Unmatched Matched	Mean Insured (1)	Uninsured (2)	%bias (3)	% reduct. —bias— (4)	t-test t (5)	p> $\|t\|$ (6)
		.	.	.	.	.	.
UC14	U	0.0473	0.0469	0.200	.	0.0200	0.981
	M	0.0477	0.0415	2.900	-1532	0.460	0.649
		.	.	.	.	.	.
UC15	U	0.0495	0.0614	-5.200	.	-0.690	0.488
	M	0.0499	0.0457	1.800	64.70	0.300	0.765
		.	.	.	.	.	.
UC16	U	0.0237	0.0433	-10.90	.	-1.500	0.135
	M	0.0239	0.0254	-0.900	92.20	-0.150	0.880
		.	.	.	.	.	.
UC17	U	0.0495	0.0397	4.700	.	0.610	0.540
	M	0.0499	0.0682	-8.800	-87.50	-1.180	0.239
		.	.	.	.	.	.
UC18	U	0.0602	0.0505	4.200	.	0.550	0.582
	M	0.0607	0.0507	4.400	-3.500	0.660	0.508
		.	.	.	.	.	.
UC19	U	0.0301	0.0758	-20.50	.	-2.850	0.00400
	M	0.0304	0.0201	4.600	77.50	0.990	0.320
		.	.	.	.	.	.
UC20	U	0.0452	0.0433	0.900	.	0.120	0.907
	M	0.0456	0.0410	2.200	-147.6	0.340	0.734
		.	.	.	.	.	.
UC21	U	0.0387	0.0397	-0.500	.	-0.0700	0.946
	M	0.0369	0.0445	-3.900	-665.1	-0.590	0.557
		.	.	.	.	.	.
UC22	U	0.0215	0.0505	-15.60	.	-2.170	0.0310
	M	0.0217	0.0269	-2.800	82.10	-0.510	0.609
		.	.	.	.	.	.
UC23	U	0.0366	0.0397	-1.600	.	-0.220	0.828
	M	0.0369	0.0345	1.200	24.60	0.190	0.846
		.	.	.	.	.	.
HH size	U	8.021	8.285	-10.10	.	-1.350	0.177
	M	8.022	7.941	3.100	69.40	0.500	0.617
		.	.	.	.	.	.
% mem. w/o education in HH	U	0.384	0.407	-10	.	-1.330	0.184
	M	0.385	0.375	4.200	57.70	0.660	0.511
		.	.	.	.	.	.
HH with >=1 mem. w.sec. education	U	0.357	0.325	6.800	.	0.890	0.375
	M	0.360	0.331	6.200	8.100	0.940	0.347
		.	.	.	.	.	.
HH w/o mem. having completed primary school	U	0.226	0.267	-9.600	.	-1.270	0.203
	M	0.228	0.237	-2.100	77.80	-0.330	0.742
		.	.	.	.	.	.
% children under 6 years within HH	U	0.110	0.114	-3.100	.	-0.410	0.685
	M	0.111	0.114	-2.700	11	-0.410	0.684
		.	.	.	.	.	.
% elderly over 65 years within HH	U	0.0232	0.0250	-2.500	.	-0.320	0.746
	M	0.0232	0.0159	10.20	-304.9	1.680	0.0930
		.	.	.	.	.	.
% female mem. within HH	U	0.482	0.471	6.400	.	0.850	0.398
	M	0.482	0.478	2.100	67.60	0.320	0.750
		.	.	.	.	.	.
% wage earning mem. within HH	U	0.224	0.235	-8.200	.	-1.100	0.271
	M	0.224	0.232	-6	26.60	-0.970	0.331
		.	.	.	.	.	.
poverty score	U	12.52	12.30	6	.	0.790	0.428
	M	12.54	12.49	1.500	75.10	0.220	0.823
		.	.	.	.	.	.
avg. monthly HH income (PKR)	U	20188	19613	2.700	.	0.340	0.732
	M	20265	19900	1.700	36.50	0.260	0.798
		.	.	.	.	.	.
receiving transfers	U	0.688	0.635	11.20	.	1.480	0.140
	M	0.688	0.672	3.400	69.70	0.520	0.603
		.	.	.	.	.	.
receiving remittances	U	0.110	0.108	0.400	.	0.0600	0.954
	M	0.111	0.127	-5.200	-1087	-0.760	0.445
		.	.	.	.	.	.
loans (PKR, win95)	U	65561	62069	3	.	0.390	0.696

continued

Table B.4: Balancing before and after matching - Household level

Variable	Unmatched Matched	Mean Insured (1)	Uninsured (2)	%bias (3)	% reduct. —bias— (4)	t-test t (5)	p> $\|t\|$ (6)
	M	65605	51631	11.90	-300.1	1.920	0.0550
		.	.	.	.	.	.
savings (PKR, win95)	U	503.1	515.5	-0.500	.	-0.0700	0.947
	M	507.5	470.0	1.600	-202.8	0.240	0.809
		.	.	.	.	.	.
wealth index	U	-0.595	-0.545	-3.100	.	-0.420	0.676
	M	-0.597	-0.566	-2	36.60	-0.310	0.755
		.	.	.	.	.	.
hygiene index	U	0.180	0.141	2.700	.	0.350	0.723
	M	0.179	0.205	-1.800	33.20	-0.280	0.780
		.	.	.	.	.	.
number of HH in village	U	1361	1273	8.300	.	1.090	0.277
	M	1363	1312	4.800	41.70	0.740	0.457
		.	.	.	.	.	.
% eligible HH in village	U	0.322	0.302	17.90	.	2.360	0.0180
	M	0.321	0.324	-2.500	86.20	-0.390	0.700
		.	.	.	.	.	.
reported distance to next hosp. (minutes)	U	45.27	42.47	10.70	.	1.390	0.166
	M	45.40	44.67	2.800	74	0.410	0.681
		.	.	.	.	.	.
linear distance to next empanelled hosp. (km)	U	10.26	10.59	-4.100	.	-0.580	0.560
	M	10.26	10.27	-0.100	96.50	-0.0300	0.977
		.	.	.	.	.	.
Strongest disagreement with: would never go to govt. hosp.	U	0.477	0.487	-2	.	-0.260	0.793
	M	0.477	0.413	12.80	-545.4	1.960	0.0500
		.	.	.	.	.	.
willingness to take fin. risk <= 5 (scale 1 to 10)	U	0.516	0.394	24.80	.	3.260	0.00100
	M	0.514	0.519	-0.900	96.30	-0.140	0.892
		.	.	.	.	.	.
having heard of insurance	U	0.0129	0.0144	-1.300	.	-0.180	0.861
	M	0.0130	0.00468	7.200	-442.3	1.350	0.177
		.	.	.	.	.	.
negl. health care due to fin. reasons	U	0.932	0.913	6.800	.	0.370	0.714
	M	0.932	0.918	5.200	24.70	0.300	0.762
		.	.	.	.	.	.
citing health shock as a risk	U	0.961	0.942	8.900	.	1.200	0.231
	M	0.961	0.967	-2.700	69.50	-0.470	0.637
		.	.	.	.	.	.
citing health shock as main risk	U	0.682	0.686	-0.900	.	-0.120	0.905
	M	0.681	0.706	-5.400	-503.4	-0.830	0.404
		.	.	.	.	.	.
ladder of life rating <= 5 (scale 1/10)	U	0.733	0.751	-4	.	-0.530	0.598
	M	0.735	0.739	-0.800	79.20	-0.130	0.900
		.	.	.	.	.	.
fin. satisfaction <= 5 (scale 1/10)	U	0.705	0.675	6.500	.	0.870	0.387
	M	0.705	0.704	0.200	97.30	0.0300	0.978
		.	.	.	.	.	.
dif'ty finding money for health care >=8 (scale 1/10)	U	0.488	0.498	-2	.	-0.260	0.792
	M	0.486	0.453	6.600	-230	1.010	0.315
		.	.	.	.	.	.
use of inp. care (ext. margin, HH-level)	U	0.252	0.300	-10.70	.	-1.430	0.154
	M	0.249	0.203	10.50	2.600	1.700	0.0900
		.	.	.	.	.	.
inp. care (no. of mem. per HH)	U	0.312	0.379	-10.60	.	-1.420	0.155
	M	0.310	0.254	8.900	16.60	1.480	0.140
		.	.	.	.	.	.
case of neglected health care	U	0.157	0.166	-2.500	.	-0.330	0.745
	M	0.158	0.134	6.600	-166.5	1.040	0.299
		.	.	.	.	.	.
case of outpatient care	U	0.817	0.769	11.90	.	1.590	0.113
	M	0.816	0.861	-11.20	5.800	-1.880	0.0610
		.	.	.	.	.	.
Min. of health status within hh (scale of 1/ 5)	U	3.112	2.978	11.80	.	1.550	0.121

continued

Table B.4: Balancing before and after matching - Household level

Variable	Unmatched Matched	Mean Insured (1)	Uninsured (2)	%bias (3)	% reduct. —bias— (4)	t-test t (5)	p> $\|t\|$ (6)
	M	3.111	3.263	-13.40	-14	-2.060	0.0390

► *Note*: This table shows the balancing of UCs and baseline covariate in the unmatched (U) and matched (M) samples. Variables on the left, statistics on top.
► Sample: Household-level PSM sample (N=795).
► Source: Baseline survey (2015).
► Columns (1) and (2) display the mean of the insured and uninsured households in the matched and unmatched sample respectively. Column (3) and (4) show the percentage bias before and after matching, together with the achieved percentage in reduction in the absolute value of the bias. The standardized percentage bias is the percent-difference of the sample means in the insured and the uninsured subsamples as a percentage of the square root of the average of the sample variances in the insured and uninsured groups (formula from Rosenbau and Rubin, 1985). Columns (5) and (6) display the t-static and the p-value testing for equality of means int he two samples. T-tests are based on a regression of the variable on a treatment indicator.
► The suffix *win99* indicates that we winsorized the variable at the 99th percentile level.

Table B.5: Balancing before and after matching - Member level

Variable	Unmatched Matched	Mean Insured (1)	Uninsured (2)	%bias (3)	% reduct. —bias— (4)	t-test t (5)	p> $\|t\|$ (6)
age (win99)	U	22.02	22.24	-1.300	.	-0.470	0.637
	M	21.97	21.88	0.500	61.70	0.210	0.830
		.	.	.	.	.	.
school-aged (6 to 16)	U	0.391	0.375	3.200	.	1.160	0.246
	M	0.392	0.395	-0.600	80	-0.270	0.788
		.	.	.	.	.	.
female	U	0.483	0.482	0.200	.	0.0600	0.952
	M	0.483	0.479	0.600	-289.9	0.270	0.785
		.	.	.	.	.	.
prim. school not comp'd	U	0.633	0.644	-2.400	.	-0.890	0.373
	M	0.632	0.636	-0.800	66.30	-0.350	0.727
		.	.	.	.	.	.
comp'd sec. educ. or higher	U	0.0836	0.0777	2.200	.	0.790	0.432
	M	0.0819	0.0855	-1.300	39.20	-0.550	0.582
		.	.	.	.	.	.
worked for salary in previous month	U	0.179	0.188	-2.300	.	-0.860	0.391
	M	0.177	0.185	-1.900	18.50	-0.820	0.414
		.	.	.	.	.	.
working in agriculture	U	0.0210	0.0270	-4	.	-1.470	0.141
	M	0.0206	0.0215	-0.600	85.80	-0.250	0.799
		.	.	.	.	.	.
in schooling	U	0.296	0.282	3.100	.	1.130	0.259
	M	0.296	0.293	0.700	77.40	0.300	0.767
		.	.	.	.	.	.
main occupation: HH/ child care	U	0.228	0.222	1.400	.	0.490	0.621
	M	0.228	0.233	-1.200	11	-0.510	0.611
		.	.	.	.	.	.
resp'ble for fin. decisions	U	0.384	0.392	-1.800	.	-0.660	0.512
	M	0.382	0.383	-0.200	88.50	-0.0900	0.930
		.	.	.	.	.	.
resp'ble for health decisions	U	0.271	0.274	-0.500	.	-0.200	0.841
	M	0.270	0.276	-1.300	-141.4	-0.560	0.573
		.	.	.	.	.	.
total cost of last treatment (PKR, win99)	U	24990	29651	-9.900	.	-0.750	0.456
	M	25252	22309	6.200	36.80	0.560	0.576
		.	.	.	.	.	.
usage of inpatient care	U	0.0343	0.0474	-6.600	.	-2.470	0.0130
	M	0.0335	0.0360	-1.200	81.20	-0.570	0.567
		.	.	.	.	.	.

continued

Table B.5: Balancing before and after matching - Member level

Variable	Unmatched Matched	Mean Insured (1)	Mean Uninsured (2)	%bias (3)	% reduct. —bias— (4)	t-test t (5)	t-test p> $\|t\|$ (6)
use of pub. hos. for inp. care	U	0.0231	0.0303	-4.500	.	-1.660	0.0970
	M	0.0225	0.0249	-1.500	66.50	-0.670	0.501
		.	.	.	.	.	.
health step of HH mem. (scale: 1/5)	U	4.385	4.315	7.300	.	2.680	0.00700
	M	4.383	4.386	-0.300	95.80	-0.130	0.895

► *Note*: This table shows the balancing of UCs and baseline covariates in the unmatched (U) and matched (M) samples. Variables on the left, statistics on top.

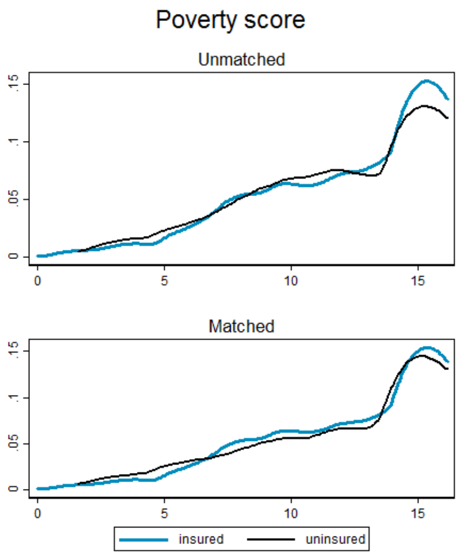
► Sample: Member-level PSM sample (panel, N=6,000).

► Source: Baseline survey (2015).

► Columns (1) and (2) display the mean of the insured and uninsured households in the matched and unmatched sample respectively. Column (3) and (4) show the percentage bias before and after matching, together with the achieved percentage in reduction in the absolute value of the bias. The standardized percentage bias is the percent-difference of the sample means in the insured and the uninsured subsamples as a percentage of the square root of the average of the sample variances in the insured and uninsured groups (formula from Rosenbaum and Rubin, 1985). Columns (5) and (6) display the t-static and the p-value testing for equality of means in the two samples. T-tests are based on a regression of the variable on a treatment indicator.

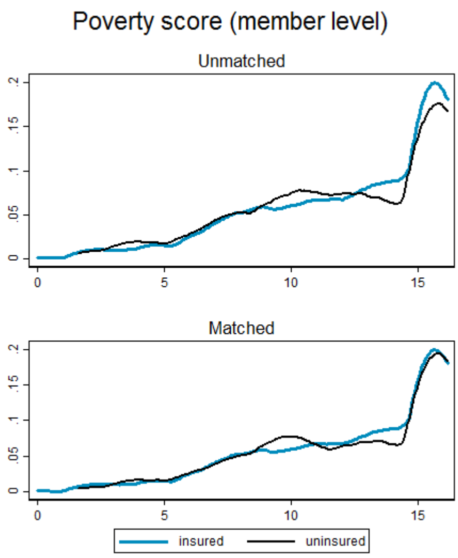
B.3 Distribution of poverty score among insured and matched uninsured sample

Figure B.2: Distribution of poverty score in households of PSM sample by insurance status



► *Note*: This figure shows the distribution of the poverty score among insured and uninsured households in the unmatched (top) and matched (bottom) samples.
► Sample: Household-level PSM sample (panel, N=795).
► Source: Baseline survey (2015) and endline survey (2017).

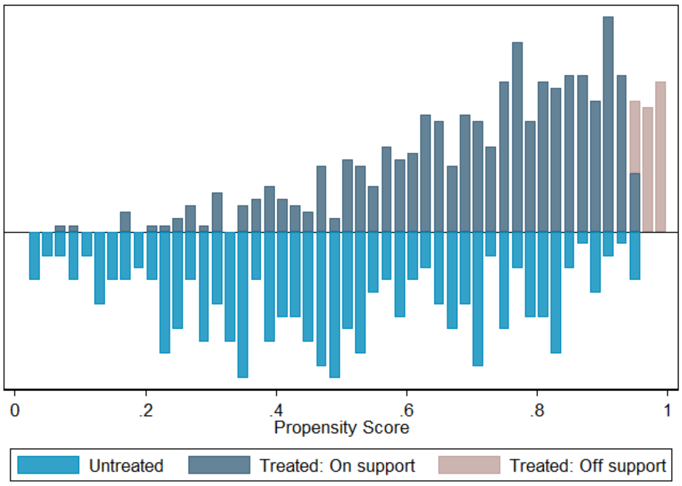
Figure B.3: Distribution of poverty score in household members of PSM sample by insurance status



► *Note*: This figure shows the distribution of the poverty score among members of insured and uninsured households in the unmatched (top) and matched (bottom) samples.
► Sample: Member-level PSM sample (panel, N=6,000).
► Source: Baseline survey (2015) and endline survey (2017).

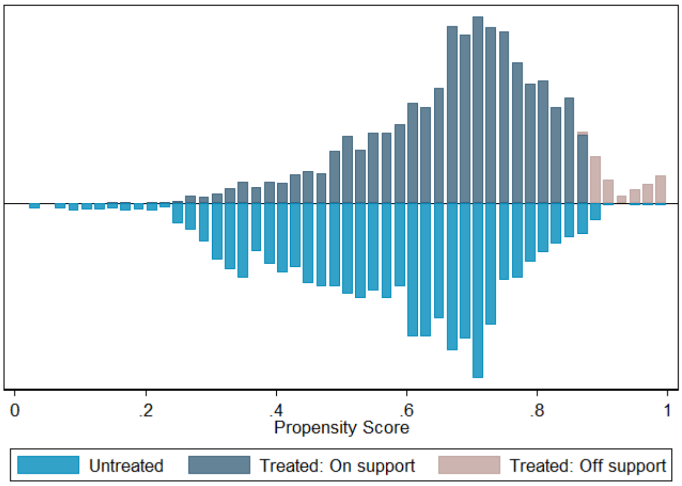
B.4 Common support

Figure B.4: Distribution of common support in households of PSM sample by insurance status



► *Note*: This figure shows the distribution of the common support among insured (top) and uninsured (bottom) households. The blue bars indicate common support, whereas rose bars are off the common support.
► Sample: Household-level PSM sample (panel, N=795).
► Source: Propensity scores calculated from baseline survey (2015) and insurance status in endline survey (2017).

Figure B.5: Distribution of common support in household members of PSM sample by insurance status

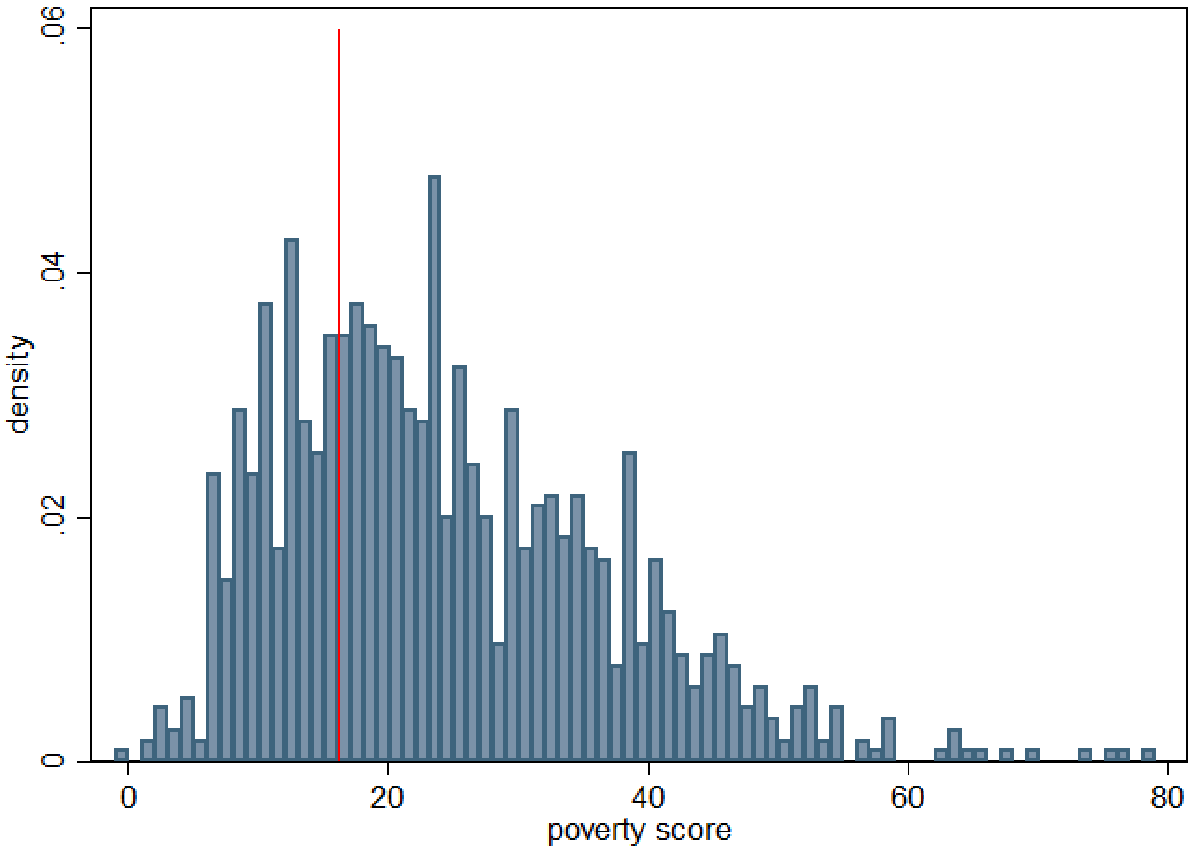


► *Note*: This figure shows the distribution of the common support among individuals in insured (top) and uninsured (bottom) households. The blue bars indicate common support, whereas rose bars are off the common support.
► Sample: Member-level PSM sample (panel, N=6,000).
► Source: Propensity scores calculated from baseline survey (2015) and insurance status in endline survey (2017).

B.5 McCrary density test

To assess manipulative sorting more formally, graph B.6 depicts the distribution of the poverty score in our sample, restricted to those households which were sampled at random and not for oversampling around the cut-off score. One indication of precise sorting would be a bunching of households just below the cut-off poverty score, implying a discontinuity in the density of the poverty score at this point (McCrary 2008). Figure B.6 suggests that comparable proportions of households in the respective districts fall on both sides. This visual evidence is confirmed by the result of density tests as proposed in McCrary (2008) which fail to detect a significant discontinuity, see Figure B.7.[37]

Figure B.6: Distribution of poverty score in random sample

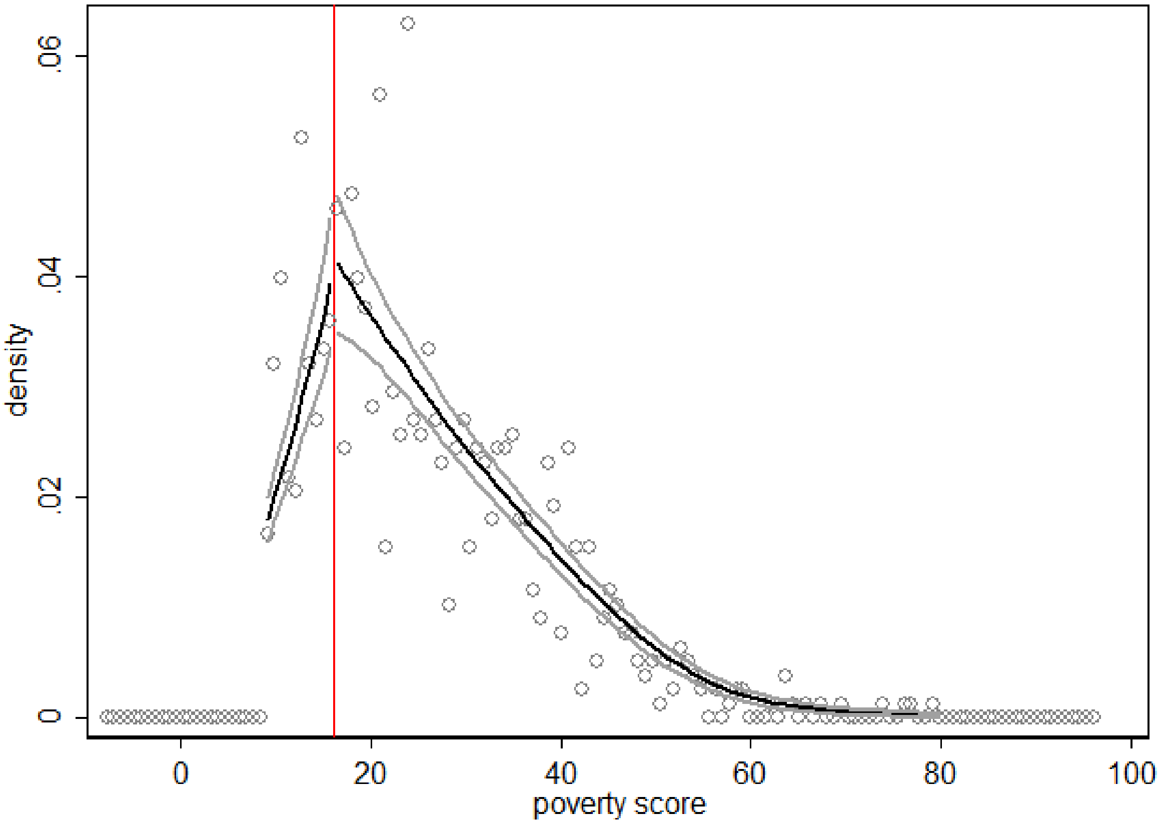


► *Notes:* This figure shows the distribution of the poverty score in the random panel sample. The red vertical line indicates the cut-off score of 16.17.
► Sample: Household-level random sample (N=1,152).
► Source: Endline survey (2017).

Additionally, note that it is also highly unlikely that households could influence their poverty score. This score was assigned based on a PMT in a targeting survey in 2010, which used 23 household-level indicators (OLeary et al. 2011). Households may have seized the opportunity to manipulate their poverty score by exaggerating their poverty status in order to obtain social benefits in the future. However, the fear of perfect manipulation is mitigated by the fact that each variable is weighted differently in the PMT (Uddin, Rafi, Uddin, and Khurshid 2013). Moreover, to the best of our knowledge, there is no official household questionnaire available online, and no list with all variables including their weighting can be found. Pakistani

[37] This result is also supported by the report on the initial targeting survey of the BISP program, which also finds no sharp break in the density of the poverty score and no significant jump at the threshold for baseline covariates and outcome variables (OLeary, Cheema, Hunt, Carraro, and Pellerano 2011).

Figure B.7: McCrary density test



► *Notes:* This figure illustrates the result of the McCrary density test around the cut-off 16.17. See McCrary (2008) for details.
► Sample: Household-level random sample (N=1,152).
► Source: Endline survey (2017).

households and the enumerating survey staff could not have known the exact method of how the poverty score was computed and therefore failed to select into treatment.

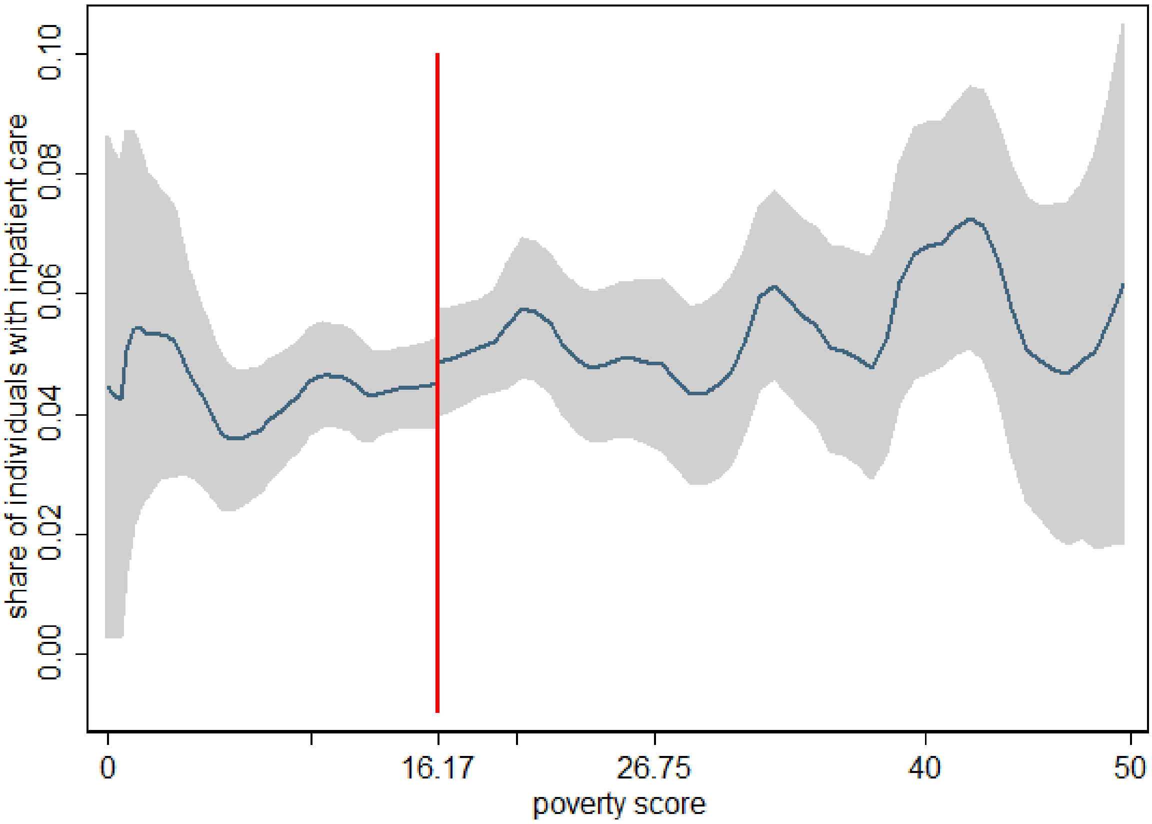
B.6 Continuity of baseline variables across the cut-off score

Table B.6: Pseudo-effects on inpatient care consumption from baseline

Outcome	RDD β_{LATE} (1)	S.E. (2)
Individual outcomes		
Usage of inpatient care	**0.001**	0.008
N (left/ right of cut-off)	*3,217/ 2,557*	
conditional on usage of inpatient care		
Usage of private versus public hospitals	**0.021**	0.079
N (left/ right of cut-off)	*161/ 132*	
Household outcomes		
Neglected health care	**-0.016**	0.030
N (left/ right of cut-off)	*434/ 364*	

- *Note*: This table shows results from RDD estimates on baseline variables, main outcome variables on the left, the different statistics on top.
- Sample: RDD sample (varying N).
- Source: Baseline survey (2015).
- Columns (1) and (2) show the coefficient and standard errors for the pseudo-local average treatment effect for households just below the cut-off score, estimated using a sharp regression discontinuity design. Estimated using local linear regression models on both sides of the cutoff and reported analytic S.E. are based on the regressions. Optimal bandwidth from Imbens and Kalyanaram (2009) minimizing the mean squared error.
- The statistical significance is given as follows: * indicates p < 0.1, ** p < 0.05, and *** p < 0.01, with the null hypothesis being a zero effect size.

Figure B.8: Inpatient care by poverty score at baseline



► *Notes:* 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)