Selection in Health Insurance: An evaluation of India's RSBY*

Cynthia Kinnan[†], Anup Malani[‡], Alessandra Voena[§],
Gabriella Conti[¶], and Kosuke Imai[‖]

August, 2022

Abstract

The question of how price affects insurance takeup is of both theoretical and policy relevance. We use a multi-armed RCT that varies insurance premiums to study selection of above poverty line households into an at-scale health insurance program in India. Demand is downward-sloping but even at the actuarially fair price, over 58% of households take up insurance. Testing for adverse selection using baseline measures of health or health spending yields no evidence of selection. When we instead use data-driven predictions of health and health spending, we find some evidence of adverse selection by predicted spending, but its magnitude is small. Cost sharing for middle-income households may increase the fiscal sustainability of public insurance.

JEL Classification: O12, I13, H51, D91

*We thank Kate Baicker, Amitabh Chandra, Pascaline Dupas, Johannes Haushofer, Rick Hornbeck, Seema Jayachandran, Edward Miguel, Joe Newhouse, and seminar participants at Brown University, University of Chicago, Harvard, Stanford and Tufts for comments. Tianyu Zheng provided excellent research assistance. All errors are ours. This study was funded by the Department for International Development in the UK Government; the Tata Trusts through the Tata Centre for Development at the University of Chicago; the MacLean Center, the Becker-Friedman Institute, the Neubauer Collegium, and the Law School at the University of Chicago; SRM University; and the International Growth Centre. Trial registered at ClinicalTrials.gov (NCT03144076) and the American Economic Association Registry (AEARCTR-0001793).

[†]Tufts University, cynthia.kinnan@tufts.edu.
[‡]University of Chicago, amalani@uchicago.edu.
[§]Stanford University, avoena@stanford.edu.
[¶]University College London, gabriella.conti@ucl.ac.uk.
[‖]Harvard University, imai@harvard.edu.

1 Introduction

Health shocks are a significant sources of risk in low and middle income countries (LMICs). Over 150 million households are pushed into poverty due to health expenditures each year (Shahrawat and Rao, 2011).[1] In response, many LMICs have stated a goal of achieving universal health insurance coverage (Lagomarsino et al., 2012). The COVID-19 pandemic has further laid bare the need for a health safety net.

Weak state capacity, however, limits the feasibility of universal health insurance coverage. Administrative constraints make it difficult to enforce private purchase of health insurance (Banerjee et al., 2021), while low fiscal capacity is a barrier for the government to provide free insurance for everyone. LMICs may therefore attempt to ease their fiscal burden via cost sharing: charging households premiums to enroll or imposing deductibles or copays for insurance. Cost-sharing for insurance, however, raises the specter of adverse selection, namely that, when participation is not free, higher-cost individuals may be more likely to participate (Akerlof, 1970; Einav and Finkelstein, 2011; Cliff et al., 2021; MacLeod, 2021; Gruber, 2022). This, in turn, can raise unit costs, further threatening the goal of universal coverage. On the other hand, greater enrollment by those who are higher-cost can be seen to reflect efficient targeting, with cost-sharing screening in out those with lower expected benefits from the program (Alatas et al., 2016; Ashraf et al., 2010).

The extent (if any) of adverse selection is also relevant for applied welfare analysis (Hendren and Sprung-Keyser, 2020; Finkelstein and Hendren, 2020). As noted by Finkelstein and Hendren (2020), "empirical economists interested in translating the benefits of the 'credibility revolution' into progress on applied welfare analysis should focus their efforts on estimating behavioral responses that have fiscal externalities on the government budget, not on behavioral responses whose costs are (approximately) fully internalized by the responding individuals." Adverse selection is such a behavioral response, and estimating its extent is important for calculating the marginal value of public funds (MVPF) of policies that may induce a selection response.

[1] In India alone, as many as 63 million households may fall into poverty each year due to health spending (Berman et al., 2010; Shahrawat and Rao, 2011).

In contrast to a vast literature in rich countries[2], to our knowledge adverse selection in health insurance markets has been studied in low and middle income countries (LMICs) by only a few papers, including Banerjee et al. (2014b), Fischer et al. (2018), Asuming et al. (2019) and Banerjee et al. (2021). These studies have yielded varying results, with adverse selection seen in some contexts but not others. Consistent with adverse selection, Banerjee et al. (2021) find evidence that subsidies (i.e., lower cost sharing) select in lower-cost households in Indonesia. Asuming et al. (2019) find a similar pattern in Ghana. Fischer et al. (2018) find that, in Pakistan, adverse selection is seen when the enrollment decision is at the individual level but not when enrollment is at the household or credit group level. Banerjee et al. (2014b) find no adverse selection in the market for a product bundling credit and health insurance in India, driven in that case by very low demand for the insurance. These findings suggest that the target population and the structure of the insurance product may matter crucially for the nature and extent of selection into insurance. There is significant value in identifying settings/products with little adverse selection in terms of informing policy design elsewhere.

Our paper contributes to this literature by leveraging a large-sample experiment in Central and South India that gave participants access to an at-scale government-run hospital insurance plan (Rashtriya Swasthya Bima Yojana, or RSBY).[3] We randomized the price at which households had access to RSBY.[4] Using a two-stage randomization, we also varied the premiums that *neighbors* faced, holding constant the premium a given household faced, allowing investigation of spillover effects. We collected detailed data at three points in time: at baseline as well as two waves of followup data. To our knowledge, our ∼11,000 household (50,000 person) sample is the largest thus far collected in the context of evaluating health insurance in a LMIC. The large sample, design, and rich data allow us to study selection and its mechanisms with a relatively high degree of power to precisely estimate effects, including those which may be small or null. Moreover, the product we study is a low-cost "no frills" insurance product covering only inpatient care.[5] This type of product is likely to play an

[2]See Einav and Finkelstein (2011) and Handel and Ho (2021) for overviews.
[3]The name translates to "National Health Insurance Scheme."
[4]The control group in essence faced an infinite price of enrolling through our experiment.
[5]And some outpatient "day surgeries"; see section 2.

important role as countries attempt to provide coverage against serious health shocks at low cost, and may achieve higher levels of takeup and different patterns of selection than other, higher-cost programs.

To examine whether cost-sharing changes the composition of households who enroll in insurance, we use two approaches. First, we use backward-looking (i.e., baseline) measures of health and health spending as proxies of forward-looking predicted health needs. Second, we use *predicted* measures of health and health spending generated from a machine-learning exercise using LASSO on the control group to form predictions of midline health and health spending. Using baseline measures of health and spending leads to essentially no evidence of adverse selection. Using predicted measures does reveal some statistically detectable adverse selection on predicted medical spending; however, the magnitude is small.

We explore the robustness of these results in several ways. We adjust for multiple testing and verify that the (small) degree of adverse selection seen with predicted spending remains significant. We also show that the findings are robust to controls for measures of wealth, cognitive performance, education and risk aversion interacted with price. Next, to understand why some selection is found with predicted (ML-based) medical spending but not with baseline medical spending, we estimate, for each set of measures, a correlation between the prediction and realized post-intervention outcomes. The ML-based predicted values are more strongly correlated with post-intervention outcomes than are the baseline values, and the difference is highly significant. These results demonstrate that machine learning methods can be used to construct stronger predictors than can be obtained by using the baseline value itself in settings where outcomes (such as health and health spending) can display complex correlations over time. While the potential of LASSO methods in predicting health care costs has been noted (e.g., Kan et al. 2019), to our knowledge the ability to generate sharper tests for adverse selection has not been demonstrated.

Our findings have a number of important implications. We find that the extent of adverse selection, while statistically detectable, is small. Methodologically, we demonstrate that ML-based methods can be better powered to detect or rule out modest degrees of adverse selection, as opposed to using baseline levels of health or health spending. Given the interest in exploring adverse selection a wide array of markets for both insurance and credit (Einav

and Finkelstein, 2011; Ahlin et al., 2020), this approach may be useful in many other contexts as well.

Moreover, our companion paper (Malani et al., 2022) has shown that charging actuarially fair premiums does not cause takeup to fall to very low levels: the opportunity to purchase insurance led to 59% uptake while access to free insurance led to 78% uptake.[6] Further, Malani et al. (2022) shows that charging increases utilization of the program. In combination with the findings of this paper on (absent to modest) selection, these results suggest that cost-sharing, for households who are above the poverty line, is a viable strategy to increase the financial sustainability of low-cost, no-frills public health insurance programs. These results also suggest that, methodologically, a finding that higher cost-sharing increases unit costs does not necessarily imply adverse selection. Rather that driven by changes in selection, the increase in utilization may be driven by other mechanisms such as a sunk-cost effect or price as a signal of value.

The rest of the paper proceeds as follows. Section 2 describes our setting and context, Section 3 details the setup of our experiment, and Section 4 explains our approaches to measuring health risk. Section 5 presents our results and robustness checks, Section 6 discusses potential explanations and mechanisms and Section 7 concludes.

2 Setting

2.1 Health care in India

Indians have access to subsidized government health care facilities and non-subsidized private medical providers. The government operates a large number of facilities, from Primary Health Centres and Sub-Centres to District Hospitals. These government facilities largely offer free care, though they may not cover all populations, they often lack supplies, and their quality has been questioned (Comptroller and Auditor General of India, 2019). Public facilities may also be subject to overcrowding (Andrews and Vera Hernandez, 2022) and absenteeism (Banerjee et al., 2008). Thus, households often turn to the unsubsidized private

[6] We reproduce these results in Table 1.

health sector, which includes private hospitals, clinics and doctor's offices. Private facilities will often request at least a down payment before providing service and, in some cases, do not release the patient until the bill is paid in full. Nevertheless, the private sector provides 80% of all outpatient treatment and 60% of all inpatient treatment in the country (Ministry of Health and Family Welfare, 2014). Although total medical expenditures are roughly 4% of GDP, government expenditures amount to only 1.15% of GDP. Overall, India faces a shortfall in supply in providers. For instance, 47% of children are estimated to live in villages without any health facility at all (Ma and Sood, 2008).

Aside from RSBY (at the time of the study) and PMJAY (which replaced it in 2018, as we discuss below), there are a limited set of insurance options in India. Some state governments have provided insurance programs that cover tertiary care (e.g., Arogyashri Vajpayee in Karnataka). In addition there are private insurance options, often provided by employers, including the government for its employees. However, private insurance covers less than 4% of all medical expenditures and this too is concentrated in urban centers. Although RSBY and PMJAY are ambitious public insurance scheme, insurance coverage presently accounts for just 8% of government health expenditures (Gupta and Chowdhury, 2014). As a result, 69% of all expenditures remain out-of-pocket.

In the areas of Karnataka where this study took place, at the time of the study there were few insurance options aside from RSBY, Arogyashri Vajpayee, and a plan called Yeshasvini, which is only available to members of certain occupational cooperative societies (i.e., trade associations).

2.2 Rashtriya Swasthya Bima Yojana insurance scheme

RSBY was introduced in 2008 to provide hospitalization insurance to India's poor. Like Medicaid in the U.S., it is largely free to enrollees and is designed and principally funded by the national government, but administered by the state governments.[7] As discussed below, in 2018 RSBY was replaced by PMJAY.

RSBY Eligibility. All households carrying BPL ration cards or those with members in

[7]For more discussion of RSBY's implementation in Karnataka, see Rajasekhar et al. (2011); Berg et al. (2019).

certain occupations[8] were eligible for RSBY. In addition states could expand eligibility to other groups so long as they (as opposed to the central government) paid the full cost to cover these groups. The scheme covered up to five members of each enrolled household: the head of household, the spouse and up to three dependents[9]. The threshold to define a household as BPL was set at approximately INR 900/month in rural areas, and INR 1,100/month in urban areas in Karnataka.[10]

Coverage. RSBY covered up to INR 30,000 per year per household for over 700 procedures at empaneled hospitals. The covered procedures largely include those that require an overnight stay at a hospital, though there are a number of so-called day surgeries that are also covered[11]. Child delivery is also included. There are no deductibles or co-pays. RSBY covered all pre-existing diseases and there was no age limit for beneficiaries. The rates of most surgical procedures were fixed[12]. Transportation charges were also covered at a rate of INR 100 per hospitalization up to a maximum of INR 1,000 per year. The coverage lasts one year starting the month after the first enrollment in a particular district, but is often extended without cost to beneficiaries.

Administration. RSBY was a completely paperless program which used biometric-enabled smart cards as a vehicle of delivery. Empaneled hospitals included both private hospitals and government hospitals that meet certain criteria and sign MOUs with the state agency running the scheme; not all public hospitals are included. Insurance was provided by

[8]These include: (1) building and other construction workers registered with the welfare boards; (2) licensed railway porters; (3) street vendors; (4) MNREGA workers who have worked for more than 15 days during the preceding financial year; (5) beedi workers; (6) domestic workers; (7) sanitation workers; (8) mine workers; (9) rickshaw pullers; (10) rag pickers; and (11) auto/taxi driver. See `http://www.rsby.gov.in/about_rsby.aspx`.

[9]An exception is in the case of childbirth: the newborn is always covered even if five members of the household are already covered. This coverage continues until the renewal date, at which point the newborn is only covered if the household chooses to include them among the five that are covered. See `http://www.rsby.gov.in/faq_medical.aspx`.

[10]Using the INR 17.52/dollar purchasing-power-parity exchange rate (OECD, 2020) these figures are roughly USD 51 and USD 63, respectively.

[11]These include: haemo-dialysis; parenteral chemotherapy; radiotherapy; eye surgery; lithotripsy (kidney stone removal); tonsillectomy; D&C; dental surgery following an accident; surgery of hydrocele; surgery of prostrate; few gastrointestinal surgery; genital surgery; surgery of nose; surgery of throat; surgery of ear; surgery of urinary system; treatment of fractures/dislocation (excluding hair line fracture), contracture releases and minor reconstructive procedures of limbs which otherwise require hospitalization; laparoscopic therapeutic surgeries that can be done in day care; identified surgeries under general anesthesia; and any disease/procedure mutually agreed upon. See `http://www.rsby.gov.in/faq_medical.aspx`.

[12]They can be found at `http://www.rsby.gov.in/Documents.aspx?ID=4`.

private companies, but the premium was paid for by the government. Government funding was shared by the central and the state government in a 3:1 ratio. The insurance premium was determined at the state-level based on an open-tender process. (The premium cost was approximately INR 200 in the state of Karnataka during our study period). The only cost to the beneficiary was that of a registration charge of INR 30 to obtain the smart card. (For households in our study's "free insurance" arm, we covered this cost.)

Replacement by PM-JAY. In September of 2018, RSBY was subsumed by a scheme called Pradhan Mantri Jan Arogya Yojna (PMJAY). While several states opted out, PMJAY took effect in Karnataka and was available to all below-poverty-line households and some above the poverty line. This affected households in treatment and control symmetrically.

We next detail the experimental design used to understand how cost-sharing affects selection into RSBY.

3 Experimental Design

We carried out a randomized controlled trial (RCT) to test the impacts of different methods of expanding eligibility for RSBY hospital insurance to APL households.[13]

3.1 Treatment arms

We evaluate three methods of accessing RSBY:

A. Free insurance: Households obtain access to RSBY for no charge. The total cost (the premium[14] and the INR 30 typically charged to obtain a biometric, smart card that functions as the insurance card) is paid by the study.

B. Cost-sharing and unconditional cash transfer: Households receive the right to purchase RSBY – for the premium the government pays for RSBY in their district plus the

[13]This study was registered before outcomes were measured in the first post-treatment survey at 18 months. A pre-analysis plan was posted prior to the last follow-up survey at 4 years under American Economic Association Registry Identifier AEARCTR-0001793.

[14]In the two districts of Karnataka in our study, the premium in 2015 was INR 133 ($7.59) per household in Gulbarga and INR 173 ($9.87) per household in Mysore. We employ the 17.52 INR/USD purchasing-power-parity exchange rate (OECD, 2020).

cost of an insurance (smart) card – within a 3 week window. Households receive an unconditional cash transfer equal to the RSBY premium plus the cost of the smart card.

C. Cost-sharing: Households receive the right to buy RSBY (for the same price as in condition B), but no cash transfer, unconditional or otherwise.

We compare outcomes under these conditions versus a control condition:

D. No intervention.

In this paper, in order to understand the effect of changes in cost-sharing on the level and composition of enrollees, we primarily focus on the effect of premium on uptake and utilization of insurance, i.e., we focus on comparisons of arm C (cost-sharing) to arm A (free insurance).

3.2 Sampling Strategy

For a household to be eligible for the study, its members had to meet the following inclusion criteria at the start of the study in 2013:

1. The household resided in a village in Gulbarga or Mysore districts;
2. The household's village was within 25 km of at least one (private or public) hospital empaneled in RSBY; and
3. A member of the household held an Above Poverty Line (APL) ration card.

The trial was conducted in Gulbarga and Mysore districts because they are representative of central and southern India, respectively. We focused on APL households because they were treatment naive: that is, they were not otherwise eligible for RSBY. We use the distance-to-hospital restriction to ensure that insurance would have some value; hospital insurance is not useful in the absence of hospitals at which to use it.

The exclusion criteria for the study were:

1. the household resided in a village with ≤ 10 eligible households (villages with small populations were excluded because surveying in those villages was not cost-effective.);
2. The possession of a BPL card by a member of the household;
3. Having a member working in one of the occupations that made the household eligible for RSBY regardless of BPL status; or
4. Having a member with insurance that covered secondary hospital care (most commonly, Yeshasvini).

3.3 Sample size

The target enrollment for each of the 4 household-level arms was 4,500 households for arm A and 2,250 households for each of conditions B, C and D. The 2,250 target for conditions B to D ensured the RCT was powered to detect a 25% change in hospitalization rate across study arms, allowing for a 10% attrition rate. The sample size for condition A was doubled because we predicted (correctly) that free insurance would be the likely approach used by the government to expand eligibility for public insurance. Due to some attrition between listing and completion of baseline, our final randomized sample was 11,089 households in 424 villages across the two districts. (Attrition is discussed in more detail in Section 3.7.)

3.4 Data Collection

We conducted 6 rounds of data gathering: (1) a listing exercise; (2) a baseline survey; (3) an enrollment survey; (4) a 12 month follow-up survey, which uses a novel design we call a Post-Health Event Survey (PHES); (5) an 18 month follow-up survey (midline survey); and (6) a 4 year follow-up survey (endline survey). Here we report results using data from our baseline, midline and endline surveys. The surveys themselves are described in Appendix C.

3.5 Treatment assignment

We designed a two-stage randomization process to identify both direct treatment and spillover effects of different methods of providing access to health insurance. The two stages of ran-

domization provide us with two instruments, one to study direct effects and one to study indirect or spillover effects. In this paper we are interested in the effect of (own) price on (own) enrollment and so we focus on the direct effects; we control for overall village-level treatment saturation throughout. As documented in Malani et al. (2022) we do not find evidence of spillovers at the enrollment stage (in contrast to the utilization stage).

A schematic of the randomization process appears in Figure A.1. In a first stage, we randomly assigned villages to one of five village-level arms; these are defined by the percentage of households within the village assigned to each of the four household-level study arms defined in section 3.1. The percentage allocations to the 4 household-level arms are given in the last 4 columns of Table A.1. The percent of villages assigned to each of the 5 village-level arms are in the second column. Villages were matched before this first stage randomization.[15] In a second stage, we randomly assigned households within a village to the four arms according to the allocation probabilities assigned to the village. Households were matched before this second stage randomization.[16]

3.6 Balance Tests

We conduct balance tests to validate that assignment to treatment was indeed random. We do this in 3 steps. First, we gather baseline measurements on a range of variables on demographics, financial status and health (number of persons and of children in household, age and education of head of household, distance to nearest town, number of rooms and of concrete rooms in home, annual household budget and good expenditure, major sickness) and on a subset of baseline health outcomes (visited health care facility, annual hospital and non-hospital expenditures). Second, we estimate multinomial logit models predicting

[15]Specifically, we first stratified villages into quintiles of # eligible hhds per village. Within each quintile, we created blocks of 20 villages. Using data from our listing exercise on average values for certain variables (among eligible hhds in a village), create blocks using Mahalanobis matching on: education, age of household head, # children, # rooms in house; and caliper matching on binary variables: major illness, unemployment in household. Within a village block, we randomly assign villages to 5 village-level conditions (without replacement).

[16]Specifically, within each village, we first created blocks of 10 eligible and consented households. Using household-level data from listing survey, create blocks using Mahalanobis matching on: education, age of household head, # children, # rooms; and caliper matching on binary variables: major illness, unemployment in household. Within each block of households, we randomly allocate households to the 4 household-level conditions in accordance with the village-level assignment probability condition (without replacement).

household treatment assignments for each household (A/B/C/D) as a function of outcomes measured at baseline, one outcome at a time. Third, we conduct likelihood ratio tests where the null model is the same multinomial model without the baseline covariate, to determine if we can reject the null that these two models are statistically equivalent, i.e., that the baseline covariate has no explanatory power. We collect the p-values from these LR tests.

If the randomization is successful, then the p-values from these tests should stochastically dominate the uniform distribution. Figure C.6 plots these p-values and the CDF of the uniform distribution. Due to stratification, the allocation is highly balanced, even more so than we would expect due to chance (without stratification): a one-sided Kolmogorov–Smirnov test confirms that our p-values stochastically dominate a uniform.

3.7 Sample Attrition

Randomization was based on a listing of households conducted prior to baseline. Attrition between randomization and baseline can be attributed to household non-response, inability to locate sample households, households no longer meeting eligibility criteria for the study, and missing baseline data. Attrition between baseline and midline and then endline can be attributed to household non-response and the inability to locate sample households, including their movement outside the village. Some households with missing baseline data were able to be surveyed at midline, and endline. Table A.2 provides numbers on response rates and attrition over the course of the study. Attrition is balanced across study arms.

4 Empirical approach: Who takes up insurance?

The key implication of adverse selection is that demand for (i.e., enrollment in) insurance will be less price-elastic among those with higher expected costs. We test this hypothesis by examining how take up of insurance varies between the free and charged treatment conditions (group A vs. group C) as a function of proxies for expected costs. Enrollment is measured by whether a household enrolled in RSBY during the 1-2 days that we brought a mobile

enrollment truck to their village or town during our 2015 enrollment drive.[17]

Before discussing how enrollment varies with household characteristics, it is relevant to understand enrollment *levels*: how high was enrollment across treatment arms? This is relevant since, as pointed out by Banerjee et al. (2014a), there can be no adverse selection if (almost) no one takes up insurance — or, similarly, if (almost) everyone does. Table 1 shows that demand for RSBY is high — but not universal — and downward-sloping in price. Almost four-fifths (78.3%) of households take up insurance when it is free, 71.6% purchase at an actuarially fair premium when they receive an unconditional cash transfer equal to the premium, and 58.8% purchase at an actuarially fair premium (no cash transfer). The implied price elasticity for insurance is -0.33 with linear demand, which is in the mid-range of estimates from the health insurance literature (e.g., Pendzialek et al. (2016), Gruber and Lettau (2004)). The fact that demand for RSBY in our setting is bounded well away from zero and one makes this a good setting to look for evidence of adverse (or advantageous) selection.

4.1 Measuring propensity to use insurance

A natural dimension along which the expected value of insurance, and hence willingness to pay, may differ is the latent propensity of a household to experience health shocks.[18] Since this variable is latent, it must be estimated. We use two approaches to estimate variation in households' expected propensity to experience health shocks: one using baseline measures of health status as a proxy for the likelihood of health shocks and one using machine learning (LASSO) to select baseline variables which best predict subsequent midline health in the control group.[19] Specifically, we use the following measures: general health of the male (female) survey respondent, derived from a ladder question measuring the subjective health

[17] In theory households in arms A, B and C also had the option to enroll by visiting the RSBY office in their district at any time, but we have no evidence that enrollment at the district office, which in general is very uncommon, occurred in our sample. Nor did the government conduct any RSBY enrollment drive after 2015 for any group. Group D households were ineligible to enroll in RSBY via any channel unless they subsequently became eligible, e.g., by obtaining a BPL card.

[18] RSBY covers inpatient care and some day surgeries as described above. However, we consider proxies of both the narrow measure of inpatient care and the broader measure of health care. It may be difficult for households to forecast *ex ante* which conditions would require care of the type covered by RSBY; predicting the need for health care more generally may be more straightforward.

[19] Both approaches are specified in our pre-analysis plan.

on a 1-5 scale; a binary indicator of the male (female) survey respondent having visited a medical facility in the year prior to the survey; and total household medical expenditure in the year prior to the survey. For each of these 5 variables, we consider both its baseline value (baseline method) and a prediction of its midline value, generated from a LASSO-selected set of baseline predictors.

Whether the respondent visited a medical facility in the year prior to the survey captures, on the extensive margin, characteristics associated with having a condition serious enough to warrant seeking health care. Total household medical expenditure captures the intensive margin of spending on health care.[20]

4.2 Baseline measures

Our first approach is to use baseline values of the five health measures described above: general health of the male (female) survey respondent; a binary indicator of the male (female) survey respondent having visited a medical facility in the year prior to the baseline survey; and total household medical expenditure in the year prior to the baseline survey. This parallels the approach taken in much of the recent literature (Asuming et al., 2019; Fischer et al., 2018; Banerjee et al., 2021).

4.3 LASSO procedure

To complement the baseline data approach, we predict households' likelihood to experience covered health event using machine learning methods (LASSO). This approach avoids the need for researcher discretion in choosing what baseline characteristics are most predictive of subsequent usage/valuation of health insurance. It also allows us to, in principle, form a more detailed prediction of midline health if to the extent that health at midline can

[20]Spending is not conditional on seeking care: it is zero for those who did not seek care. For both visits and spending, we focus on any medical care rather than inpatient care. Measuring hospital utilization was challenging because there is no agreed upon definition of a hospital. At baseline and midline we asked the male and the female respondents if they had visited a hospital in the last year, as has been done on other surveys, including the National Statistical Survey (NSS). However, our examination of the resulting hospital utilization rate and separate but concurrent ethnographic work we did in each district suggests that respondents call even small clinics hospitals. Therefore, we interpret hospital use at baseline and midline as health care facility use, not just hospital use. We do not explicitly restrict to visits involving an overnight stay since outpatient day surgeries are an important margin of utilization.

be predicted by baseline characteristics beyond the five measures above. Midline outcomes are predicted using baseline variables selected from the full set of available baseline health-related variables. To generate the prediction, we fit 10-fold cross-validated LASSO models on the control group. The resulting model is then used to generate predictions for the full sample. See Appendix B for the set of baseline variables and the LASSO procedure details. Histograms of the predicted values appear in Appendix B.2.

4.4 Estimating equation

We test for differential selection using the following regression specification:

$$enroll_{iv} = \alpha + \sum_{h=2}^{3} \beta_h d_{iv}^h + \gamma X_i + \sum_{h=2}^{3} \delta_h d_{iv}^h \times X_i + s_v + \epsilon_{iv} \tag{1}$$

where $enroll_{iv}$ is the enrollment decision of household i in village v, d_{iv}^h is an indicator for assignment to treatment arm $h \in A, B, C$, X_i is a measure of the household's expected cost under insurance (discussed below), and s_v is treatment saturation of village (i.e., the share of the village sample assigned to arms A, B, C as opposed to the control arm, D). In some specifications we also control for additional baseline covariates interacted with treatment status.

Variation in likelihood to experience a covered health event is captured by the term X_i, which is constructed using either baseline measures or LASSO-based measures as described above. We scale X to have unit standard deviation and normalize the sign so that higher values imply better health and hence *lower* predicted likelihood to experience a covered health event. The omitted category is group A (free insurance). Therefore the coefficients δ_2, δ_3 identify differences in the differential propensity to enroll as health changes in groups B, C relative to group A.[21]

Our key test for (lack of) selection is $\delta_3 = 0$. If we reject the null in favor of $\delta_3 < 0$, implying that those with better (baseline or predicted) health are less likely to enroll under cost-sharing relative to free insurance, this suggests that adverse selection is present to a

[21] Because group D could not enroll in RSBY during the enrollment drive, group D does not appear in the sample for these regressions.

detectable degree. Conversely, $\delta_3 > 0$ would indicate advantageous selection (Fang et al., 2008). We also estimate and present δ_2 coefficients, which measure differential selection (if any) under the "paid insurance plus cash" arm relative to the free arm. To the extent that selection is seen in the paid arm to a different extent than the paid+cash arm, this helps to rule out concerns that health risk is being confounded with liquidity constraints. We also show below specifications controlling for baseline assets interacted with treatment assignment.

5 Results: Does cost sharing change the enrolled pool?

Table 3 presents estimates of equation 1. Inasmuch as RSBY insurance covers visits to empanelled facilities, the likelihood of visiting a health care facility is a natural predictor to use when forecasting differential incentives to enroll. Columns 1 to 4 use the baseline realization or predicted midline probability of visiting a medical facility for the male and female respondent.[22] Using baseline values (cols 1 and 2) yields very small and insignificant estimates of the degree of selection (i.e., δ_3); moreover the sign of the (negligible) coefficient estimate flips sign between specifications using the male vs female respondent. One way to benchmark the magnitude of selection (if any) is to compare the degree of selection implied by the point estimate relative to mean enrollment: using the (insignificant) point estimate for the effect of male facility visits, a one standard deviation (SD) reduction (i.e., lower risk) is associated with a 1.4% increase in the enrollment rate; the number for female health is even smaller in magnitude and opposite in sign (-0.8%). Using the predicted midline values (cols 3 and 4) yields even smaller (and likewise insignificant) estimates of the degree of selection, with a one SD reduction in predicted probability for males leading to a 0.9% reduction in predicted enrollment; and a 0.0% effect associated with changing predicted probability for females. In summary, we find no evidence of selection using either baseline facility visits or predicted midline facility visits. For comparison, the relative average enrollment rates in arm B vs arm C (paid+cash vs. paid) demonstrate that providing a INR. 220 unconditional cash

[22] We flip the sign so that a higher number corresponds to better health. We do the same for expenditure, below.

transfer causes an increase in enrollment of 22% (71.6pp vs 58.8pp), an order of magnitude larger than any of the (insignificant) selection effects.

Columns 5 to 8 focus on the overall health of the male and female survey respondents. Using baseline health of either (cols 5 and 6), there is no statistically significant selection; δ_3 is not significantly different from zero. In addition to being statistically indistinguishable from zero, the point estimates imply very small degrees of selection relative to mean enrollment: using the (insignificant) point estimate for the effect of male health, a one standard deviation improvement in baseline health is associated with only a 2.1% reduction in the enrollment rate; the number for female health is similar (2.0%).

We next turn to predicted midline health (cols 7 and 8). Using male health (col 7) shows no detectable selection. In the case of female health (col 8), we do observe marginally statistically significant selection; the sign of $delta_3$ is positive, implying *advantageous* selection. However, the point estimate is small, implying that a one standard deviation improvement in predicted midline female health is associated with only a 3.6% increase in the enrollment rate. This is less than one fifth of the 22% effect of the unconditional cash transfer. In addition, as we discuss below, once we control for additional baseline characteristics and/or correct for multiple testing, the estimate is no longer significantly different from zero at conventional levels.

Next we examine total household medical expenditure (columns 9 and 10). As with visiting a medical facility, this is a potentially important variable which households might forecast and use to determine their willingness to pay to enroll in insurance; it is also arguably the key variable from the perspective of the government forecasting its costs. When we examine baseline medical spending, we find a very small and statistically insignificant δ_3 coefficient of -0.005, which would imply that a 1 SD decrease in baseline medical spending is associated with only a 0.8% decrease in enrollment under cost-sharing. Turning to *predicted* midline health spending, the estimated δ_3 is -0.04, significantly different from zero at the 1% level. Its sign implies that, consistent with adverse selection, those with lower predicted spending are differentially less likely to enroll under cost sharing. The magnitude, however, remains relatively small, with a 1 SD decrease in predicted spending associated with a 10.1% differential fall in enrollment under cost-sharing. To benchmark this, recall that providing a

INR. 220 unconditional cash transfer causes an increase in enrollment of 22%. The standard deviation of (winsorized) predicted health spending is approximately INR. 7000. The fact that the effect associated with a change in predicted health spending of INR. 7000 is less than half the magnitude of the effect of a much smaller INR. 220 cash transfer speaks to the fact that, while adverse selection is present to a statistically detectable extent, that extent is small.

5.1 Robustness

We explore the robustness of our key results in several ways.

Multiple testing

We test for adverse selection using many (10, across the baseline and predicted values) measures of utilization propensity. Out of these 10 measures, we find significant evidence of adverse selection in one case, and one marginally significant case suggesting advantageous selection. This raises the question of whether these findings are robust to multiple testing concerns. To this end, we report sharpened q-values (Anderson, 2008) at the bottom of Table 3. The specification for which we found evidence of a statistically significant (albeit small) degree of adverse selection, using predicted midline health spending, retains its significance with this correction (col. 10, $q = 0.03$). The evidence of advantageous selection, using predicted female midline health, is no longer significant (col. 8, $q = 0.45$). Thus our finding of (a small degree of) adverse selection is robust to the multiple testing adjustment. Without a large sample, we would likely have been unable to reject that this effect was equal to zero.

Is health risk proxying for something else?

Health status is not randomly assigned; it is potentially correlated with other characteristics. This may be important for tests of differential selection, as documented by Fang et al. (2008). In Table 4, we present a version of equation 1 which controls for several additional household characteristics (both their main effects and interactions with treatment arm). Specifically,

we control for education, Raven score[23], baseline assets and risk aversion[24] (main effects and interactions with treatment status). Wealth is one particularly salient possible confounder as those with worse health may also have differential amounts of assets and this may directly affect their willingness and/or ability to purchase insurance (Gropper and Kuhnen, 2021). Therefore we report the treatment arm-asset interactions; the other interactions are included but not reported in the table in the interest of readability.[25]

Examining the δ_3 coefficients in this specification shows that the marginally significant evidence of advantageous selection using predicted female midline health is no longer significant with the added controls, even before accounting for multiple testing, further suggesting that this finding may be spurious. Using predicted midline health spending we continue to observe a statistically significant (albeit small) degree of adverse selection. After adjusting for multiple testing, the coefficient is significant at the 10% level ($p = 0.06$). As before, using the other eight measures reveals no detectable evidence of either adverse or advantageous selection. Comparing the corresponding columns in Table 3 vs 4 shows that the magnitudes of the point estimates are not changed markedly by the inclusion of the additional controls: for instance, the significant δ_3 associated with predicted medical spending is -0.040 without additional controls and -0.047 with controls. The fact that the inclusion of observable controls does not lead to economically or statistically significant changes in coefficients indicates that there is unlikely to be a large role for *un*observed controls in confounding our interpretation of the health risk variables (Altonji et al., 2005; Oster, 2019).

The role of asset levels is also of interest in its own right. The main effect of assets identifies the role of assets in predicting takeup in arm A, when insurance is free. A higher level of assets is associated with *higher* takeup even under no cost-sharing. While perhaps surprising, this echoes a recent finding for the US context (Gropper and Kuhnen, 2021). In our context, one explanation may be that, even if care is free, time and transportation costs associated with seeking care serve as *de facto* co-pays which differentially affect poorer

[23]The important potential role of cognitive ability for selection into insurance is highlighted by Fang et al. (2008).

[24]Adjei-Mantey and Horioka (2022) show that, in cross-sectional data, both risk aversion and assets predict enrollment into insurance in Ghana; due to the nature of their data they do not have randomized variation in price, however.

[25]The table with all coefficients reported appears as Appendix Table A.3.

households.

Turning to the differential effect of assets under cost-sharing, we first note that the interactions between arm C (cost sharing, no cash transfer) and asset levels are consistently negative in sign, but uniformly smaller in magnitude than the main effect and not consistently significantly different from zero. This suggests that those with higher asset levels enroll at (generally insignificantly) higher levels under cost-sharing in arm C compared to those with lower asset holdings (because the total effect of assets remains positive). However, the negative sign on the $C \times asset$ term reveals that the total effect of assets in arm C is (always insignificantly) less than in A, the free arm. It is notable that, in this context, cost-sharing does not differentially screen out those who are poorer — if anything, the opposite pattern is seen, with higher asset-holding individuals leaving the risk pool under cost sharing. However, the relatively low precision of this effect makes this conclusion somewhat suggestive.

The differential effect of assets under arm B (cost sharing + unconditional cash transfer) relative to A (free insurance) is very similar to, and statistically indistinguishable from, the differential effect of assets under C (cost sharing, no cash transfer). In this context, providing an unconditional cash transfer does not change the wealth-enrollment gradient under cost-sharing.

5.2 Validating the approach to selection

Baseline measures versus predicted measures

We find a statistically detectable, albeit small, extent of selection using our LASSO-generated prediction of medical spending. One the other hand, we find no statistically detectable evidence of selection using baseline values (the approach commonly taken in the literature) of medical spending or of health. Why is (a small degree of) selection apparent with predicted values but not baseline values?

To investigate this question, we estimate the following regressions:

$$y_i^M = \alpha_B + \beta_B \times y_i^B + \epsilon_{i,B} \tag{2}$$

$$y_i^M = \alpha_P r + \beta_{Pr} \times \hat{y}_i^M + \epsilon_{i,Pr} \tag{3}$$

where y_i^M is an outcome (medical spending, the male household head's health, etc.) at midline, y_i^B is the corresponding outcome at baseline, and $\hat{y}_i^M$ is the *predicted* midline value, where the prediction is formed based on baseline characteristics chosen by LASSO as described above in section 4.3.

These regressions allow us to compare the extent to which y_i^B and $\hat{y}_i^M$ can statistically predict y_i^M by comparing β_B vs. $\beta_P r$. The results appear in Appendix A tables A.4-A.8. For all the outcomes (General health, male; General health, female; Visiting medical facility, male; Visiting medical facility, female; and medical expenditure), the predicted value is the better predictor: it is more correlated with the midline value than is the baseline value. The null hypothesis that the coefficients are equal is rejected at the 1% level in all cases.

These results demonstrate that machine learning methods can be used to construct stronger predictors than can be obtained by using the baseline value itself. Given that outcomes such as health (and consequently health spending) can display complex correlations over time, which can be positive or negative; and that medical needs at time t may be predicted by many variables from time $t-1$ or before, not just medical needs at $t-1$, this is perhaps intuitive. But, to our knowledge this point has not been demonstrated before in the context of testing for adverse selection.

6 Discussion

6.1 The role of predicted vs. baseline risk measures

A methodological point raised by our findings is that, had we used only, say, baseline general health to predict subsequent health, or only baseline spending to predict subsequent spending (the approach described in section 4.2), we would have concluded that there was *no* evidence for adverse selection, rather than a small but detectable amount as detected using the machine-learning-based approach described in section 4.3.[26] We reconcile these

[26]It it noteworthy that Fischer et al. (2018) use a related but distinct approach to identifying risk compared to our ML approach, namely, by directly collecting information on all inpatient health events and calculating the maximum claimable amount, $\bar{C_{i1}}$ based on these events, then using an OLS regression to predict $\bar{C_{i1}}$ as a function of baseline characteristics X_{i0}. Like us, they examine inpatient-only insurance and find a relatively small, but statistically detectable, extent of adverse selection when insurance must be purchased at the household level (compared to a much larger extent of selection when insurance can be purchased for

findings by showing that the ML-based predictions are more predictive of midline outcomes than are the baseline outcomes.

Given the growth of "big data" contexts where the number of available predictors may outstrip the number of observations (particularly in RCTs, where the number of observations is typically limited by cost and other considerations, but with administrative or otehr data there may be many possible Xs), the ML-based approach to identifying risk may be promising in other contexts including credit and other insurance products (annuities, crop insurance, auto insurance, etc.). To our knowledge, this approach has not previously been used in tests of adverse selection in health insurance markets.

6.2 Why do we find little selection?

Even with the ML method, the degree of selection we find is small; we can only detect it due to the large sample size. Why do we find little evidence of selection, while other studies of health insurance in LMICs such as Banerjee et al. (2021); Asuming et al. (2019) and Fischer et al. (2018) do? A key explanation appears to be differences in product design. In the Indonesian setting of Banerjee et al. (2021), a key driver of selection appears to be the ability to strategically add and drop coverage so that coverage is only carried during times of high expenditure. This is not possible in our context, where the enrollment decision had to be made during a 3-week sign-up window and was then locked in.

In the Pakistani context of Fischer et al. (2018), adverse selection was observed to a much greater extent when the household could flexibly choose which members to enroll, paying a per-person premium. When households were required to enroll their whole household (or no one), the extent of adverse selection fell significantly. A treatment in which 50% or more of a larger community unit (a loan group or community organization) had to sign up in order for anyone to be able to enroll eliminated selection altogether.

In the Ghanaian context of Asuming et al. (2019), the authors find that partial subsidies attract higher-cost individuals than full subsidies. A difference with our setting is the comprehensive nature of the insurance, covering both in- and outpatient procedures, medications, and emergency services. Accordingly, costs were higher than for RBSY, with a

each household member individually).

premium of USD 5.46 per adult, plus a processing fee. (In contrast, RSBY is a less generous program, and the premium of USD 9.87 covered up to five household members.)

In sum, RSBY's combination of fixed enrollment window, household-level enrollment system, and "no-frills," low-cost structure may have contributed to the high takeup rates and very modest amount of adverse selection we document. As other LMICs consider expanding healthcare access, such policies, offered at at the unit of the household and priced with actuarially fair premiums, may have important potential to achieve relatively high coverage while maintaining a sustainable risk pool.

7 Conclusion

Our findings show that, in our context (APL households in South and Central India), there is robust demand for health insurance even at actuarially fair prices. Moreover, we find little evidence of adverse selection, and even when it is detected, it is small in magnitude. Furthermore, charging for insurance is associated with increased utilization, a phenomenon which may be due to sunk cost effects and/or price signalling value.

These pieces of evidence, taken together, suggest that implementing cost sharing for above-poverty-line households may be a valuable way for developing countries to increase access to insurance in a fiscally sustainable manner.

Our results also provide methodological guidance for future work. The fact that predictions based on a data-driven selection over a large set of baseline outcomes appear better-powered to test selection, compared to directly using the baseline analog of the variable in question, may have relevance in other contexts such as credit and other insurance products. The finding of a small degree of selection, which was detectable thanks to the large sample, underscores the importance of—to the extent feasible—using large samples in settings where it is important to distinguish between no selection vs. a small extent of selection.

References

Adjei-Mantey, K. and Horioka, C. Y. (2022). Determinants of health insurance enrollment and health expenditure in Ghana: An empirical analysis. *ISER DP*, (1165).

Ahlin, C., Gulesci, S., Madestam, A., and Stryjan, M. (2020). Loan contract structure and adverse selection: Survey evidence from Uganda. *Journal of Economic Behavior & Organization*, 172:180–195.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Alatas, V., Purnamasari, R., Wai-Poi, M., Banerjee, A., Olken, B. A., and Hanna, R. (2016). Self-targeting: Evidence from a field experiment in Indonesia. *Journal of Political Economy*, 124(2):371–427.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy*, 113(1):151–184.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Andrews, A. and Vera Hernandez, A. (2022). Incentivizing demand for supply-constrained care: Institutional birth in india. *The Review of Economics and Statistics*.

Ashraf, N., Berry, J., and Shapiro, J. M. (2010). Can higher prices stimulate product use? evidence from a field experiment in zambia. *The American Economic Review*, 100(5):2383–2413.

Asuming, P., Kim, H. B., Sim, A., et al. (2019). Long-run consequences of health insurance promotion when mandates are not enforceable: Evidence from a field experiment in Ghana. Technical report.

Banerjee, A., Duflo, E., and Hornbeck, R. (2014a). Bundling health insurance and microfinance in India: There cannot be adverse selection if there is no demand. *American Economic Review*, 104(5):291–97.

Banerjee, A., Duflo, E., and Hornbeck, R. (2014b). Bundling health insurance and microfinance in india: There cannot be adverse selection if there is no demand</p>. *American Economic Review: Papers and Proceedings*.

Banerjee, A., Finkelstein, A., Hanna, R., Olken, B. A., Ornaghi, A., and Sumarto, S. (2021). The challenges of universal health insurance in developing countries: Evidence from a large-scale randomized experiment in Indonesia. *American Economic Review*, forthcoming.

Banerjee, A. V., Duflo, E., and Glennerster, R. (2008). Putting a band-aid on a corpse: incentives for nurses in the indian public health care system. *Journal of the European Economic Association*, 6(2-3):487–500.

Berg, E., Ghatak, M., Manjula, R., Rajasekhar, D., and Roy, S. (2019). Motivating knowledge agents: Can incentive pay overcome social distance? *The Economic Journal*, 129(617):110–142.

Berman, P., Ahuja, R., and Bhandari, L. (2010). The impoverishing effect of healthcare payments in India: new methodology and findings. *Economic and Political Weekly*, pages 65–71.

Cliff, B. Q., Miller, S., Kullgren, J. T., Ayanian, J. Z., and Hirth, R. (2021). Adverse selection in medicaid: Evidence from discontinuous program rules. Technical report, National Bureau of Economic Research.

Comptroller and Auditor General of India (2019). Hospital management in Uttar Pradesh for the year ended 31 march 2018. Report.

Einav, L. and Finkelstein, A. (2011). Selection in insurance markets: Theory and empirics in pictures. *Journal of Economic Perspectives*, 25(1):115–38.

Fang, H., Keane, M. P., and Silverman, D. (2008). Sources of advantageous selection: Evidence from the medigap insurance market. *Journal of political Economy*, 116(2):303–350.

Finkelstein, A. and Hendren, N. (2020). Welfare analysis meets causal inference. *Journal of Economic Perspectives*, 34(4):146–67.

Fischer, T., Frölich, M., and Landmann, A. (2018). Adverse selection in low-income health insurance markets: Evidence from a RCT in Pakistan.

Gropper, M. and Kuhnen, C. M. (2021). Wealth and insurance choices: Evidence from us households. *Available at SSRN 3888422*.

Gruber, J. (2022). Financing health care delivery. Technical report, National Bureau of Economic Research.

Gruber, J. and Lettau, M. (2004). How elastic is the firm's demand for health insurance? *Journal of Public Economics*, 88(7-8):1273–1293.

Gupta, I. and Chowdhury, S. (2014). Public financing for health coverage in India. *Economic Political Weekly*, 49(35).

Handel, B. R. and Ho, K. (2021). Industrial organization of health care markets. Technical report, National Bureau of Economic Research.

Hendren, N. and Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics*, 135(3):1209–1318.

Kan, H. J., Kharrazi, H., Chang, H.-Y., Bodycombe, D., Lemke, K., and Weiner, J. P. (2019). Exploring the use of machine learning for risk adjustment: A comparison of standard and penalized linear regression models in predicting health care costs in older adults. *PloS one*, 14(3):e0213258.

Lagomarsino, G., Garabrant, A., Adyas, A., Muga, R., and Otoo, N. (2012). Moving towards universal health coverage: health insurance reforms in nine developing countries in africa and asia. *The Lancet*, 380(9845):933–943.

Ma, S. and Sood, N. (2008). *A comparison of the health systems in China and India*. RAND, Santa Monica, CA.

MacLeod, W. B. (2021). A welfare analysis of competitive insurance markets with vertical differentiation and adverse selection. Technical report, National Bureau of Economic Research.

Malani, A., Holtzman, P., Imai, K., Kinnan, C., Miller, M., Shah, A., Swaminathan, S., Voena, A., Woda, B., and Conti, G. (2022). Effect of health insurance in India: A randomized controlled trial. *Working paper*.

Ministry of Health and Family Welfare (2014). National health policy 2015 draft.

OECD (2020). "purchasing power parities (ppp) (indicator).

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Pendzialek, J. B., Simic, D., and Stock, S. (2016). Differences in price elasticities of demand for health insurance: a systematic review. *The European Journal of Health Economics*, 17(1):5–21.

Rajasekhar, D., Berg, E., Ghatak, M., Manjula, R., and Roy, S. (2011). Implementing health insurance: the rollout of Rashtriya Swasthya Bima Yojana in Karnataka. *Economic and Political Weekly*, pages 56–63.

Shahrawat, R. and Rao, K. D. (2011). Insured yet vulnerable: out-of-pocket payments and India's poor. *Health policy and planning*, 27(3):213–221.

Tables

Table 1: Enrollment.

Free RSBY (A)	0.783***
	(0.008)
Sale of RSBY + Cash (B)	0.716***
	(0.012)
Sale of RSBY (C)	0.588***
	(0.014)
A = C (P-value)	0.000
B = C (P-value)	0.000
Mean enrollment	0.576
SD of enrollment	0.494
N	10879

All models estimated with OLS. Standard errors were clustered at the village level. Significance levels: * 10%, ** 5%, *** 1%.

Table 2: Summary Statistics of predicted health outcomes

	Obs.	Mean	Std. Dev.	Min	Max
General health of respondent, 5-scale, male, predicted	10,874	0.45	0.05	0.09	0.97
General health of respondent, 5-scale, female, predicted	10,871	0.45	0.05	0.01	0.97
Pr(rsp hospitalized last healthcare facility visit, female)	10,879	0.18	0.07	0.17	0.98
Pr(rsp hospitalized last healthcare facility visit, male)	10,752	0.17	0.17	0.00	1.00
Pr(rsp visited medical facility in past year, female)	10,727	0.67	0.19	0.00	1.00
Pr(rsp visited medical facility in past year, male)	10,594	0.43	0.24	0.00	1.00
Medical expenditures, predicted	10,879	28,465.39	4,832.42	-45,692.75	21,450.38

All variables are predicted using the LASSO procedure described in Appendix B. The general health variables are ladder questions measuring the subjective health of the respondent on the 1-5 ladder scale (1="Very good", 2="Good", 3="Fair", 4="Poor", 5="Very poor"). Predicted medical expenditure is winsorized at the 0.5th and 99.5 percentile.

Table 3: Enrollment by health

	(1) Enrollment	(2) Enrollment	(3) Enrollment	(4) Enrollment	(5) Enrollment	(6) Enrollment	(7) Enrollment	(8) Enrollment	(9) Enrollment	(10) Enrollment
Pay + Cash (B)	-0.057***	-0.053***	-0.063***	-0.064***	-0.057***	-0.054***	-0.065***	-0.065***	-0.062***	-0.067***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Pay (C)	-0.18***	-0.17***	-0.19***	-0.19***	-0.18***	-0.17***	-0.19***	-0.19***	-0.18***	-0.19***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)
shareABC	0.097**	0.083*	0.092*	0.085*	0.092*	0.081*	0.087*	0.086*	0.095*	0.086*
	(0.048)	(0.048)	(0.048)	(0.047)	(0.048)	(0.047)	(0.047)	(0.047)	(0.050)	(0.046)
Health	-0.018**	0.0048	0.011*	-0.027***	-0.015**	0.0034	-0.0054	0.0056	-0.0024	-0.012
	(0.0073)	(0.0061)	(0.0067)	(0.0066)	(0.0071)	(0.0071)	(0.0077)	(0.0049)	(0.0068)	(0.0074)
B ×	0.029**	-0.028**	-0.0075	0.030**	0.017	0.015	0.0065	0.017*	-0.026**	-0.033***
health	(0.012)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)	(0.011)	(0.0090)	(0.010)	(0.011)
C ×	0.0085	-0.0049	0.0041	-0.0072	0.013	0.012	0.0093	0.020*	-0.0050	-0.040***
health	(0.013)	(0.013)	(0.013)	(0.013)	(0.015)	(0.014)	(0.013)	(0.011)	(0.014)	(0.013)
mean dep var	0.733	0.728	0.719	0.719	0.733	0.728	0.718	0.718	0.725	0.718
SD dep var	0.442	0.445	0.450	0.450	0.442	0.445	0.450	0.450	0.446	0.450
N	7611	7917	8488	8675	7619	7903	8727	8727	7630	8727
health variable	- Visited medical facility, male	- Visited medical facility, female	- Pr(visiting medical facility, male)	- Pr(visiting medical facility, female)	General health, male, baseline	General health, female, baseline	General health, male, midline predicted	General health, female, midline predicted	- Medical expenditure, baseline (winsorized)	- Medical expenditure, predicted (winsorized)
CXHealth × SD/Mean enrollment	0.014	-0.008	0.007	-0.012	0.021	0.020	0.016	0.036	-0.008	-0.068
Sharpened q-value	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.45	1.00	0.03

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at village level. Health variables are standardized by the sample mean and standard deviation.

Table 4: Enrollment by health (education, raven scores and risk aversion controlled)

	(1) Enrollment	(2) Enrollment	(3) Enrollment	(4) Enrollment	(5) Enrollment	(6) Enrollment	(7) Enrollment	(8) Enrollment	(9) Enrollment	(10) Enrollment	(11) Enrollment
Pay + Cash (B)	-0.15***	-0.14***	-0.15***	-0.13**	-0.14***	-0.16***	-0.13**	-0.13**	-0.14**	-0.14***	-0.15***
	(0.052)	(0.052)	(0.052)	(0.053)	(0.053)	(0.052)	(0.051)	(0.053)	(0.054)	(0.052)	(0.056)
Pay (C)	-0.19***	-0.19***	-0.19***	-0.19***	-0.17***	-0.20***	-0.19***	-0.20***	-0.17***	-0.19***	-0.20***
	(0.058)	(0.058)	(0.058)	(0.059)	(0.057)	(0.059)	(0.059)	(0.059)	(0.058)	(0.057)	(0.062)
Health		0.0079	-0.0069	0.014**	-0.0022	0.0039	-0.020***	-0.018**	0.0035	-0.0058	-0.00045
		(0.0081)	(0.0046)	(0.0071)	(0.0073)	(0.0068)	(0.0070)	(0.0073)	(0.0061)	(0.0080)	(0.0069)
B ×		-0.0070	-0.012	-0.012	-0.014	-0.0023	0.032**	0.028**	-0.027**	0.026**	0.025**
health		(0.012)	(0.0088)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.011)	(0.010)
C ×		-0.012	-0.017	-0.012	-0.013	0.0097	0.0044	0.0079	-0.0038	0.047***	0.0059
health		(0.014)	(0.011)	(0.015)	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)	(0.017)	(0.014)
Asset	0.062***	0.065***	0.063***	0.063***	0.078***	0.065***	0.055***	0.066***	0.078***	0.064***	0.065***
	(0.014)	(0.015)	(0.014)	(0.014)	(0.015)	(0.016)	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)
B × asset	-0.033	-0.035	-0.034	-0.036	-0.041*	-0.034	-0.023	-0.040*	-0.044*	-0.042*	-0.049**
	(0.022)	(0.023)	(0.022)	(0.022)	(0.023)	(0.023)	(0.022)	(0.022)	(0.023)	(0.023)	(0.024)
C × asset	-0.034	-0.039	-0.035	-0.040	-0.049*	-0.038	-0.031	-0.042	-0.050*	-0.051*	-0.030
	(0.026)	(0.027)	(0.026)	(0.026)	(0.027)	(0.028)	(0.026)	(0.026)	(0.027)	(0.027)	(0.028)
shareABC	0.076	0.077	0.076	0.091*	0.072	0.082*	0.075	0.097**	0.074	0.074	0.086*
	(0.047)	(0.047)	(0.047)	(0.048)	(0.047)	(0.047)	(0.047)	(0.048)	(0.048)	(0.047)	(0.051)
mean dep var		0.718	0.718	0.733	0.728	0.719	0.719	0.733	0.728	0.718	0.725
SD dep var		0.450	0.450	0.442	0.445	0.450	0.450	0.442	0.445	0.450	0.446
N	8727	8727	8727	7619	7903	8488	8675	7611	7917	8727	7630
health variable		General health, male, midline predicted	General health, female, midline predicted	General health, male, baseline	General health, female, baseline	- Pr(visiting medical facility, male)	- Pr(visiting medical facility, female)	- Visited medical facility, male	- Visited medical facility, female	Medical expendi-ture, predicted	Medical expendi-ture, baseline
CXHealth × SD/Mean enrollment		-0.021	-0.031	-0.020	-0.022	0.016	0.007	0.013	-0.006	0.079	0.010
Sharpened q-value		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.06	1.00

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at village level. Health variables are standardized by the sample mean and standard deviation. Controls for levels and interactions with education, raven score, and risk aversion are included but not reported. Education is a categorical variable recording the educational attainment of household head (1 = Never attended class 1, 2 = Class 1-5, 3 = Class 6-8, 4 = Class 9-10, 5 = Class 11-12, 6 = Graduate and above). Risk aversion is a dummy variable that equals 1 if respondent had a certain equivalence smaller than 180 in a gamble game. Missing indicators of assets, raven scores, risk aversion and education are included.

A Appendix: Supplemental tables

A.1 Randomization Design

Table A.1: Two-Stage Randomization Design

Village-level arms (%)		Household-level arms (%)			
Arm	Village allocation	Group A	Group B	Group C	Group D
I	15	30	50	10	10
II	15	30	10	50	10
III	15	30	10	10	50
IV	35	70	10	10	10
V	20	10	30	30	30
Total	100	40	20	20	20

A.2 Attrition

Table A.2: Randomized sample, response rate and attrition by survey

	Insurance (access) arm					
	(A) Free insurance	(B) Sale of insurance + transfer	(C) Sale of insurance	(D) No intervention	Total	Attrition
Randomized sample	4,401	2,180	2,146	2,152	10,879	
Biometric sample	1,363	718	737	833	3,651	
Baseline survey						
Main survey	4,155	2,080	2,022	2,033	10,290	589
	(94%)	(95%)	(94%)	(94%)	(95%)	(5%)
Biometric survey	1,324	652	702	748	3,426	225
	(97%)	(91%)	(95%)	(90%)	(94%)	(6%)
Midline survey (18 mo.)						
Main survey	4,091	2,036	2,007	1,977	10,111	768
	(93%)	(93%)	(94%)	(92%)	(93%)	(7%)
Biometric survey	1,195	607	632	667	3,101	550
	(88%)	(85%)	(86%)	(80%)	(85%)	(15%)
Endline survey (4 yr.)						
Main survey	3,879	1,902	1,855	1,874	9,510	1,369
	(88%)	(87%)	(86%)	(87%)	(87%)	(13%)

A.3 Reproducing Table 3 with all coefficients

Table A.3: Enrollment by health (education, raven scores and risk aversion controlled)

	(1) Enrollment	(2) Enrollment	(3) Enrollment	(4) Enrollment	(5) Enrollment	(6) Enrollment	(7) Enrollment	(8) Enrollment	(9) Enrollment	(10) Enrollment	(11) Enrollment
Pay + Cash (B)	-0.15***	-0.14***	-0.15***	-0.13**	-0.14***	-0.16***	-0.13**	-0.13**	-0.14**	-0.14***	-0.15***
	(0.052)	(0.052)	(0.052)	(0.053)	(0.053)	(0.052)	(0.051)	(0.053)	(0.054)	(0.052)	(0.056)
Pay (C)	-0.19***	-0.19***	-0.19***	-0.19***	-0.17***	-0.20***	-0.19***	-0.20***	-0.17***	-0.19***	-0.20***
	(0.058)	(0.058)	(0.058)	(0.059)	(0.057)	(0.059)	(0.059)	(0.059)	(0.058)	(0.057)	(0.062)
Health		0.0079	-0.0069	0.014**	-0.0022	0.0039	-0.020***	-0.018**	0.0035	-0.0058	-0.00045
		(0.0081)	(0.0046)	(0.0071)	(0.0073)	(0.0068)	(0.0070)	(0.0073)	(0.0061)	(0.0080)	(0.0069)
B ×		-0.0070	-0.012	-0.012	-0.014	-0.0023	0.032**	0.028**	-0.027**	0.026**	0.025**
health		(0.012)	(0.0088)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.011)	(0.010)
C ×		-0.012	-0.017	-0.012	-0.013	0.0097	0.0044	0.0079	-0.0038	0.047***	0.0059
health		(0.014)	(0.011)	(0.015)	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)	(0.017)	(0.014)
Asset	0.062***	0.065***	0.063***	0.063***	0.078***	0.065***	0.055***	0.066***	0.078***	0.064***	0.065***
	(0.014)	(0.015)	(0.014)	(0.014)	(0.015)	(0.016)	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)
B × asset	-0.033	-0.035	-0.034	-0.036	-0.041*	-0.034	-0.023	-0.040*	-0.044*	-0.042*	-0.049**
	(0.022)	(0.023)	(0.022)	(0.022)	(0.023)	(0.023)	(0.022)	(0.022)	(0.023)	(0.023)	(0.024)
C × asset	-0.034	-0.039	-0.035	-0.040	-0.049*	-0.038	-0.031	-0.042	-0.050*	-0.051*	-0.030
	(0.026)	(0.027)	(0.026)	(0.026)	(0.027)	(0.028)	(0.026)	(0.026)	(0.027)	(0.027)	(0.028)
Raven score	0.0044	0.0046	0.0044	0.0046	0.0046	0.0038	0.0040	0.0045	0.0051	0.0044	0.0053
	(0.0036)	(0.0036)	(0.0037)	(0.0038)	(0.0038)	(0.0037)	(0.0037)	(0.0038)	(0.0038)	(0.0036)	(0.0036)
B × raven	0.0015	0.0013	0.0017	0.000012	0.0026	0.0027	0.0026	0.00025	0.0016	0.0013	-0.00062
score	(0.0058)	(0.0059)	(0.0058)	(0.0060)	(0.0061)	(0.0059)	(0.0058)	(0.0060)	(0.0060)	(0.0058)	(0.0061)
C × raven	-0.0032	-0.0035	-0.0038	-0.0039	-0.0042	-0.0019	-0.0034	-0.0040	-0.0040	-0.0036	-0.00047
score	(0.0069)	(0.0069)	(0.0069)	(0.0074)	(0.0072)	(0.0070)	(0.0070)	(0.0074)	(0.0073)	(0.0068)	(0.0071)
Risk aversion	-0.026	-0.026	-0.028	-0.024	-0.033	-0.033	-0.023	-0.025	-0.033	-0.026	-0.027
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
B × risk	0.032	0.031	0.031	0.031	0.031	0.043	0.028	0.029	0.028	0.030	0.041
aversion	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.039)	(0.040)	(0.041)	(0.040)	(0.042)
C × risk	0.021	0.021	0.020	0.020	0.016	0.020	0.018	0.023	0.014	0.017	0.013
aversion	(0.045)	(0.045)	(0.045)	(0.045)	(0.044)	(0.046)	(0.045)	(0.046)	(0.044)	(0.045)	(0.048)
Education	0.0013	0.0021	0.00095	0.0033	0.0017	0.0016	0.0032	0.0025	0.0019	0.0013	0.0014
	(0.0044)	(0.0045)	(0.0044)	(0.0045)	(0.0044)	(0.0044)	(0.0045)	(0.0045)	(0.0044)	(0.0044)	(0.0046)
B ×	0.017**	0.016**	0.016**	0.014*	0.014*	0.016**	0.012	0.014*	0.015**	0.016**	0.016**
education	(0.0073)	(0.0074)	(0.0073)	(0.0074)	(0.0074)	(0.0075)	(0.0074)	(0.0074)	(0.0074)	(0.0073)	(0.0076)
C ×	0.0074	0.0061	0.0069	0.0081	0.0051	0.0091	0.0065	0.0089	0.0059	0.0065	0.0056
education	(0.0075)	(0.0076)	(0.0075)	(0.0078)	(0.0078)	(0.0078)	(0.0076)	(0.0079)	(0.0077)	(0.0075)	(0.0080)
shareABC	0.076	0.077	0.076	0.091*	0.072	0.082*	0.075	0.097**	0.074	0.074	0.086*
	(0.047)	(0.047)	(0.047)	(0.048)	(0.047)	(0.047)	(0.047)	(0.048)	(0.048)	(0.047)	(0.051)
mean dep var		0.718	0.718	0.733	0.728	0.719	0.719	0.733	0.728	0.718	0.725
SD dep var		0.450	0.450	0.442	0.445	0.450	0.450	0.442	0.445	0.450	0.446
N	8727	8727	8727	7619	7903	8488	8675	7611	7917	8727	7630
health variable		General health, male, midline predicted	General health, female, midline predicted	General health, male, baseline	General health, female, baseline	- Pr(visiting medical facility, male)	- Pr(visiting medical facility, female)	- Visited medical facility, male	- Visited medical facility, female	Medical expendi-ture, predicted	Medical expendi-ture, baseline
CXHealth × SD/Mean enrollment		-0.021	-0.031	-0.020	-0.022	0.016	0.007	0.013	-0.006	0.079	0.010
Sharpened q-value		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.06	1.00

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at village level. Health variables are standardized by the sample mean and standard deviation. This table reproduces the results of Table 3 but reports the coefficients on controls for levels and interactions with education, raven score, and risk aversion. Education is a categorical variable recording the educational attainment of household head (1 = Never attended class 1, 2 = Class 1-5, 3 = Class 6-8, 4 = Class 9-10, 5 = Class 11-12, 6 = Graduate and above). Risk aversion is a dummy variable that equals 1 if respondent had a certain equivalence smaller than 180 in a gamble game. Missing indicators of assets, raven scores, risk aversion and education are included.

A.4 Correlations between predicted and realized midline outcomes

Table A.4: Correlation: General health, male

	(1) General health, male, baseline	(2) General health, male, midline predicted
General health, male, midline	0.061***	0.11***
	(0.011)	(0.010)
N	8293	9244
chi2 test p-value	0.00	0.00

*** p<0.01, ** p<0.05, * p<0.1. The bottom row displays the p value from the chi2 test of the difference between the two coefficients.

Table A.5: Correlation: General health, female

	(1) General health, female, baseline	(2) General health, female, midline predicted
General health, female, midline	0.10***	0.17***
	(0.011)	(0.010)
N	8883	9739
chi2 test p-value	0.00	0.00

*** p<0.01, ** p<0.05, * p<0.1. The bottom row displays the p value from the chi2 test of the difference between the two coefficients.

Table A.6: Correlation: Visiting medical facility, male

	(1) visited med fac, male, baseline	(2) visited med fac, male, midline predicted
Pr(visiting med fac), male, midline	0.041***	0.20***
	(0.011)	(0.010)
N	8313	9011
chi2 test p-value	0.00	0.00

*** p<0.01, ** p<0.05, * p<0.1. The bottom row displays the p value from the chi2 test of the difference between the two coefficients.

Table A.7: Correlation: Visiting medical facility, female

	(1) visited med fac, female, baseline	(2) visited med fac, female, midline predicted
Pr(visiting med fac), female, midline	0.029***	0.15***
	(0.011)	(0.0089)
N	8850	9633
chi2 test p-value	0.01	0.00

*** p<0.01, ** p<0.05, * p<0.1. The bottom row displays the p value from the chi2 test of the difference between the two coefficients.

Table A.8: Correlation: Medical expenditure

	(1) Medical expenditure (w), baseline	(2) Medical expenditure (w), midline predicted
Medical expenditure (w), midline	0.056***	0.12***
	(0.010)	(0.0098)
N	8779	10030
chi2 test p-value	0.00	0.00

*** p<0.01, ** p<0.05, * p<0.1. The bottom row displays the p value from the chi2 test of the difference between the two coefficients.The hospital expenditure is winsorized at 0.5% and 99.5%. The medical expenditure is winsorized at 0.5% and 99.5%.

B Appendix: LASSO details

B.1 Predictive models

We generated the following predicted midline outcomes through training a LASSO model on the control group sample.

- Predicted male health at midline. Predicted variable is a ladder question measuring the subjective health of the male respondent on the 1-5 ladder scale. Fitted with a linear model.

- Predicted female health at midline is a ladder question measuring the subjective health of the female respondent on the 1-5 ladder scale. Fitted with a linear model.

- Predicted probability of visiting medical facility for male respondent at midline. Predicted variable is a dummy for visiting medical facility in the past year for the male respondent. The model is fitted with logistic regression.

- Predicted probability of visiting medical facility for female respondent at midline. Predicted variable is a dummy for visiting medical facility in the past year for the female respondent. The model is estimated with logistic regression.

- Predicted medical spending at midline. Predicted variable is medical expenditure at midline, winsorized at 0.5% and 99.5%. The model is estimated with linear regression.

  To generate the predictions, we fit 10-folds cross-validated LASSO on the control group of 2152 (or smaller due to missing outcomes), and then predicted the outcomes in treatment group with the trained models. The penalty parameter is selected under minimum rule.

B.2 Histograms of predicted outcomes

Figure B.1: Histogram: General health of respondent, 5-scale, male, predicted

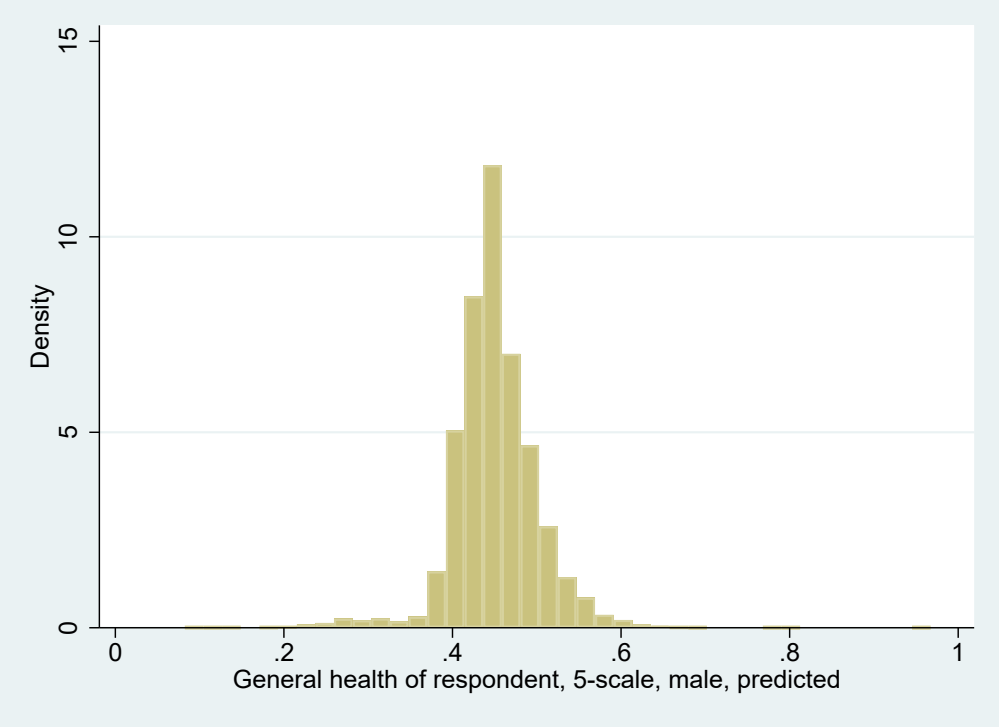


Figure B.2: Histogram: General health of respondent, 5-scale, female, predicted

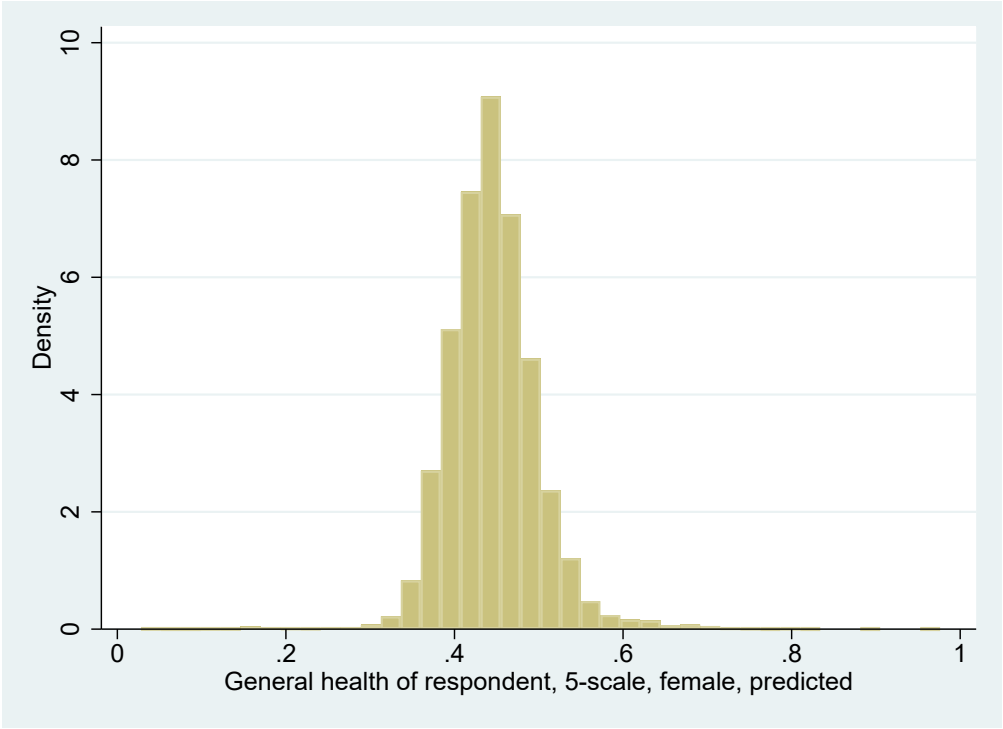


Figure B.3: Histogram: Pr(rsp visited medical facility in past year, male))

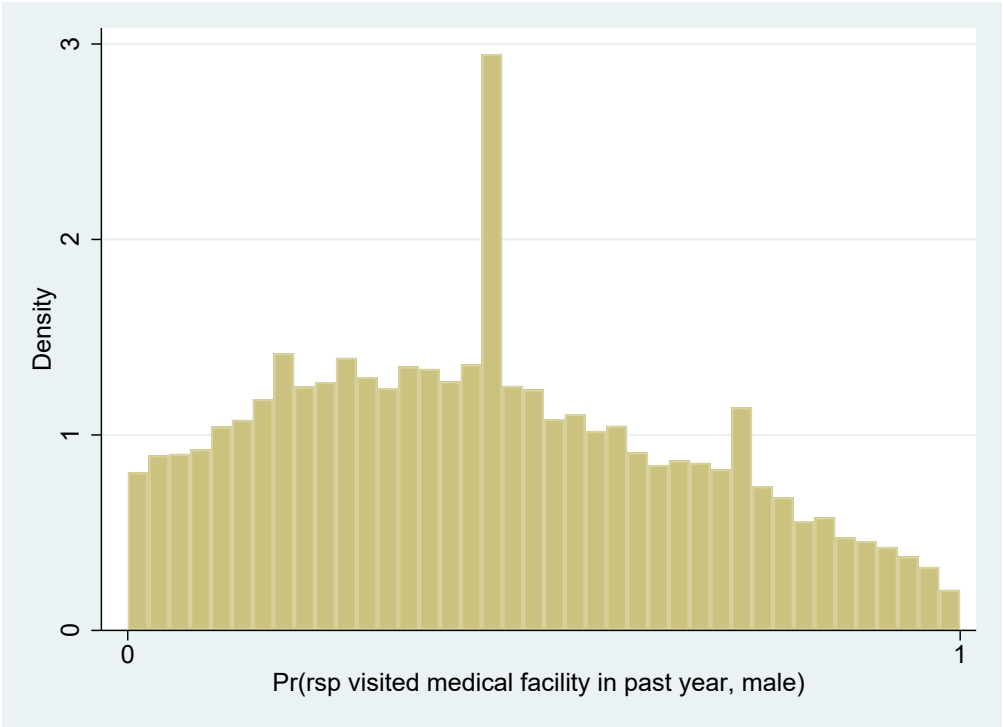


Figure B.4: Histogram: Pr(rsp visited medical facility in past year, female))

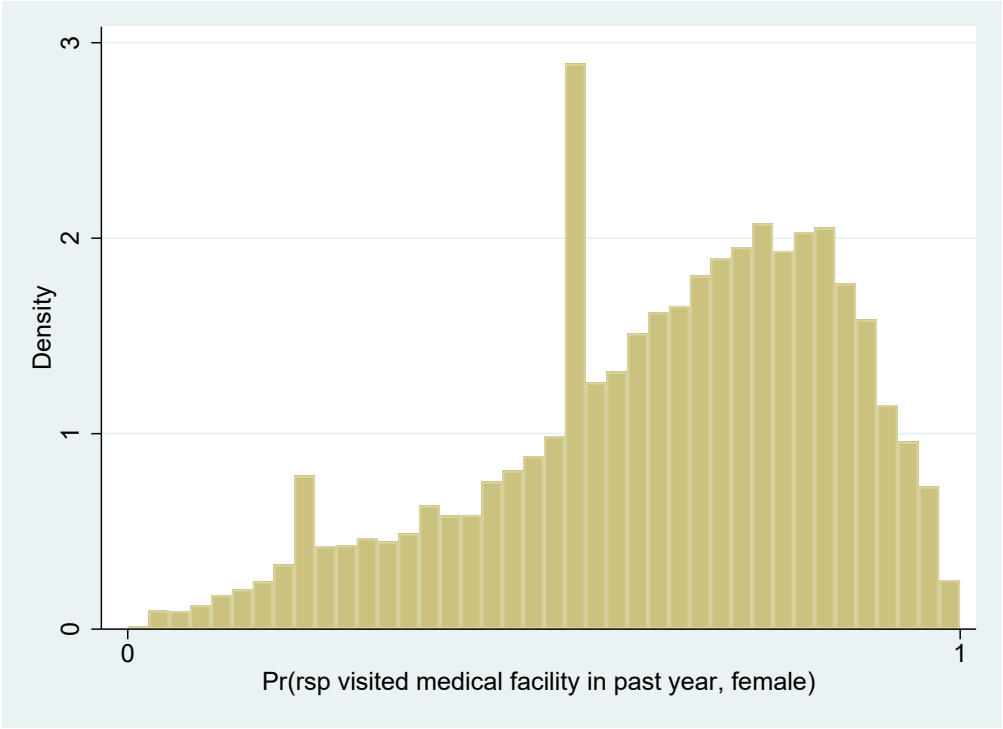
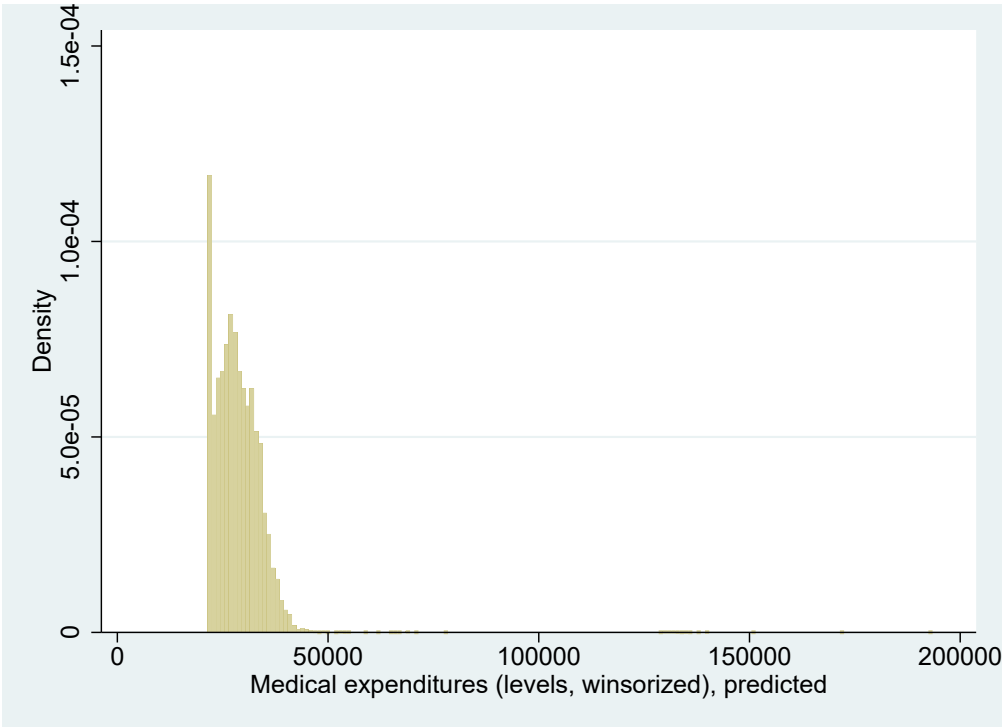


Figure B.5: Medical expenditures (levels, winsorized), predicted



C Description of surveys

C.1 Baseline Survey

The baseline survey took place in August 2013 – February 2014. This round included a consent form for the overall study as well as the survey. We administered surveys to up to 3 distinct members of each household (the female and male most knowledgeable about household finances and a female of childbearing age) for the entire sample. These individuals were asked modules about subjective health status, health care consumption and financial status. In addition, for a subsample of roughly 4000 households, we also conducted an anthropometric survey that gathered objective health status (e.g., BP, body fat, weight, lung capacity), on up to 3 members of the household (the male most knowledgeable about household affairs, a female with childbearing capacity, and a child under the age of 5. Households were paid INR 250 as a participation incentive for completing major sections of the survey. We performed back checks on 10% of households, a rate known to surveyors *ex ante*.

C.2 Midline Survey

In November 2016 – February 2017, we conducted an 18 month follow-up survey. The format was nearly identical to the baseline survey, including the anthropometric survey. Households were paid INR 250 as a participation incentive for completing major sections of the survey. We performed backchecks on approximately 15% of households,[1] a rate known to surveyors *ex ante*.

C.3 Endline Survey

In March - May 2019, we conducted a 4 year follow-up survey. We surveyed 1 member (our first priority was to interview the female most knowledgeable from baseline (or midline for households with missing baseline data), followed by the current female most knowledgeable, and then the male most knowledgeable) of each household in the sample. Respondents were asked about subjective health status, health care consumption and financial status. House-

[1]The actual rate is 4 households per village on each module of the the survey.

holds were paid a participation incentive comprised of bars of soap and tubes of toothpaste valued at approximately INR 50 for completing major sections of the survey. We performed backchecks on 10% of households, a rate known to surveyors *ex ante*.

C.4 Balance Tests

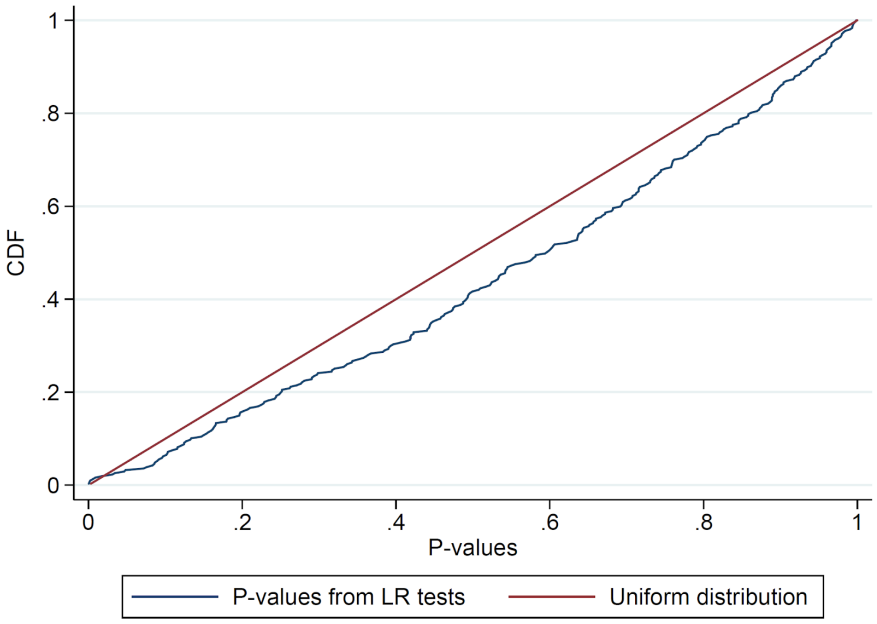


Figure C.6: Test of balance.

We conduct balance tests to validate that assignment to treatment was indeed random. We do this in 3 steps. First, we gather baseline measurements on a range of variables on demographics, financial status and health (number of persons and of children in household, age and education of head of household, distance to nearest town, number of rooms and of concrete rooms in home, annual household budget and good expenditure, major sickness) and on a subset of outcomes (visited health care facility, annual hospital and non-hospital expenditures). Second, we estimate multinomial logit models predicting household treatment assignments for each household (A/B/C/D) as a function of outcomes measured at baseline, one outcome at a time. Third, we conduct likelihood ratio tests where the null model is the same multinomial model without the baseline covariate, to determine if we can reject the null that these two models are statistically equivalent, i.e., that the baseline covariate has no explanatory power. We collect the p-values from these LR tests.

If the randomization is successful, then the p-values from these tests should stochastically dominate the uniform distribution. Figure C.6 plots these p-values and the CDF of the uniform distrobution. A one-sided Kolmogorov–Smirnov test confirms that our p-values stochastically dominate a uniform ($P = 1.000$).