UNITED NATIONS
UNIVERSITY
UNU-WIDER



This paper is a draft submission to the

WIDER Development Conference

# Human capital and growth

6-7 June 2016 Helsinki, Finland

# From Awareness to Adverse Selection: Cardiovascular Disease Risk and Health Insurance Decisions

Berber Kramer*

May 18, 2016

**Abstract**

This paper tests for adverse selection on cardiovascular disease (CVD) risk in a health insurance scheme in rural Nigeria. Cardiovascular diseases are an increasing burden in developing countries. Although they can be largely prevented by treatment of CVD risk factors like hypertension and diabetes, private insurance rarely covers such treatment for fear of adverse selection. We test whether this fear is warranted using panel survey data collected around a health insurance program that does not restrict treatment of hypertension and diabetes. We link *measured* total CVD risk and *reported* cardiovascular (CV) health problems to subsequent enrollment and find that awareness leads to adverse selection. Initially, few individuals report CV health problems and measured CVD risk does not predict enrollment. Over time, more individuals report CV health problems, and their privately observed CVD risk is predictive of enrollment. Thus, when individuals become increasingly aware of whether they are at risk of developing CVDs, asymmetric information will hinder private health insurance markets' ability to provide affordable treatment for CVD risk factors.

**JEL:** D82, D83, and I13.
**Keywords:** Microinsurance, Cardiovascular disease, Awareness.

*Markets, Trade and Institutions Division of the International Food Policy Research Institute (IFPRI), b.kramer@cgiar.org. I gratefully acknowledge the Amsterdam Institute for International Development, Amsterdam Institute for Global Health and Development, and University of Ilorin Teaching Hospital for providing access to the data. I would like to thank Kate Ambler, Michael Anderson, Chris Elbers, Paul Gertler, Jan Willem Gunning, Glenn Harrison, Wendy Janssens, David Levine, Maarten Lindeboom, Jeremy Magruder, Jacques van der Gaag and Elisabeth Sadoulet for helpful suggestions, as well as participants in CSAE 2014 in Oxford, PacDev 2015 in San Diego, Annual Conference on Behavioral Health Economics 2015 in Berkeley, International Microinsurance Conference 2015 in Casablanca, and seminars at the VU University Amsterdam, University of California Berkeley and Center for Economic Analysis of Risk, Georgia State University.

## 1 Introduction

Cardiovascular diseases (CVDs) are among the most pressing development challenges in low- and middle-income countries (LMICs). Each year, an estimated 17.5 million people die from heart failure, heart attacks and strokes. This is 31 percent of all deaths worldwide, with more than 75 percent occuring in LMICs.[1] The economic burden of heart failure in these countries is estimated at $15 billion per annum and will continue to rise (Cook et al., 2014). Further, CVDs arise at a younger age in developing nations compared to developed countries, and are hence associated not only with high and often catastrophic healthcare costs, but also with long-lasting productivity losses due to mortality, disability, and caregiving (Kengne et al., 2013).

In sub-Saharan Africa, most CVDs are due to strokes and can be prevented by treatment of hypertension and other risk factors (Lemogoum, Degaute and Bovet, 2005; Gaziano, 2008). This requires increased availability of technologies and medicines in a region where healthcare spending per capita is among the lowest in the world, despite a large burden of malnutrition and infectious disease (Gaziano, 2007; Kengne et al., 2013). Expanding health insurance coverage is often suggested as a solution to improve the availability and affordability of treatment (WHO, 2014). Insurance indeed improved utilization of diabetes and hypertension treatment in Mexico and Nigeria (Bleich et al., 2007; Sosa-Rubí, Galárraga and López-Ridaura, 2009; Hendriks et al., 2014, 2016).

This paper tests whether the provision of such health insurance is potentially hindered by adverse selection on CVD risk. Insurance providers generally hesitate to cover treatment of major CVD risk factors for fear of adverse selection. When information about CVD risk is asymmetric and insurance providers cannot observe an individual's risk prior to enrollment, adverse selection will raise overall premiums. Alternatively, providers will separate high-

[1] See the WHO Media Center, Fact sheet N° 317, for more information on the burden of CVDs.

and low-risk types by offering different levels of insurance coverage, which in theory results in suboptimal coverage for both types (Rothschild and Stiglitz, 1976). This warrants public intervention in insurance markets, for instance mandatory enrollment or subsidies.

We test for adverse selection in a voluntary health insurance scheme for a rural population in Nigeria that does provide treatment of hypertension and diabetes. Previous literature has found mixed evidence on whether adverse selection in comparable health insurance schemes occurs. In China, individuals with worse self-reported health are more likely to enroll in future survey rounds (Zhang and Wang, 2008). In Cambodia, individuals revealing a higher willingness to pay for insurance use more costly health care (Polimeni and Levine, 2012). By contrast, Parmar et al. (2012) do not find adverse selection in a program in Burkina Faso. Banerjee, Duflo and Hornbeck (2014) show that demand for a product in India is too low for selection to be adverse.

None of these studies test for selection on CVD risk. Because CVD risk is chronic, it is predictive of future health expenses. This may exacerbate adverse selection compared to prior studies that focus on health in general. Alternatively, selection on CVD risk may be less adverse for at least two reasons. First, individuals lack knowledge of whether they are at risk or not. For instance, over 60 percent of hypertensive adults in sub-Saharan Africa and China are unaware of their condition (Addo, Smeeth and Leon, 2007; Zhao, Konishi and Glewwe, 2013). They will not consider hypertension treatment costs in enrollment decisions. Second, CVD risk can be reduced through preventive behaviors like diets, exercise, and reducing tobacco or alcohol use. In theory, factors such as education or risk aversion are associated with both prevention and enrollment, which limits adverse selection (Finkelstein and McGarry, 2006; Doiron, Jones and Savage, 2008).

Whether adverse selection on CVD risk occurs is hence an empirical question. This paper aims to answer this question and analyze the two mechanisms above. To assess

an individual's total CVD risk, we use rich panel survey data with measurements of age, gender, smoking status, BMI, blood pressure and glucose. Combined, these indicators allow calculating the 10-year risk of developing any CVD (D'Agostino et al., 2008). We relate this risk score to enrollment two years later, controlling for age and gender in order to focus on asymmetric information regarding CVD risk factors. Unlike prior literature, we distinguish individuals who do not self-report cardiovascular (CV) health problems from those who do report such problems (including hypertension, diabetes and CVDs), and test whether adverse selection is stronger among the latter group.

To study whether selection on CVD risk is weakened by selection on other factors that are associated with risk-reducing behaviors, we build on a unique feature of the insurance scheme. In contrast to many other schemes, households are not required to enroll the entire family. This allows a test for adverse selection *within* partially enrolling households, controlling for selection on unobserved household-level dimensions associated with lower risks. To our best knowledge, only Wang et al. (2006) have previously done so using the case of a Chinese insurance scheme. They find evidence of adverse selection in partially enrolling households, but study a context in which individual-based enrollment was formally not allowed. This could have induced stronger adverse selection within households.[2]

We find that initially, few individuals report CV health problems, and measured CVD risk does not predict enrollment in the full sample. To the extent that households do selectively enroll higher-risk family members, this is solely driven by selection on age, which is not a source of asymmetric information. Over time, more individuals report CV health

[2] A few other studies with intra-household variation in enrollment correlate cross-sectional measures of health and enrollment and hence cannot disentangle selection from incentive effects (moral hazard) (Chiappori and Salanié, 2000). In a Vietnamese student health insurance scheme, self-reported health did not relate to whether a child is insured Nguyen and Knowles (2010). Witman (2015) find that an older spouse becoming eligible for Medicare crowds out insurance for younger spouses only when healthy, indicative of adverse selection. Other studies that do have panel data, notably Zhang and Wang (2008) and Parmar et al. (2012), use individual rather than household fixed effects. These studies hence analyze whether variation in family members' health over time influences enrollment. They do not capture whether variation among family members influences enrollment if confounding sources of unobserved heterogeneity vary over time as well.

problems, and less observable total CVD risk becomes predictive of subsequent enrollment. Private awareness of CV health problems most likely increased because individuals learned about their health status through blood pressure measurements in the surveys.

We conclude that adverse selection on total CVD risk can pose a threat to program sustainability in the presence of increasing awareness. LMICs rarely impose mandatory enrollment in health insurance, and with treatment of CVD risk factors such as hypertension and diabetes being a major cost driver, private insurance providers prefer restricting coverage. This is because asymmetric information about CVD risk prevents them from pricing the increased costs into the insurance premium for high-risk individuals. Expanding insurance coverage for treatment of CVD risk factors to optimal levels requires mandatory enrollment. Until that is feasible, insurance providers will need to limit adverse selection through solutions such as family-based or even group-based enrollment.

The remainder of this paper is structured as follows. Section 2 introduces the study context. Section 3 describes our data, econometric strategy and indicators of CV health and enrollment. Section 4 discusses the main findings and potential mechanisms driving these results. Section 5 concludes.

## 2 Study context

This study analyzes enrollment during the pilot phase of the 'Hygeia Community Health Care' (HCHC) program, currently named the Kwara State Health Insurance (KSHI) program. This health insurance scheme was introduced in 2009 in Asa, a Local Government Area (LGA) in a rural part of Kwara State, Nigeria. Kwara State is the fourth poorest state in Nigeria in terms of consumption, with a large share of the population relying on subsistence farming. Of its 2.5 million people, a majority - 61.8 percent - is estimated to

live below the poverty line of one dollar a day.[3]

Kwara's health system is not exceptional for sub-Saharan Africa with low health care utilization, poor health infrastructure and a high share of health expenditures paid out of pocket. Prior to the introduction of the HCHC, health insurance coverage was virtually zero in the program area. In addition, awareness of health risks was low. For instance, in the panel surveys that we will be using for our analyses, 21.0 percent of adults was diagnosed with raised blood pressure in 2009, but the vast majority (92 percent) did not know their blood pressure was too high (Hendriks et al., 2012).

To strengthen the health system, the PharmAccess Foundation and Nigerian HMO Hygeia Ltd. launched the HCHC in 2007. In 2009, funded jointly by the Health Insurance Fund and Kwara State government, the scheme expanded to Asa LGA. The HCHC aims at breaking through a cycle of low supply and demand for quality health care. By pre-paying for health care through insurance, partnering health facilities can invest in health care, improving the *supply* of quality health care. This, in turns, is expected to improve the willingness to pre-pay, i.e. *demand*, for quality health care. PharmAccess' approach to pro-poor health care delivery has received international recognition through among others the OECD Health Innovation prize, commendation from the United Nations Secretary General and the Bill & Melinda Gates Foundation (Bonfrer et al., 2015).

To improve the *supply* of quality health care, the program funds upgrading of equipment in health facilities where the health insurance scheme will cover health services. Further, the program conducts a baseline assessment in these facilities, formulates a quality improvement plan, and plans follow-up visits. Examples of quality improvement interventions include implementation of treatment guidelines, upgrading of laboratory equipment, assurance of continuous essential drug supplies, adequate medical file keeping, waste man-

[3]Source: National Bureau of Statistics, Nigeria Living Standard Survey 2009/2010, using PPP exchange rates.

agement protocols and hospital infection control protocols. Finally, the program provides relevant training for health care staff at the partnering health facilities. Individuals do not need to be enrolled in the HCHC to use these services, but non-enrollees will pay for services out-of-pocket. In Asa LGA, the HCHC partners with two health facilities that are located in two different towns.

To increase the *demand* for health care, the HCHC offers subsidized health insurance, which pre-pays for health care in the upgraded health facilities. The insurance provider pays the facilities directly for healthcare services provided to HCHC enrollees so that enrollees do not incur out-of-pocket expenses when using health care. The insurance provider pays facilities a capitation fee of 240 Naira per enrollee per month to cover all primary healthcare services and a fee-for-service for any other service. The total yearly premium is community-rated at a fixed 4,000 Naira per person (US $26.67).[4] Households themselves pay only 300 Naira (US $2) for every insured family member. This is 7.5 percent of the total yearly premium and 23.1 percent of pre-program annual per capita health expenditures.

Enrollees register in one of the two facilities, usually in the nearest town. Enrollment is individual-based, meaning that households do not need to enroll their entire family, and one enrollee's policy does not cover other family members. All individuals living in the program area are eligible for enrollment in the HCHC. Although there is in principle no exclusion restriction for chronic diseases, some form of screening may have been introduced in 2011. The local program management indicates that in that year, enrollment agents started requesting elderly individuals with high blood pressure to not enroll by themselves, but also enroll younger household members.[5]

Trained agents go door-to-door to explain and offer the insurance scheme to the population. Also one of the two partner facilities is actively engaged in enrollment by employing

[4] At the time of program design, this was equal to the per capita cost of the existing National Health Insurance Scheme available to formal and public sector employees.

[5] Personal communication with Dr. Opowoye, program manager of the HCHC.

its own team of enrollment agents.[6] Interested individuals can enroll with the agents, at dedicated kiosks in several town centers, and during several events organized for the community. Consistent with an availability bias (Tversky and Kahneman, 1973), individuals often do sign up or renew an expired policy after experiencing a health shock (Janssens and Kramer, 2016). Nonetheless, they face a waiting period of six to 36 days after signing up, and only 10 to 16 percent has a first visit within a month after enrolling (Bonfrer et al., 2015). Thus, individuals who enroll after falling ill do not receive immediate coverage.

The insurance policy is comprehensive and covers primary and outpatient healthcare services, minor and intermediate surgeries, as well as hospital admissions and inpatient healthcare, including inpatient healthcare for CVD patients in need of acute health care. The scheme covers both inpatient and outpatient care (including follow-up treatment) for hypertension and diabetes mellitus (excluding dialysis). Stroke management is unavailable in the program facilities and is not covered. The cost of CVD prevention care is US$ 144 per patient per year, and the cost is US$ 118 per year per patient with hypertension (Hendriks et al., 2015), the most common CVD risk factor. Appendix Table A1 summarizes covered services for hypertension treatment: consultations, laboratory tests, and monthly drug collections, as well as a specialist visit for patients with complex hypertension.

Households can save substantially by enrolling in the HCHC. During the second year of the program, they spent on average 1,030 Naira or US$ 6.87 (more than three times the premium) less on health care compared to households in an uninsured control area, despite increased health care utilization (Gustafsson-Wright, Tanovic and Van der Gaag, 2013). Given these large cost savings, one would a priori expect that the majority of households enroll, without adverse selection being a concern. Indeed, a relatively large share of the

[6]Because the premium is lower than the capitation fee, this partner facility faces an incentive to increase enrollment (and limit adverse selection). This facility started offering households the option to pay in installments when enrolling the entire family instead of just a few members. The second partner facility, a government hospital, does not face such an incentive because the Kwara State Ministry of Health administers and controls its capitation fees. Partial enrollment is more common among households enrolling in this facility.

households in our sample, 63.9 percent, enrolls at least one adult. For comparison, Acharya et al. (2012) document enrollment of up to 60 percent when insurance premiums are *fully* subsidized, and enrollment in community-based health insurance in Burkina Faso with a comparable insurance premium lies between 5.2 and 6.3 percent (Gnawali et al., 2009). The 36.1 percent of households who did not enroll any household member reported difficulties in enrollment procedure as the main reason (36.7%) for not enrolling, followed by the perception that they did not need health insurance or could enroll when falling ill (13.7%), and a lack of knowledge of, or trust in, the insurance provider (12.6%). Households did don opt out because of financial reasons; rather, logistics appear the main barrier to enrolling at least one household member.

However, because enrollment was individual-based, households did not need to enroll the entire family. Among households with at least one enrolled adult, a mere 57.2 percent enrolled all adults; the remaining 42.8 percent enrolled only half of all adults.[7] Participants in informal focus group discussions attributed this to large families being unable to pay the insurance premium for all family members in one go. In the endline survey, partially enrolling households reported for individuals not enrolled they did not register them because they would wait until they wree ill (27.2%), logistical reasons (20.6%), or financial reasons (10.3%). These families may have enrolled their least healthy members, for instance those with CV health problems. The analyses will therefore not only analyze selection in the total sample, but also zoom in on partially enrolling households, who reveal a more binding budget constraint.

[7]The percentage of households enrolling all family members, including children, is substantially lower than 57.2 percent. CVDs are relevant mainly for the adult sample. Further, non-enrolled children in otherwise enrolled households may not have been restricted from utilizing the benefits offered by the insurance program. This paper therefore focuses on adverse selection among adult family members.

## 3 Methods

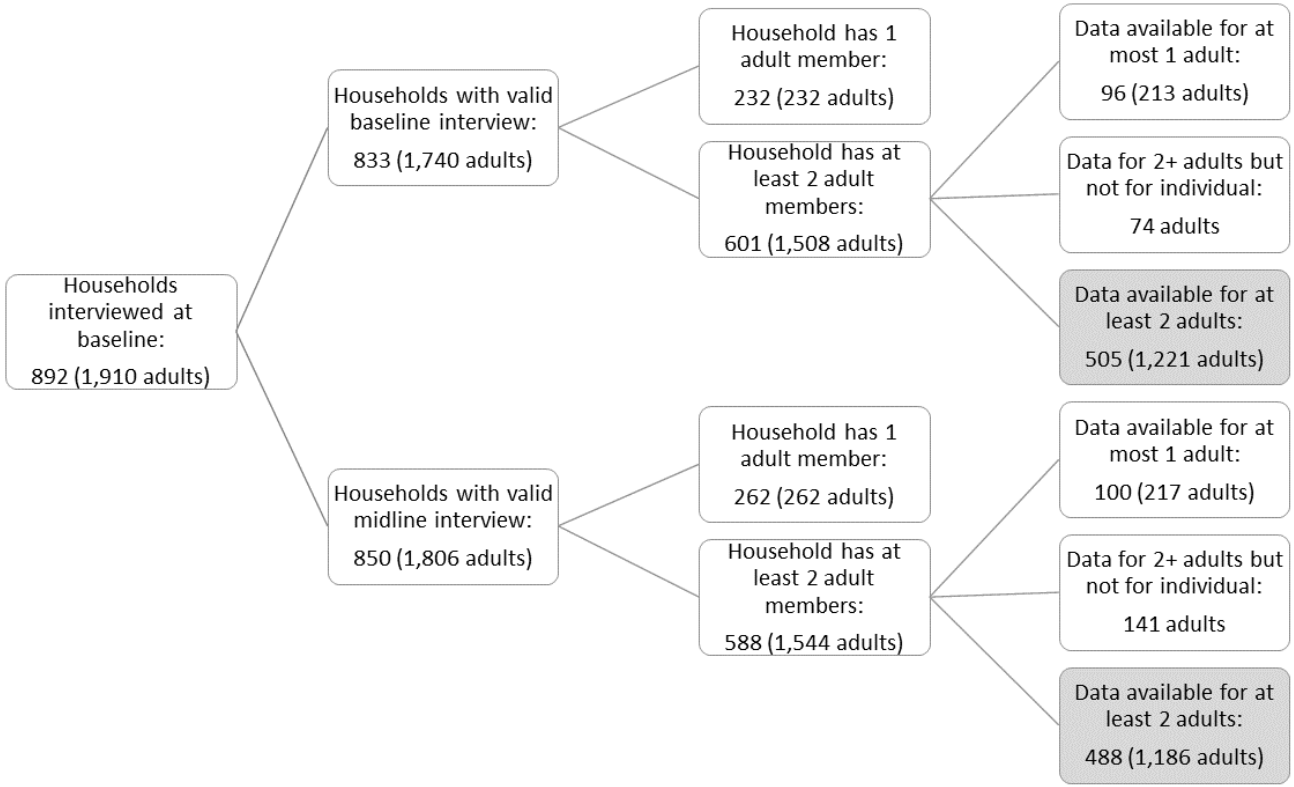
### 3.1 Data

To test for adverse selection on CVD risk, the analyses employ a panel dataset with three waves of household surveys: A baseline completed before the launch of the HCHC in 2009 ($t = 0$), a midline in 2011 ($t = 1$), and an endline in 2013 ($t = 2$). Prior to baseline, 900 households were sampled from 60 census enumeration areas within 10 kilometers of the two program facilities. Households were sampled proportional to census area population size so that the sample was representative of the population. The survey also included 600 households from 40 census areas in a neighboring district where health insurance was unavailable. The main analyses focus on the program area, where households could enroll in the HCHC.

All surveys included household-level questions on consumption and wealth, as well as individual-level questions for all household members' demographics, employment, health, health care utilization, and enrollment in health insurance. We restrict the sample to adult household members above 18 years of age, for whom nurses performed a health exam including (among others) measurements of height, weight, blood pressure, and non-fasting glucose. The resulting dataset links intra-household variation in CVD risk measured at time $t$ to enrollment in the HCHC between time $t$ and the next survey round, $t+1$. Because our primary interest is to identify selection within partially enrolling households, we focus on households with at least two adult members in the analyses; that is, households in which we observe current CVD risk and subsequent enrollment for at least two adults.

Figure 1 describes how we construct the final sample included in the analyses. Among the sample of 900 households in the program area, 892 were interviewed at baseline. The analyses omit 59 households (6.6%) due to errors in data collection, 232 households (26.2%)

Figure 1: Construction of analysis sample




*Notes:* 'Data available' in the last column means that data from both current and follow-up surveys are available; referring to availability of both baseline and midline data (midline and endline data) in the top (bottom) panel. Valid midline interviews include X households without valid baseline interview and valid baseline interviews include X households without valid midline interview.

with only one adult member, and another 96 households (10.8%) with complete data on baseline CV health and enrollment before midline for at most one adult.[8] A final 74 adults *without* complete data are dropped from households that have at least two other adult members *with* complete data on baseline CV health and enrollment before midline. The remaining baseline sample includes 505 households with in total 1,221 adults.

At midline, 786 of the 892 households interviewed at baseline (88.1%) were traced. Some had split into at least two different households, yielding a total number of 850 household interviews. Among them, the analyses omit 262 households (30.8%) with only one adult member, 100 households (11.8%) with complete data on CVD risk factors and subsequent enrollment for at most one adult, and a final 141 adults without complete data from households with complete data for at least two other adult members. The remaining midline sample includes 488 households with in total 1,186 adults. Among them, 966 adults are included in both samples (79.1% and 81.5% of the baseline and midline samples, respectively).

### 3.2 Econometric strategy

This section describes how we test for selection on CVD risk, and how we analyze the mechanisms that are driving our results. Our main outcome variable is enrollment between the current and next survey round, $d_{ih}$ for individual $i \in \{1, ..., N\}$ in household $h \in \{1, ..., H\}$. The analyses relate this variable to current CV health, $CV_{ih}$, using the following equation, estimated separately for enrollment from baseline to midline, and from midline

[8] For a number of individuals, some but not all CVD risk factors are missing. To avoid dropping their entire household, we impute by values from the previous or following round. If these are unavailable as well, we use the median value calculated by round, gender, age category, wealth (either above- or below-median), nearest program facility, and location type (town versus village). Appendix Table A5 drops individuals with imputed values, yielding similar results as Tables 4 and 5.

to endline:

$$d_{ih} = CV_{ih}^{M}\beta_0 + \mathbf{z}_{ih}\gamma_0 + u_{ih} \tag{1}$$

where $CV_{ih}^{M}$ is the measured ('M') log odds of developing a CVD; $\mathbf{z}_{ih}$ a $k_1 \times 1$ vector of individual-level variables that may directly influence enrollment through screening and adverse selection on non-CVD-related health risks, i.e. gender, log age, past illnesses and injuries, upcoming pregnancies, and current enrollment status[9]; and $u_{ih}$ an individual-specific residual.[10]

We infer adverse selection from testing whether the coefficient on CVD risk, $\hat{\beta}_0$, is strictly positive. Note that the model relates *future* instead of *current* enrollment status to current CVD risk. Individuals may take more risks once covered by health insurance. Such moral hazard increases the correlation between risk and current coverage (Chiappori and Salanié, 2000; Chiappori et al., 2006). At the same time, the HCHC improved utilization of hypertension treatment, leading to a reduction in CVD risk (Hendriks et al., 2014, 2016). By relating *future* enrollment to *current* CVD risk (conditional on current enrollment), our estimate of adverse selection, $\hat{\beta}_0$, is unbiased by moral hazard or treatment effects.[11]

We estimate Equation (1) for the entire sample, but also for households that never enroll partially, and for households that enroll partially at least once. Partially enrolling households decide to pay the premium for some but not all family members. Within these households, factors such as inertia, unfamiliarity with the scheme and not trusting (or

[9] We control for age and gender because these two risk factors are typically observed symmetrically. In theory, insurance premiums or targeted subsidies could incorporate actuarial cost differences for different age and gender groups, so that insurance providers face similar profit margins independent of demographics, and have no incentive to screen and discourage enrollment among symetrically observed high-risk individuals.

[10] To give every household an equal weight independent of the number of adult members, the analyses weigh observations by the inverse number of adults in the household. Standard errors are clustered by enumeration area.

[11] If households are offered different insurance premiums, and the variation in the insurance premium is exogenous, an alternative approach to circumvent such biases is to compare health care utilization among households that enrolled at a high premium versus utilization among households that enrolled at a low premium (Polimeni and Levine, 2012). Because HCHC premiums did not randomly vary, we cannot adopt a similar approach in this paper.

being visited by) enrollment agents cannot explain why some family members do not enroll. More likely, these households - unwilling to pay the insurance premium for some but not all members of the household - reveal a more binding budget constraint. Consistent with this line of thought, Table 1 compares all households and households that enroll partially at least once. The latter type of household has more household members, which means that the cost of enrolling all family members is higher. As such, these households will be more selective in whom they enroll, and their decision-making is perhaps more representative for other health insurance schemes in which premium subsidies are lower, and out-of-pocket premiums are higher.

We hypothesize that adverse selection on CVD risk is limited, and perhaps even absent, $\hat{\beta}_0 = 0$, for two reasons. First, individuals in low-income countries typically lack awareness of personal CVD risks (addo2007hypertension,zhao2013does). To test this hypothesis, we estimate a second model, including a binary variable indicating whether an individual reports CV health problems, $CV_{ih}^R$:

$$d_{ih} = CV_{ih}^M \beta_1 + CV_{ih}^R \beta_2 + CV_{ih}^M \times CV_{ih}^R \beta_3 + \mathbf{z}_{ih}\gamma_1 + u_{ih} \tag{2}$$

Hypothesis (1) is that selection is limited among individuals who do not report CV health problems, $\hat{\beta}_1 = 0$, whereas adverse selection does occur among individuals who do report CV health problems, $\beta_1 + \beta_3 > 0$.

Second, even if individuals are aware and select on the basis of CVD risks, observed selection may not be adverse if individuals also select on the basis of characteristics associated with precautionary behaviors and lower CVD risks, such as wealth, education or risk aversion. Several studies find that not only high-risk individuals but also individuals who take precautionary and risk-reducing actions are more likely to enroll in insurance (De Meza and Webb, 2001; Finkelstein and McGarry, 2006; Doiron, Jones and Savage,

2008). Such actions occur to some extent at the household level, for instance through diets. We control for such unobserved heterogeneity *between* households by comparing CV health among enrolling and non-enrolling adults *within* a household. This may increase estimates of adverse selection.

To test this hypothesis, we expand Equation (1) as follows:

$$d_{ih} = CV_{ih}^{M}\beta_1' + CV_{ih}^{R}\beta_2' + CV_{ih}^{M} \times CV_{ih}^{R}\beta_3' + \mathbf{z}_{ih}\gamma_1' + \mathbf{x}_{ih}\gamma_2' + \eta_h + u_{ih} \tag{3}$$

where $\mathbf{x}_{ih}$ is a $k_2 \times 1$ vector of individual characteristics that are in theory associated with both enrollment and preventive behavior, including an individual's risk aversion, rank within the household and personal savings as a proxy for bargaining power or ability to purchase health insurance, and personal income as a proxy for the loss from sickness absenteeism (given that a household may prefer to enroll the breadwinner in order to prevent health costs when this person falls ill and cannot generate income); and $\eta_h$ a household fixed effect to capture any unobserved characteristics at the household level, such as risk preferences of the household decision-maker or wealth, that may well be correlated with both enrollment and CVD risk.

### 3.3 Measuring total cardiovascular disease (CVD) risk

To measure our main risk indicator, $CV_{ih}^{M}$, we build on the Framingham Heart Study. This longitudinal cohort study collected medical indicators for 8,491 participants over a period of 12 years. Using these data, D'Agostino et al. (2008) estimate the total 10-year risk of developing a CVD as a function of multiple risk factors, and we use their risk functions to predict this 10-year risk for our study sample.[12] A major advantage of the total CVD

[12] D'Agostino et al. include the following cardiovascular events in their outcome measure of developing a CVD: coronary death; myocardial infarction; coronary insufficiency; angina, ischemic or hemorrhagic strokes; transient ischemic attack; peripheral artery disease; and heart failure.

Table 1: Description of household-years included in the analysis sample

| | From baseline to midline | | | From midline to endline | | | | |
|---|---|---|---|---|---|---|---|---|
| | All households | Ever partly enroll HH | Diff. never partly enroll | All households | Ever partly enroll HH | Diff. never partly enroll | Differences between rounds (1)-(4) | (2)-(5) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Household lives in a town with health facility | 0.471 | 0.491 | 0.036 | 0.477 | 0.491 | 0.025 | 0.006 | -0.000 |
| Lives in area with private program facility | 0.440 | 0.446 | 0.012 | 0.445 | 0.446 | 0.003 | 0.005 | 0.000 |
| Number of household members | 5.176 | 5.674 | 0.895** | 5.482 | 5.893 | 0.760** | 0.305** | 0.219† |
| Number of adults in analysis sample | 2.418 | 2.638 | 0.396** | 2.430 | 2.594 | 0.302** | 0.013 | -0.045 |
| Yearly per capita consumption (1,000 NGN '09) | 79.26 | 74.44 | -8.653 | 71.97 | 68.35 | -6.750 | -7.287 | -6.092 |
| Wealth index (standardized) | -0.165 | -0.113 | 0.093 | 0.171 | 0.211 | 0.074 | 0.336** | 0.324** |
| Household head is literate | 0.428 | 0.393 | -0.063 | 0.470 | 0.411 | -0.110[*] | 0.043† | 0.018 |
| Household head has no education | 0.584 | 0.634 | 0.089† | 0.545 | 0.616 | 0.131** | -0.039 | -0.018 |
| Farming main income source of household head | 0.511 | 0.504 | -0.012 | 0.439 | 0.469 | 0.056 | -0.072** | -0.036 |
| Trading main income source of household head | 0.069 | 0.076 | 0.012 | 0.086 | 0.098 | 0.022 | 0.017 | 0.022 |
| Household head is female | 0.073 | 0.089 | 0.029 | 0.074 | 0.085 | 0.020 | 0.001 | -0.004 |
| Households | 505 | 224 | | 488 | | 224 | | |

Notes: Sample includes all households with current health and subsequent enrollment observed for at least two adult family members. Partially enrolling HH: Households that enrolled partially at least once (some but not all family members enrolled between baseline and midline, or between midline and endline). Diff. all/none enroll: Difference in means between partially enrolling households and households in which either all or zero adults enroll between baseline and midline, and between midline and endline. Exchange rate: NGN 1,000 is USD 6.67. Significance levels calculated after clustering standard errors by census area. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

risk assessment approach is that it allows several slightly elevated risk factors to result in a higher total risk than a single, more strikingly raised factor (Bitton and Gaziano, 2010; Cooney et al., 2010). This is important because CVD risk is not a linear function of independent risk factors. Rather, risk factors interact in increasing the risk.

D'Agostino et al. (2008) provide two sets of risk functions for the 10-year risk of developing a CVD. One set requires costly laboratory measures of cholesterol, which were not collected in the health exams. We therefore calculate individual risk scores using the second set of Framingham risk functions, which replace cholesterol by body mass index (BMI). A study among urban-dwelling Black South Africans finds a high correlation between the two scores, making the non-laboratory-based score a useful measure of CVD risk in a context of resource constraints (Peer et al., 2014).

The analyses will use the log odds of this 10-year risk score. For individual $i$ of sex $s \in F, M$, the score is defined as

$$p_{ihs} = 1 - \alpha_s^{\exp(\beta_s \mathbf{x}_{ihs} - \gamma_s)}, \tag{4}$$

where $\beta_s \mathbf{x}_{ihs}$ is a linear function of the following CVD risk factors: log age, log BMI, log systolic blood pressure for patients who are and who are not on treatment for high blood pressure, respectively, whether the individual currently smokes cigarettes and/or tobacco, and whether the individual has diabetes.[13] D'Agostino et al. (2008) use Cox proportional-hazards regressions to estimate the baseline survival rate, $\alpha_s$, the constant, $\gamma_s$, and the coefficients on risk factors, $\beta_s$. Appendix Table A2 presents the estimated coefficients and hazard ratios.

The Framingham risk scores have been validated across different populations, including

[13]The Framingham risk functions apply to individuals between 30 and 74 years of age. We predict risk scores for all individuals 18 years and above, but results are robust to censoring age to 30 for individuals younger than 30 at baseline (23.2 percent) or midline (19.2 percent), and to censoring age to 74 for individuals older than 74 at baseline (5.3 percent) or midline (5.7 percent).

ethnic minorities, and are recommended by numerous international guideline committees for CVD prevention. The World Health Organization and International Society of Hypertension (WHO/ISH) developed an alternative risk score for low- and middle-income countries where cohort data on CVD risks are not readily available (Mendis et al., 2007). We do not use this score because it does not consider variation in BMI or account for the effects of hypertension treatment, which expanded after the launch of the HCHC (Hendriks et al., 2014, 2016). In addition, it discretizes blood pressure and age. This reduces observed variation in CVD risk and the ability to identify adverse selection.[14]

### 3.4 Summary statistics

Table 2 summarizes the main individual-level variables included in the analyses. Columns (1)-(3) describe baseline characteristics for the sample included in analyses of enrollment between baseline and midline, whereas Columns (4)-(6) present midline characteristics for the sample included in analyses of enrollment between midline and endline. Columns (1) and (4) present sample means for all individuals. Columns (2) and (5) restrict the sample to individuals from partially enrolling households. Columns (3) and (6) present the difference in means between individuals from households that ever versus never enroll partially. Columns (7) and (8) summarize differences in baseline and midline characteristics for the full and restricted sample, respectively.

Our dependent variable, enrollment between two survey waves $t$ and $t+1$, is constructed from start and expiry dates of past insurance policies. These were reported for all family

[14] Appendix Figure A1 compares the two types of risk scores. At baseline, the average Framingham risk score is linearly increasing in the WHO/ISH risk score. In addition, the risk scores fall within the same range. At midline, this is true only for the group with the lowest risk; the Framingham estimates a higher risk for individuals in the middle WHO/ISH risk class, and a lower risk for individuals in the higher risk class. Nonetheless, Appendix Table A3 shows that the two scores are strongly correlated. Because we are not interested in the absolute risk but in the relation between risk and enrollment, which requires sufficient variation in the risk score, the Framingham risk score suffices for the purposes of this paper.

Table 2: Description of individuals in the analysis sample

| | From baseline to midline | | | From midline to endline | | | | |
|---|---|---|---|---|---|---|---|---|
| | All households | Ever partly enroll HH | Diff. never partly enroll | All households | Ever partly enroll HH | Diff. never partly enroll | Differences between rounds (1)-(4) | (2)-(5) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **Outcome variable** | | | | | | | | |
| Enrolls before follow-up survey | 0.567 | 0.563 | -0.006 | 0.464 | 0.498 | 0.062 | -0.103** | -0.065* |
| **Cardiovascular disease risk and indicators** | | | | | | | | |
| Reports CV health problem | 0.025 | 0.025 | 0.000 | 0.093 | 0.091 | -0.005 | 0.069** | 0.066** |
| Framingham 10-year CVD risk score | 0.055 | 0.054 | -0.001 | 0.062 | 0.062 | 0.000 | 0.007** | 0.008** |
| BMI (weight in kg / height in meters sq) | 23.16 | 23.23 | 0.117 | 22.72 | 22.83 | 0.214 | -0.445† | -0.391 |
| Overweight or obese (BMI >= 25) | 0.225 | 0.239 | 0.025 | 0.229 | 0.232 | 0.005 | 0.004 | -0.007 |
| Mean systolic blood pressure | 120.604 | 119.5 | -1.938† | 121.0 | 120.5 | -1.005 | 0.441 | 0.963 |
| High systolic blood pressure (>= 140 mm Hg) | 0.140 | 0.123 | -0.031 | 0.151 | 0.133 | -0.032† | 0.011 | 0.011 |
| Is diabetic (self-reported/measured) | 0.013 | 0.015 | 0.004 | 0.022 | 0.031 | 0.016 | 0.009† | 0.016† |
| Currently smokes cigarettes/tobacco | 0.085 | 0.087 | 0.003 | 0.080 | 0.074 | -0.012 | -0.005 | -0.013 |
| **Control variables** | | | | | | | | |
| Age | 45.42 | 45.55 | 0.235 | 47.10 | 47.95 | 1.535 | 1.682** | 2.399** |
| Female | 0.544 | 0.551 | 0.012 | 0.546 | 0.557 | 0.020 | 0.001 | 0.006 |
| Ill or injured in past 12 months | 0.316 | 0.301 | -0.027 | 0.304 | 0.293 | -0.020 | -0.012 | -0.008 |
| Pregnant in next 2 years | 0.118 | 0.103 | -0.028 | 0.109 | 0.108 | -0.001 | -0.010 | 0.005 |
| Income last 7 days (100 NGN) | 24.32 | 25.75 | 2.584 | 40.09 | 34.90 | -9.389 | 15.77** | 9.148* |
| Rank within household | 0.500 | 0.500 | -0.000 | 0.499 | 0.507 | 0.013 | -0.001 | 0.007 |
| Willingness to take risks | 2.036 | 2.004 | -0.057 | 2.713 | 2.677 | -0.065 | 0.677** | 0.673** |
| Personal savings (1,000 NGN) | 8.841 | 7.559 | -2.305 | 12.806 | 11.544 | -2.284 | 3.965* | 3.985* |
| Education level | 0.648 | 0.600 | -0.087 | 0.799 | 0.684 | -0.208** | 0.150** | 0.084* |
| Observations | 1221 | 591 | | 1186 | 567 | | | |

Notes: Sample includes all households with current health and subsequent enrollment observed for at least two adult family members. Means are weighted by the inverse number of observations in a household. None/fully/partially enrolling: Zero/all/some but not all family members enrolling between current survey and follow-up round. Exchange rate: NGN 100 is USD 0.67. Significance levels calculated after clustering standard errors by census area. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

members at endline.[15] At midline, households only reported current insurance status at the individual level. This variable is used for individuals who were not part of the endline survey. From baseline to midline, 56.7 percent of the analysis sample enrolls. From midline to endline, the enrollment rate reduces to 46.4 percent. From baseline to midline, enrollment in partially enrolling households is significantly lower ($p < 0.10$), but this is not the case from midline to endline, and differences between the two rounds do not significantly differ.

As indicator of reported CV health problems ($CV_{ih}^{R}$), we construct a binary variable that takes on value one if a respondent reports (ever) having (been diagnosed with) heart disease, heart failure, heart attack, stroke, diabetes or hypertension.[16] At baseline, 2.5 percent of individuals reports CV health problems. At midline, this percentage increases to 9.3 percent. Reported CV health problems are most often hypertension (82.5%), followed by diabetes (12.1%), and heart disease or cardiovascular events in the last 12 months (5.4%). The increase from baseline to midline in the percentage of individuals reporting CV health problems is statistically significant ($p < 0.01$). We find very similar patterns in the subsample of partially enrolling households.

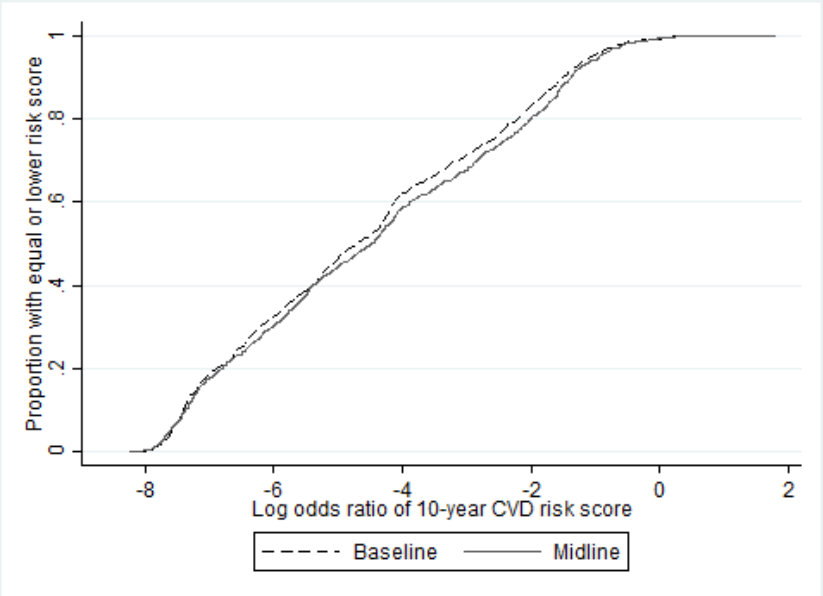
Measured risk also increases over time. The 10-year risk of developing a CVD is on average 0.055 at baseline, and 0.062 at midline. A very similar pattern arises in the subsample of individuals from households that ever enroll partially. However, the distribution of the log odds of the CVD risk score does not significantly differ from baseline to midline (Figure 2). We will later explore whether the increase in reporting of CV health problems from baseline to midline can be explained by the increase in measured risk.

The increase in CV reporting and CVD risk cannot be due to underlying CVD risk

[15] To facilitate recall, enumerators used old insurance cards, which many respondents kept in their house.

[16] At midline and endline, this variable takes on value one also if the respondent reported one of these events or risk factors in a previous survey round. Thus, this variable can be interpreted as an individual ever having reported hypertension, diabetes or a CVD, either in the current or a previous survey round.

Figure 2: CDF of log odds ratio of total 10-year CVD risk score




*Notes:* Sample includes individuals from the program area, including individuals from both baseline and midline. Only members from households with health and subsequent enrollment observed for at least two adult family members are included. The data is weighted by the inverse number of adult family members. Kolmogorov-Smirnov test statistic for equality of distribution functions is 0.0350 ($p = 0.454$).

factors, which remain fairly stable over time. The average BMI is 23.2 at baseline and 22.7 at midline. Although this falls in the normal or healthy weight range, a substantial 22.4 - 22.9 percent of individuals are overweight or obese. An individual's systolic blood pressure is the average of three measurements within an interval of 15 minutes. This measure is on average 120.8 mm Hg, which is 0.8 mm Hg above the maximum level to be considered normal. Most individuals have pre-hypertensive levels of 120-140 mm Hg, and blood pressure is beyond pre-hypertensive levels for 13.9 percent at baseline and 14.9 percent at midline. We classify 1.3 percent of baseline individuals and 2.2 percent of midline individuals as diabetic, meaning that the individual either reports having diabetes or tests positive for high non-fasting glucose. On average, 8.5 percent and 8.0 percent smokes cigarettes or tobacco at baseline and midline, respectively.

The remaining variables are included as control variables. The main reason why measured CVD risk increases is because individuals become older from baseline to midline; the average age increases from 45.4 years at baseline to 47.1 at midline. Other health-related variables do not vary significantly over time. Around 54.5 percent is female in both rounds. 31.6 percent was ill or injured in the past 12 months at baseline, and 30.4 percent reported an illness or injury at midline. From baseline to midline, 11.8 percent of adults (20.3 percent of women) becomes pregnant, and from midline to endline, 10.9 percent of adults (18.6 percent of women) reports a pregnancy.

The final set of variables are potentially related to both enrollment in health insurance and precautionary behaviors. Real income in the last 7 days is on average 2,400 NGN at baseline, and increases significantly to 4,009 NGN at midline ($p < 0.01$). The average rank, which is a proxy for bargaining power within the household, does not change over time. Risk aversion is measured as the person's willingness to take risks on a scale from one (extremely unwilling to take risks) to seven (extremely willing to take risks). The

average respondent is moderately unwilling to take risks but the willingness to take risks increases by about 0.6 points on this scale ($p < 0.01$). Personal savings are on average 884 NGN at baseline and increase significantly to 1281 NGN at midline ($p < 0.05$). The eduation level is measured on an ordinal scale taking on values 0 (no or some primary education), 1 (primary education completed), 2 (secondary education completed) and 3 (tertiary education completed). The majority of individuals has completed some primary education and from baseline to midline, education levels improve.

### 3.5 Validation: Relating CVD risk to health-seeking behavior

Table 3 validates our measure of CVD risk as a relevant health indicator, relating current CVD risk to health expenditures reported in the next survey round. In the control district, health insurance is unavailable so that health expenditures are not confounded by the decision to enroll in health insurance. Odd columns estimate Equation (1), and even columns Equation (2), using health expenditures instead of enrollment as dependent variable. Columns (1)-(2) focus on acute health expenditures, whereas Columns (3)-(4) focus on chronic health expenditures.[17] We apply an inverse hyperbolic sine transformation because the distribution of health expenditures is right-skewed and includes zeros.[18]

In Columns (1)-(2), total CVD risk does not significantly predict follow-up expenditures on the treatment of illnesses or injuries. Predicted expenditures are 17.3 percent higher for individuals who report having developed hypertension, diabetes or a CVD than for individuals who do not report one of these cardiovascular health problems, but this difference is not statistically significant, and the model in Column (2) explains only 2.4 percent of the variation in acute health expenditures at follow-up.

[17] In the control district, we observe very few instances of expenditures on maternal health care. Hence, although pregnancies and deliveries in program facilities are a major cost driver in the HCHC, we cannot estimate the effects of becoming pregnant on maternal healthcare expenditures in the control district.

[18] Except for very small values of $y$, the inverse sine is approximately equal to $log(2) + log(y)$, and coefficients can hence be interpreted in exactly the same way as log transformations.

Columns (3)-(4) estimate the same model for expenditures on the treatment of chronic diseases. Total CVD risk is an important predictor of these health expenditures. In Column (3), a one percent increase in the odds of developing a CVD increases chronic health expenditures by 25.4 percent. This effect is largest for individuals who report cardiovascular health problems; for individuals who do not report such health problems, a one percent increase in the odds of developing a CVD increases health expenditures by 12.4 percent, whereas this effect is 17.1 percentage points stronger for individuals who do report cardiovascular health problems. Further, individuals who report cardiovascular health problems with an average risk score have 68.5 percent higher health expenditures. Thus, CVD risk is a significant predictor of future chronic health expenditures when health insurance is unavailable. In the program area, individuals can enroll in the HCHC to cover their health expenditures. Adverse selection on CVD risk will hence result in substantially higher program costs.

On a final note, in Columns (1)-(2), past illnesses and injuries increase acute health expenditures two years later by 16.5 to 17.5 percent ($p < 0.05$). Selection into the HCHC on past illnesses and injuries may hence pose substantial costs to the HCHC as well. Administrative data indicate that also pregnances are an important cost driver. Thus, if there is no adverse selection on past health shocks or pregnancies, we would also expect limited selection on CVD risk. In that case, the premium subsidy may be large enough to rule out adverse selection, or households may not sufficiently understand the insurance concept to realize they can benefit more by enrolling only high-risk memebers. Alternatively, if selection on past illnesses or pregnancies is adverse, the premium subsidy is apparently not large enough to eliminate adverse selection, and households do realize that it is beneficial to enroll only high-risk members. In that case, limited selection on CVD risk could be due to something more specific to CVDs, for instance limited awareness or a correlation

between risk-reducing behaviors and enrollment.

## 4 Results

This section first describes selection on total CVD risk. We then explore whether selection on total CVD risk depends on whether an individual knows he or she has a cardiovascular health problem, and whether observed selection is potentially limited by selection on characteristics or behaviors associated with lower CVD risk. Our analyses indicate that improved knowledge of a person's cardiovascular health status results in stronger selection on total CVD risk. The section concludes with a number of robustness checks and a discussion of what may have caused an increase in such knowledge.

### 4.1 Adverse selection on total CVD risk

Table 4 estimates Equation (1) for enrollment between baseline and midline in Columns (1)-(3), and enrollment between midline and endline in Columns (4)-(6). Columns (1) and (4) include all households with at least two household members, whereas Columns (2) and (5) restrict the sample to households who always enroll either all or no adult family members. Columns (3) and (6) only include households who enroll some but not all adult family members at least once.

Column (1) shows that conditional on age, gender and health care needs unrelated to CVDs, a one-percent increase in the baseline odds of developing a CVD reduces enrollment between baseline and midline by 4.2 percentage points ('pps'). This is not statistically significant, meaning that selection on asymmetrically observed CVD risk is neither adverse nor advantageous. This is not to say that there is no adverse selection related to other healthcare needs; enrollment is higher among the elderly ($p < 0.01$), individuals who were ill or injured ($p < 0.10$), and women who are or will become pregnant before midline

Table 3: Health and follow-up expenditures in uninsured control district

| | Follow-up expenses on treatment most recent illness or injury in past 12 months (inv. hyper. sine) | | Follow-up expenses on chronic disease treatment (cumulative over past 12 months (inv. hyper. sine) | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Log odds total CVD risk | 0.041 | 0.029 | 0.254** | 0.124* |
| | (0.048) | (0.053) | (0.061) | (0.049) |
| Reports CV problem | | 0.173 | | 0.685** |
| | | (0.124) | | (0.159) |
| ..$X$ Log odds total CVD risk | | -0.059 | | 0.171* |
| | | (0.051) | | (0.068) |
| Log age | 0.005 | 0.030 | -0.173 | 0.092 |
| | (0.170) | (0.175) | (0.148) | (0.128) |
| Female | 0.163 | 0.107 | 0.833** | 0.427* |
| | (0.178) | (0.189) | (0.205) | (0.170) |
| Had acute illness | 0.175* | 0.165* | 0.032 | 0.011 |
| | (0.075) | (0.075) | (0.044) | (0.041) |
| Gets pregnant | 0.052 | 0.053 | -0.007 | -0.030 |
| | (0.074) | (0.073) | (0.041) | (0.038) |
| Midline to endline | 0.109+ | 0.094 | 0.008 | -0.041 |
| | (0.060) | (0.061) | (0.049) | (0.050) |
| $p$-val. Log odds risk \| Reports problem | | 0.653 | | 0.001 |
| Location effects | Yes | Yes | Yes | Yes |
| Observations | 1122 | 1122 | 1122 | 1122 |
| R-squared | 0.021 | 0.024 | 0.100 | 0.134 |
| Mean expenses (inv. hyper sine) | 0.388 | 0.388 | 0.247 | 0.247 |

Notes: Analysis sample includes all households with baseline (midline) CVD risk factors and health expenses prior to midline (endline) observed for at least two adult family members. Someone reporting a cardiovascular (CV) problem was ever diagnosed with hypertension, diabetes or heart disease. Estimated using linear least squares, with the inverse number of baseline (midline) adult family members included as a weight. All analyses control for location effects. Standard errors in parentheses are clustered by census area. $^{\dagger}$ $p < 0.1$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$.

Table 4: Total CVD risk and subsequent enrollment in health insurance

| | Dependent variable: Individual will enroll | | | | | |
|---|---|---|---|---|---|---|
| | Enrollment between baseline and midline | | | Enrollment between midline and endline | | |
| | All households | Never partially enrolling | Ever partially enrolling | All households | Never partially enrolling | Ever partially enrolling |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Log odds total CVD risk | -0.042 | 0.006 | -0.086* | 0.042* | 0.018 | 0.084** |
| | (0.028) | (0.036) | (0.039) | (0.019) | (0.031) | (0.025) |
| Log age | 0.236** | -0.009 | 0.474** | -0.130+ | -0.078 | -0.214* |
| | (0.083) | (0.121) | (0.111) | (0.076) | (0.115) | (0.090) |
| Female | -0.122 | -0.002 | -0.231 | 0.160* | 0.034 | 0.358** |
| | (0.100) | (0.127) | (0.142) | (0.066) | (0.103) | (0.091) |
| Had acute illness | 0.061+ | 0.054 | 0.060 | 0.115** | 0.055 | 0.167** |
| | (0.035) | (0.047) | (0.040) | (0.032) | (0.043) | (0.045) |
| Gets pregnant | 0.138** | 0.098+ | 0.179* | 0.208** | 0.185** | 0.206* |
| | (0.047) | (0.054) | (0.072) | (0.049) | (0.059) | (0.082) |
| Enrolled before midline | | | | 0.235** | 0.288* | 0.220** |
| | | | | (0.047) | (0.112) | (0.046) |
| HH member was enrolled | | | | 0.058 | 0.089 | -0.165** |
| | | | | (0.055) | (0.126) | (0.048) |
| Location effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1221 | 630 | 591 | 1186 | 619 | 567 |
| Households | 505 | | | | | |
| R-squared | 0.090 | 0.171 | 0.069 | 0.169 | 0.284 | 0.101 |
| Mean enrollment | 0.567 | 0.569 | 0.563 | 0.464 | 0.436 | 0.498 |

Notes: Columns (1)-(3) include observations of enrollment from baseline to midline. Columns (4)-(6) include observations of enrollment from midline to endline. Columns (1) and (4) include all households with current CVD risk and enrollment between the current and next survey observed for at least two adult family members. Columns (2) and (5) restrict the sample to households that either enroll all or no adult household members from baseline to midline, and from midline to endline. Columns (3) and (6) restrict the sample to households in which some but not all family members will enroll between baseline and midline, or between midline and endline. Estimated using a linear probability model, with the inverse number of adult family members in the current survey included as a weight. Standard errors in parentheses are clustered by census area. $^{\dagger}$ $p < 0.1$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$.

($p < 0.01$).

Columns (2) and (3) split the sample of individuals by household type, focusing on households that never enroll partially versus households that enroll partially at least once. The latter household type may have opted to enroll some but not all adult members for financial reasons and we would expect stronger adverse selection in this sample. Nonetheless, selection on CVD risk is advantageous: a one-percent increase in the odds of developing a CVD reduces enrollment by 8.6 pps ($p < 0.05$). By contrast, selection on characteristics associated with non-CVD-related healthcare needs, including age and pregnancies, is primarily concentrated within partially enrolling households. Only selection on acute health shocks does not vary between the two household types.

Columns (4) to (6) estimate the same models for enrollment from midline to endline. In Column (4), a one-percent increase in the odds of developing a CVD increases enrollment by 4.2 pps ($p < 0.05$). Columns (5) and (6) show that this effect is concentrated within partially enrolling households, where a one-percent increase in midline risk raises enrollment between midline and endline by 8.4 pps ($p < 0.01$). These are large effects, given that a one-percent increase in the odds of developing a CVD raises chronic health expenditures by 25.4 percent. Interestingly, selection on age is no longer adverse, but advantageous. In response to adverse selection on age from baseline to midline, the HCHC may have started discouraging enrollment of older family members. Program officers confirm introducing such screening at midline, but not as a formal restriction.

### 4.2 Adverse selection and a lack of awareness

Limited selection - at times advantageous - on total CVD risk could be due to a lack of knowledge regarding a person's CV health status. To test this hypothesis, Table 5 estimates Equation (2), estimating adverse selection for individuals who do versus do not report CV

health problems. Following the structure of Table 4, we estimate this model for enrollment from baseline to midline in Columns (1)-(3), and for enrollment from midline to endline in Columns (4)-(6). Equation (2) includes the same control variables as Equation (**??**). However, for brevity, we do not report the estimated coefficients for these variables from Table 5 onwards.

Only few individuals (2.5 percent, or 31 individuals) report CV health problems at baseline. Thus, from baseline to midline, Equations (1) and (2) - estimated in Tables 4 and 5, respectively - yield similar estimates of the coefficients for the log odds total CVD risk score. Reporting CV health problems and its interaction with the log odds CV risk score are unrelated to subsequent enrollment. This may in part be due to a low sample size. Given that the number of individuals reporting CV health problems is very low, privately observed CVD risk cannot be a major consideration in households' health insurance decisions.

Columns (4) to (6) present a different pattern for enrollment between midline and endline. In Column (4), among individuals who *do not* report CV health problems at midline, CVD risk and subsequent enrollment are uncorrelated. Among individuals who *do* report such health problems, the correlation between CV risk and subsequent enrollment is significantly higher. Increasing their odds of developing a CVD by one percent raises enrollment by an additional 5.2 pps ($p < 0.05$). Further, evaluated at the average log odds risk score, individuals reporting CV health problems are 11.9 percentage points more likely to enroll compared to individuals not reporting CV health problems ($p < 0.01$). Columns (5) and (6) show that these differences are concentrated in partially enrolling households.[19]

Table 6 tests whether adverse selection on total CVD risk between midline and endline is

[19] Appendix Table A4 formally tests whether selection patterns between baseline and midline differ from selection patterns between midline and endline. Selection on CVD risk is significantly more adverse from midline to endline only in partially enrolling households (a difference of 10.5 pps). In this round, selection is more adverse for both individuals who do not and individuals who do report CV health problems (a difference of 12.6 pps and 20.7 pps, respectively).

Table 5: Total CVD risk, reporting CV health problems, and subsequent enrollment

| | Dependent variable: Individual will enroll | | | | | |
|---|---|---|---|---|---|---|
| | Enrollment between baseline and midline | | | Enrollment between midline and endline | | |
| | All households | Never partially enrolling | Ever partially enrolling | All households | Never partially enrolling | Ever partially enrolling |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Log odds total CVD risk | -0.050 | -0.004 | -0.086* | 0.010 | -0.005 | 0.044 |
| | (0.030) | (0.039) | (0.040) | (0.027) | (0.037) | (0.034) |
| Reports CV health problem | 0.003 | 0.094 | -0.134 | 0.119** | 0.066 | 0.194** |
| | (0.079) | (0.082) | (0.149) | (0.040) | (0.063) | (0.072) |
| ...$X$ Log odds total CVD risk | 0.015 | 0.012 | 0.006 | 0.052* | 0.033 | 0.091* |
| | (0.039) | (0.044) | (0.063) | (0.020) | (0.022) | (0.036) |
| $p$-val. Log odds CVD risk \| Reports CV health problem | 0.455 | 0.884 | 0.270 | 0.068 | 0.538 | 0.002 |
| Controls 1 | Yes | Yes | Yes | Yes | Yes | Yes |
| Location effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1221 | 630 | 591 | 1186 | 619 | 567 |
| R-squared | 0.090 | 0.172 | 0.069 | 0.175 | 0.287 | 0.113 |
| Mean enrollment | 0.567 | 0.569 | 0.563 | 0.464 | 0.436 | 0.498 |

Notes: Columns (1)-(3) include observations of enrollment from baseline to midline. Columns (4)-(6) include observations of enrollment from midline to endline. Columns (1) and (4) include all households with current CVD risk and enrollment between the current and next survey observed for at least two adult family members. Columns (2) and (5) restrict the sample to households that either enroll all or no adult household members from baseline to midline, and from midline to endline. Columns (3) and (6) restrict the sample to households in which some but not all family members will enroll between baseline and midline, or between midline and endline. Controls 1: Female, Log age, Had acute illness, Gets pregnant. Estimated using a linear probability model, with the inverse number of adult family members in the current survey included as a weight. Standard errors in parentheses are clustered by census area. $^{\dagger}$ $p < 0.1$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$.

driven by selection in first-time enrollment, selective renewal of insurance policies, or both. One might worry that past enrollment influences both awareness and future enrollment, creating a spurious estimate of adverse selection. The table therefore estimate Equations (1) in Panel A and (2) in Panel B for two groups of individuals: those who did not enroll before midline in Columns (1)-(3), versus those who did enroll before midline in Columns (4)-(6). Selection between midline and endline among the first type of individual can be interpreted as selection in first-time enrollment, whereas for the second type, this can be interpreted as selection in renewal decisions.

Columns (1)-(3) show that total CVD risk does not significantly affect first-time enrollment, independent of whether the individual reports CV health problems. Panel B, Column (3), focusing on partially enrolling households, is the only model to find (weakly significant) higher enrollment among individuals reporting CV health problems. Thus, first-time enrollment faces some adverse selection on self-reported CVD risk, but no significant selection on measured CVD risk.

Columns (4)-(6) test for selective renewal among individuals who enrolled between baseline and midline. In Column (4) Panel A, a one-percent increase in the log odds of developing a CVD increases enrollment by 4.1 pps ($p < 0.10$). Panel B shows that this effect is concentrated among individuals who report CV health problems. At the average risk score, they are 11.4 percentage points more likely to enroll ($p < 0.05$), and a one-percent increase in the log odds of developing a CVD increases enrollment by an additional 5.8 percentage points ($p < 0.01$). This effect is again concentrated within partially enrolling households.

Nonetheless, comparing the estimated coefficients in Columns (3) and (6), we do not observe significantly stronger selection in renewal compared to first-time enrollment. Both first-time enrollment and renewal contribute to the estimated adverse selection in partially

enrolling households, but through different channels; selection in first-time enrollment is mainly driven by self-reported CV health problems, whereas selective renewal is due to an interaction of self-reported CV health problems and measured total CVD risk. These findings suggest a direct causal link from awareness to selection, instead of past enrollment influencing renewal and awareness at the same time.

### 4.3 Risk-reducing characteristics and subsequent enrollment

A second hypothesis is that individuals select into health insurance on characteristics associated with risk-reducing behaviors, for instance risk aversion, wealth or education. Alternatively, these characteristics are associated with increased awareness and increased enrollment. Table 7 therefore estimates Equation (3), controlling for potential confounds at the individual level in Panel A, and at the household level in Panel B. Panel A controls for observed individual characteristics that are potentially related to risk-reducing behaviors, including income, savings, willingness to take risk, the rank within the household and the level of education. The second panel controls for household fixed effects, in order to control for unobserved risk-reducing behaviors that occur at the household level.

Columns (1)-(3) show that these potential sources of advantageous selection cannot for limited adverse selection from baseline to midline. Selection on measured CVD risk is zero in both Panels A and B, irrespective of whether the individual reports cardiovascular health problems. Further, Columns (4)-(6) show that the estimated coefficients on the log odds of the CVD risk score between midline and endline are robust to the inclusion of individual characteristics as control variables in Panel A, and the inclusion of household fixed effects in Panel B. Including household fixed effects reduces precision in the interaction between the CVD risk and reporting CV health problems, but the size of the coefficient does not change substantially.

Table 6: Total CVD risk and first-time enrollment versus renewal of insurance

| | Dependent variable: Individual will enroll between midline and endline | | | | | |
|---|---|---|---|---|---|---|
| | Individual did not enroll before midline | | | Individual did enroll before midline | | |
| | All households | Never partially enrolling | Ever partially enrolling | All households | Never partially enrolling | Ever partially enrolling |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Total CVD risk and subsequent enrollment** | | | | | | |
| Log odds total CVD risk | 0.015 | -0.009 | 0.079 | 0.041$^{+}$ | 0.012 | 0.072$^{*}$ |
| | (0.048) | (0.053) | (0.080) | (0.023) | (0.036) | (0.031) |
| R-squared | 0.104 | 0.235 | 0.107 | 0.077 | 0.112 | 0.072 |
| **B. Total CVD risk, reported CV health problems, and subsequent enrollment** | | | | | | |
| Log odds total CVD risk | -0.003 | -0.003 | 0.010 | 0.002 | -0.027 | 0.040 |
| | (0.049) | (0.055) | (0.086) | (0.036) | (0.045) | (0.044) |
| Reports CV health problem | 0.084 | -0.041 | 0.361$^{+}$ | 0.114$^{*}$ | 0.069 | 0.118 |
| | (0.107) | (0.082) | (0.212) | (0.043) | (0.074) | (0.081) |
| ...$X$ Log odds total CVD risk | 0.032 | -0.011 | 0.002 | 0.058$^{**}$ | 0.042 | 0.116$^{**}$ |
| | (0.048) | (0.023) | (0.093) | (0.021) | (0.025) | (0.033) |
| R-squared | 0.107 | 0.236 | 0.128 | 0.085 | 0.118 | 0.092 |
| $p$-val. Log odds CVD risk \| Reports CV health problem | 0.678 | 0.805 | 0.919 | 0.152 | 0.780 | 0.002 |
| Controls 1 | Yes | Yes | Yes | Yes | Yes | Yes |
| Location effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 446 | 235 | 211 | 740 | 384 | 356 |
| Mean enrollment | 0.260 | 0.159 | 0.387 | 0.583 | 0.601 | 0.563 |

Notes: Analyses only include observations of enrollment from midline to endline. Columns (1) and (4) include all households with midline CVD risk and enrollment between midline and endline observed for at least two adult family members. Columns (2) and (5) restrict the sample to households that either enroll all or no adult household members from baseline to midline, and from midline to endline. Columns (3) and (6) restrict the sample to households in which some but not all family members will enroll between baseline and midline, or between midline and endline. Controls 1: Female, Log age, Had acute illness, Gets pregnant. Estimated using a linear probability model, with the inverse number of adult family members in the current survey included as a weight. Standard errors in parentheses are clustered by census area. $^{\dagger}$ $p < 0.1$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$.

Table 7: Controlling for individual- and household-level confounds of adverse selection

| | Dependent variable: Individual will enroll | | | | | |
|---|---|---|---|---|---|---|
| | Enrollment between baseline and midline | | | Enrollment between midline and endline | | |
| | All households | Never partially enrolling | Ever partially enrolling | All households | Never partially enrolling | Ever partially enrolling |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Controlling for individual characteristics ('Controls 2')** | | | | | | |
| Log odds total CVD risk | -0.039 | 0.004 | -0.071$^{\dagger}$ | 0.012 | -0.003 | 0.039 |
| | (0.030) | (0.039) | (0.042) | (0.027) | (0.038) | (0.033) |
| Reports CV health problem | 0.122 | 0.143 | -0.005 | 0.249$^{**}$ | 0.185$^{\dagger}$ | 0.387$^{**}$ |
| | (0.170) | (0.143) | (0.309) | (0.077) | (0.096) | (0.133) |
| ... $X$ Log odds total CVD risk | 0.017 | 0.017 | 0.002 | 0.051$^{*}$ | 0.035 | 0.090$^{*}$ |
| | (0.041) | (0.046) | (0.066) | (0.020) | (0.022) | (0.036) |
| R-squared | 0.101 | 0.189 | 0.078 | 0.177 | 0.296 | 0.123 |
| **B. Controlling for household fixed effects** | | | | | | |
| Log odds total CVD risk | -0.020 | | -0.057 | 0.022 | | 0.055 |
| | (0.024) | | (0.047) | (0.020) | | (0.045) |
| Reports CV health problem | -0.194 | | -0.379 | 0.110 | | 0.272 |
| | (0.122) | | (0.265) | (0.081) | | (0.182) |
| ... $X$ Log odds total CVD risk | -0.020 | | -0.047 | 0.031 | | 0.076 |
| | (0.030) | | (0.085) | (0.019) | | (0.050) |
| R-squared | 0.091 | | 0.164 | 0.082 | | 0.136 |
| Controls 1 | Yes | Yes | Yes | Yes | Yes | Yes |
| Location effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1221 | 630 | 591 | 1186 | 619 | 567 |
| Number of housholds | 505 | 281 | 224 | 488 | 277 | 224 |
| Mean enrollment | 0.567 | 0.569 | 0.563 | 0.464 | 0.436 | 0.498 |

Notes: Columns (1)-(3) include observations of enrollment from baseline to midline. Columns (4)-(6) include observations of enrollment from midline to endline. Columns (1) and (4) include all households with current CVD risk and enrollment between the current and next survey observed for at least two adult family members. Columns (2) and (5) restrict the sample to households in which either all or no family members will enroll between baseline and midline, and between midline and endline. Columns (3) and (6) restrict the sample to households in which some but not all family members will enroll between baseline and midline, or between midline and endline. Controls 1: Female, Log age, Had acute illness, Gets pregnant. Controls 2: Personal income, Personal savings, Willingness to take risks, Rank within household, Highest level of education completed. Estimated using a linear probability model, with the inverse number of adult family members in the current survey included as a weight. Standard errors in parentheses are clustered by census area. $^{\dagger}$ $p < 0.1$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$.

### 4.4 Robustness checks

We also analyze whether adverse selection between midline to endline for individuals reporting CV health problems is driven by one specific risk factor, for instance hypertension or diabetes. To that end, Table 8 estimates Equation (2), controlling for blood pressure, diabetic status and BMI in Panels A, B and C, respectively. We also interact these risk factors with the reported CV health problem indicator. Because few individuals reporting CV health problems are currently smoking, we do not include a panel estimating a similar model for current smoking status.

Each of the panels A, B and C find adverse selection on total CVD risk only between midline and endline, and especially among individuals who report CV health problems. This implies that our main finding is highly robust to including separate risk factors underlying the Framingham risk score and underscore the value of using a total CVD risk assessment approach. On the separate risk factors, we would not have observed a robust pattern of adverse selection. We conclude that our findings are not solely driven by selection on high blood pressure, diabetic status, or BMI alone, but result from an interaction of the different risk factors.

As a final step, Appendix Tables A5 and A6 provide a number of robustness checks for selection from baseline to midline in Columns (1)-(3), and selection from midline to endline in Columns (4)-(6). These tables presents estimates when including all households in the sample, including partially enrolling households but also households in which either all or zero family members will enroll.

Appendix Table A5 tests whether omitting observations for whom we impute some CVD risk factors affects the main results. From baseline to midline in Columns (1)-(3), we cannot reject the hypothesis of zero selection on total CVD risk (or advantageous selection in partially enrolling households), consistent with the findings in Table 4. From midline to

Table 8: Selection by CVD risk factor

| | Enrollment baseline-midline | | | Enrollment midline-endline | | |
|---|---|---|---|---|---|---|
| | All households | Never partially enrolling | Ever partially enrolling | All households | Never partially enrolling | Ever partially enrolling |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Blood pressure** | | | | | | |
| Log systolic blood pressure | 0.154 | 0.003 | 0.323 | -0.014 | 0.139 | -0.198 |
| | (0.125) | (0.154) | (0.205) | (0.114) | (0.164) | (0.179) |
| ... $X$ Reports CV problem | -0.025 | -0.568 | 0.641 | -0.299 | -0.130 | -0.510 |
| | (0.562) | (0.685) | (0.915) | (0.298) | (0.405) | (0.388) |
| Log odds total CVD risk score | -0.078* | -0.003 | -0.150* | 0.014 | -0.023 | 0.073† |
| | (0.037) | (0.042) | (0.057) | (0.033) | (0.048) | (0.037) |
| ... $X$ Reports CV problem | 0.021 | 0.029 | 0.003 | 0.055** | 0.040† | 0.081* |
| | (0.039) | (0.034) | (0.068) | (0.020) | (0.021) | (0.035) |
| **B. Diabetic status** | | | | | | |
| Is diabetic | -0.096 | 0.043 | -0.326 | -0.027 | -0.225 | 0.115 |
| | (0.180) | (0.250) | (0.302) | (0.161) | (0.152) | (0.215) |
| ... $X$ Reports CV problem | -0.087 | -0.261 | 0.046 | 0.160 | 0.682** | -0.224 |
| | (0.311) | (0.355) | (0.445) | (0.202) | (0.192) | (0.274) |
| Log odds total CVD risk score | -0.049 | -0.005 | -0.079† | 0.007 | -0.008 | 0.039 |
| | (0.031) | (0.039) | (0.043) | (0.027) | (0.037) | (0.032) |
| ... $X$ Reports CV problem | 0.019 | 0.010 | 0.028 | 0.047* | 0.020 | 0.091* |
| | (0.037) | (0.040) | (0.068) | (0.021) | (0.021) | (0.036) |
| **C. Body-Mass Index** | | | | | | |
| Log BMI | 0.096 | 0.118† | -0.006 | 0.061 | 0.106 | -0.025 |
| | (0.060) | (0.064) | (0.080) | (0.102) | (0.130) | (0.134) |
| ... $X$ Reports CV problem | 0.097 | 0.540† | -0.271 | 0.143 | -0.183 | 0.397 |
| | (0.344) | (0.295) | (0.735) | (0.240) | (0.286) | (0.276) |
| Log odds total CVD risk score | -0.064* | -0.026 | -0.082† | -0.004 | -0.009 | 0.021 |
| | (0.030) | (0.040) | (0.044) | (0.028) | (0.037) | (0.040) |
| ... $X$ Reports CV problem | 0.019 | 0.022 | 0.004 | 0.056** | 0.033 | 0.106* |
| | (0.038) | (0.038) | (0.062) | (0.021) | (0.023) | (0.040) |
| Observations | 1221 | 630 | 591 | 1186 | 619 | 567 |

Notes: Sample includes all households with current CVD risk and subsequent enrollment observed for at least two adult family members. Estimated using a linear probability model (without household fixed effects and individual controls), with the inverse number of adult family members at midline included as a weight. All analyses control for location effects, our health-related control variables, and a binary variable indicating whether the individual reports a CV health problem. Standard errors in parentheses are clustered by census area. † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

endline in Columns (4)-(6), adverse selection on total CVD risk does occur, in particular among individuals who self-report CV health problems, consistent with our findings in Table 5.

Appendix Table A6 tests whether selection from baseline to midline differs from selection between midline and endline because of selective attrition. To that end, this table estimates our main equations including only individuals with baseline, midline and endline data available, so that the sample from baseline to midline is the same as the sample from midline to endline. Again the table finds results that are qualitatively comparable to our previous estimates; we only find evidence of adverse selection on total CVD risk from midline to endline, in particular among individuals who self-report CV health problems.

### 4.5 Determinants of increased awareness

The question, then, is why the percentage of individuals reporting CV health problems increases over time. Section 3.4 already showed that age is the main risk factor that increases over time. The resulting increase in CVD risk may explain some of the increase in awareness. This section explores three alternative explanations for increased awareness, distinguishing between three groups of individuals: *enrollees* in the program area, *non-enrollees* in the program area, and individuals in the *control* district where health insurance was unavailable.

First, health insurance coverage may improve awareness by removing barriers to seek health care and providing access to a diagnosis. Prior to enrollment, individuals were largely unaware of their high CVD risk (Hendriks et al., 2012). Having health insurance coverage increased their health care utilization (Bonfrer et al., 2015). If they were diagnosed with hypertension, diabetes or other CV health problems during one of these health visits, their awareness will have improved. This will affect only *enrollees* in the program area.

Second, the HCHC upgraded health care in the program area and organized health outreach events as part of its marketing strategy. Some events involved blood pressure readings to improve awareness. Upgraded health facilities and outreach events were accessible for non-enrollees, and may have affected knowledge of both *enrolleess and non-enrollees* in the program area. In the control district, the HCHC did not upgrade any health facilities or organize health outreach events. Thus, this channel cannot have affected individuals in the control area.

Third, individuals learned about their health status through the survey measurements. Individuals with a systolic blood pressure above 140 mmHg or a diastolic blood pressure over 90 mmHg were informed they were at risk and referred to the nearest health facility. Further, individuals participating in blood tests were invited to visit a counsellor, who informed them of their blood test results and again referred them for further counseling in the nearest health facility. Enrollees and non-enrollees in the program area as well as individuals in the control district may have learned whether they have hypertension or diabetes through this channel.

To explore which of these mechanisms most likely accounts for increased reporting of CV health problems, Table 9 estimates a linear probability model for reporting CV health problems in the next survey round. Columns (1) and (4) include both the program area and the control district. Columns (2) and (5) include observations from the program area only, whereas Columns (3) and (6) focus on the control district. The first three columns include the log odds of the total CVD risk score as a covariate, whereas the last three columns also include each of the risk factors separately. We control for current reporting of CV health problems, in order to focus on increases in CV reporting.

Column (1) shows that non-enrolling individuals are 6.5 pps less likely to start reporting a CV health problem in the program area compared to similar individuals in the control

Table 9: Determinants of increased reporting of hypertension, diabetes and CVDs

| | Dependent var.: Reports CV health problem in next survey round | | | | | |
|---|---|---|---|---|---|---|
| | Both areas (1) | Program area (2) | Control area (3) | Both areas (4) | Program area (5) | Control area (6) |
| Program area | -0.065** | | | -0.054** | | |
| | (0.014) | | | (0.014) | | |
| ...$X$ Will enroll | 0.056** | 0.058** | | 0.056** | 0.057** | |
| | (0.010) | (0.010) | | (0.010) | (0.010) | |
| ...$X$ Was enrolled | -0.033 | -0.031 | | -0.029 | -0.028 | |
| | (0.027) | (0.027) | | (0.027) | (0.027) | |
| Log odds total CVD risk | 0.087** | 0.075** | 0.114** | 0.011 | 0.002 | 0.031† |
| | (0.009) | (0.009) | (0.017) | (0.009) | (0.010) | (0.016) |
| Log systolic blood pressure | | | | 0.447** | 0.425** | 0.483** |
| | | | | (0.046) | (0.055) | (0.083) |
| Currently smokes | | | | -0.010 | 0.003 | -0.060 |
| | | | | (0.018) | (0.018) | (0.047) |
| High glucose | | | | 0.106* | 0.110* | 0.087 |
| | | | | (0.048) | (0.049) | (0.094) |
| Log BMI | | | | 0.048* | 0.050* | 0.043 |
| | | | | (0.023) | (0.023) | (0.058) |
| Log age | -0.114** | -0.098** | -0.150** | 0.014 | 0.030 | -0.025 |
| | (0.021) | (0.025) | (0.038) | (0.020) | (0.022) | (0.040) |
| Female | 0.341** | 0.301** | 0.426** | 0.085** | 0.053 | 0.158* |
| | (0.034) | (0.037) | (0.066) | (0.032) | (0.036) | (0.059) |
| Had acute illness | 0.004 | 0.020† | -0.031 | 0.007 | 0.023* | -0.025 |
| | (0.010) | (0.011) | (0.018) | (0.010) | (0.011) | (0.018) |
| Gets pregnant | -0.045** | -0.058** | -0.016 | -0.039** | -0.048** | -0.017 |
| | (0.011) | (0.011) | (0.023) | (0.010) | (0.011) | (0.022) |
| Observations | 3529 | 2407 | 1122 | 3529 | 2407 | 1122 |
| R-squared | 0.469 | 0.452 | 0.500 | 0.490 | 0.472 | 0.523 |
| Mean dep. var. | 0.147 | 0.134 | 0.175 | 0.147 | 0.134 | 0.175 |

Notes: Includes all households with current health and future enrollment (if program area) or health expenditures (if control area) observed for at least two adult family members, excluding individuals who currently report ever having been diagnosed with heart disease, hypertension or diabetes. Model is estimated using a linear probability model with weights for the inverse number of adult family members. Controls included in the regressions are a dummy indicating whether the individual reports CV health problems in the current survey round, observations from midline to endline, location effects, and an indicator for whether at least one household member was enrolled. Standard errors in parentheses are clustered by census area. † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

district ($p < 0.01$). This difference does not exist for individuals who do enroll in the program area; compared to individuals who never enroll, they are 5.6 pps more likely to start reporting a CV health problem ($p < 0.01$). Individuals who enroll between baseline and midline are also not significantly more likely to report a CV health problem at endline. A one-percent increase in the odds of developing a CVD increases the probability of reporting a CV health problem in the next survey round by 8.7 pps ($p < 0.01$). Acute illnesses and upcoming pregnancies are not significantly associated with increased reporting.

Columns (2) and (3) estimates the same model separately for the program area and control district, respectively. Coefficients are relatively stable. In both areas, total CVD risk increases the probability of reporting CV health problems in the next round. Increased health care utilization for acute illnesses and injuries in the program area appears to have weakly improved knowledge of CV health problems, but we do not replicate this effect for increased utilization associated with pregnancies. Thus, improved access to health care in general and outreach activities does not appear the main channel through which knowledge improved, rejecting the second hypothesis stated above.

Columns (4) to (6) estimate to what extent each CVD risk factor separately predicts improved CV reporting. In Column (5), the main determinants of improved reporting are systolic blood pressure, high glucose, and BMI, and combined, these risk factors eliminate the effect of total CVD risk on future awareness. Each of these risk factors were measured during the baseline and midline surveys. This suggests that the health exams and associated counseling improved awareness, providing evidence of the third channel discussed above. Regarding the first channel, we cannot disentangle whether individuals become aware because they enroll and are diagnosed with a CVD (treatment effect of insurance), or whether they enroll in health insurance because they become aware (adverse selection).

In sum, increased reporting of CV health problems from baseline to midline, and further

increases from midline to endline appear associated with diagnostic tests performed during the baseline and midline surveys. Increased health care access in the program area cannot explain improved awareness. Enrollees may have become more aware by utilizing more health care, or they may have enrolled because they became more aware.

## 5 Conclusion

This paper tested for adverse selection on cardiovascular disease (CVD) risk in a voluntary subsidized health insurance program. We hypothesized that observed selection may be limited for three reasons: limited awareness of cardiovascular health problems such as hypertension, diabetes or CVDs, selection on other characteristics that are associated with a lower risk, and high premium subsidies. We test these hypotheses building on a health insurance program in Nigeria, in which enrollment was individual-based, allowing us to control for both individual characteristics and unobserved household fixed effects potentially correlated with both enrollment and risk-reducing behaviors; and enabling us to zoom in on partially enrolling households, who reveal a more binding budget constraint and are more exemplary of household decisions in the face of reducing premium subsidies.

We find that initially, awareness is low and although households selectively enroll older household members, they are not more likely to enroll individuals with a higher total CVD risk score. Over time, total CVD risk becomes predictive of health insurance decisions, in particular among individuals who report having hypertension, diabetes or a CVD, and within households deciding to enroll only a few household members. We do not find stronger selection when controlling for either individual characteristics or household fixed effects. Thus, initially limited selection on CVD risk may well be related to a lack of awareness. Increased awareness - mostly due to health exams in household surveys - appears to result in more adverse selection. Finally, households who reveal a more binding budget constrained

by enrolling only part of the household are more selective in whom they enroll. This raises concerns with the sustainability of voluntary health insurance aimed at covering treatment of CVD risk factors in a context of increasing awareness and reducing premium subsidies.

Insurance programs can however potentially limit adverse selection by introducing alternative enrollment policies. The program studied in this paper allowed households to select individuals into the program. Health insurance programs typically impose family-based or group-based enrollment, engage in screening, or exclude coverage for pre-existing conditions out of fear for adverse selection when households have the option to enroll a few family members rather than the entire family (see Cutler and Zeckhauser, 2000, for a review). Our findings show that they may well be right in doing so to prevent adverse selection on CV risk.

Enrollment restrictions such as family-based or group-based enrollment can however make insurance premiums bulky, especially for poorer and larger households, even when premiums are heavily subsidized or discounted (Fink et al., 2013; Kusi et al., 2015). For such households, individual-based enrollment potentially enhances currently low take-up rates, but at the cost of increased adverse selection. Such barriers can potentially be removed by combining health insurance with improved credit or savings facilities. In the HCHC, one health facility offered households the option to pay the premium on credit, and reduced partial enrollment. A direction for future research is to explore the optimal design of partnering facilities' incentives to limit adverse selection and maximize enrollment. In addition, one can explore the effects of providing health insurance through MFIs.

To conclude, this study finds evidence of increasing adverse selection over time on CV risk factors beyond age. CVDs are increasingly a burden in low- and middle-income countries, but based on our findings, voluntary individual health insurance may not be the best strategy to finance treatment of CV risk factors such as hypertension and diabetes.

In addition, reducing selection on age over time may reflect increased screening on part of the insurance provider, in order to control costs. Health savings accounts and group-based health insurance programs, provided through existing institutions like producer groups or microfinance institutions, may be a more viable strategy to prevent CVDs among the informal poor.

## References


**Acharya, A., S. Vellakkal, F. Taylor, E. Massett, et al.** 2012. "The impact of health insurance schemes for the informal sector in low-and middle-income countries: A systematic review." *The World Bank Research Observer*, doi:10.1093/wbro/lks009.

**Addo, J., L. Smeeth, and D. A. Leon.** 2007. "Hypertension in sub-Saharan Africa: A systematic review." *Hypertension*, 50(6): 1012–1018.

**Banerjee, Abhijit, Esther Duflo, and Richard Hornbeck.** 2014. "Bundling Health Insurance and Microfinance in India: There Cannot be Adverse Selection if There is No Demand." *American Economic Review: Papers & Proceedings*, 104(5): 291–297.

**Bitton, Asaf, and Thomas Gaziano.** 2010. "The Framingham Heart Study's impact on global risk assessment." *Progress in cardiovascular diseases*, 53(1): 68–78.

**Bleich, Sara N, David M Cutler, Alyce S Adams, Rafael Lozano, and Christopher JL Murray.** 2007. "Impact of insurance and supply of health professionals on coverage of treatment for hypertension in Mexico: population based study." *Bmj*, 335(7625): 875.

**Bonfrer, Igna, Ellen Van de Poel, Emily Gustafsson-Wright, and Eddy Van Doorslaer.** 2015. "Effects of a subsidized voluntary health insurance on insured and uninsured in Nigeria." *BMG Working Paper 2015.01.*

**Chiappori, P., and B. Salanié.** 2000. "Testing for asymmetric information in insurance markets." *Journal of Political Economy*, 108(1): 56–78.

**Chiappori, P., B. Jullien, B. Salanie, and F. Salanie.** 2006. "Asymmetric information in insurance: General testable implications." *RAND Journal of Economics*, 37(4): 783–798.

**Cook, Christopher, Graham Cole, Perviz Asaria, Richard Jabbour, and Darrel P Francis.** 2014. "The annual global economic burden of heart failure." *International journal of cardiology*, 171(3): 368–376.

**Cooney, Marie Therese, Alexandra Dudina, Ralph D'Agostino, and Ian M Graham.** 2010. "Cardiovascular risk-estimation systems in primary prevention do they differ? Do they make a difference? Can we see the future?" *Circulation*, 122(3): 300–310.

**Cutler, D.M., and R.J. Zeckhauser.** 2000. "The anatomy of health insurance." In *Handbook of Health Economics*. Vol. 1, , ed. A. J. Culyer and J. Newhouse, Chapter 11, 563–643. Elsevier.

**D'Agostino, Ralph B, Ramachandran S Vasan, Michael J Pencina, Philip A Wolf, Mark Cobain, Joseph M Massaro, and William B Kannel.** 2008. "A General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study." *Circulation*, 117(6): 743–753.

**De Meza, D., and D.C. Webb.** 2001. "Advantageous selection in insurance markets." *RAND Journal of Economics*, 249–262.

**Doiron, D., G. Jones, and E. Savage.** 2008. "Healthy, wealthy and insured? The role of self-assessed health in the demand for private health insurance." *Health Economics*, 17(3): 317–334.

**Finkelstein, A., and K. McGarry.** 2006. "Multiple dimensions of private information: Evidence from the long-term care insurance market." *American Economic Review*, 96(4): 938.

**Fink, G., P. J. Robyn, A. Sié, and R. Sauerborn.** 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.

**Gaziano, Thomas A.** 2007. "Reducing the growing burden of cardiovascular disease in the developing world." *Health affairs*, 26(1): 13–24.

**Gaziano, Thomas A.** 2008. "Economic burden and the cost-effectiveness of treatment of cardiovascular diseases in Africa." *Heart*, 94(2): 140–144.

**Gnawali, D. P., S. Pokhrel, A. Sié, M. Sanon, M. De Allegri, A. Souares, H. Dong, and R. Sauerborn.** 2009. "The effect of community-based health insurance on the utilization of modern health care services: Evidence from Burkina Faso." *Health Policy*, 90(2): 214–222.

**Gustafsson-Wright, E., Z. Tanovic, and J. Van der Gaag.** 2013. "The impact of subsidized private health insurance and clinic upgrades on health care utilization and financial protection in rural Nigeria." *Amsterdam Institute for International Development Working Paper*.

**Hendriks, Marleen E, Nicole TA Rosendaal, Ferdinand WNM Wit, Oladimeji A Bolarinwa, Berber Kramer, Daniëlla Brals, Emily Gustafsson-Wright, Peju Adenusi, Lizzy M Brewster, Gordon K Osagbemi, et al.** 2016. "Sustained effect of health insurance and facility quality improvement on blood pressure in adults with hypertension in Nigeria: A population-based study." *International Journal of Cardiology*, 202: 477–484.

**Hendriks, Marleen E, Oladimeji A Bolarinwa, Heleen E Nelissen, Alexander C Boers, Gabriela B Gomez, Siok Swan Tan, William Redekop, Peju Adenusi, Joep MA Lange, Kayode Agbede, et al.** 2015. "Costs of cardiovascular disease prevention care and scenarios for cost saving: A micro-costing study from rural Nigeria." *Journal of Hypertension*, 33(2): 376–684.

**Hendriks, M.E., F.W.N.M. Wit, M.T.L. Roos, L.M. Brewster, T.M. Akande, et al.** 2012. "Hypertension in sub-Saharan Africa: Cross-sectional surveys in four rural and urban communities." *PLoS ONE*, 7(3): e32638.

**Hendriks, M. E., F.W.N.M. Wit, T.M. Akande, B. Kramer, G.K. Osagbemi, Z. Tanović, E. Gustafsson-Wright, L.M. Brewster, J.M.A. Lange, and C. Schultsz.** 2014. "Effect of health insurance and facility quality improvement on blood pressure in adults with hypertension in Nigeria: A population-based study." *JAMA Internal Medicine*, 174(4): 555–563.

**Janssens, Wendy, and Berber Kramer.** 2016. "Why wealthy, healthy and insured? Liquidity constraints and health-seeking behavior in Nigeria." *Mimeo.*

**Kengne, Andre Pascal, Zandile June-Rose Mchiza, Albert George Baidoe Amoah, and Jean-Claude Mbanya.** 2013. "Cardiovascular diseases and diabetes as economic and developmental challenges in Africa." *Progress in cardiovascular diseases*, 56(3): 302–313.

**Kusi, A., U. Enemark, K. S. Hansen, and F. A. Asante.** 2015. "Refusal to enrol in Ghana's National Health Insurance Scheme: Is affordability the problem?" *International Journal for Equity in Health*, 14(2).

**Lemogoum, Daniel, Jean-Paul Degaute, and Pascal Bovet.** 2005. "Stroke prevention, treatment, and rehabilitation in sub-saharan Africa." *American journal of preventive medicine*, 29(5): 95–101.

**Mendis, Shanthi, Lars H Lindholm, Giuseppe Mancia, Judith Whitworth, Michael Alderman, Stephen Lim, and Tony Heagerty.** 2007. "World Health Organization (WHO) and International Society of Hypertension (ISH) risk prediction charts: Assessment of cardiovascular risk for prevention and control of cardiovascular disease in low and middle-income countries." *Journal of Hypertension*, 25(8): 1578–1582.

**Nguyen, H., and J. Knowles.** 2010. "Demand for voluntary health insurance in developing countries: The case of Vietnams school-age children and adolescent student health insurance program." *Social Science & Medicine*, 71(12): 2074–2082.

**Parmar, D., A. Souares, M. de Allegri, G. Savadogo, and R. Sauerborn.** 2012. "Adverse selection in a community-based health insurance scheme in rural Africa: Implications for introducing targeted subsidies." *BMC Health Services Research*, 12(1): 181.

**Peer, Nasheeta, Carl Lombard, Krisela Steyn, Thomas Gaziano, and Naomi Levitt.** 2014. "Comparability of total cardiovascular disease risk estimates using laboratory and non-laboratory based assessments in urban-dwelling South Africans: The CRIBSA study." *SAMJ: South African Medical Journal*, 104(10): 621–696.

**Polimeni, R., and D.I Levine.** 2012. "Prices matter: Comparing two tests of adverse selection in health insurance." *UC Berkeley Institute for Research on Labor and Employment Working Paper 135-12*.

**Rothschild, M., and J. Stiglitz.** 1976. "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information." *Quarterly Journal of Economics*, 90(4): 629–649.

**Sosa-Rubí, Sandra G, Omar Galárraga, and Ruy López-Ridaura.** 2009. "Diabetes treatment and control: the effect of public health insurance for the poor in Mexico." *Bulletin of the World Health Organization*, 87(7): 512–519.

**Tversky, Amos, and Daniel Kahneman.** 1973. "Availability: A heuristic for judging frequency and probability." *Cognitive psychology*, 5(2): 207–232.

**Wang, H., L. Zhang, W. Yip, and W. Hsiao.** 2006. "Adverse selection in a voluntary Rural Mutual Health Care health insurance scheme in China." *Social Science & Medicine*, 63(5): 1236–1245.

**WHO.** 2014. "Global target 8: At least 50% of eligible people receive drug therapy and counselling (including glycaemic control) to prevent heart attacks and strokes." In *Global status report on noncommunicable diseases 2014.* , ed. World Health Organization, Chapter 8, 95–104. World Health Organization: Geneva.

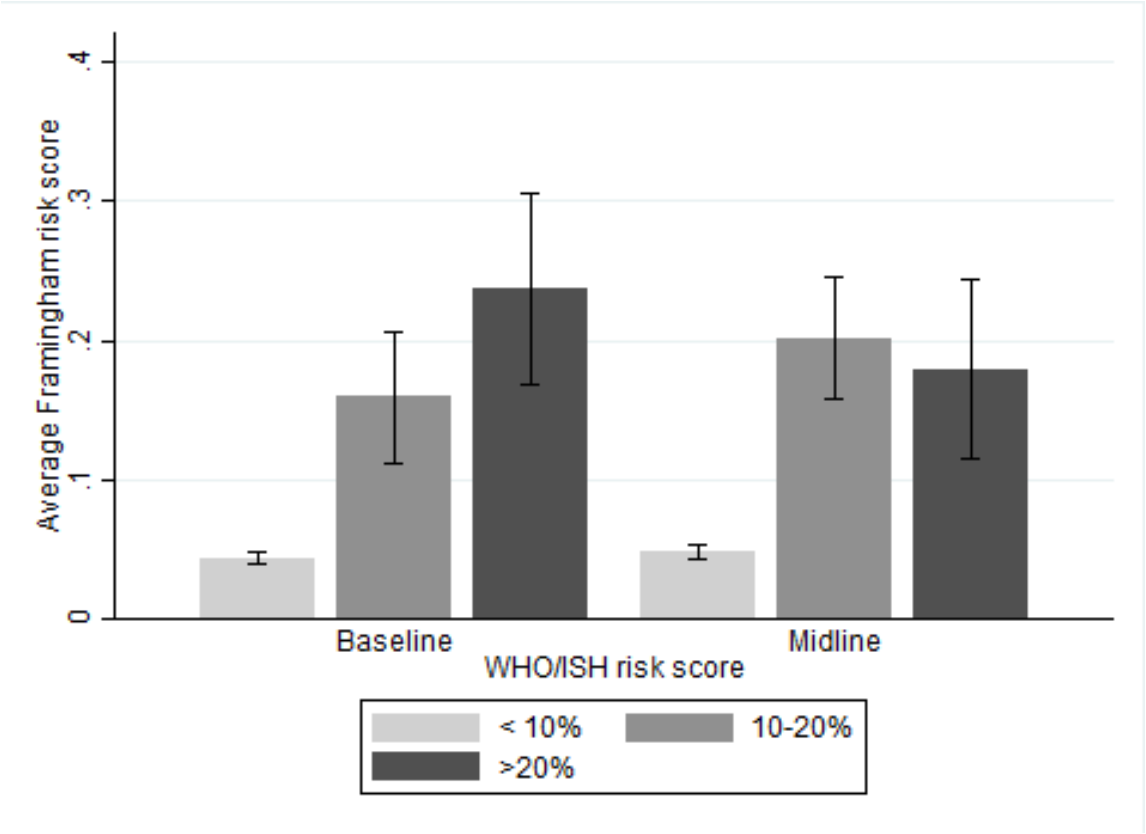
**Witman, A.** 2015. "Public health insurance and disparate eligibility of spouses: The Medicare eligibility gap." *Journal of Health Economics*, 40: 10–25.

**Zhang, L., and H. Wang.** 2008. "Dynamic process of adverse selection: Evidence from a subsidized community-based health insurance in rural China." *Social Science & Medicine*, 67(7): 1173–1182.

**Zhao, M., Y. Konishi, and P. Glewwe.** 2013. "Does information on health status lead to a healthier lifestyle? Evidence from China on the effect of hypertension diagnosis on food consumption." *Journal of Health Economics*, 32(2): 367–385.

# A Appendix

Figure A1: Framingham versus WHO/ISH risk scores




*Notes:* Sample includes individuals from the program area, including individuals from both baseline and midline. Only members from households with health and subsequent enrollment observed for at least two adult family members are included. The data is weighted by the inverse number of adult family members.

Table A1: Annual coverage of hypertension-related health care services

| | Complex hypertension | Non-complex hypertension | Initial / follow-up |
|---|---|---|---|
| # of doctor visits | 4 | 3 | 1 |
| # of specialist consultations | 1 | 0 | 0 |
| # of laboratory tests | 2 | 1 | 1 |
| # of drug collections | 12 | 12 | N/A |

*Notes:* Drugs for complex hypertension: Hydrochlorothiazide (25mg, 1 dly), Nifedipine (30mg, 3 dly), Atenolol 50mg (2 dly), to be purchased at a fee of 1,000 Naira. Drugs for non-complex hypertension: Hydrochlorothiazide (25mg, 1 dly), Nifedipine (30mg, 1.5 dly), Atenolol 50mg (1 dly), to be purchased at a fee of 700 Naira.

Table A2: Risk functions for 10-year risk of developing any CVD

| Variable | Beta | $p$-value | Hazard Ratio | 95% CI |
|---|---|---|---|---|
| **Men** (*10-year baseline survival:* $S_0(10) = 0.88431$) | | | | |
| Log of Age | 3.11296 | <.0001 | 22.49 | (14.8, 34.2) |
| Log of Body Mass Index | 0.79277 | <.0066 | 2.21 | (1.25, 3.91) |
| Log of SBP if not treated | 1.85508 | <.0001 | 6.39 | (3.61, 11.3) |
| Log of SBP if treated | 1.92672 | <.0001 | 6.87 | (3.90, 12.1) |
| Smoking | 0.70953 | <.0001 | 2.03 | (1.75, 2.37) |
| Diabetes | 0.53160 | <.0001 | 1.70 | (1.37, 2.11) |
| **Women** (*10-year baseline survival:* $S_0(10) = 0.94833$) | | | | |
| Log of Age | 2.72107 | <.0001 | 15.20 | (8.59, 26.9) |
| Log of Body Mass Index | 0.51125 | <.0609 | 1.67 | (0.98, 2.85) |
| Log of SBP if not treated | 2.81291 | <.0001 | 16.66 | (8.27, 33.5) |
| Log of SBP if treated | 2.88267 | <.0001 | 17.86 | (8.97, 35.6) |
| Smoking | 0.61868 | <.0001 | 1.86 | (1.53, 2.25) |
| Diabetes | 0.77763 | <.0001 | 2.18 | (1.63, 2.91) |

Source: Included from the Framingham Heart Study (D'Agostino et al., 2008). The 10-year risk for women can be calculated as $1 - 0.94833 \exp(\Sigma\beta X - 26.0145)$ where $\beta$ is the regression coefficient and $X$ is the level for each risk factor; the risk for men is given as $1 - 0.88431 \exp(\Sigma\beta X - 23.9388)$.

Table A3: Relation WHO and Framingham risk scores

| | Dependent variable: WHO/ISH 10-year total CVD risk score (ranges from 0 to 4) | | | | | |
|---|---|---|---|---|---|---|
| | Baseline | | | Midline | | |
| | All households | Never partially enrolling | Ever partially enrolling | All households | Never partially enrolling | Ever partially enrolling |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Total CVD risk and WHO/ISH score** | | | | | | |
| Log odds total CVD risk | 0.375** | 0.380** | 0.373** | 0.290** | 0.318** | 0.246** |
| | (0.040) | (0.059) | (0.069) | (0.038) | (0.037) | (0.063) |
| R-squared | 0.296 | 0.290 | 0.322 | 0.261 | 0.298 | 0.222 |
| **B. Total CVD risk, reported CV health problems, and WHO/ISH score** | | | | | | |
| Log odds total CVD risk | 0.384** | 0.405** | 0.362** | 0.252** | 0.280** | 0.203** |
| | (0.046) | (0.069) | (0.074) | (0.047) | (0.055) | (0.067) |
| Reports CV problem | 0.434** | 0.326* | 0.582** | 0.547** | 0.550** | 0.530** |
| | (0.103) | (0.125) | (0.180) | (0.074) | (0.086) | (0.122) |
| ... *X* Log odds total CVD risk | 0.021 | -0.000 | 0.056 | 0.060 | 0.077 | 0.028 |
| | (0.047) | (0.067) | (0.060) | (0.039) | (0.056) | (0.047) |
| *p*-value Log odds CVD risk \| Reports CV problem | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.009 |
| R-squared | 0.299 | 0.299 | 0.324 | 0.269 | 0.309 | 0.230 |
| Location effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls 1 | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1221 | 630 | 591 | 1186 | 619 | 567 |
| Mean WHO/ISH score | 0.111 | 0.123 | 0.096 | 0.142 | 0.151 | 0.130 |

Notes: Columns (1)-(3) include households with baseline health and enrollment between baseline and midline observed for at least two adult family members. Columns (4)-(6) include households with midline health and enrollment between midline and endline observed for at least two adult family members. The analyses omit individuals with data on health or subsequent enrollment missing. Estimated using a linear regression model, with the inverse number of adult family members included as a weight. Standard errors in parentheses are clustered by census area. $^{\dagger}$ $p < 0.1$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$

Table A4: Change in estimated coefficients over time

| | Dependent variable: Individual will enroll between current and follow-up round | | | | | |
|---|---|---|---|---|---|---|
| | All households | Never partially enrolling | Ever partially enrolling | All households | Never partially enrolling | Ever partially enrolling |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Log odds total CVD risk | -0.047 | -0.006 | -0.078$^{+}$ | -0.050 | -0.004 | -0.086$^{*}$ |
| | (0.029) | (0.038) | (0.041) | (0.030) | (0.039) | (0.040) |
| ... $X$ Midline to endline | 0.055 | 0.005 | 0.105$^{+}$ | 0.058 | -0.002 | 0.126$^{*}$ |
| | (0.043) | (0.058) | (0.054) | (0.044) | (0.058) | (0.054) |
| Reports CV health problem | | | | 0.003 | 0.094 | -0.134 |
| | | | | (0.079) | (0.082) | (0.149) |
| ... $X$ Midline to endline | | | | 0.122 | -0.028 | 0.340$^{+}$ |
| | | | | (0.091) | (0.109) | (0.173) |
| Log risk $X$ Reports problem | | | | 0.015 | 0.012 | 0.006 |
| | | | | (0.039) | (0.044) | (0.063) |
| ... $X$ Midline to endline | | | | 0.036 | 0.020 | 0.081 |
| | | | | (0.048) | (0.054) | (0.072) |
| Location, Controls 1 | Yes | Yes | Yes | Yes | Yes | Yes |
| ... X Endline | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2407 | 1249 | 1158 | 2407 | 1249 | 1158 |
| R-squared | 0.137 | 0.240 | 0.083 | 0.141 | 0.242 | 0.095 |
| Mean enrollment | 0.516 | 0.504 | 0.531 | 0.516 | 0.504 | 0.531 |

Notes: Sample includes all households with current CVD risk factors and enrollment between current and follow-up round observed for at least two adult family members. Someone reporting a cardiovascular (CV) problem was ever diagnosed with hypertension, diabetes or heart disease. Estimated using a linear probability model, with the inverse number of adult family members at baseline included as a weight. All analyses control for location effects. All variables are interacted with a dummy indicating observations from midline to endline. Standard errors in parentheses are clustered by census area. $^{\dagger}$ $p < 0.1$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$.

Table A5: Selection among individuals without imputed CV risk score

| | Dependent variable: Individual will enroll between current and follow-up round | | | | | |
|---|---|---|---|---|---|---|
| | Enrollment between baseline and midline | | | Enrollment between midline and endline | | |
| | All households | Never partially enrolling | Ever partially enrolling | All households | Never partially enrolling | Ever partially enrolling |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Total CVD risk and subsequent enrollment** | | | | | | |
| Log odds total CVD risk | -0.047 | -0.011 | -0.075$^{\dagger}$ | 0.011 | 0.001 | 0.031 |
| | (0.029) | (0.038) | (0.041) | (0.028) | (0.039) | (0.035) |
| R-squared | 0.103 | 0.187 | 0.071 | 0.173 | 0.289 | 0.097 |
| **B. Total CVD risk, reported CV health problems, and subsequent enrollment** | | | | | | |
| Log odds total CVD risk | -0.049 | -0.010 | -0.083$^{*}$ | 0.007 | -0.006 | 0.037 |
| | (0.030) | (0.039) | (0.041) | (0.028) | (0.038) | (0.035) |
| Reports CV health problem | -0.002 | 0.079 | -0.130 | 0.104$^{*}$ | 0.046 | 0.191$^{*}$ |
| | (0.078) | (0.079) | (0.148) | (0.040) | (0.063) | (0.074) |
| ... $X$ Log odds total CVD risk | 0.017 | 0.014 | 0.006 | 0.054$^{*}$ | 0.031 | 0.098$^{**}$ |
| | (0.039) | (0.045) | (0.063) | (0.020) | (0.023) | (0.036) |
| R-squared | 0.103 | 0.188 | 0.072 | 0.181 | 0.291 | 0.122 |
| $p$-val. Log odds CVD risk \| Reports CV health problem | 0.477 | 0.942 | 0.288 | 0.080 | 0.584 | 0.002 |
| Location effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls 1 | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1174 | 607 | 567 | 1112 | 586 | 526 |
| Mean enrollment | 0.569 | 0.573 | 0.563 | 0.473 | 0.445 | 0.507 |

Notes: Columns (1)-(3) include households with baseline cardiovascular (CV) health and enrollment before midline observed for at least two adult family members. Columns (4)-(6) include households with midline cardiovascular (CV) health and enrollment before endline observed for at least two adult family members. Someone reporting a CV health problem was ever diagnosed with hypertension, diabetes or heart disease. Estimated using a linear probability model, with the inverse number of adult family members in the current survey included as a weight. All analyses control for location effects. Individual controls include personal income in the last 7 days, personal savings, risk aversion on a 7-point Likert scale, rank within the household, and a categorical variable indicating highest level of completed education (measured in the current, not follow-up, survey round). Standard errors in parentheses are clustered by census area. $^{\dagger}$ $p < 0.1$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$.

Table A6: Selection among individuals observed in all survey rounds

| | Dependent variable: Individual will enroll between current and follow-up round | | | | | |
|---|---|---|---|---|---|---|
| | All households | Never partially enrolling | Ever partially enrolling | All households | Never partially enrolling | Ever partially enrolling |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Total CVD risk and subsequent enrollment** | | | | | | |
| Log odds total CVD risk | -0.020 | -0.002 | -0.027 | 0.019 | 0.017 | 0.038 |
| | (0.029) | (0.034) | (0.046) | (0.033) | (0.044) | (0.039) |
| R-squared | 0.105 | 0.219 | 0.032 | 0.179 | 0.314 | 0.089 |
| **B. Total CVD risk, reported CV health problems, and subsequent enrollment** | | | | | | |
| Log odds total CVD risk | -0.032 | -0.008 | -0.046 | 0.011 | 0.011 | 0.032 |
| | (0.029) | (0.033) | (0.052) | (0.033) | (0.042) | (0.039) |
| Reports CV health problem | -0.109 | -0.033 | -0.227 | 0.135** | 0.110† | 0.174* |
| | (0.080) | (0.086) | (0.152) | (0.045) | (0.063) | (0.078) |
| ... $X$ Log odds total CVD risk | 0.036 | 0.028 | 0.024 | 0.061* | 0.046† | 0.082† |
| | (0.046) | (0.052) | (0.069) | (0.023) | (0.026) | (0.043) |
| R-squared | 0.107 | 0.220 | 0.038 | 0.191 | 0.322 | 0.107 |
| $p$-val. Log odds CVD risk \| Reports CV health problem | 0.939 | 0.766 | 0.787 | 0.066 | 0.273 | 0.012 |
| Location effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls 1 | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 966 | 498 | 468 | 966 | 498 | 468 |
| Mean enrollment | 0.631 | 0.635 | 0.627 | 0.454 | 0.422 | 0.492 |

Notes: Columns (1)-(3) include households with baseline cardiovascular (CV) health and enrollment before midline observed for at least two adult family members. Columns (4)-(6) include households with midline cardiovascular (CV) health and enrollment before endline observed for at least two adult family members. Someone reporting a CV health problem was ever diagnosed with hypertension, diabetes or heart disease. Estimated using a linear probability model, with the inverse number of adult family members in the current survey included as a weight. All analyses control for location effects. Individual controls include personal income in the last 7 days, personal savings, risk aversion on a 7-point Likert scale, rank within the household, and a categorical variable indicating highest level of completed education (measured in the current, not follow-up, survey round). Standard errors in parentheses are clustered by census area. $^{\dagger}$ $p < 0.1$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$.