

TI 2026-079/IV
Tinbergen Institute Discussion Paper

The Impact of Health Insurance: Evidence from High-frequency Data in Kenya

*Clotilde Mahe*¹

*Richard de Groot*²

*Amanuel Abajobir*³

*Caroline Wainaina*⁴

*Estelle Monique Sidze*⁵

*Menno Pradhan*⁶

*Wendy Janssens*⁷

¹ Vrije Universiteit Amsterdam, Health Economics Research Amsterdam (HERA), and Amsterdam Institute for Global Health and Development (AIGHD), Netherlands

² United Nations Children's Fund Innocenti – Global Office of Research and Foresight, Italy

³ African Population and Health Research Center (APHRC), Kenya

⁴ African Population and Health Research Center (APHRC), Kenya

⁵ African Population and Health Research Center (APHRC), Kenya

⁶ Vrije Universiteit Amsterdam and University of Amsterdam

⁷ Vrije Universiteit Amsterdam

The Impact of health insurance: Evidence from high-frequency data in Kenya*

Clotilde Mahé Richard de Groot Amanuel Abajobir
Caroline Wainaina Estelle Monique Sidze Menno Pradhan
Wendy Janssens

September 28, 2026

Abstract

Health insurance in low- and middle-income countries increases financial protection and utilization, but health improvements have not been consistently found, prompting debate about the value of subsidized coverage. We revisit this question using a cluster-randomized experiment of a fully subsidized, mobile-based health insurance program in Kenya, combined with weekly surveys collected over eighteen months. Full subsidies combined with trusted, hands-on registration support drove take-up near universal coverage, addressing the weak first-stage that has limited prior evaluations, while high-frequency measurement captures health events that standard recall surveys miss. We find that insurance reduces catastrophic health expenditure, shifts responses to illness from foregone care and self-medication toward formal care, and speeds care-seeking after illness or injury. We find suggestive evidence of health improvements through fewer days unable to perform daily activities. These effects would likely go undetected in standard endline surveys, underscoring the value of high-frequency measurement combined with strong take-up.

JEL classifications: I13, I15, I18, O12

Keywords: Health insurance, Randomized control trial
High-frequency data, Kenya

*Mahé: Corresponding author, c.n.mahe@vu.nl. Abajobir, Sidze, Wainaina: African Population and Health Research Center, Kenya. de Groot: United Nations Children's Fund Innocenti – Global Office of Research and Foresight, Italy. Janssens and Mahé: Vrije Universiteit Amsterdam and Amsterdam Institute for Global Health and Development, Netherlands. Pradhan: Universiteit van Amsterdam, Vrije Universiteit Amsterdam and Amsterdam Institute for Global Health and Development, Netherlands. This study was funded by the Dutch National Postcode Lottery, the Joep Lange Institute and the Dutch Ministry of Foreign Affairs through the Health Insurance Fund. Ethical approval, granted by the Amref Health Africa Ethics and Scientific Review Committee (P679-2019, August 8, 2019), was amended for the Covid-19-induced switch to phone interviews. The trial was registered at AEARCTR-0006089.

1 Introduction

Despite international commitments to universal health coverage (UHC), more than half of the global population lacks access to quality health care, and roughly 100 million people fall into extreme poverty each year due to ill health, particularly in low- and middle-income countries (LMICs) (World Health Organization and World Bank, 2021; Wagstaff et al., 2018). While health insurance schemes in these settings have consistently provided financial protection and increased health care utilization, they have not systematically led to improvements in health outcomes (Das and Do, 2023). This disconnect has raised questions about whether financial protection alone justifies expanding subsidized health insurance (Finkelstein et al., 2019b).

Several factors may explain the absence of clearly identified health improvements. Insurance might not improve health in these settings, if, for instance, quality of care is low, or the covered benefit package is limited (Das et al., 2008; Kruk et al., 2018). Also, formal insurance might merely substitute for pre-existing informal risk-sharing arrangements, rather than adding to households' effective coverage (Hillebrecht et al., 2021). Alternatively, providers' responses to insurance could weaken its health benefits. Health care providers have been found to respond to insurance-induced changes in relative prices by providing suboptimal care to insured patients (Dalinjong and Laar, 2012), opaque billing (Jain, 2024), or inappropriate treatments (Lu, 2014). Such behaviors lower perceived quality of care (Duku et al., 2018), erode trust in providers and insurance, and ultimately diminish the value of coverage for enrollees (Das and Do, 2023).

A second reason is that weak or null health effects reflect low uptake instead of a genuine absence of impacts. Voluntary take-up of insurance is often limited in LMICs, despite substantial exposure to health expenditure risk (Thornton et al., 2010; Acharya et al., 2013; Banerjee et al., 2021). Methodologically, low take-up leads to a weak 'first-stage,' making it difficult to detect effects, even when they exist. Low take-up may itself stem from little perceived value for money (Finkelstein et al., 2019a), or from enrollment barriers unrelated to value, such as affordability or liquidity constraints (Cole et al., 2013; Finkelstein et al., 2019b), cumbersome administrative and payment procedures (Banerjee et al., 2021), low geographical access to covered

care (Dupas and Jain, 2024), or limited trust (Dercon et al., 2019). Few health insurance evaluations achieve take-up rates sufficiently high to rule out confidently a weak first-stage as an explanation for null health effects.

Third, detecting health effects may be hampered by measurement challenges. Existing studies tend to be underpowered to detect significant effect sizes (Acharya et al., 2013; Levine et al., 2016). They often rely on a limited set of health indicators, and may not capture behavioral responses such as delays in care-seeking or self-medication, plausibly missing dimensions most responsive to insurance. Health surveys are also prone to recall bias, particularly among the poorest (Das et al., 2012). The longer the recall period, the more likely respondents are to report only severe illnesses or large expenditures, while minor but frequent health events, which may represent substantial portions of annual out-of-pocket (OOP) health spending, go unreported. Standard surveys, that typically collect a single or few follow-ups after roll-out using long recall periods, could miss half or more of health events and corresponding expenditures (Nelissen et al., 2020).

In this study, we address the latter two empirical challenges directly. We study the effects of health insurance using a matched-pair cluster-randomized controlled trial (cRCT), combined with uniquely detailed high-frequency data. The intervention, embedded in Kenya’s Innovative Partnership for Universal Sustainable Healthcare (i-Push), provided mobile phone-based, fully subsidized, 12-month access to the National Health Insurance Fund (NHIF) SupaCover premium for targeted women and their families, as well as hands-on support in the required administrative procedures. The experimental design overcomes endogeneity issues that plague observational studies of health insurance (e.g. Spenkuch, 2012).¹ Combining full subsidy and administrative support permits individual take-up to reach 80% in treated areas, far above low voluntary take-up rates commonly documented across LMICs, which strengthens our ability to detect health effects where they exist.

¹ Identifying health insurance effects is typically confounded by adverse selection, since take-up and outcomes of interest are correlated with ex-ante health risk, and moral hazard, whereby anticipated access to insurance could alter pre-enrollment health behaviors. Our matched-pair cRCT design addresses such concerns by randomizing access to insurance.

To address measurement concerns, we conducted weekly surveys (diaries) among 240 households, over 18 months – six months before and 12 months after roll-out – allowing us to study individual and household responses at a level of detail rarely feasible in LMICs. We track all health events occurring within a household each week, together with corresponding care-seeking decisions, provider choices, treatments and related financial transactions. Such high-frequency data limit recall bias, and capture minor and infrequent health events that standard surveys tend to miss, while compensating for a relatively small sample size. We also measure outcomes rarely captured in the literature, including the time elapsed before care is sought and the incidence of self-medication. Where possible, we compare high-frequency estimates with those obtained from a single endline using 12-month recall variables to assess whether a standard ex-post survey would capture the effects we identify, or, instead, misestimate the benefits of health insurance enrollment.

We first show that the intervention successfully increased NHIF enrollment: individual take-up rose by 56.3 percentage points (p.p.) from a baseline of 19.1% in control communities, reaching approximately 80% individual enrollment in treated villages by October 2020, while the share of households with at least one member insured reached 95%, almost universal coverage. The success of the intervention likely reflects its comprehensive approach, combining subsidized premia with hands-on registration support delivered by trusted community health workers, that addresses the administrative, logistical and trust barriers that have limited enrollment in other settings. Having successfully encouraged take-up implies a first-stage that should be ‘strong’ enough to identify the impacts of health insurance, if present, a strength compared to many studies.

We next aggregate individual weekly data at the month level, and run a series of Analysis of Covariance (ANCOVA) specifications, to estimate intention-to-treat (ITT) and Local Average Treatment Effects (LATE) of the intervention, using village treatment assignment as instrumental variable (IV). Assignment to the intervention significantly increases coverage of formal consultation costs: it reduces the probability of not being insured for any formal consultation costs by 10.8 p.p. and increases partial or full NHIF coverage by 16.7 p.p., which translates into a 18.7 p.p. decrease and 26.9 p.p. increase, among NHIF enrollees, respectively.

This leads to substantial improvements in financial protection: total weekly out-of-pocket (OOP) health expenditures (unconditional on illness or injury) fall by KES 6.6 on average at the individual level, which, compared with a post-intervention mean of KES 12.6 in the control group, amounts to a reduction of 52.0%. Reductions are driven by laboratory, medication, formal and curative care spending – those types of expenses covered by the subsidized scheme. At the household level, health expenditure as a share of capacity to pay declines by 4.1 p.p. from a post-intervention mean in the control group of 16.6%; the probability of catastrophic health expenditure drops by 10.3 p.p. – 18.1 p.p. among compliers – from a post-intervention mean of 45.6% in control villages. Endline survey data with 12-month recall also indicate that health care expenditures halved as a result of the intervention.

Using individual illness episodes, we find that assignment to the intervention shifts responses to illness: self-medication decreases by 8.7 p.p. (from 35.3% post-intervention in control), while formal care-seeking increases by 5.7 p.p. (from 38.9% post-intervention in control), even though the latter estimate loses statistical significance after small-cluster corrections. Among those ill who seek care, the intervention reduces the number of days between illness and care seeking by 0.8 days, which, compared to a post-intervention control mean of 3.3 days, amounts to a reduction of 23%. These illness-episode responses do not translate into corresponding significant changes in unconditional health service utilization. Still, when allowing treatment effects to vary with distance to the closest health facility, we do find that the intervention increases care-seeking, primarily through formal and curative care, with effects diminishing with greater distance. The absence of an effect on preventive care is consistent with the intervention subsidizing primarily curative services, effectively increasing the relative price of preventive care. Endline data collected through traditional 12-month recall surveys also fail to detect significant effects on formal care utilization, in spite of identifying a significant reduction in foregone care among compliers.

While we find a negative but imprecisely estimated effect on being ill or injured, the intervention significantly reduces the number of days that individuals are unable to do daily activities on average in a week by 36.4%. This reduction in severity is consistent with the faster

care-seeking and lower foregone care documented above: insurance accelerates consultation when symptoms first arise, potentially preventing complications. These health effects are not detectable using traditional self-reported health status collected at endline, in line with prior studies that failed to identify health improvements using standard recall-based instruments. The discrepancy between our weekly and endline results suggests that existing doubts about whether insurance improves health may reflect measurement challenges, insufficient sample sizes, or indicators that do not capture the relevant outcomes, rather than genuine absence of health effects.

Our findings first contribute to a large literature on the demand for health insurance in LMICs. The magnitude of take-up we identified is consistent with a few studies finding that full subsidies substantially increase enrollment (e.g. Asuming et al., 2024; Chemin, 2018), but differs from the more modest effects obtained from (full or partial) subsidies, information provision or registration assistance alone (e.g. Thornton et al., 2010; Banerjee et al., 2021; Wagstaff et al., 2016; Capuno et al., 2016). We believe that pairing subsidies with hands-on enrollment support was central to achieving near-universal take-up in our setting. The mobile-based design of the intervention, reflecting its origins in a broader digital innovation agenda, may have further facilitated this support. Our findings suggest that similar investments in accessible, hands-on registration processes could be a requisite for achieving the health benefits of coverage elsewhere.

Second, our findings speak to a large literature on the health impacts of health insurance. Several studies document reduced health expenditures and increased health care utilization, yet fail to identify health improvements (e.g. Bauhoff et al., 2011; Fink et al., 2013; King et al., 2009; Levine et al., 2016). The few studies that do identify health gains rely on administrative data with very large sample sizes, and exploit exogenous variation in nationwide programs – Sood and Wagner (2018) link insurance to reduced cardiac and cancer mortality in India, while Camacho and Conover (2013) and Balsa and Triunfo (2021) find infant health improvements in Colombia and Uruguay, respectively. That these health effects are only detected with large administrative data sets suggests that empirical issues, such as sample size and measurement errors, are central to the difficulty in precisely identifying health effects. Our study combines a strong first-

stage with high-frequency measurement, allowing us to detect health improvements that standard evaluations of this scale would likely miss.

Last, our weekly survey data contribute to a growing literature using high-frequency data in impact evaluations, in three ways. First, collecting data weekly over 18 months generates a sufficiently large estimation sample to detect effects that would be underpowered in a standard ex-post survey. More broadly, collecting data at high frequency can substitute for large sample sizes when outcomes are noisy or rare (McKenzie, 2012), offering a practical alternative where increasing sample size is costly or infeasible.

Second, high-frequency measurement captures health events, care-seeking decisions and financial transactions that respondents are likely to forget or misreport in retrospective surveys with long recall periods – a concern well documented for health shocks, care-seeking and health expenditures (Das et al., 2012; Nelissen et al., 2020; Geng et al., 2018), sexual behaviors (Cassidy et al., 2021), agricultural labor (Arthi et al., 2018; Deininger et al., 2012) and microenterprise profits (de Mel et al., 2009). Our setup captures all, or at least substantially more health events than traditional surveys. We observe the full set of health problems in the sample population, together with corresponding decisions of whether and when care was sought, at which provider, how much was paid, and whether formal visits were covered by insurance. For financial protection outcomes, our estimates obtained with standard endline data using 12-month recall are qualitatively and quantitatively similar to those from weekly surveys. This is consistent with the existing literature documenting expenditure reductions across diverse insurance products and recall periods. However, the advantage of high-frequency data becomes clear for care-seeking and health outcomes, where endline data with long recall periods fail to detect effects that weekly measurements reveal.

Finally, weekly data enable measuring outcomes that exploit the very granularity of the diary format – an approach with applications beyond our setting, including dynamic effects of cash transfers on agricultural productivity (Aggarwal et al., 2019), or labor supply responses to household shocks (Heath et al., 2022). In our context, we use individual illness episodes to

identify a decrease in self-medication, and matched illness episodes to corresponding care-seeking to estimate an increase in the speed of seeking care following illness onset. Weekly health data confirm a decrease in the severity of illnesses or injuries, as captured by the number of days ill or injured respondents are unable to perform their daily activities.

In sum, our paper demonstrates that frequently collected surveys, combined with successfully incentivized take-up, can detect health improvements that more standard, less frequent surveys are likely to miss. We conclude that health improvements, in addition to financial protection, can justify expanding subsidized health insurance. Future evaluations of health insurance, and of other interventions with delayed health effects, may want to consider this approach where feasible.

The remainder of this article is organized as follows. Section 2 presents background information on the study area and the intervention we study. Section 3 describes data sources and variables of interest. Section 4 presents the empirical strategy, and Section 5, results. Section 6 concludes.

2 Context and intervention

2.1 Setting

Despite major improvements over the past decades, Kenya has been experiencing persistently high geographical and socio-economic inequities in health and access to care (e.g. Achoki et al., 2019). Not only do poorer populations display worse health than wealthier groups, but they also face stronger barriers in accessing a host of needed care services (Ilinca et al., 2019), and have a lower propensity to seek care at formal health facilities when experiencing health problems (Awiti, 2014) – facilities that tend to be of lower quality (Toda et al., 2012). The incidence of catastrophic health payments is also greater for the most vulnerable in rural areas, mainly because of outpatient services (Salari et al., 2019).

As an essential step to address such persistent disparities and ensure care is provided to those in need, Kenya’s Government initiated efforts in 2017 to improve affordability of quality care for all in

order to reach universal health coverage (UHC) by 2022. In fact, in 2014, most Kenyans aged 15 to 49 years old did not have any health insurance (82% of women and 79% of men); among those who did, the National Health Insurance Fund (NHIF) was the most common scheme, at a low 14% for women and 18% for men (Kenya National Bureau of Statistics et al., 2015). It mainly covered civil servants, other formal sector workers, and a limited number of low-income households through the health insurance subsidy for the poor (HISP) program, targeting orphans and other vulnerable children. Most informal sector workers did not enroll in the NHIF voluntary insurance scheme with monthly household premium of KES 500 (Barasa et al., 2018). UHC was therefore included as one of Kenya's Government 'Big Four Agenda' action points (Wangia and Kandie, 2019).

2.2 The Innovative Partnership for Universal Sustainable Healthcare (i-Push) program

To support the Kenyan government in its efforts towards reaching UHC, the non-governmental organizations Amref Health Africa and PharmAccess Foundation implemented the Innovative Partnership for Universal Sustainable Health care (i-Push) program since 2017. Being a multi-component initiative, i-Push aimed at empowering low-income women of reproductive age and their family members by enhancing their access to health care through innovative digital tools (Abajobir et al., 2021).

Its first component, which is the focus of our study, consisted of a fully subsidized digital health insurance product. The program provided free access to the SupaCover scheme of the NHIF for 12 months – a comprehensive package of out- and inpatient health care services at NHIF-empaneled public and private providers.² Households were eligible if they were living in a low-income community and included at least one woman of reproductive age, who was either pregnant or with a child of 4 years old or less. Program staff first identified households belonging to the

² NHIF SupaCover covers outpatient and inpatient curative care, maternal health services, immunizations and drugs at providers contracted by the NHIF. For outpatient services, there is no co-payment. For inpatient services, co-payments depend on the status of the contracted hospital. Government hospitals never charge co-payments. Coverage extends to NHIF members, their declared spouse and children below 18 years of age (<https://hosi.co.ke/nhif-guide/>).

target population with the aid of community health workers (CHWs), who have good knowledge of their assigned catchment area of approximately 100 households. Women in identified households were subsequently visited by an i-Push agent, accompanied by their respective CHW to facilitate introductions. Agents would give information about the digital insurance to each woman, i.e. the offer of a 12-month fully subsidized NHIF SupaCover for her and her family members, and assist interested women with the digital registration procedure. The insurance was registered on women's own SIM cards with the aim of enhancing women's agency. If a woman needed more time to consider the offer, to get a SIM card, or to collect the necessary documents, a second visit was planned. As an NHIF requirement, an ID-card or birth certificate was necessary to register adult and child household members, respectively. Since a considerable number of households did not have a birth certificate for their child, the program staff also assisted in obtaining such certificates from the county authorities. The agent uploaded the required documents, took photographs and completed the registration process via mobile phone during their visit to the households. This eliminated the need for women and their family members to engage in complex paperwork and repeatedly visit insurance offices to submit necessary documents, thereby considerably reducing logistical and administrative barriers to enrollment. Once NHIF had approved their digital registration, coverage was activated to start on the first day of the following month for a duration of one year. Participants also received a physical NHIF card. To use care at the NHIF-empaneled health care facilities, program enrollees could show their registration on their mobile phone or present their NHIF card. Women enrolling in NHIF were requested to choose one of the NHIF-empaneled providers as their preferred facility.³

The second component of i-Push was targeted at CHWs. The program enhanced their training using a mobile phone-based training tool that focused on raising awareness about health insurance and improving reproductive, maternal, newborn and child health knowledge. It provided information, for instance, on the importance of timely antenatal care, danger signs during

³ NHIF enrollees are required to attend their preferred health care provider for outpatient care, but are allowed to visit other NHIF-empaneled providers for inpatient or emergency care or upon referral. Every quarter, enrollees can update their preferred provider.

pregnancy, nutrition requirements for children under five and family planning. The improved CHW training in turn was expected to enhance health knowledge and behavior of targeted women through their contacts with the CHWs.

The third component focused on health care facilities. The program upgraded the quality of care in selected health care providers following the SafeCare approach, a standards-based stepwise quality improvement methodology (Johnson et al., 2016), accredited by the International Society for Quality in Health Care External Evaluation Association (ISQua-EEA). SafeCare consists of a professional quality assessment, after which facilities receive technical assistance in the development and implementation of a quality improvement plan (QIP). Health care facilities did not receive resources directly. They were enrolled on a digital health platform to facilitate the collection of data on health care services and reimbursements from NHIF. Program agents encouraged women to select as preferred provider one of the facilities that had participated in the SafeCare quality improvement program, although this was not compulsory.

Our study focuses solely on the effects of the 12-month subsidized, digital health insurance coverage combined with enrollment support (i.e. the first component, henceforth the intervention), since the improved training of CHWs was rolled out in the entire study area, and all households – regardless of their treatment or insurance status – could benefit from the quality upgrade in the selected clinics. The latter two components should thus be viewed as the context in which the intervention took place.⁴

3 Study design and methodology

3.1 Experimental design and sampling

Our study focuses on Khwisero, a sub-county of Kakamega County, Western Kenya, where i-Push was expanding in 2019. The program invited four health care facilities in Khwisero, which were

⁴ The program also sought to stimulate savings for co-payment of next year's health insurance premium through a digital savings component ('health wallet'), running on people's mobile phone, using the M-TIBA platform. In our study area, the health wallet was not promoted to participant households; it is therefore not included in our study.

empaneled on NHIF, and served low-income communities, to participate in the SafeCare quality improvement process.⁵ We conducted a matched-pair cluster-randomized controlled trial (cRCT) to evaluate the impact of the mobile phone-based health insurance coverage on health outcomes, health care utilization and health expenditures. Randomization occurred at the village level. The study employed a two-stage sampling methodology.

First, we randomly selected 24 villages, six each from the catchment areas of the four selected health care facilities. Second, in each village we conducted a complete listing of households. From this list, eligible households were identified in line with the program target criteria. Households eligible for the study included those with at least one woman of reproductive age (aged 15-49),⁶ who (i) had at least one child of 4 years old or below living with her; or (ii) was pregnant. In each village, ten eligible households were randomly selected to be included in the study sample to reach a baseline sample size of 240 households – 120 treatment and 120 control households – corresponding to more than 1,300 individuals.^{7,8} Selected health care providers and location of sampled household are displayed in Figure 1.⁹

The study started with a baseline survey conducted among all sampled households, as well as a village survey with key informants. This information was used to match villages in 12 pairs as similar as possible to each other (Imai et al., 2009), on village and aggregated household

⁵ Two were private primary health centers and two were public secondary hospitals. While public health care at primary facilities was provided for free, public secondary facilities were allowed to charge fees.

⁶ Women below 18 years old were excluded as target women from the study sampling frame because of ethical requirements, except pregnant girls aged 15 years old and above who were considered emancipated minors by the Kenyan government and ethics bodies.

⁷ Additional eligible households were sampled as replacement in case of refusals or dropouts in each village.

⁸ Initially, the study sought an even allocation between households with a pregnant woman and households with a child under 4 years old. After the household listing exercise, it became clear that there were too few pregnant women in each village to fulfill this criterion. As a result, all pregnant women were included in the study sample up to five per village, and we randomly sampled additional households with children under 4 years old until the cluster size (10 households per village) was reached.

⁹ Contrary to original plans, only two of the four selected health care providers enrolled in the SafeCare quality improvement program and the M-TIBA digital health platform. One private facility was not included because of organizational challenges; another private facility was dropped because it was located 500 meters away from the public facility, with overlapping catchment areas (see Figure 1). Households in the treatment communities were therefore encouraged to choose one of the two remaining program-selected health care providers as their preferred NHIF provider. The possible role geographical accessibility plays in using health care services is explored when conducting heterogeneity effect analyses in Annex Table D.

characteristics.¹⁰ Within pairs, villages were randomly assigned to either treatment or control. Randomization was blocked at the health facility level to ensure that each health facility served an equal number of treatment and control villages. Each village was thus matched with one of the other five villages in the catchment area of its nearest health facility.

Assignment to the treatment or the control group was carried out by the research team during a public ceremony in the presence of key stakeholders, local liaison persons and village representatives on March 12, 2020, a few days before the onset of the Covid-19 pandemic in Kenya.¹¹ The roll-out of the digital insurance component, initially scheduled to start shortly after the public randomization ceremony, was delayed until June 2020 due to the stringent Covid-19 lockdown measures implemented from March 15, 2020, onwards. The 12 treatment villages were visited sequentially, starting in June 2020. Enrollment of individual households in treatment villages was completed by the end of September 2020.

After the baseline survey, but before the roll-out of the intervention, 40 households (27 in the treatment and 13 in the control group) had dropped out, mainly due to Covid-19-induced relocation. In addition, 11 target women in the treatment group did not have a valid national ID-card, a requirement to obtain NHIF coverage. These women and their household members as well as the attrited households were replaced with households from the replacement list.¹² Both treatment and control samples were thereby replenished in August 2020; the newly included households participated in a short version of the baseline survey. Mean differences between attritors, replacement and retained households in August 2020 are minimal (see Annex Tables A1 and A2). Households that dropped out of the sample before August 2020 are excluded from

¹⁰ Pairs were constructed based on Mahalanobis distance matching. Matching indicators included village (demographic, infrastructure, health service availability) and baseline household indicators aggregated at the village level (female educational attainment, share of women earning income, membership of savings groups for women, mobile phone ownership for women, wealth, health care utilization, health expenditures, health insurance).

¹¹ Consent for the procedures was obtained from local government officials before the random assignment. During the randomization ceremony, papers with paired village names were folded and put in a bag; two village representatives from each paired village discussed who would pick the paper. After the other group members verified that the names could not be seen, one paper was picked, and a coin was flipped to decide which group the selected village belonged to. The process of choosing the folded paper and flipping the coin was repeated for each village pair.

¹² While the i-Push team assisted households in getting birth certificates for their children to enroll in NHIF, it was not possible to assist adults in getting a national ID card.

the analysis. After replenishment, the total study sample included 243 households (123 treatment and 120 control households). A detailed description of the sampling methodology and sample size is provided in Annex Figure A.

3.2 Data collection

This study uses three data sources collected over the study period from October 2019 until June 2021: (1) a baseline household survey, (2) weekly financial and health surveys (diaries), and (3) an endline household survey (see Figure 2). The baseline survey was collected in October-November 2019, gathering information on household demographics, socio-economic indicators, food consumption, financial inclusion, participation in community networks, as well as self-reported health status, health care utilization and health expenditures of all household members in the year prior to the survey. The endline survey was conducted in June 2021, including the same information as the baseline survey.

Two weeks after baseline data collection was completed, we began collecting weekly financial and health diaries for a total of 18 months, from December 2019 until May 2021. The financial diaries recorded all financial transactions in the seven days prior to each interview – income, expenditures, loans, gifts and remittances given and received and (dis)savings. Respondents to financial diaries encompassed all economically active adult household members able to participate in the interviews, thus excluding very old-aged or disabled members and young adults still in school and not working for income. Men and women were interviewed separately and in private about their own financial transactions.

Health diaries collected detailed information on all illnesses and injuries, as well as health care utilization in the seven days prior to each interview. Collected data included date of onset of the symptoms, symptoms of illnesses and injuries, whether any (curative or preventive) health care was sought, which (formal or informal) health provider was visited, what health services were received, out-of-pocket health expenditures and date of provider visit(s). Financial diaries respondents reported on their own health as well as on other adult respondents who were absent

during the interview. An informed adult, usually the mother, responded for children, the old-aged and the disabled. Thus, we have complete health data for a household if at least one adult respondent was present during the interview.

Each interviewer visited the same twenty households every week over the entire study period. The in-person weekly financial and health diaries interviews lasted on average 15 minutes per respondent.¹³ To allow for ‘on-the-job’ learning for interviewers and respondents, we do not include the first six weeks of weekly data collection. The analysis uses weekly data from February 2020 onwards.¹⁴

3.3 Key variables

We use four sets of outcome variables relating to: (i) enrollment in health insurance, (ii) financial protection, (iii) health care-seeking behaviors, and (iv) health. Below, we describe the construction of these variables.

Enrollment in NHIF is recorded at the individual level at baseline and endline, specifying whether an individual is currently insured, and if so, since when. We transform this information into a monthly individual-level indicator taking on unity if an individual was enrolled in NHIF in a certain month.¹⁵ This is our preferred enrollment variable. We also construct a monthly household-level enrollment variable equal to one if at least one household member was enrolled in NHIF in that month, and zero otherwise.

Financial protection against health risk is measured using two groups of indicators. We start by exploring to what extent the intervention changed the percentage of formal consultations that were partially or fully covered by insurance (either by NHIF or other schemes), acknowledging

¹³ After the first case of Covid-19 was registered in Kenya in March 2020, the government implemented stringent movement restrictions. Fieldwork switched from in-person to phone interviews. Within two weeks, all households without mobile phone had received a basic feature phone from the research team. Due to the built-up rapport and trust between interviewers and respondents during the three months preceding the switch, this transition occurred with minimal disruptions (Janssens et al., 2021).

¹⁴ Diaries collected in December 2020 are excluded from the analysis as many households were absent due to end-of-the-year festivities.

¹⁵ We cannot explicitly distinguish between treated households who enrolled in NHIF themselves or through i-Push.

that the probability of a formal consultation itself may have changed due to the intervention. Next, we analyze a set of financial indicators: individual-level out-of-pocket expenditures (OOPs) for health as recorded in health diaries, our primary outcome indicator for health expenditures. All financial figures are expressed in KES as weekly averages per month, which we calculate across the four diary weeks within each month. This individual-level weekly average per month is our preferred measure of OOPs. We also disaggregate this measure, distinguishing between expenses for preventive versus curative care, and by expenses at formal versus informal providers. In the endline survey, respondents reported the total individual-level health OOPs in the past 12 months. To ease comparison, the endline amount is divided by 52 (weeks) to obtain the average weekly individual-level OOPs.

We then turn to the probability that a household incurs catastrophic health expenditures (CHE). To calculate this measure, we first calculate the household's weekly average OOP expenditures by aggregating individual average weekly OOPs over all household members, recorded in the health diaries, and adding any other health-related spending household members reported in the financial diaries. Next, we define the household's capacity to pay (CTP) as the average weekly non-subsistence spending, which we obtain by subtracting the national poverty line, corrected for household size, from monthly total household expenditures and dividing by four to obtain average weekly figures. It captures the money a household is left to spend after taking care of basic necessities, such as food (Xu et al., 2003). Finally, we define CHE as health OOPs exceeding the 10% threshold of a household's CTP, which corresponds to the Sustainable Development Goal (SDG) target indicator 3.8.2 (World Health Organization and World Bank, 2023; United Nations Statistics Division, 2025).

Health-care seeking behavior is measured as individual-level consultations in the weekly health diaries, both for preventive and curative reasons, and both at formal and informal providers. Preventive care includes, for instance, health check-ups and immunizations; curative care is in response to an illness or injury. Formal providers encompass primary health centers, secondary/tertiary clinics, private doctors or nurses, and pharmacies, while informal providers

include traditional healers and informal drug vendors. Consultations related to reproductive health such as antenatal care visits are studied elsewhere (Abajobir et al., 2024), and excluded from our analyses.

The first health-care seeking behavior measures exploit the granularity of the diaries. For each reported occurrence of being ill or injured, an individual is asked whether they sought care, self-medicated, or forwent care. We also analyze delays in seeking care through the number of days between the first symptom of an illness or injury and the moment an individual sought care for this issue. We continue by calculating, at the individual level, the average weekly number of consultations for any reason at any health care provider within a month. We distinguish between consultations for preventive versus curative care, and at informal versus formal providers. In the endline survey, we collected the total number of consultations at any provider, and disaggregated by formal and informal, as well as the number of times an individual did not seek care when ill or injured, measured over the past 12 months. These measures are divided by 52 (weeks) to obtain the average weekly number of consultations and foregone care for comparison with diaries.

Finally, we measure the impact on *health* by constructing an individual-level indicator of the weekly number of illnesses or injuries averaged over four weeks in a month. We also calculate the number of days per week, averaged over a month, that an individual was unable to do his or her daily activities because of health problems. We do not have comparable data from the endline survey; we only measure whether the individual self-reported to be in (very) good health as a binary variable.

3.4 Description and balance of study population

Table 1.1 presents means and standard deviations of household characteristics measured at baseline for the total analysis sample and by treatment and control group, as well as differences in means, p-values of corresponding t-tests and normalized differences. Table 1.2 displays similar statistics for the values of the above listed outcome variables in pre-intervention (February-May 2020) months, measured with weekly survey data.

Table 1.1 indicates that, at baseline, households had on average 4.9 members, of whom 1.2 were below five years old. Household heads were 37.1 years old on average, and 25.0% of them were women. Most heads, 91.4%, were married, and two thirds had completed primary education or more. Among women, 37.2% had completed at least primary education. While one third of all household members earned income from work, 12.5% of female members did. The vast majority of households, 94.8%, owned a mobile phone at baseline, and 76.7% of female weekly survey respondents had their own phone. Individuals reported a total of 2.5 hospital admissions in the past 12 months, 15.3 outpatient consultations at formal providers, and 15.3 informal visits. Individuals reported having foregone care 3.6 times while being ill or injured. Total reported household annual health care expenditures were KES 8,134, on average.¹⁶

Treatment arms were largely comparable at baseline with similar demographic characteristics, education and employment levels, phone ownership, savings group membership and health indicators. We observe few imbalances. On average, households in the control group have slightly more old-aged household members and lower wealth index. While they were more likely to have foregone care, they were equally likely to have consulted health care providers or incurred OOPs. Normalized differences remain below Imbens and Wooldridge's (2009) 0.25 threshold. In the 12 months prior to baseline, control households reported a significantly higher likelihood of NHIF enrollment (26.3%) than treatment households (12.7%), which reflects an imbalance in coverage between November 2018 and October 2019. While this variable was not used in the experimental design, i.e. to generate matched village pairs, we do not have a good explanation for this sizable difference. Reassuringly, Table 1.2 confirms that, in the months defining the pre-intervention period for the analysis, that is, from February to May 2020, the sample is no longer imbalanced on insurance status at conventional levels, with a difference in enrollment between control and treatment households of 23.9% versus 18.4%, respectively (p-value of 0.065, normalized difference of -0.096). Our estimation approach, detailed below, in which we include lagged outcome variables as explanatory variables, reduces potential remaining bias as a result of baseline imbalances.

¹⁶ As of July 2026, the official exchange rate is USD 1 to KES 129.53 (<https://www.centralbank.go.ke/>).

3.5 Attrition and survey non-response

Of the 243 households in the study sample (123 and 120 in the treatment and control group, respectively), 232 households were retained in the sample until endline (118 and 114 in the treatment and control group). Attrition rates were thus low at 3.3% (1.7% and 5.0% in the treatment and control group, respectively). The difference between treatment arms is not statistically significant, nor selective by baseline characteristics (see Annex Tables A1 and A2). The 232 retained households encompass 1,276 individual household members. They constitute our balanced analysis sample, for which we have baseline and endline data, as well as weekly survey data.

Non-response rates to weekly diaries, reported in Annex Table F, were generally low. At the level of the surveyed household, health diaries had a non-response rate of 2.0% among control households, and 2.5% among treatment households. Similarly, financial diaries had a non-response rate of 1.1% in the control group, and 1.4% in the treatment group. At the individual level, non-response rates were higher, from 15.4 to 19.9%. The large difference between individual and household non-response rates across financial and health diary modules suggests that, mostly, someone in the household answered to weekly survey interviews, but that, often, not all members were present at the time of interview. It is important to note that weekly health diaries were recorded for all household members if at least one adult member was present during the weekly interview. However, financial diary respondents were not asked to substitute for missing household members. As explained above, financial – specifically, expenditure – diaries are used as inputs to calculate household-level health expenditures, capacity to pay and catastrophic health expenditures. While non-response at the individual level might imply mismeasurement of the resulting household-level, Table F confirms that differential non-response between treatment arms is unlikely to threaten the validity of our estimates.¹⁷

¹⁷ While differential non-response to the health and financial modules are statistically significant, this reflects the large number of observations rather than any economically meaningful differences: corresponding normalized differences are far below the 0.25 threshold recommended by Imbens and Wooldridge (2009).

4 Estimation strategy

We first estimate intention-to-treat (ITT) effects by running the following ANCOVA specification:¹⁸

$$Y_{ivm} = \alpha_1 T_v + \alpha_2 \bar{Y}_{i,t=0} + \sum_v \delta_v + \sum_m \gamma_m + v_{ivm} \quad (1)$$

where Y_{ivm} is an outcome of interest for the individual (or household) i residing in village v in a given month m in post-intervention months, from June 2020 to May 2021. The unit of analysis is thus the individual-month (or household-month). T_v is a binary variable taking on unity if individual (household) i is located in a treatment village, and 0, otherwise. $\bar{Y}_{i,t=0}$ is the mean of outcome Y_{ivm} in pre-intervention months, from February 2020 to May 2020, included when available.¹⁹ When running specifications with endline survey data, Y_{ivm} is (mostly) measured over the past 12 months at endline in June 2021, referring to the period from July 2020 to June 2021, and $\bar{Y}_{i,t=0}$ is defined as lagged outcome, measured at baseline, in October-November 2019. $\sum_v \delta_v$ and $\sum_m \gamma_m$ stand for village pair (the randomization stratum) and month-year fixed-effects, respectively. When relevant, distance to the nearest upgraded i-Push clinic, in kilometers, is included as additional regressor, and interacted with the treatment indicator to explore heterogeneity. Standard errors are clustered at the village-level, the level of treatment assignment, throughout the analysis.

The coefficient of interest in specification (1), α_1 , estimates the ITT effect of the 12-month fully subsidized health insurance scheme on outcome Y_{ivm} . It captures the reduced-form, causal effect of living in a village assigned to treatment, i.e. being offered insurance, irrespective of actual NHIF enrollment. To estimate the causal effect of NHIF enrollment on financial protection, care seeking and health outcomes, we use village treatment assignment T_v as an instrument

¹⁸ Following McKenzie (2012), we run ANCOVA specifications that, in our setting, offer more efficiency and statistical power than a difference-in-differences approach.

¹⁹ Baseline values are available for most individual and household-level outcomes, except for seeking care when ill or injured, self-medicating, and the number of days between reported health problems and (matched) seeking care, analyzed in Table 5.

for enrollment in an instrumental variable (IV) framework. In practice, $NHIF_{ivm}$ is a binary variable equal to one if individual i , residing in village v , was enrolled in NHIF in month m , and is instrumented by T_v . In household-level specifications, $NHIF_{ivm}$ equals one if at least one household member was enrolled in month m . The following specifications yield local average treatment effect (LATE) estimates, with β_1 capturing the causal effect of NHIF enrollment among compliers:

$$NHIF_{ivm} = \pi_1 T_v + \pi_2 \bar{Y}_{i,t=0} + \sum_v \delta_v + \sum_m \gamma_m + u_{ivm} \quad (2)$$

$$Y_{ivm} = \beta_1 \widehat{NHIF}_{ivm} + \beta_2 \bar{Y}_{i,t=0} + \sum_v \delta_v + \sum_m \gamma_m + \epsilon_{ivm} \quad (3)$$

Estimating in levels allows us to retain zero-valued observations, and to express the ITT estimate as a fraction of the control mean. For weakly positive continuous variables, such as health spending, however, the proportional effect is often the more policy-relevant measure. Log and inverse hyperbolic sine transformations are problematic – they are not unit-invariant – in the presence of zero values (Chen and Roth, 2024). Instead, we estimate the average proportional treatment effect via Poisson Pseudo- (Quasi-) Maximum Likelihood (PPML), consistent for the conditional mean under weak assumptions and accommodates zeros directly (Santos Silva and Tenreiro, 2006), as below:

$$Y_{ivm} = \exp(\alpha_1 T_v + \alpha_2 \bar{Y}_{i,t=0} + \sum_v \delta_v + \sum_m \gamma_m) v_{ivm} \quad (4)$$

which permits estimating an unit-invariant average proportional treatment effect as:

$$\hat{\theta}_{ITT\%} = \exp(\hat{\alpha}_1) - 1 \quad (5)$$

The advantage of this approach is that the model expresses the treatment effect as a proportional rather than an additive change in expected expenditures. In the next section, we discuss estimates of the average proportional effects based on the ITT estimates using both approaches. We do not run PPML specifications to obtain LATE estimates, as IV estimation in nonlinear models requires that the exclusion restriction holds within the specific functional form (here, exponential), which conflates identification with functional form assumptions in a way that linear IV avoids (Wooldridge, 2014).

The number of clusters in our study, 24, is relatively small, and the cluster-robust variance estimator is likely biased, leading to over-rejection of the null hypothesis. We address this caveat, first, via wild cluster bootstrapping (WCB), by resampling residuals within the regression structure, preserving the village-level clustering. Canay et al. (2021) establish the validity of this procedure with a small number of clusters, provided that the model is correctly specified. Second, we exploit the known treatment randomization to assess robustness via randomization inference (RI), which constructs the reference distribution by enumerating from all possible treatment assignments that could have occurred under our design (Young, 2019). In our setting, within each of the 12 matched village pairs, one village was assigned to treatment and one to control, which implies $2^{12} = 4,096$ possible randomizations. For each possible randomization, we compute the test statistics of interest, e.g. the coefficient on the treatment indicator or on the interaction between treatment and distance to closest i-Push clinic. The corresponding RI p-value is the share of those randomizations that produce a test statistic as extreme as the one we observed with the true treatment assignment.

Given that we estimate effects on many outcomes, we control for the false discovery rate (FDR) within families of related outcomes (Benjamini et al., 2006), and report Anderson’s (2008) sharpened q-values. We group outcomes into families when they constitute distinct but related tests of the same broad hypothesis, so that each rejection is a separate potential discovery. Our families are: formal consultation coverage, health expenditures, health care utilization and health status. We do not apply corrections in two cases. First, for standalone outcomes that constitute a single test of a distinct hypothesis, where no multiplicity arises: NHIF enrollment, total individual OOP

health expenditure, days to consultation and weekly average number of consultations. Second, for sets of outcomes that are mechanically linked, where one is a deterministic function of the others, and rejections are therefore not independent: illness-episode variables and household-level financial-protection measures.^{20,21} When exploring heterogeneity in distance to the closest health facility, we correct the treatment coefficient and its interaction with distance separately, as they test conceptually distinct hypotheses (the effect of treatment assignment versus whether the effect varies with distance).

Our analysis sample focuses on the 193 households out of the 232 with completed endline surveys who have diaries data for the baseline period from February to May 2020. That is, we exclude replacement households who started only in weeks 38/39 of weekly survey data collection. To assess robustness to this choice, we run estimation for the full set of 232 households, imputing zero for missing lagged outcome values, and interacting this variable with an indicator taking on unity if this variable is not missing (and 0, otherwise). Results, reported in Annex E tables, are very similar, so we do not discuss them in the next section.

5 Results

This section presents results estimating the above equations for the outcomes of interest. We first establish that the intervention increased NHIF enrollment. We then turn to financial protection, where effects are most expected. We continue with care-seeking, and end with estimating the impact of the intervention on health. Throughout, we report ITT and (IV) LATE estimates, with standard errors clustered at the village level, corresponding wild cluster bootstrapped p-values,

²⁰ For each reported health event, respondents indicated whether they sought a consultation, self-medicated, or took no action. We report ITT and LATE estimates on the first two responses, that, together with the omitted category (no action), fully characterize the behavioral response to illness. Because these outcomes are derived from a single three-way classification, they are mechanically linked (shifts into consultation necessarily come from self-medication or inaction); as a result, we do not apply multiple hypothesis testing corrections.

²¹ That is household-level total OOP health expenditure, expenditure over capacity to pay, and catastrophic health expenditure, the latter two being transformations of the first.

and sharpened q-values, when applicable. Randomization inference p-values for ITT estimates are presented in Annex Table B.

5.1 Is health insurance take-up successfully encouraged?

Figure 3 shows the share of individuals and households in Panels (a) and (b), respectively, who were enrolled in NHIF in each month during the study period by treatment arm. Individual enrollments in the treatment group increased steadily during the roll-out of the intervention from June to September 2020, reaching a maximum of about 80% by the end of 2020. Enrollment rates in the control group increased slightly from 19 to 23%. This translated into a household enrollment rate of nearly 95% in the treatment group from October 2020 onwards, compared with just above 20% in the control group.

Covid-19 reached the study area during the period of analysis – the intervention was rolled out from June to September 2020, during the pandemic. Importantly, in Figure 3, enrollment trends in control villages show no evidence of pandemic-driven enrollment into NHIF. Control enrollment is essentially flat, rising very gradually from March 2020 till the end of the study period; there is no spike when Covid-19 arrived. Enrollment in treated villages rises sharply from June 2020 onwards. This confirms that Covid-19 did not ‘contaminate’ the first stage, i.e. the first-stage (NHIF enrollment) difference between treatment and control is driven by the intervention we study rather than by differential responses to Covid-19. Our identification relies on within-pair, within-month comparisons, with month fixed effects absorbing common time-varying shocks, including any possible pandemic-related disruptions. To the extent that the pandemic might have increased the value of insurance for treated households, for instance, by lowering barriers to care-seeking during a health crisis, our estimates capture the effect of health insurance during a period of heightened health risk and potential disruption to health services. We view this as a relevant

policy setting rather than a limitation, as health emergency conditions represent a ‘stress test’ for health insurance.²²

Table 2 confirms high compliance with the experimental design: the intervention was clearly successful in encouraging NHIF enrollment in targeted villages. Column (1) of Table 2 indicates that the intervention significantly increased individual take-up by 56.3 percentage points (p.p.) compared to a mean enrollment rate in the control group of 23.3% post-intervention. Household-level enrollment in NHIF similarly responded to the intervention: assignment to treatment led to a significant 60.1 p.p. increase in NHIF take-up, compared to a mean enrollment in the control group of 27.7% post-intervention. Estimates are robust to wild cluster bootstrapping and randomization inference corrections.

Having successfully encouraged take-up implies a first-stage that should be sufficiently strong to identify the impacts of health insurance, if present, which stands in contrast to existing studies. Our results are consistent with Asuming et al. (2024) and Chemin (2018) who find that the provision of full subsidies increased enrollment by, respectively, 51.7 p.p. in Ghana’s national health insurance scheme, from a baseline of 19.7% among control households, and by 45% in Kenya. However, they stand in contrast to interventions that provide partial or full subsidies, information or registration assistance. In Nicaragua, subsidizing insurance for informal sector workers only attracted 20.3% new enrollments (Thornton et al., 2010). Banerjee et al. (2021) studied various measures to increase insurance coverage in Indonesia. They find that subsidies and registration assistance increased enrollment; providing information did not. Specifically, a full subsidy – one of the three treatment arms – increased enrollment by less than 20 p.p., compared to 7.8% of the no subsidy, status-quo registration group, which is far from universal coverage. Partially subsidizing insurance or providing information also proved to be of little impact on enrollment in Vietnam (Wagstaff et al., 2016), even when offering handholding support, such as in the Philippines (Capuno et al., 2016). As these studies suggest that there are substantial non-price barriers to enrollment and health insurance across LMICs, not only does i-Push reduce costs,

²² Note that infectious disease incidence among children in our sample during the Covid-19 pandemic has been reported elsewhere (Gómez-Pérez et al., 2023).

but it also works towards eliminating various – administrative, logistical and trust – barriers to enrollment. The introduction of the initiative and its agents by trusted local community health workers, hands-on assistance with the digital registration procedures at women’s homes, and support in retrieving the necessary documentation such as children’s birth certificates, likely all contributed to this success.

5.2 Does health insurance improve financial protection?

Table 3 presents ITT and LATE estimates for insurance coverage of medical costs for formal consultations.²³ Assignment to the intervention leads to a 10.8 p.p. decrease in not being insured for any of the formal medical costs, compared to a post-intervention mean of 90.1% in control communities, and a 16.7 p.p. increase in costs that were partially or fully covered by NHIF, compared to a control group mean of 10.5%, without altering costs covered by other insurance schemes. Estimates are robust to wild cluster bootstrapping, randomization inference and multiple hypothesis testing corrections. Corresponding LATE estimates indicate a 18.7 p.p. decrease in not being insured for any costs, and a 26.9 p.p. increase in costs being fully or partially covered by NHIF among compliers.

Table 4 presents results for OOP health expenditures, using weekly averages in a month from diaries in Panel A, and weekly averages over a year from the endline survey in Panel B. Corresponding PPML ITT estimates and average proportional treatment effect, $\hat{\theta}_{ITT\%}$, are displayed in Annex Table C. Panel A1 of Table 4 indicates that treatment assignment reduces individual-level weekly average total OOP health expenditures by KES 6.6, an effect robust to small cluster number corrections. Compared to a post-intervention mean of KES 12.6, this implies a reduction of 52.4%. When we account for the prevalence of zeros in expenditures, PPML reduced-form estimates, in Annex Table C, confirm that assignment to the intervention about halves (-51.38%) the average weekly OOP health expenditures. This reduction is driven by reductions in spending in formal and curative care. The estimates in levels for these subcomponents are

²³ This is because only consultations at formal providers could be covered by health insurance.

marginally significant after wild cluster bootstrapping and randomization inference corrections, but do not survive multiple hypothesis testing. However, PPML ITT estimates for formal and curative care are robust to these corrections.

Panel A2 of Table 4 presents estimates for financial protection at the household level using weekly data. It indicates that, household-level average weekly OOP health expenditures significantly decrease due to the intervention. Compared to a post-intervention mean of KES 108.0 among control households, assignment to the intervention decreases spending by KES 48.1, a decrease of 44.5%, equivalent to a 44.6% reduction in household health expenditures when looking at PPML estimates in Annex Table C. The intervention does not significantly alter household CTP – its stability validates interpreting changes in the expenditures/CTP ratio as solely driven by the numerator. Compared to a post-intervention average of 16.1% in the control group, the intervention decreases household-level OOP health expenditures over household CTP by 4.1 p.p.. As a result, the probability of incurring CHE, defined at the 10% of household CTP threshold, drops by 10.3 p.p., from an average of 45.5% post-intervention among control households. The LATE estimate indicates a significant 18.1 p.p. reduction in the probability of incurring CHE.

Panel B of Table 4 (Annex Table C) reports ITT and LATE (PPML ITT) estimates at the individual and household level using standard, 12-month recall data collected at endline, averaged over 52 weeks. Levels of health expenditures are generally higher than those captured by diaries, which suggests that, with a one-year recall, and after having responded to diaries for a year, respondents tend to overestimate their health expenditures. This leads to a greater absolute magnitude of ITT and LATE estimates. Relative declines are, however, similar. Assignment to the intervention decreases individual (household) weekly health expenditures by KES 14.3 KES (KES 68.2) from a post-intervention control mean of KES 27.6 (KES 133.1). This corresponds to a decline of 51.1% (52.4%), similar to those reported for the PPML estimates with weekly data in Panel A of Annex Table C.

PPML estimates obtained with endline survey data are qualitatively similar to those obtained with weekly diaries, which is reassuring. That health insurance is shown to strengthen financial

protection using weekly and endline survey data is in line with existing findings showing a decline in OOP health expenditures using single ex-post data with recall periods ranging from 30 days to 12 months prior to the day of the interview, for various types of insurance products, for instance in Georgia (Bauhoff et al., 2011), Burkina Faso (Fink et al., 2013), Mexico (King et al., 2009) or Cambodia (Levine et al., 2016).

5.3 Does health insurance alter care-seeking behaviors?

In Table 5, we examine care-seeking behaviors. First, in Panel A, we look at the level of the individual illness episodes, where we observe immediate response to each reported illness or injury event, and explore the unique granularity of our data. Second, in Panel B, we look at the individual-month level, where we observe average weekly reported health service utilization. Finally, in Panel C, we provide corresponding estimates using standard, 12-month recall data collected at endline, averaged over 52 weeks.

Using individual illness episodes, Panel A1 indicates that insurance affects the very response to illness in the first place. Assignment to treatment increases seeking care when sick by 5.7 p.p., and 11.1 p.p. among compliers, although statistical significance does not survive wild cluster bootstrapping nor randomization inference corrections. On the other hand, we precisely identify an 8.7 p.p. decrease (17.0 among compliers) in self-medicating, robust to these adjustments.

Panel A2 presents estimates using illness episodes matched with corresponding care-seeking to inform on timing. Conditional on someone reporting to be sick, the intervention reduces the number of days until seeking care by 0.8 days from a post-intervention control mean of 3.3 days. While this estimate does not survive wild cluster bootstrapping correction, PPML estimates in Annex Table C show that the intervention reduces the estimated number of days to seek care by 23%, *ceteris paribus*,²⁴ confirming that insurance enhances timely care-seeking.

²⁴ As a significant proportion of respondents consult on the same day, we present PPML ITT estimates in Annex Table C, to account for the outcome having substantial mass at zero. PPML estimates are our preferred specification for this outcome, as it respects the count-like structure of the data. OLS estimates, reported for comparison, yield point estimates in the same direction, but do not reach statistical significance at conventional levels, consistent with the known inefficiency of linear models for highly skewed, zero-inflated outcomes.

The episode-level results show that insurance changes how individuals respond to health problems. In Panel B, we examine whether these behavioral responses translate into changes in health care service utilization. On average, the intervention does not significantly affect utilization at any type of provider for individuals living in treated communities. We even observe a negative coefficient estimate for preventive care. A priori, this stands at odds with the few studies that find a causal, positive effect of health insurance on preventive care utilization with concomitant improvements in health.²⁵ The negative sign on preventive consultations we observe is likely explained by the fact that the subsidized health insurance premium, NHIF SupaCover, covered curative, rather than preventive, care services – the intervention might have increased the relative price of preventive compared to curative care (Cabral and Cullen, 2017).

Interestingly, in Annex Table D, we show that health care service utilization depends on how far beneficiaries live from the closest i-Push clinic. Assignment to the intervention does significantly increase the weekly average number of consultations in a month, but this effect diminishes with distance to the closest facility. This is likely driven by monetary and non-monetary time and transportation costs (Dupas and Jain, 2024).²⁶

When running specifications with endline survey data, in Table 5 Panel C, we cannot precisely identify an effect on seeking care at formal providers. Consultations at informal providers decreased significantly, by 0.04, compared to a post-intervention control group mean of 0.10 consultations on average in a week. We note that the estimated coefficients on consultations all have a negative sign, contrasting with estimates obtained with weekly diaries. Still, looking at foregone care, we obtain effects that are qualitatively similar to the increased probability of seeking care when ill or injured in Panel A1. Assignment to the intervention reduces the weekly average number of times not seeking care when ill or injured by 0.01, which is robust to wild cluster bootstrapping, randomization inference and multiple hypothesis testing adjustments.

²⁵ See, for instance, identified reduced cardiac and cancer deaths in India (Sood and Wagner, 2018), or improved infant health in Colombia (Camacho and Conover, 2013; Miller et al., 2013) and Uruguay (Balsa and Triunfo, 2021), explained by greater use of preventive care resulting from access to health insurance.

²⁶ We do not observe such patterns for the remaining primary outcome variables.

5.4 Does health insurance improve health outcomes?

Last, we turn to the most ‘disputed’ claim – whether health insurance improves health per se. Panel A of Table 6 reports ITT and LATE estimates using diaries for the weekly average number of reported illnesses or injuries in a month, and for the weekly average number of days a respondent was not able to do daily activities because of such health issues in a month. While, on average, the intervention does not significantly affect being ill or injured, the intervention leads to a statistically significant reduction of 0.04 days per week being unable to do daily activities, from a post-intervention average of 0.11 days in control communities. This corresponds to a drop of 34%, suggesting that illness episodes have become less severe, with less impact on activities such as work or chores. This is likely explained by the reduced time lapse between symptom onset and seeking care, as well as the decline in foregone care. Thus, health insurance in our setting leads to faster curative care consultations soon after individuals feel ill or got injured, when symptoms may still be minor, potentially preventing them from turning into more severe complications.

Using endline observations, in Panel B, we do not find any significant health effects, as measured by self-reported (very) good health, even though coefficient estimates are positive. Our inability to precisely identify health improvements with standard endline data is in line with existing studies relying on traditional ex-post data, finding no health impact despite significant improvements in financial protection and care seeking. Examples are Bauhoff et al. (2011) who use ex-post data on outpatient and inpatient services with recall periods ranging from 30 days to 12 months; King et al. (2009) who explore nine distinct health self-assessments based on a 10-month recall period; Fink et al. (2013) and Levine et al. (2016) who look at acute illness episodes in the 30 days prior to the day of survey interview. The results obtained with weekly collected data are not consistent with those produced with ‘one-time’ ex-post measures dependent on substantial recall periods. This suggests that existing doubts regarding the presence of health impacts are, at least to some extent, explained by empirical challenges. Insufficiently large sample sizes leave studies underpowered; and the limited set of common indicators collected in surveys might not adequately measure the (health) outcomes that are expected to improve. Collecting data on health events,

care-seeking and health outcomes through more frequent instruments, such as our weekly diaries, lowers measurement error due to recall bias, particularly present among the poorest (Das et al., 2012).

6 Conclusion

We study the effects of enrollment in Kenya’s National Health Insurance Fund using a cluster-randomized control trial, combined with weekly survey data collected over 18 months. The intervention offered low-income women and their families fully subsidized, mobile phone-based health insurance for 12 months, with hands-on registration assistance delivered by trusted community health workers and local agents. The intervention achieved near-universal coverage in treated villages: individual (household) enrollment increased significantly by 56.3 p.p. (60.1 p.p.), from a pre-intervention control mean of 19.1% (23.9%). This far exceeds the take-up achieved by either subsidy or registration assistance alone in Indonesia (Banerjee et al., 2021), Vietnam (Wagstaff et al., 2016) or the Philippines (Capuno et al., 2016). The success of this intervention likely reflects the combination of full subsidies with active support in navigating administrative and logistical barriers to enrollment.

High enrollment translated into substantial impacts on a broad range of outcomes. Insurance reduced out-of-pocket health expenditures by 51.4%, lowered the probability of catastrophic health spending by 10.3 p.p., shifted responses to illness from foregone care and self-medication toward formal care, accelerated timely consultation after illness onset, and reduced the impact of health problems on people’s ability to perform their daily activities. The estimated care-seeking and health effects are observed in weekly diaries, but not in traditionally collected endline surveys with 12-month recall, consistent with the idea that standard surveys tend to miss health events and care-seeking dynamics due to recall bias, insufficient sample sizes, and indicators that do not capture the relevant outcomes.

We draw three implications from our findings. First, full subsidies combined with trusted, hands-on registration support are effective in overcoming the multiple barriers to health insurance enrollment in low-income settings. Second, not only does health insurance strengthen financial protection, but it also improves health, providing further justification for the expansion of subsidized coverage. Third, high-frequency data collection throughout the study period may be essential to detect impacts that standard surveys are likely to miss, with direct consequences for how program effectiveness is evaluated, and policy decisions are informed.

An important question beyond the scope of this study is whether the effects we document are sustainable. Our intervention provided fully subsidized coverage for 12 months. Whether households renew their enrollment once subsidies end is critical for policy. The ‘experience’ of being insured – faster access to care, lower spending, less severe health episodes – may change how people value coverage, with insurance functioning as an experience good, increasing their willingness to pay for insurance premia. Whether the behavioral changes we document persist beyond the intervention period, and whether they translate into longer-term health gains, are open questions that longitudinal follow-up studies could address.

Acknowledgments

We thank the continued support and many fruitful discussions with PharmAccess Foundation and Amref Health Africa and their i-Push program staff, and the high-quality data collection efforts of the field research team in Khwisero. Our study participants, who offered us their valuable time and provided so many insights into their lives, deserve our special gratitude. We thank Madi Mangan, Sally Odunga, Sergio Parra Cely, and participants of the 2022 Universitat de Barcelona ‘Workshop on Public Policies: Education and health in resilient times,’ the 2021 Centre for the Study of African Economies (CSAE) Conference, the 2023 Lowlands Health Economists’ Study Group (LolaHESG) Conference, the Rotterdam Global Health Initiative, the 2026 COunterfactual Methods for Policy Impact Evaluation (COMPIE) and the Lausanne Center for Health Economics, Behavior and Policy (LCHE) for useful comments and suggestions.

References

- Abajobir, A., de Groot, R., Wainaina, C., Njeri, A., Maina, D., Njoki, S., Mbaya, N., Donfouet, H., Pradhan, M., Janssens, W., and Sidze, E. (2021). The Impact of i-Push on maternal and child health care utilization, health outcomes, and financial protection: Study protocol for a cluster randomized controlled trial based on financial and health diaries data. *Trials*, 22(1):629.
- Abajobir, A., de Groot, R., Wainaina, C., Pradhan, M., Janssens, W., and Sidze, E. (2024). The Impact of digital interventions on health insurance coverage for reproductive, maternal, newborn and child health services utilization in Kakamega, Kenya: A Cluster randomized controlled trial. *Health Policy and Planning*, 39(10):1007–1021.
- Acharya, A., Vellakkal, S., Taylor, F., Masset, E., Satija, A., Burke, M., and Ebrahim, S. (2013). The Impact of health insurance schemes for the informal sector in low- and middle-income countries: A Systematic review. *World Bank Research Observer*, 28(2):236–266.
- Achoki, T., Miller-Petrie, M., Glenn, S., et al. (2019). Health disparities across the counties of Kenya and implications for policy makers, 1990–2016: A Systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Global Health*, 7(1):e81–e95.
- Aggarwal, S., Aker, J. C., Jeong, D., Kumar, N., Park, D. S., Robinson, J., and Spearot, A. (forthcoming). The Dynamic effects of cash transfers to agricultural households. *American Economic Journal: Applied Economics*.
- Anderson, M. (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.
- Arthi, V., Beegle, K., de Weerdt, J., and Palacios-López, A. (2018). Not your average job: Measuring farm labor in Tanzania. *Journal of Development Economics*, 130:160–172.

- Asuming, P., Kim, H., and Sim, A. (2024). Selection and behavioral responses of health insurance subsidies in the long run: Evidence from a field experiment in Ghana. *Health Economics*, 33(5):992–1032.
- Awiti, J. O. (2014). Poverty and health care demand in Kenya. *BMC Health Services Research*, 14:560.
- Balsa, A. I. and Triunfo, P. (2021). The Effects of expanded social health insurance on young mothers: Lessons from a pro-choice reform in Uruguay. *Health Economics*, 30(3):603–622.
- Banerjee, A., Finkelstein, A., Hanna, R., Olken, B. A., Ornaghi, A., and Sumarto, S. (2021). The Challenges of universal health insurance in developing countries: Experimental evidence from Indonesia’s national health insurance. *American Economic Review*, 111(9):3035–3063.
- Barasa, E., Rogo, K., Mwaura, N., and Chuma, J. (2018). Kenya National Hospital Insurance Fund reforms: Implications and lessons for universal health coverage. *Health Systems & Reform*, 4(4):346–361.
- Bauhoff, S., Hotchkiss, D. R., and Smith, O. (2011). The Impact of medical insurance for the poor in Georgia: A Regression discontinuity approach. *Health Economics*, 20(11):1362–1378.
- Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.
- Cabral, M. and Cullen, M. R. (2017). The Comprehensive effect of insurance coverage on preventive care. *Economic Inquiry*, 55(3):1452–1467.
- Camacho, A. and Conover, E. (2013). Effects of subsidized health insurance on newborn health in a developing country. *Economic Development and Cultural Change*, 61(3):633–658.
- Canay, I. A., Santos, A., and Shaikh, A. M. (2021). The Wild bootstrap with a “small” number of “large” clusters. *The Review of Economics and Statistics*, 103(2):346–363.

- Capuno, J. J., Kraft, A. D., Quimbo, S., Tan Jr, C. R., and Wagstaff, A. (2016). Effects of price, information, and transactions cost interventions to raise voluntary enrollment in a social health insurance scheme: A Randomized experiment in the Philippines. *Health Economics*, 25(6):650–662.
- Cassidy, R., Groot Bruinderink, M., Janssens, W., and Morsink, K. (2021). The Power to protect: Household bargaining and female condom use. *Journal of Development Economics*, 153:102745.
- Chemin, M. (2018). Informal groups and health insurance take-up evidence from a field experiment. *World Development*, 101:54–72.
- Chen, J. and Roth, J. (2024). Logs with zeros? Some problems and solutions. *Quarterly Journal of Economics*, 139(2):891–936.
- Cole, S., Giné, X., Tobacman, J., Topalova, P., and Townsend, R. Vickery, J. (2013). Barriers to household risk management: Evidence from India. *American Economic Journal: Applied Economics*, 5(1):104–135.
- Dalinjong, P. A. and Laar, A. S. (2012). The National health insurance scheme: Perceptions and experiences of health care providers and clients in two districts of Ghana. *Health Economics Review*, 2:13.
- Das, J. and Do, Q.-T. (2023). The Prices in the crises: What we are learning from 20 years of health insurance in low- and middle-income countries. *Journal of Economic Perspectives*, 37(2):123–152.
- Das, J., Hammer, J., and Leonard, K. (2008). The Quality of medical advice in low-income countries. *Journal of Economic Perspectives*, 22(2):93–114.
- Das, J., Hammer, J., and Sánchez-Paramo, C. (2012). The Impact of recall periods on reported morbidity and health seeking behavior. *Journal of Development Economics*, 98(1):76–88.

- de Mel, S., McKenzie, D. J., and Woodruff, C. (2009). Measuring microenterprise profits: Must we ask how the sausage is made? *Journal of Development Economics*, 88(1):19–31.
- Deininger, K., Carletto, C., Savastano, S., and Muwonge, J. (2012). Can Diaries help in improving agricultural production statistics? Evidence from Uganda. *Journal of Development Economics*, 98(1):42–50.
- Dercon, S., Gunning, J. W., and Zeitlin, A. (2019). The Demand for insurance under limited trust: Evidence from a field experiment in Kenya. Working Paper 2019-06, Centre for the Study of African Economies. Oxford: University of Oxford.
- Duku, S. K. O., Nketiah-Amponsah, E., Janssens, W., and Pradhan, M. (2018). Perceptions of healthcare quality in Ghana: Does health insurance status matter? *PLoS One*, 13(1):e0190911.
- Dupas, P. and Jain, R. (2024). Women left behind: Gender disparities in utilization of government health insurance in India. *American Economic Review*, 114(10):3345–3385.
- Fink, G., Robyn, P. J., Sié, A., and Sauerborn, R. (2013). Does health insurance improve health? Evidence from a randomized community-based insurance rollout in rural Burkina Faso. *Journal of Health Economics*, 32(6):1043–1056.
- Finkelstein, A., Hendren, N., and Luttmer, E. F. P. (2019a). The Value of Medicaid: Interpreting results from the Oregon health insurance experiment. *Journal of Political Economy*, 127(6):2836–2874.
- Finkelstein, A., Hendren, N., and Shepard, M. (2019b). Subsidizing health insurance for low-income adults: Evidence from Massachusetts. *American Economic Review*, 109(4):1530–1567.
- Geng, X., Janssens, W., Kramer, B., and van der List, M. (2018). Health insurance, a friend in need? Impacts of formal insurance and crowding out of informal insurance. *World Development*, 111:196–210.

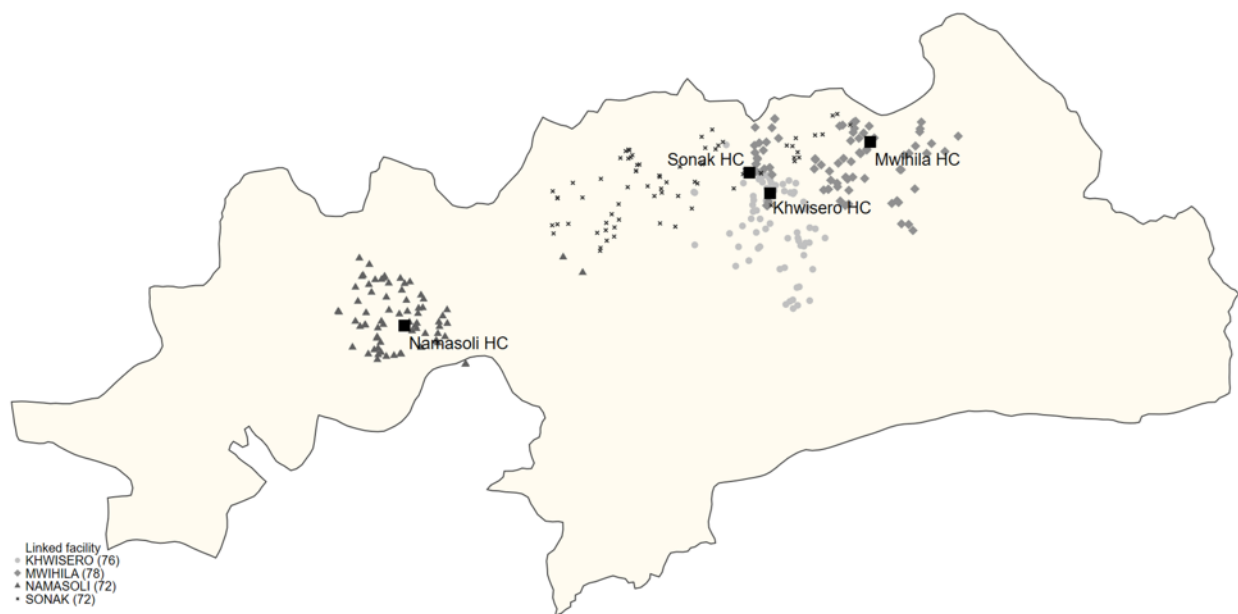
- Gómez-Pérez, G. P., de Groot, R., Abajobir, A. A., Wainaina, C. W., Rinke de Wit, T. F., Sidze, E., Pradhan, M., and Janssens, W. (2023). Reduced incidence of respiratory, gastrointestinal and malaria infections among children during the COVID-19 pandemic in western Kenya: An Analysis of facility-based and weekly diaries data. *Journal of Global Health*, 13:06024.
- Heath, R., Mansuri, G., and Rijkers, B. (2022). Labor supply responses to health shocks. *Journal of Human Resources*, 57(1):143–177.
- Hillebrecht, M., Klonner, S., Sauerborn, R., Sié, A., and Souares, A. (2021). The Demand for health insurance in a poor economy: Evidence from Burkina Faso. *Economic Development and Cultural Change*, 69(4):273–300.
- Ilinca, S., Di Giorgio, L., Salari, P., and J., C. (2019). Socio-economic inequality and inequity in use of health care services in Kenya: Evidence from the fourth Kenya household health expenditure and utilization survey. *International Journal for Equity in Health*, 18:196.
- Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.
- Jain, R. (2024). Private hospital behavior under government insurance: Evidence from reimbursement changes in India. mimeo.
- Janssens, W., Pradhan, M., de Groot, R., Sidze, E., Donfouet, H. P. P., and Abajobir, A. A. (2021). The Short-term economic effects of COVID-19 on low-income households in rural Kenya: An Analysis using weekly financial household data. *World Development*, 138:105280.
- Johnson, M. C., Schellekens, O., Stewart, J., van Ostenberg, P., de Wit, T. R., and Spieker, N. (2016). SafeCare: An Innovative approach for improving quality through standards, benchmarking, and improvement in low- and middle-income countries. *The Joint Commission Journal on Quality and Patient Safety*, 42(8):350–360.

- Kenya National Bureau of Statistics, Ministry of Health/Kenya, National AIDS Control Council/Kenya, Kenya Medical Research Institute, National Council for Population and Development/Kenya, and ICF International (2015). Kenya Demographic and Health Survey 2014. Technical report, Kenya National Bureau of Statistics and others.
- King, G., Gakidou, E., Imai, K., Lakin, J., Moore, R. T., Nall, C., Ravishankar, N., Vargas, M., T.-R. M. M., and Ávila, J. E. H. (2009). Public policy for the poor? A Randomised assessment of the Mexican universal health insurance programme. *The Lancet*, 373(9673):1447–1454.
- Kruk, M. E., Gage, A. D., Arsenault, C., Jordan, K., Leslie, H., Roder-DeWan, S., Adeyi, O., Barker, P., Daelmans, B., Doubova, S., English, M., Garcí-Elorrio, E., Guanais, F., Gureje, O., Hirschhorn, L. R., Jiang, L., Kelley, E., Lemango, E., Liljestrand, J., Malata, A., Marchant, T., Matsoso, M., Meara, J., Mohanan, M., Ndiaye, Y., Norheim, O. F., Reddy, K., Rowe, A. K., Salomon, J., Thapa, G., Twum-Danso, N. A. Y., and Pate, M. (2018). High-quality health systems in the Sustainable Development Goals era: Time for a revolution. *The Lancet Global Health*, 6(11):e1196–e1252.
- Levine, D., Polimeni, R., and Ramage, I. (2016). Insuring health or insuring wealth? An Experimental evaluation of health insurance in rural Cambodia. *Journal of Development Economics*, 119:1–15.
- Lu, F. (2014). Insurance coverage and agency problems in doctor prescriptions: Evidence from a field experiment in China. *Journal of Development Economics*, 106:156–167.
- McKenzie, D. (2012). Beyond baseline and follow-up: The Case for more T in experiments. *Journal of Development Economics*, 99(2):210–221.
- Miller, G., Pinto, D., and Vera-Hernández, M. (2013). Risk protection, service use, and health outcomes under Colombia’s health insurance program for the poor. *American Economic Journal: Applied Economics*, 5(4):61–91.

- Nelissen, H. E., Brals, D., Ameen, H. A., van der List, M., Kramer, B., Akande, T. M., Janssens, W., and van't Hoog, A. H. (2020). The Prominent role of informal medicine vendors despite health insurance: A Weekly diaries study in rural Nigeria. *Health Policy and Planning*, 35(3):354–363.
- Salari, P., Di Giorgio, L., Ilinca, S., and Chuma, J. (2019). The Catastrophic and impoverishing effects of out-of-pocket healthcare payments in Kenya, 2018. *BMJ Global Health*, 4:e001809.
- Santos Silva, J. M. C. and Tenreyro, S. (2006). The Log of gravity. *Review of Economics and Statistics*, 88(4):641–658.
- Sood, N. and Wagner, Z. (2018). Social health insurance for the poor: Lessons from a health insurance programme in Karnataka, India. *BMJ Global Health*, 3:e000582.
- Spenkuch, J. (2012). Moral hazard and selection among the poor: Evidence from a randomized experiment. *Journal of Health Economics*, 31(1):72–85.
- Thornton, R. L., Hatt, L. E., Field, E. M., Islam, M., Solís Diaz, F., and González, M. A. (2010). Social security health insurance for the informal sector in Nicaragua: A Randomized evaluation. *Health Economics*, 19(S1):181–206.
- Toda, M., Opwora, A., Waweru, E., Noor, A., Edwards, T., Fegan, G., Molyneux, C., and Goodman, C. (2012). Analyzing the equity of public primary care provision in Kenya: Variation in facility characteristics by local poverty level. *International Journal for Equity in Health*, 11:75.
- United Nations Statistics Division (2025). Sustainable Development Goals indicators, metadata repository. <https://unstats.un.org/sdgs/metadata/>. Accessed December 15, 2025.

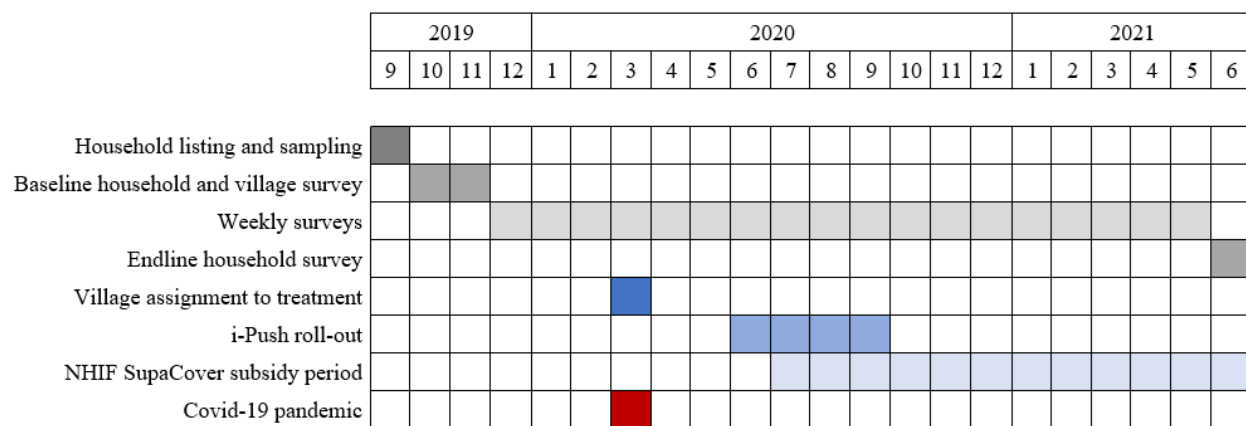
- Wagstaff, A., Flores, G., Hsu, J., Smits, M.-F., Chepynoga, K., Buisman, L. R., van Wilgenburg, K., and Eozenou, P. (2018). Progress on catastrophic health spending in 133 countries: A Retrospective observational study. *The Lancet Global Health*, 6(2):e169–e179.
- Wagstaff, A., Nguyen, H. T. H., Dao, H., and Bales, S. (2016). Encouraging health insurance for the informal sector: A Cluster randomized experiment in Vietnam. *Health Economics*, 25(6):663–674.
- Wangia, E. and Kandie, C. (2019). Refocusing on quality of care and increasing demand for services; Essential elements in attaining universal health coverage in Kenya. Technical report, Ministry of Health, UNFPA, WHO, World Bank, Government of Japan.
- Wooldridge, J. M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics*, 182(1):226–234.
- World Health Organization and World Bank (2021). Tracking universal health coverage: 2021 global monitoring report. Technical report. Washington, DC: The World Bank.
- World Health Organization and World Bank (2023). Tracking universal health coverage: 2023 global monitoring report. Technical report. Geneva: World Health Organization and International Bank for Reconstruction and Development/The World Bank.
- Xu, K., Evans, D., Kawabata, K., Zeramdini, R., and Klavus, J., M. C. J. (2003). Household catastrophic health expenditure: A Multicountry analysis. *The Lancet*, 362(9378):111–117.
- Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, 134(2):557–598.

Figure 1: Study site



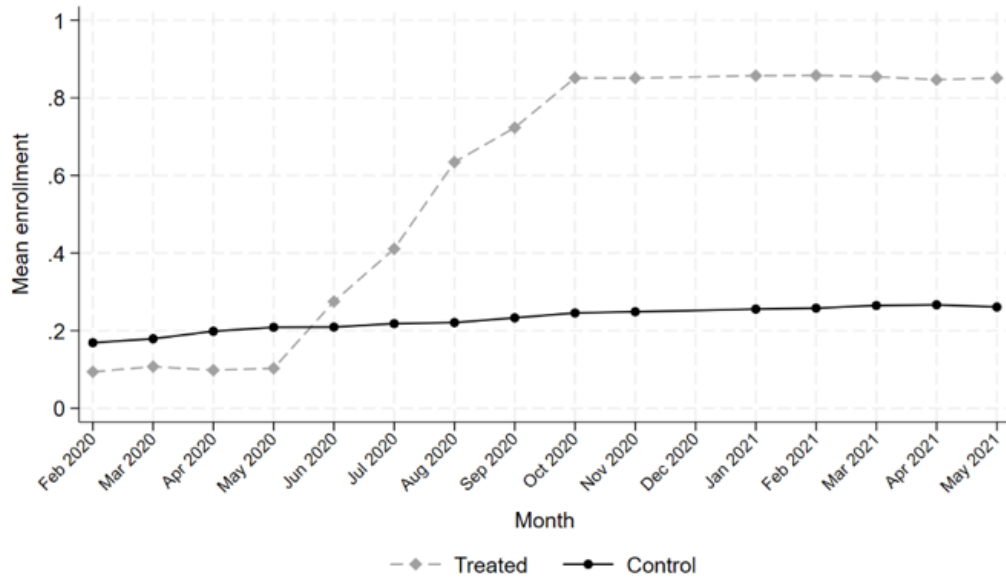
Notes: This figure displays the map of the study site, with GPS locations of study households and originally selected health care providers. Private Sonak Health Center (HC) and private MWhila HC were initially invited but ultimately not included in the i-Push intervention. Private Namasoli HC and public Khwisero HC were included.

Figure 2: Timeline

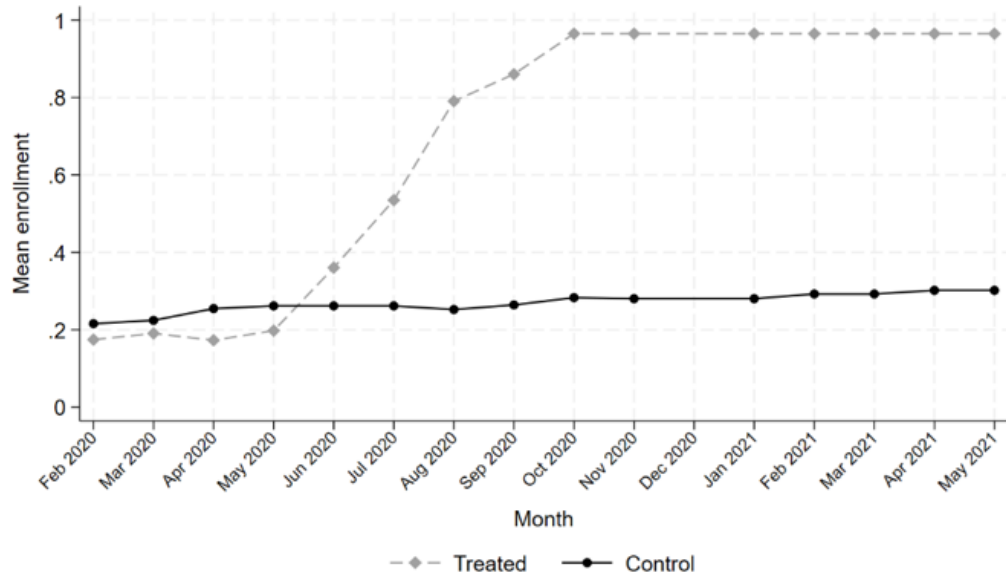


Notes: This figure provides a timeline of the data collection and intervention.

Figure 3: Average NHIF enrollment by treatment status



(a) Individual take-up



(b) Household take-up

Notes: Sub-figures (a) and (b) display the average share of individuals enrolled in NHIF per month, and the average share of households with at least one member enrolled in NHIF by month, respectively, by village treatment status.

Table 1.1: Baseline household covariates by treatment assignment

	Total		Control		Treatment		Difference in means		Normalized differences
	Mean	SD	Mean	SD	Mean	SD	(3)-(5)	p-value	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Number of household members	4.8793	1.732	4.9649	1.804	4.7966	1.662	0.1683	0.4604	-0.0686
Number of members aged 0-4 years old	1.1638	0.58	1.1491	0.598	1.178	0.564	-0.0288	0.7059	0.0351
Number of members aged 5-17 years old	1.6767	1.434	1.6667	1.479	1.6864	1.394	-0.0198	0.9166	0.0097
Number of members aged 18-64 years old	2.0086	0.69	2.0965	0.787	1.9237	0.572	0.1728	0.0564	-0.1776
Number of members aged 65 years old or more	0.0302	0.171	0.0526	0.224	0.0085	0.092	0.0442	0.0496	-0.1821
Head: Age in years	37.0776	10.454	38.0877	11.843	36.1017	8.851	1.986	0.1484	-0.1343
Head: Female	0.25	0.434	0.2368	0.427	0.2627	0.442	-0.0259	0.6509	0.0421
Head: Married	0.9138	0.281	0.9211	0.271	0.9068	0.292	0.0143	0.7001	-0.0358
Head: Completed primary education or higher	0.6379	0.482	0.6228	0.487	0.6525	0.478	-0.0297	0.6393	0.0436
Female members: Completed primary education or higher	0.3723	0.484	0.3509	0.479	0.3932	0.491	-0.0423	0.5084	0.0617
Share of members earning income from work	0.3191	0.314	0.3483	0.337	0.2908	0.289	0.0575	0.164	-0.1295
Share of female members earning income from work	0.1253	0.227	0.1331	0.24	0.1177	0.214	0.0155	0.605	-0.048
Standardized wealth index (PCA)	0	1	-0.1302	0.867	0.1258	1.102	-0.2559	0.0511	0.1825
Household owns any mobile phone	0.9483	0.222	0.9298	0.257	0.9661	0.182	-0.0363	0.214	0.1154
Household owns any mobile phone (diary female respondents)	0.7656	0.425	0.7453	0.438	0.7907	0.409	-0.0454	0.4627	0.0758
Members have savings at savings groups	0.6114	0.489	0.6262	0.486	0.593	0.494	0.0331	0.6408	-0.0478
Female members have savings at savings groups	0.4611	0.5	0.4673	0.501	0.4535	0.501	0.0138	0.8494	-0.0195
At least one member enrolled in NHIF in past 12 months	0.194	0.396	0.2632	0.442	0.1271	0.335	0.136	0.0087	-0.2453
Number of hospital admissions in past 12 months	2.456	4.81	2.2991	4.581	2.6512	5.101	-0.3521	0.6145	0.0514
Number of outpatient visits in past 12 months	15.3316	10.621	16.2056	11.205	14.2442	9.8	1.9614	0.203	-0.1318
Number of informal visits in past 12 months	17.2228	20.114	17.3832	16.55	17.0233	23.92	0.3599	0.902	-0.0124
Number of times forewent care in past 12 months	3.5699	5.657	4.2991	6.236	2.6628	4.719	1.6363	0.0455	-0.2092
Total health expenditures in past 12 months	8133.5603	20016.661	9732.9561	26157.654	6588.3814	11195.035	3144.5748	0.2324	-0.1105
Observations	232		114		118				

Notes: Means and standard deviations (SD) of listed household variables at baseline for impact evaluation sample, control and treated groups, in columns (1)-(2), (3)-(4) and (5)-(6), respectively. Column (7) presents difference in means, and column (8), p-value of the corresponding t-test. Column (9) presents normalized differences.

Table 1.2: Baseline values of outcome variables in pre-intervention (February-May 2020) months by treatment assignment

	Total		Control		Treatment		Difference in means		Normalized
	Mean	SD	Mean	SD	Mean	SD	(3)-(5)	p-value	differences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Individual level variables									
Enrolled (in NHIF)	0.1525	0.36	0.1912	0.393	0.1005	0.301	0.0907	0	-0.183
Not insured for any of the costs*	0.9058	0.292	0.9123	0.284	0.8947	0.308	0.0175	0.5832	-0.042
Costs partially or fully covered by NHIF*	0.0776	0.268	0.0833	0.277	0.0677	0.252	0.0157	0.5927	-0.042
Costs partially or fully covered by other insurance*	0.0388	0.193	0.0395	0.195	0.0376	0.191	0.0019	0.9291	-0.007
Total health OOP expenditures, excluding transportation	7.2696	71.507	6.8198	66.37	7.8725	77.882	-1.0527	0.6776	0.01
Formal consultation	5.2544	70.265	4.2295	64.307	6.6281	77.538	-2.3986	0.335	0.024
Informal consultation	2.0151	13.446	2.5903	16.307	1.2443	8.111	1.3459	0.0047	-0.074
Preventive consultation	1.5355	12.232	1.3494	9.859	1.785	14.829	-0.4356	0.3146	0.024
Curative consultation	5.734	70.205	5.4704	65.06	6.0875	76.583	-0.6171	0.804	0.006
Transportation (for health care)	1.1719	8.634	1.208	8.536	1.1235	8.767	0.0845	0.7822	-0.007
Sought care for reported health problem**	0.4379	0.496	0.4482	0.498	0.4216	0.494	0.0266	0.3144	-0.038
Self-medicated for reported health problem**	0.2726	0.445	0.2566	0.437	0.2979	0.458	-0.0413	0.0821	0.065
Number of days between reported health problems and seeking care***	2.306	4.432	2.5683	4.93	1.7177	2.965	0.8506	0.0755	-0.148
Consulted	0.0451	0.128	0.047	0.125	0.0426	0.132	0.0043	0.3367	-0.024
Consulted at informal provider	0.0138	0.069	0.0159	0.073	0.0109	0.064	0.0051	0.0386	-0.052
Consulted at formal provider	0.0335	0.111	0.0342	0.107	0.0326	0.116	0.0015	0.6938	-0.01
Consulted for preventive care	0.0194	0.083	0.0184	0.075	0.0207	0.092	-0.0022	0.4481	0.019
Consulted for curative care	0.0282	0.101	0.0322	0.104	0.0227	0.096	0.0095	0.0079	-0.067
Illness or injury	0.0906	0.188	0.0969	0.193	0.0823	0.18	0.0146	0.028	-0.055
Number of days unable to do daily activities because of illness or injury	0.07	0.248	0.0659	0.23	0.0754	0.271	-0.0095	0.279	0.027
Observations	3260		1867		1393				
Panel B: Household level variables									
At least one household member is enrolled	0.2148	0.411	0.2393	0.427	0.184	0.388	0.0554	0.0651	-0.096
Total health OOP expenditures	67.5331	234.996	70.8877	246.327	63.3325	220.279	7.5552	0.6602	-0.023
Capacity to pay	559.2368	865.29	585.062	881.544	526.8978	844.686	58.1642	0.3579	-0.048
Total/capacity to pay	0.1188	0.231	0.1098	0.206	0.1301	0.259	-0.0202	0.2312	0.061
Catastrophic health expenditure	0.2688	0.444	0.2773	0.448	0.2582	0.438	0.0191	0.5562	-0.03
Observations	759		422		337				

Notes: Means and standard deviations (SD) of listed outcome variables in pre-intervention (February-May 2020) months as defined in running ANCOVA specifications, in columns (1)-(2), (3)-(4) and (5)-(6), respectively. Column (7) presents difference in means, and column (8), p-value of the corresponding t-test. Column (9) presents normalized differences.

*Limited to 361 observations (228 control and 133 treated group observations).

**Individual ill- or injured-health events (908 control and 574 treated group observations).

***Matched ill- or injured-health episodes with corresponding care-seeking (278 control and 124 treated group observations).

Table 2: Enrollment in NHIF

	Pre-intervention control group mean	Post-intervention control group mean	ITT	Observations
	(1)	(2)	(3)	(4)
Enrolled	0.1912	0.2326	0.5633 (0.0295) [0.0000]	10,237
At least one household member is enrolled	0.2393	0.2774	0.6007 (0.0288) [0.0000]	2,117

Notes: Estimates obtained using individual observations in row (1) and household observations in row (2). Column (1) presents the pre-intervention mean of row-specified dependent variables among control villages; column (2) post-intervention mean. Column (3) presents intention-to-treat (ITT) estimates. February-May 2020 is defined as pre-intervention period; June 2020-May 2021 as post-intervention period. Estimated specifications include village-pair and month-year fixed effects. Standard errors are clustered at the village level in parentheses; corresponding p-values obtained via wild cluster bootstrapping are in brackets.

Table 3: Insurance coverage

	Pre-intervention control group mean	Post-intervention control group mean	ITT	LATE	Observations	KP F-statistic
	(1)	(2)	(3)	(4)	(5)	(6)
Not insured for any of the costs	0.9123	0.901	-0.1081 (0.0337) [0.0093] {0.0100}	-0.1874 (0.0577) [0.0088] {0.0090}	719	212.84
Costs partially or fully covered by NHIF	0.0833	0.1048	0.1673 (0.0333) [0.0001] {0.0010}	0.2694 (0.0589) [0.0010] {0.0040}	719	167.23
Costs partially or fully covered by other insurance	0.0395	0.02	-0.0002 (0.0139) [0.9929] {0.4950}	-0.0003 (0.0232) [0.9929] {0.4950}	719	173.34

Notes: Monthly aggregated indicators for formal consultations using weekly individual observations. Column (1) presents the pre-intervention mean of row-specified dependent variables among control villages; column (2) post-intervention. Column (3) presents ITT estimates. Column (4) presents local average treatment effect (LATE) estimates instrumenting NHIF enrollment with treatment assignment. F-statistics of weak IV test corresponding to the specifications estimated in column (4) are presented in column (6). February-May 2020 is defined as pre-intervention period; June 2020-May 2021 as post-intervention period. Estimated specifications include pre-intervention outcome mean as additional covariate village-pair and month-year fixed effects. Standard errors are clustered at the village level in parentheses; corresponding p-values obtained via wild cluster bootstrapping in brackets; sharpened q-values in braces.

Table 4: Health out-of-pocket (OOP) expenditures

	Pre-intervention control group mean	Post-intervention control group mean	ITT	LATE	Observations	KP F-statistic
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Diaries, weekly average in a month</i>						
<i>Panel A1: Individual-level OOPs</i>						
Total OOPs, excluding transportation	6.8198	12.6427	-6.6202 (2.6445) [0.0594]	-13.6589 (5.4717) [0.0588]	10,237	243.04
Transportation	1.208	1.1265	0.0405 (0.2504) [0.9462]	0.0833 (0.5164) [0.9464]	10,237	250.18
Formal consultation	4.2295	8.6719	-5.7186 (2.7158) [0.1097]	-11.8235 (5.6526) [0.1109]	10,237	243.21
Informal consultation	2.5903	3.9708	-0.5522 (0.2891) [0.2056]	-1.1404 (0.5691) [0.1806]	10,237	236.81
Preventive consultation	1.3494	1.8966	-0.5805 (0.3678) [0.3323]	-1.1982 (0.7687) [0.3409]	10,237	249.84
Curative consultation	5.4704	10.7461	-6.0856 (2.6201) [0.0781]	-12.5495 (5.4087) [0.0751]	10,237	243.21
<i>Panel A2: Household-level health expenditures</i>						
Total OOPs	70.8877	108.0007	-48.0976 (12.3994) [0.0024]	-83.318 (21.7459) [0.0027]	2,117	927.51
Capacity to pay	585.062	559.441	16.4071 (49.3661) [0.8254]	28.3719 (85.6317) [0.8262]	2,117	840.11
Total OOPs/capacity to pay	0.1098	0.1606	-0.0405 (0.0102) [0.0108]	-0.0708 (0.0178) [0.0102]	2,117	708.38
Catastrophic health expenditure	0.2773	0.455	-0.1033 (0.0262) [0.0073]	-0.1808 (0.0467) [0.0088]	2,117	798.08
<i>Panel B: Endline survey, OOPs in past 12 months, averaged over 52 weeks</i>						
Individual-level OOPs	25.3591	27.6073	-14.327 (2.5962) [0.0002]	-22.4443 (3.4722) [0.0000]	857	253.84
Household-level OOPs	179.9887	133.1341	-68.205 (14.7369) [0.0013]	-98.1132 (19.3865) [0.0009]	193	387.9

Notes: Estimates are obtained using unique health events in Panel A; weekly average in a month using individual observations in Panel B; and endline yearly data of individual observations in Panel C. Column (1) presents the pre-intervention mean of row-specified dependent variables among control villages; column (2) post-intervention. Column (3) presents ITT estimates. Column (4) presents LATE estimates instrumenting NHIF enrollment with treatment assignment. F-statistics of weak IV test corresponding to the specifications estimated in column (4) are presented in column (6). February-May 2020 is defined as pre-intervention period; June 2020-May 2021 as post-intervention period. Estimated specifications include pre-intervention outcome mean as additional covariate village-pair and month-year fixed-effects except in Panel A specifications that include village-pair and month-year fixed-effects as well as health problem or consultation day-of-the-week fixed-effects. Standard errors are clustered at the village level in parentheses; corresponding p-values obtained via wild cluster bootstrapping in brackets; sharpened q-values in braces.

Table 5: Care-seeking behaviors

	Pre-intervention control group mean	Post-intervention control group mean	ITT	LATE	Observations	KP F-statistic
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Diaries, illness episodes</i>						
<i>Panel A1: Individual ill- or injured-health events</i>						
Sought care for reported health problem	0.4482	0.3885	0.0567 (0.0373) [0.2922]	0.1108 (0.0748) [0.3026]	5,411	296.591
Self-medicated for reported health problem	0.2566	0.3531	-0.087 (0.0294) [0.0529]	-0.1699 (0.0607) [0.0666]	5,411	296.591
<i>Panel A2: Matched ill- or injured-health episodes with care-seeking</i>						
Number of days between reported health problems and seeking care	2.5683	3.307	-0.7624 (0.3998) [0.2217]	-1.3817 (0.7645) [0.2507]	1,505	213.903
<i>Panel B: Diaries, weekly average number of consultations in a month</i>						
Consulted any health care provider	0.047	0.0598	0.0013 (0.0037) [0.8338]	0.0026 (0.0077) [0.8356]	10,237	248.31
Consulted for preventive care	0.0184	0.0223	-0.0049 (0.0036) [0.4269]	-0.01 (0.0074) [0.4218]	10,237	250.43
Consulted for curative care	0.0322	0.0399	{1.000} 0.0043 (0.0044) [0.5240]	{1.000} 0.0089 (0.0093) [0.5348]	10,237	254.69
Consulted at informal provider	0.0159	0.0276	{1.000} -0.0027 (0.0021) [0.3965]	{1.000} -0.0055 (0.0042) [0.3695]	10,237	243.92
Consulted at formal provider	0.0342	0.0354	{1.000} 0.0024 (0.0032) [0.6344]	{1.000} 0.005 (0.0066) [0.6385]	10,237	249.05
<i>Panel C: Endline survey, number of consultations in past 12 months, averaged over 52 weeks</i>						
Consulted any health care provider	0.1364	0.3948	-0.3458 (0.2043) [0.1046]	-0.5422 (0.3196) [0.1017]	855	254.16
Consulted at formal provider	0.0712	0.2945	-0.3073 (0.2057) [0.2985]	-0.4822 (0.3221) [0.2881]	855	254.66
Consulted at informal provider	0.0652	0.1003	{0.1110} -0.0423 (0.0124) [0.0130]	{0.1070} -0.0663 (0.0187) [0.0102]	855	250.87
Times did not seek care when ill or injured	0.0165	0.0168	{0.0410} -0.008 (0.0027) [0.0574]	{0.0320} -0.0126 (0.0037) [0.0310]	857	242.57

Notes: Estimates are obtained using unique health events in Panel A; weekly average in a month using individual observations in Panel B; and endline yearly data of individual observations in Panel C. Column (1) presents the pre-intervention mean of row-specified dependent variables among control villages; column (2) post-intervention. Column (3) presents ITT estimates. Column (4) presents LATE estimates instrumenting NHIF enrollment with treatment assignment. F-statistics of weak IV test corresponding to the specifications estimated in column (4) are presented in column (6). February-May 2020 is defined as pre-intervention period; June 2020-May 2021 as post-intervention period. Estimated specifications include pre-intervention outcome mean as additional covariate village-pair and month-year fixed-effects except in Panel A specifications that include village-pair and month-year fixed-effects as well as health problem or consultation day-of-the-week fixed-effects. Standard errors are clustered at the village level in parentheses; corresponding p-values obtained via wild cluster bootstrapping in brackets; sharpened q-values in braces.

Table 6: Health

	Pre-intervention control group mean	Post-intervention control group mean	ITT	LATE	Observations	KP F-statistic
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Diaries, weekly average in a month</i>						
Illness or injury	0.0969	0.1311	-0.0197 (0.013) [0.3761] {0.2320}	-0.0405 (0.0268) [0.3749] {0.2310}	10,237	250
Number of days unable to do daily activities because of illness or injury	0.0659	0.1149	-0.0391 (0.0125) [0.0351] {0.0760}	-0.0807 (0.0245) [0.0246] {0.0520}	10,237	243.87
<i>Panel B: Endline survey, self-reported health</i>						
(Very) good self-reported health	0.7318	0.6783	0.026 (0.0279) [0.5565]	0.0406 (0.0422) [0.5431]	857	249.76

Notes: Weekly average in a month using individual observations in Panel A; endline yearly data of individual observations in Panel B. Column (1) presents the pre-intervention mean of row-specified dependent variables among control villages; column (2), post-intervention. Column (3) presents ITT estimates. Column (4) presents LATE estimates, instrumenting NHIF enrollment with treatment assignment. F-statistics of weak IV test corresponding to the specifications estimated in column (4) are presented in column (6). February-May 2020 is defined as pre-intervention period; June 2020-May 2021 as post-intervention period. Estimated specifications include pre-intervention outcome mean as additional covariate, village-pair and month-year fixed effects. Standard errors are clustered at the village level in parentheses; corresponding p-values obtained via wild cluster bootstrapping, in brackets; sharpened q-values, in braces.

Annexes

Figure A1: Sampling flowchart

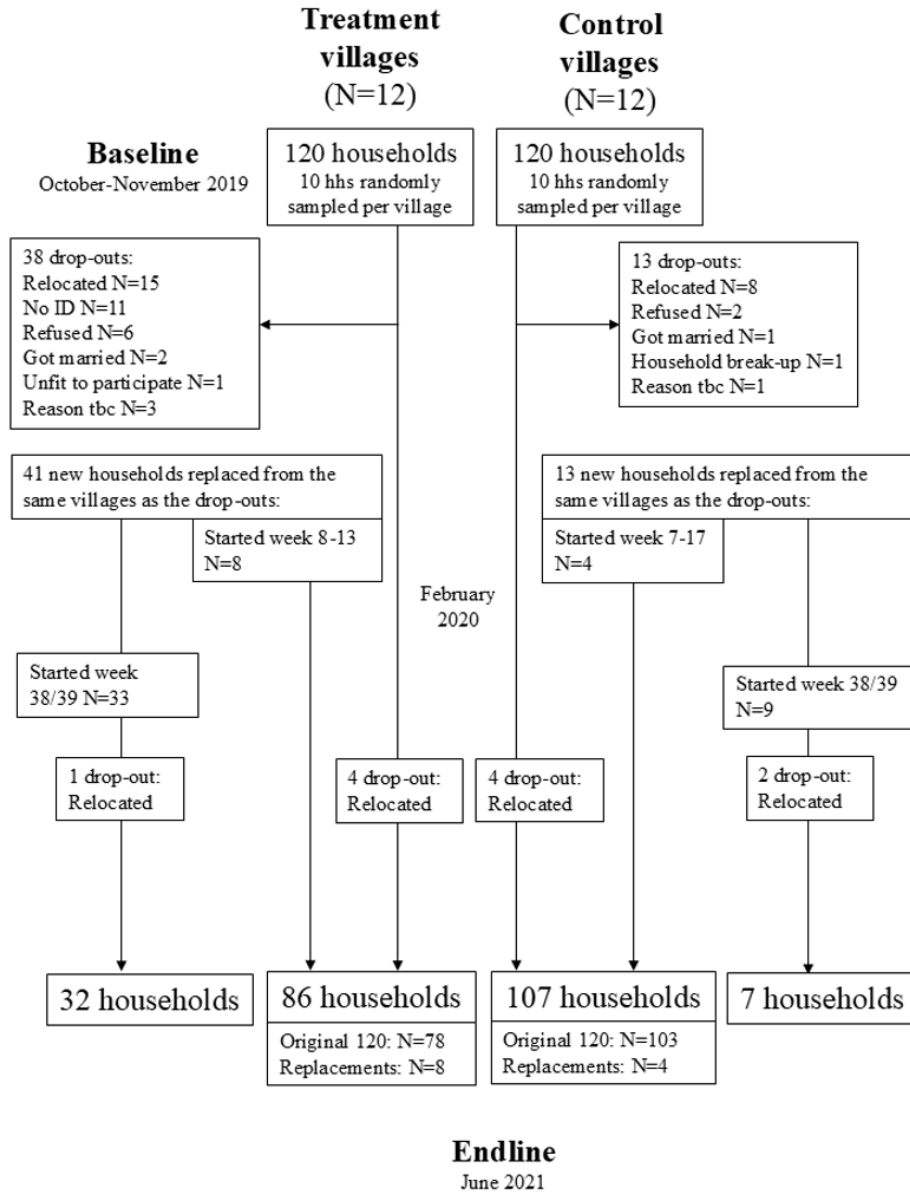


Table A1: Baseline household covariates of original (retained and drop-out) and replacement sample by August 2020

	Retained		Drop-out		Replacement		Difference in means		Normalized	Difference in means		Normalized
	Mean	SD	Mean	SD	Mean	SD	(1)-(3)	p-value	difference	(1)-(5)	p-value	difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Number of household members	4.8995	1.773	4.8824	2.179	4.5185	1.551	0.0171	0.9537	-0.0061	0.381	0.1541	-0.1617
Number of members aged 0-4 years old	1.164	0.61	1.1569	0.674	1.1296	0.436	0.0072	0.9421	-0.0079	0.0344	0.6992	-0.0459
Number of members aged 5-17 years old	1.6614	1.422	1.5686	1.526	1.4815	1.476	0.0927	0.6845	-0.0445	0.1799	0.4171	-0.0878
Number of members aged 18-64 years old	1.6614	1.422	1.5686	1.526	1.8889	0.502	0.037	0.7621	-0.0312	0.1481	0.1604	-0.1681
Number of members aged 65 or more	2.0741	0.74	2.1569	1.007	0.0185	0.136	-0.1198	0.0032	0.261	0.0185	0.5032	-0.0794
Head: Age in years	37.6614	10.811	39.4706	16.262	35.0741	9.851	-1.8092	0.3467	0.0926	2.5873	0.1153	-0.1769
Head: Female	0.2593	0.439	0.3529	0.483	0.2037	0.407	-0.0937	0.1872	0.1435	0.0556	0.4058	-0.0928
Head: Married	0.9206	0.271	0.8431	0.367	0.9074	0.293	0.0775	0.096	-0.1698	0.0132	0.7563	-0.0332
Head: Completed primary education or higher	0.6296	0.484	0.5686	0.5	0.6296	0.487	0.061	0.4286	-0.0876	0	1	0
Female members: Completed primary education or higher	0.3651	0.483	0.2745	0.451	0.434	0.5	0.0906	0.2293	-0.1371	-0.0689	0.3633	0.0991
Share of members earning income from work	0.3016	0.304	0.3482	0.369	0.3716	0.347	-0.0466	0.3558	0.0974	-0.07	0.1505	0.1516
Share of female members earning income from work	0.1116	0.218	0.1972	0.325	0.1728	0.248	-0.0857	0.0274	0.2191	-0.0613	0.0792	0.1854
Standardized wealth index (PCA)	-0.1261	0.87	0.4285	1.359	0.0388	0.926	-0.5546	0.0005	0.3438	-0.165	0.2372	0.1298
Household owns any mobile phone	0.9259	0.263	0.9412	0.238	1	0	-0.0153	0.7078	0.0431	-0.0741	0.0454	0.2821
Household owns any mobile phone (diary female respondents)	0.766	0.425	0.8627	0.348	0.8333	0.389	-0.0968	0.1357	0.1764	-0.0674	0.593	0.117
Members have savings at savings groups	0.6085	0.489	0.6275	0.488	0.5833	0.515	-0.019	0.8059	0.0275	0.0251	0.8636	-0.0354
Female members have savings at savings groups	0.455	0.499	0.4706	0.504	0.5	0.522	-0.0156	0.8439	0.0219	-0.045	0.7631	0.0622
At least one member enrolled in NHIF in past 12 months	0.2063	0.406	0.2745	0.451	0.1111	0.317	-0.0682	0.2997	0.1124	0.0952	0.113	-0.1849
Number of hospital admissions in past 12 months	2.3915	4.854	9.0196	32.038	1.9167	2.193	-6.6281	0.0065	0.2045	0.4749	0.7371	-0.0892
Number of outpatient visits in past 12 months	14.9048	10.137	18.2745	14.979	18.3333	16.751	-3.3697	0.0606	0.1863	-3.4286	0.2791	0.1751
Number of informal visits in past 12 months	16.5079	19.39	13.2549	20.379	23.1667	26.655	3.253	0.294	-0.1156	-6.6587	0.2614	0.202
Number of times forewent care in past 12 months	3.4233	5.42	1.8431	3.652	5.5	7.787	1.5801	0.0507	-0.2418	-2.0767	0.2124	0.2189
Total health expenditures in past 12 months	6088.027	9885.952	10434.76	15274.06	14491.6481	36781.253	-4346.7383	0.0149	0.2389	-8403.6217	0.0052	0.2206
Observations	189		51		54							

Notes: Means and standard deviations (SD) of listed household variables for original sample retained, drop-out and replacement households, at baseline, in columns (1)-(2), (3)-(4) and (5)-(6), respectively. Column (7) presents difference in means of retained and drop-out households, column (8), p-value of the corresponding t-test, and column (9) normalized differences; similarly, columns (10)-(12) of retained and replacement households.

Table A2: Baseline household covariates of replenished (retained and drop-out) sample by June 2021

	Retained		Drop-out		Difference in means		Normalized difference
	Mean	SD	Mean	SD	(1)-(3)	p-value	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Number of household members	4.8793	1.732	3.4545	1.036	1.4248	0.0074	-0.7061
Number of members aged 0-4 years old	1.1638	0.58	1	0.447	0.1638	0.3571	-0.2236
Number of members aged 5-17 years old	1.6767	1.434	0.4545	0.82	1.2222	0.0055	-0.74
Number of members aged 18-64 years old	2.0086	0.69	1.9091	0.539	0.0995	0.6379	-0.1136
Number of members aged 65 or more	0.0302	0.171	0.0909	0.302	-0.0607	0.2719	0.1751
Head: Age in years	37.0776	10.454	37.2727	14.677	-0.1951	0.9528	0.0108
Head: Female	0.25	0.434	0.1818	0.405	0.0682	0.6101	-0.1149
Head: Married	0.9138	0.281	1	0	-0.0862	0.3114	0.3065
Head: Completed primary education or higher	0.6379	0.482	0.4545	0.522	0.1834	0.2201	-0.2581
Female members: Completed primary education or higher	0.3723	0.484	0.5455	0.522	-0.1732	0.2495	0.2431
Share of members earning income from work	0.3191	0.314	0.275	0.343	0.0441	0.6656	-0.0948
Share of female members earning income from work	0.1253	0.227	0.125	0.212	0.0003	0.9972	-0.0008
Standardized wealth index (PCA)	0.001	1	-0.0286	1.071	0.0296	0.9346	-0.0202
Household owns any mobile phone	0.9483	0.222	0.75	0.463	0.1983	0.0186	-0.3862
Household owns any mobile phone (diary female respondents)	0.7656	0.425	0.875	0.354	-0.1094	0.4739	0.1979
Members have savings at savings groups	0.6114	0.489	0.5	0.535	0.1114	0.5297	-0.1538
Female members have savings at savings groups	0.4611	0.5	0.375	0.518	0.0861	0.6338	-0.1197
At least one member enrolled in NHIF in past 12 months	0.194	0.396	0	0	0.194	0.1065	-0.4895
Number of hospital admissions in past 12 months	2.456	4.81	0.125	0.354	2.331	0.1731	-0.4833
Number of outpatient visits in past 12 months	15.3316	10.621	9.75	9.588	5.5816	0.1455	-0.3901
Number of informal visits in past 12 months	17.2228	20.114	9.25	10.899	7.9728	0.2673	-0.3485
Number of times forewent care in past 12 months	3.5699	5.657	3	3.625	0.5699	0.7781	-0.0848
Total health expenditures in past 12 months	8133.5603	20016.661	4200	5699.174	3933.5603	0.5167	-0.189
Observations	232		11				

Notes: Means and standard deviations (SD) of listed household variables for replenished sample retained and drop-out households, in August 2020, in columns (1)-(2) and (3)-(4), respectively. Column (5) presents difference in means of retained and drop-out households, column (6), p-value of the corresponding t-test, and column (7) normalized differences.

Table A3: Baseline household covariates of estimation sample

	Total		Control		Treatment		Difference in means		Normalized
	Mean	SD	Mean	SD	Mean	SD	(3)-(5)	p-value	difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Number of household members	4.9326	1.756	5.0093	1.83	4.8372	1.665	0.1721	0.4999	-0.0696
Number of members aged 0-4 years old	1.1658	0.598	1.1402	0.59	1.1977	0.61	-0.0575	0.5083	0.0677
Number of members aged 5-17 years old	1.715	1.42	1.7009	1.481	1.7326	1.35	-0.0316	0.8783	0.0158
Number of members aged 18-64 years old	2.0207	0.729	2.1121	0.805	1.907	0.606	0.2052	0.0516	-0.2036
Number of members aged 65 years old or more	0.0311	0.174	0.0561	0.231	0	0	0.0561	0.0257	-0.2426
Head: Age in years	37.5078	10.398	38.6636	11.954	36.0698	7.89	2.5938	0.085	-0.1811
Head: Female	0.2694	0.445	0.243	0.431	0.3023	0.462	-0.0593	0.3584	0.0939
Head: Married	0.9119	0.284	0.9159	0.279	0.907	0.292	0.0089	0.8292	-0.0221
Head: Completed primary education or higher	0.6373	0.482	0.6168	0.488	0.6628	0.476	-0.046	0.5116	0.0674
Female members: Completed primary education or higher	0.3627	0.482	0.3551	0.481	0.3721	0.486	-0.017	0.8088	0.0248
Share of members earning income from work	0.3013	0.307	0.3291	0.327	0.2669	0.278	0.0622	0.1625	-0.1448
Share of female members earning income from work	0.1178	0.228	0.1278	0.239	0.1052	0.215	0.0226	0.4959	-0.0703
Standardized wealth index (PCA)	0	1	-0.1215	0.846	0.1511	1.151	-0.2726	0.0596	0.1908
Household owns any mobile phone	0.9378	0.242	0.9252	0.264	0.9535	0.212	-0.0283	0.4218	0.0834
Household owns any mobile phone (diary female respondents)	0.7656	0.425	0.7453	0.438	0.7907	0.409	-0.0454	0.4627	0.0758
Members have savings at savings groups	0.6114	0.489	0.6262	0.486	0.593	0.494	0.0331	0.6408	-0.0478
Female members have savings at savings groups	0.4611	0.5	0.4673	0.501	0.4535	0.501	0.0138	0.8494	-0.0195
At least one member enrolled in NHIF in past 12 months	0.2228	0.417	0.2617	0.442	0.1744	0.382	0.0873	0.1491	-0.1495
Number of hospital admissions in past 12 months	2.456	4.81	2.2991	4.581	2.6512	5.101	-0.3521	0.6145	0.0514
Number of outpatient visits in past 12 months	15.3316	10.621	16.2056	11.205	14.2442	9.8	1.9614	0.203	-0.1318
Number of informal visits in past 12 months	17.2228	20.114	17.3832	16.55	17.0233	23.92	0.3599	0.902	-0.0124
Number of times forewent care in past 12 months	3.5699	5.657	4.2991	6.236	2.6628	4.719	1.6363	0.0455	-0.2092
Total health expenditures in past 12 months	7411.1244	20316.257	9359.4112	26551.05	4987.093	6517.845	4372.3182	0.1377	-0.1599
Observations	193		107		86				

Notes: Means and standard deviations (SD) of listed household variables at baseline for estimation sample, control and treated groups, in columns (1)-(2), (3)-(4) and (5)-(6), respectively. Column (7) presents difference in means, and column (8), p-value of the corresponding t-test. Column (9) presents normalized differences.

Table B1: Intention-to-treat estimates: Randomization inference p-values

	Coefficient estimate	Randomization inference p-values	
		Two-sided	One-sided
	(1)	(2)	(4)
<i>Panel A: Enrollment in NHIF</i>			
Enrolled	0.5633	0.0002	0.0002
At least one household member is enrolled	0.6007	0.0005	0.0002
<i>Panel B: Insurance coverage</i>			
Not insured for any of the costs	-0.1081	0.03	0.0151
Costs (partially or fully) covered by NHIF	0.1673	0.0022	0.0012
Costs (partially or fully) covered by other insurance	-0.0002	0.9897	0.4949
<i>Panel C: Out-of-pocket health expenditures</i>			
<i>Panel C.1: Diaries, weekly average in a month</i>			
<i>Panel C.1.1: Individual-level</i>			
Total, excluding transportation	-6.6202	0.0447	0.0225
Transportation	0.0405	0.9475	0.4739
Formal consultation	-5.7186	0.0994	0.0498
Informal consultation	-0.5522	0.1863	0.0933
Preventive consultation	-0.5805	0.3401	0.1702
Curative consultation	-6.0856	0.074	0.0371
<i>Panel C.1.2: Household-level</i>			
Total	-48.0976	0.0044	0.0022
Capacity to pay	16.4071	0.823	0.4116
Total/capacity to pay	-0.0405	0.0249	0.0125
Catastrophic health expenditure	-0.1033	0.0166	0.0083
<i>Panel C.2: Endline survey, out-of-pocket health expenditures in past 12 months, averaged over 52 weeks</i>			
Individual-level	-14.327	0.0027	0.0015
Household-level	-68.205	0.0029	0.0015
<i>Panel D: Health seeking behaviors</i>			
<i>Panel D.1: Diaries, illness episodes</i>			
Sought care for reported health problem	0.0567	0.3108	0.1555
Self-medicated for reported health problem	-0.087	0.083	0.0415
<i>Panel D.2: Diaries, matched ill- or injured-health episodes with care-seeking</i>			
Number of days between reported health problems and seeking care	-0.7624	0.2241	0.1121
<i>Panel D.3: Diaries, weekly average in a month</i>			
Consulted any health care provider	0.0013	0.813	0.4065
Consulted for preventive care	-0.0049	0.4175	0.2087
Consulted for curative care	0.0043	0.4863	0.2432
Consulted at informal provider	-0.0027	0.3906	0.1953
Consulted at formal provider	0.0024	0.5981	0.2991
<i>Panel D.4: Endline survey, number of consultations in past 12 months, averaged over 52 weeks</i>			
Consulted any health care provider	-0.3458	0.0269	0.0134
Consulted at formal provider	-0.3073	0.228	0.114
Consulted at informal provider	-0.0423	0.0034	0.0017
Times did not seek care when ill or injured	-0.008	0.0698	0.0349
<i>Panel E: Health</i>			
<i>Panel E.1: Diaries, weekly average in a month</i>			
Illness or injury	-0.0197	0.3862	0.1931
Number of days unable to do daily activities because of illness or injury	-0.0391	0.0537	0.0269
<i>Panel E.2: Endline survey, self-reported health</i>			
(Very) good self-reported health	0.026	0.5212	0.2607
<i>Panel F: Heterogeneity in distance to closest i-Push facility</i>			
Enrolled in NHIF			
Treatment	0.5938	0.0288	0.0144
Treatment x Distance	-0.0169	0.8962	0.4482
Total health OOP expenditures, excluding transportation			
Treatment	-2.4033	0.4756	0.2378
Treatment x Distance	-2.3607	0.5085	0.2544
Consulted any health care provider			
Treatment	0.021	0.061	0.0305
Treatment x Distance	-0.0119	0.0181	0.009
Number of days unable to do daily activities because of illness or injury			
Treatment	-0.0007	0.9805	0.4902
Treatment x Distance	-0.0228	0.3738	0.187

Notes: Intention-to-treat (ITT) coefficient estimates, obtained from specification (1), and randomization inference (RI) p-values. RI p-values are computed by permuting treatment assignment within the 12 matched pairs (212 = 4,096 permutations, enumerated exhaustively), re-estimating the regression for each permutation, and comparing the resulting distribution of coefficients to the observed estimate. Two-sided p-values, in Column (2), count the share of permutations producing a coefficient at least as extreme (in absolute value) as the observed estimate. One-sided p-values, in Column (3), count the share of permutations producing a coefficient at least as large (in the same direction) as the observed estimate.

Table C1: Pseudo-Maximum Likelihood (PPML) intention-to-treat (ITT) estimates

	Pre-intervention control group mean (1)	Post-intervention control group mean (2)	Coefficient estimate (3)	$\hat{\theta}_{ITT\%}$ (4)	Observations (5)
<i>Panel A: Diaries, weekly average in a month</i>					
<i>Panel A1: Individual-level</i>					
Total, excluding transportation	6.8198	12.6427	-0.7211 (0.2069) [0.0044]	-0.5138 (0.1006)	10,237
Transportation	1.208	1.1265	-0.0224 (0.2423) [0.9364]	-0.0221 (0.2369)	10,237
Formal consultation	4.2295	8.6719	-1.0387 (0.3581) [0.0127] {0.0270}	-0.6461 (0.1267)	10,237
Informal consultation	2.5903	3.9708	-0.182 (0.0915) [0.0551] {0.0390}	-0.1664 (0.0763)	10,237
Preventive consultation	1.3494	1.8966	-0.3463 (0.2202) [0.1416] {0.0770}	-0.2927 (0.1558)	10,237
Curative consultation	5.4704	10.7461	-0.8164 (0.2294) [0.0075] {0.0270}	-0.558 (0.1014)	10,237
<i>Panel A2: Household-level</i>					
Total	70.5873	108.0007	-0.591 (0.1042) [0.0000]	-0.4462 (0.0577)	2,117
Total/capacity to pay	0.1098	0.1606	-0.284 (0.0785) [0.0009]	-0.2472 (0.0591)	2,117
<i>Panel B: Endline survey, out-of-pocket health expenditures in past 12 months, averaged over 52 weeks</i>					
Individual-level	25.3591	27.6073	-0.7162 (0.1299) [0.0001]	-0.5114 (0.0635)	857
Household-level	179.9887	133.1341	-0.7426 (0.1546) [0.0001]	-0.5241 (0.0736)	193
<i>Panel C: Health seeking behaviors</i>					
Number of days between matched reported health problems and seeking care	2.5683	3.30697	-0.2572 (0.1274) [0.0731]	-0.2268 (0.0985)	1,505

Notes: Estimates are obtained using monthly aggregated weekly average data of individual and household observations in Panel A; endline yearly data of individual and household observations in Panel B; and unique health events in Panel C. Column (1) presents the pre-intervention mean of row-specified dependent variables among control villages; column (2), post-intervention. Column (3) presents PPML ITT estimates. Column (4) displays the average proportional change in expenditures, with delta-method standard errors in parentheses. February-May 2020 is defined as pre-intervention period; June 2020-May 2021 as post-intervention period. Estimated specifications include pre-intervention outcome mean as additional covariate, village-pair and month-year fixed effects, except in Panel C specifications, that include village pair and month-year fixed-effects, as well as consultation day-of-the-week fixed-effects. Standard errors are clustered at the village level in parentheses; corresponding p-values obtained via wild cluster bootstrapping, in brackets; sharpened q-values in braces.

Table D1: Heterogeneity effect in distance to closest i-Push facility

	Pre-intervention	Post-intervention	Treatment	Treatment	Observations
	control group mean	control group mean		x Distance	
	(1)	(2)	(3)	(4)	(5)
Enrolled in NHIF	0.1912	0.2326	0.5938 (0.0621) [0.0002]	-0.0169 (0.0271) [0.6355]	10,237
Individual-level total OOP health expenditures, excluding transportation	6.8198	12.6427	-2.4033 (3.9847) [0.6002]	-2.3607 (2.2334) [0.2472]	10,237
Consulted any health care provider	0.047	0.0598	0.021 (0.0083) [0.0705]	-0.0119 (0.0037) [0.0167]	10,237
Number of days unable to do daily activities because of illness or injury	0.0659	0.1149	-0.0007 (0.0232) [0.9798]	-0.0228 (0.0124) [0.1563]	10,237

Notes: Estimates using monthly aggregated data of individual observations in row (1); weekly average in a month, in rows (2)-(4). Column (1) presents the pre-intervention mean of row-specified dependent variables among control villages; column (2), post-intervention. Column (3) and (4) display ITT estimates of the treatment indicator and its interaction with the distance to the closest i-Push facility in kilometers, respectively. Estimation sample distance mean (standard deviation) is equal to 1.66 (1.12) km; median, to 1.33 km. February-May 2020 is defined as pre-intervention period; June 2020-May 2021 as post-intervention period. Estimated specifications include pre-intervention outcome mean and distance as additional covariates, village-pair and month-year fixed effects. Standard errors are clustered at the village level in parentheses; corresponding p-values obtained via wild cluster bootstrapping, in brackets.

Table E1: Intention-to-treat and local average treatment effect estimates setting to 0 missing lagged outcomes

	ITT	LATE	Observations	KP F-statistic
	(1)	(2)	(3)	(4)
<i>Panel A: Individual-level observations</i>				
Enrolled	0.5682 (0.0311) [0.0000]		12,044	
Total health OOP expenditures, excluding transportation	-6.1095 (2.4399) [0.0561]	-12.7333 (5.1924) [0.0601]	12,044	193.85
Transportation	-0.002 (0.1934) [0.9962]	-0.0041 (0.402) [0.9962]	12,044	200.44
Formal consultation	-5.4295 (2.5455) [0.0916]	-11.3419 (5.4428) [0.0996]	12,044	200.44
Informal consultation	{0.2250} -0.3265 (0.2867) [0.4401]	{0.2490} -0.682 (0.5772) [0.4229]	12,044	188.11
Preventive consultation	{0.3620} -0.4332 (0.3814) [0.5310]	{0.3640} -0.9038 (0.7983) [0.5327]	12,044	197.32
Curative consultation	{0.362} -5.6987 (2.4368) [0.0686]	{0.3640} -11.8759 (5.1763) [0.0718]	12,044	194.25
Consulted any health care provider	{0.2250} 0.0049 (0.0027) [0.2303]	{0.249} 0.01 (0.006) [0.2721]	12,044	201.13
Consulted for preventive care	0.0011 (0.0023) [0.7648]	0.0022 (0.0048) [0.7674]	12,044	191.38
Consulted for curative care	{1.000} 0.0026 (0.0031) [0.6170]	{1.000} 0.0053 (0.0065) [0.6197]	12,044	203.19
Consulted at informal provider	{1.000} -0.0042 (0.0037) [0.4942]	{1.000} -0.0088 (0.0078) [0.4944]	12,044	197.16
Consulted at formal provider	{1.000} 0.0071 (0.0041) [0.2707]	{1.000} 0.0146 (0.009) [0.2953]	12,044	207.21
Illness or injury	{1.000} -0.0149 (0.0125) [0.5084]	{1.000} -0.0306 (0.0257) [0.5058]	12,044	205.62
Number of days unable to do daily activities because of illness or injury	{0.3410} -0.0362 (0.0119) [0.0398]	{0.3390} -0.0754 (0.0234) [0.0273]	12,044	195.28
	{0.0870} [0.0870]	{0.0580} [0.0580]		
<i>Panel B: Household-level observations</i>				
At least one household member is enrolled	0.6267 (0.0286) [0.0000]		2,466	
Total health OOP expenditures	-47.8722 (12.9014) [0.0021]	-81.5787 (22.1017) [0.0025]	2,466	756.47
Capacity to pay	41.8369 (43.0188) [0.5087]	70.176 (72.2896) [0.5088]	2,466	675.21
Total/capacity to pay	-0.0295 (0.0101) [0.0516]	-0.0513 (0.0171) [0.0446]	2,466	689.37
Catastrophic health expenditure	-0.0891 (0.0286) [0.0285]	-0.1542 (0.0492) [0.0276]	2,466	696.78

Notes: Estimates obtained using individual-level weekly survey observations in Panel A, and collapsed at the household level in Panel B. Column (1) presents ITT estimates. Column (2) presents LATE estimates, instrumenting NHIF enrollment with treatment assignment. F-statistics of weak IV test corresponding to the specifications estimated in column (2) are presented in column (4). February-May 2020 is defined as pre-intervention period; June 2020-May 2021 as post-intervention period. Estimated specifications include village-pair and month-year fixed effects. When missing, the pre-intervention outcome mean is set to 0, and its interaction with a binary variable taking on unity if non-missing, and 0 otherwise, is included as covariate. Standard errors are clustered at the village level in parentheses; corresponding p-values obtained via wild cluster bootstrapping, in brackets; sharpened q-values, in braces.

Table E2: Pseudo-Maximum Likelihood (PPML) intention-to-treat (ITT) estimates setting to 0 missing lagged outcomes

	ITT coefficient estimate	$\hat{\theta}_{ITT\%}$	Observations
	(1)	(2)	(3)
<i>Panel A: Individual-level observations</i>			
Total health OOP expenditures, excluding transportation	-0.656 (0.1856) [0.0012]	-0.4811 (0.0963)	12,044
Transportation	-0.0605 (0.1912) [0.7608]	-0.0587 (0.18)	12,044
Formal consultation	-0.9713 (0.3329) [0.0045] {0.0100}	-0.6214 (0.126)	12,044
Informal consultation	-0.1214 (0.0878) [0.1819] {0.1390}	-0.1143 (0.0778)	12,044
Preventive consultation	-0.2662 (0.229) [0.2791] {0.1630}	-0.2337 (0.1755)	12,044
Curative consultation	-0.7501 (0.2122) [0.0033] {0.0100}	-0.5277 (0.1002)	12,044
<i>Panel B: Household-level observations</i>			
Total	-0.6015 (0.1116) [0.0000]	-0.452 (0.0612)	2,466
Total/capacity to pay	-0.2055 (0.0723) [0.0044]	-0.1858 (0.0588)	2,466

Notes: Estimates obtained using individual-level weekly survey observations in Panel A, and collapsed at the household level in Panel B. Column (1) presents PPML ITT estimates. Column (2) presents LATE estimates, instrumenting NHIF enrollment with treatment assignment. Column (2) displays the average proportional change in expenditures, with delta-method standard errors in parentheses. February-May 2020 is defined as pre-intervention period; June 2020-May 2021 as post-intervention period. Estimated specifications include village-pair and month-year fixed effects. When missing, the pre-intervention outcome mean is set to 0, and its interaction with a binary variable taking on unity if non-missing, and 0 otherwise, is included as covariate. Standard errors are clustered at the village level in parentheses; corresponding p-values obtained via wild cluster bootstrapping, in brackets; sharpened q-values, in braces.

Table F1: Weekly survey non-response of estimation sample by treatment status

	Control		Treated		Mean difference		Normalized
	Observations	Mean	Observations	Mean	(4)-(2)	t-test p-value	difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Panel A: Health diaries</i>							
<i>Panel A.1: Individual level</i>							
Health module	34021	0.1538	27001	0.1729	-0.0191	0	0.0365
<i>Panel A.2: Household level</i>							
Health module	6602	0.0203	5324	0.025	-0.0047	0.0856	0.0223
<i>Panel B: Financial diaries</i>							
<i>Panel B.1: Individual level</i>							
Expenditure module	20752	0.1844	15397	0.1993	-0.0149	0.0004	0.0267
<i>Panel B.2: Household level</i>							
Expenditure module	6602	0.0108	5324	0.0135	-0.0028	0.1672	0.0279

Notes: Average non-response to diaries by treatment status using weekly observations.