The Impact of Health on Labor Market Outcomes: Experimental Evidence from MRFIT*

Melvin Stephens Jr.
University of Michigan and NBER

Desmond Toohey
University of Delaware

This Version: May 23, 2018

*Stephens: Department of Economics, University of Michigan, 611 Tappan St., Ann Arbor, MI, 48109, e-mail: mstep@umich.edu and National Bureau of Economic Research, Cambridge, MA. Toohey: Department of Economics, University of Delaware, 420 Purnell Hall, Newark, DE, 19716, e-mail: dtoohey@udel.edu. We thank seminar participants at the Michigan Labor Lunch, the NBER Labor Studies and Health Economics meetings, Wisconsin CDE & CDHA, and the H2D2 conference for helpful comments and suggestions. Research reported in this publication was supported by the National Institute On Aging of the National Institutes of Health under Award Number R03AG046503. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The Impact of Health on Labor Market Outcomes:

Experimental Evidence from MRFIT

Abstract

While economists have posited that health investments increase earnings, isolating the causal effect of health is challenging due both to reverse causality and unobserved heterogeneity. We examine the labor market effects of a randomized controlled trial, the Multiple Risk Factor Intervention Trial (MRFIT), which monitored nearly 13,000 men for over six years. We find that this intervention, which provided a bundle of treatments to reduce coronary heart disease mortality, increased earnings and family income. We find few differences in estimated gains by baseline health and occupation characteristics. Reductions in serious illnesses and work-limiting disabilities likely contributed to the observed gains.

1 Introduction

Economists have long recognized a strong connection between health and economic outcomes. Measures of health, both self-reported and objectively measured, are positively associated with human capital, earnings, income, and wealth. The direction of causality in these relationships is unclear. Better health can lead to higher productivity, less working time lost to illness, and lower mortality, which further incentivizes human capital investment. Higher productivity and financial resources can facilitate access to care, avoidance of harmful environmental factors, and access to higher-quality food and drugs.

In this paper, we examine the impact of health on earnings and income using data from the Multiple Risk Factor Intervention Trial (MRFIT), a randomized trial studying the combined effects of multiple health interventions aimed at reducing coronary heart disease (CHD) mortality risk in the United States. MRFIT began screening men aged 35 to 57 in 1974 with high risk for CHD mortality. After three rounds of screening, nearly 13,000 study participants were randomized into two treatment groups. A "Special Intervention" (SI) group received interventions aimed at lowering cholesterol, lowering blood pressure, and quitting smoking, while a "Usual Care" (UC) group was instructed to continue seeking standard medical care in the community. The MRFIT intervention substantially reduced CHD risk throughout the six (or more) years that each participant was monitored.

We find that the improved health of the SI group raises earnings by three percent and total family income by four percent when measured six years after enrollment. These differences are not driven by effects on employment or labor force participation as we find that the treatment has no effect on these outcomes. The results are robust to multiple forms of sample selection including survey attrition and labor force exit. Consistent with existing research on MRFIT, we find no differences in mortality at this horizon and only small, statistically insignificant differences at longer horizons (MRFIT Research Group 1982, 1990, 1996). While MRFIT has been examined extensively, particularly in the medical and epidemiological literatures, this paper offers, to the best of our knowledge, the first examination of the effects of the intervention on earnings, income, and other labor market outcomes.

A key contribution of this paper is that it isolates the impact of health on earnings and family income while avoiding concerns about reverse causality and unobserved heterogeneity. Surveys of the literature examining the relationship between health and labor market outcomes (Currie and Madrian 1999; O'Donnell, Doorslaer, and Van Ourti 2015) note that identification typically relies on estimation strategies with "dubious" exclusion restrictions or, more recently, within-person variation in health over time which also requires strong exogeneity assumptions. Recent work in developing countries uses randomized controlled trials to study the impact of health on labor market outcomes. However, while many of the interventions in developing economy settings focus on increasing nutrient or consumption levels, MRFIT is largely aimed at improving health through reducing overconsumption. In this way, our paper provides novel evidence of the effects of developed countries' health improvements on earnings.

Our analysis focuses on a key aspect of health in the United States, coronary heart disease. Although the rate of mortality due to heart disease has been in steady decline for at least fifty years, heart disease remains the leading cause of death in the United States (Kochanek et al. 2016). While MRFIT began in the 1970s, our findings are relevant for understanding how lowering CHD mortality risk affects labor market outcomes in the present day as the current advice for preventing CHD is dominated by the primary components of the MRFIT intervention: adjust your diet, stop smoking, and control your blood pressure (e.g., Torpy, Burke, and Glass 2009). Although the results of our analysis from MRFIT cannot speak directly to how *any* type of health improvement will affect labor market outcomes, understanding the impact of CHD risk on workers' outcomes is important due to the widespread prevalence of heart disease in the United States.

We investigate potential avenues through which the health improvements lead to higher earnings. We test for differential earnings effects across subgroups defined by baseline health characteristics including cholesterol, smoking, and blood pressure levels as well as CHD mortality risk and BMI. We also examine earnings effects by baseline occupation classification (white vs. blue collar) and occupational tasks (as defined by Autor, Levy, and Murnane (2003)). However, the resulting estimates are mainly suggestive as we are only able to reject the null hypothesis that the impact of the intervention on earnings is the same across subgroups when comparing groups defined by baseline smoking intensity and (marginally) when comparing white and blue collar workers.

In his seminal work, Grossman (1972) notes that investments in health can raise earnings by increasing the amount of healthy time available for market work. Using responses to annual survey questions which ascertain information on events that occurred over the past year, we find that the SI group experiences less time lost to illness and lower levels of work-preventing disabilities during the experimental period. Our results suggest that this mechanism posited by Grossman likely contributed to the increased earnings levels due to the intervention.

In our analysis, we focus on the reduced form effect of the MRFIT intervention on labor market outcomes. We do so for two reasons. First, we do not possess an ideal single measure of health which fully captures the impact of health as affected by MRFIT. For example, while the intervention lowered measured CHD risk, this metric may not fully capture all the dimensions through which the intervention improved health. In addition, a global health summary question, such as self-reported health status, was not asked of participants. Second, for nearly every potential health measure that could be used in our analysis, there likely would be concerns about the exclusion restriction. For example, as part of the smoking modification intervention, participants were encouraged to avoid behaviors, such as going to bars, that might trigger the desire to smoke. To the extent that this aspect of the intervention is not captured in a single health measure, using the intervention as an instrument for health would lead to concerns regarding the validity of the exclusion restriction. Our reduced form estimates still answer an important question: how does an intervention focused on reducing mortality due to the leading cause of death in the United States affect labor market outcomes?

Our paper contributes to an extensive literature examining the relationship between health and labor market outcomes. Recent studies, using experimental or quasi-experimental research designs, exploit the effects of short-run environmental factors (often, pollution) on productivity and labor supply, in some cases focusing on intrahousehold effects (Graff Zivin and Niedell 2012; Adhvaryu, Kala, and Nyshadham 2016; Aragón, Miranda, and Oliva, 2016). Other studies leverage the impacts of medical interventions aimed at preventing or treating diseases (Thirumurthy, Graff Zivin, and Goldstein, 2008; Lucas, 2010; Fink and Masiye, 2015; Dillon, Friedman, and Serneels, 2015). Compelling evidence of one aspect of health on productivity comes from experimental evidence of

the effects of iron supplementation, particularly on anemic workers (Thomas et al., 2006).[1] While this work provides important evidence in developing contexts, individuals in developed countries face different health challenges. Health is naturally a multidimensional characteristic, and the effects of variation in one dimension informs relatively less about effects along another dimension.[2]

Adult health interventions in the economics literature are relatively rarer in developed countries than in developing economies. Some notable studies examine the causal impact of variation in the costs of either health insurance or healthcare itself (Newhouse et al., 1993; Dow et al., 1997). Broadly speaking, the findings point to a reduced form result that lower healthcare costs are associated with better labor market outcomes. The intermediate step, in which greater access to care leads to improved health is also generally supported, but can become more complicated. Objective measures of health may improve with greater access to care, but self perceptions of health may fall with the availability of health information. The analysis presented in this paper partially sidesteps this issue because the UC group received continued (if less frequent) information about health markers. In other research, evidence on the effects of body composition, as measured by body-mass index (BMI), on labor market outcomes is summarized by Cawley and Ruhm (2011). In terms of the relevant health conditions, this work is most closely related to ours although evidence in these studies is plagued by concerns that the observed effects are not causal.[3]

This paper proceeds as follows. Section 2 describes the specifics of MRFIT, its sample, and the data. After briefly documenting the impact of the intervention on health in section 3, we estimate the effect of health on earnings and family income in section 4. Section 5 discusses and concludes.

[1] Other studies of iron supplementation on productivity include Haas and Brownlie (2001), Li et al. (1994), Edgerton et al. (1979), and Basta et al. (1979).

[2] A related literature, reviewed by Thomas (2009), examines the effects of resources and nutrients before birth or at young ages on adult economic outcomes. See Almond (2006), Roseboom et al. (2001), and Field, Robles, and Torrero (2009) for examples spanning various countries and contexts. For examples from the extensive literature on the effects of childhood nutrition or treatment against infection, some of which persist into adulthood, see Alderman, Hoddinott, and Kinsey (2006), Glewwe and Miguel (2008), Maluccio et al. (2009), and Hoddinott et al. (2008).

[3] Although our discussion focuses on the evidence regarding the impact of health on labor market outcomes, as we noted above the causality may run in the opposite direction. For example, Frijters et al. (2005) exploit income changes associated with German reunification to identify the effect of income on health. Work by Case, Lubotsky, and Paxson (2002) and Currie and Stabile (2003) finds that the strong relationship between adult income and adult health is also present for children and grows stronger as children age.

2 Experimental Design

We leverage the experimental impact of MRFIT to identify causal effects of health on labor market outcomes. MRFIT was designed to understand the impact of CHD on mortality where CHD mortality risk was experimentally manipulated by a bundle of treatments. The trial was not specifically designed to affect or measure our outcomes of interest. This section describes the background, intervention, and data created by MRFIT. The information presented below draws heavily on the eight papers making up volume 10, issue 4 of *Preventive Medicine* (1981), which detailed the implementation of MRFIT and presented its early results.

2.1 Development of MRFIT

Prior to the 1960s, epidemiological research had already shown connections between coronary heart disease and several risk factors. Existing research had found both that serum cholesterol levels were associated with incidence of myocardial infarction and that cholesterol levels could be affected by modifying dietary intake (Zukel et al. 1981). Blood pressure and smoking had also been established as risk factors for CHD. Interest grew throughout the 1960s in a large-scale demonstration of the effects of risk factor modification on CHD and mortality. While single-factor trials were considered by epidemiologists and public health officials, a 1970 task force, organized by the director of the National Heart and Lung Institute to address arteriosclerosis, recommended against trials modifying diet alone and suggested a multiple-risk-factor trial as the way forward. A multiple-factor intervention was considered most likely to produce measurable results.

In response to these recommendations, the National Heart and Lung Institute's Clinical Applications Program undertook planning for a trial addressing the multiple risk factors of serum cholesterol, blood pressure, and smoking. The Framingham Heart Study, an ongoing observational study of factors associated with heart disease, was used to determine the necessary sample sizes to measure the expected effects. Ultimately, grants were awarded to 22 clinics across 16 metropolitan areas in order to identify study participants and implement the trial. Study organizers provided detailed information to each of the clinics and ensured that information and techniques were consistent across them.

2.2 Screening and Randomization

The MRFIT clinics screened initially 361,662 men in total, of whom 12,866 were randomized into the Special Intervention and Usual Care groups (Sherwin et al., 1981). The study targeted men aged 35 to 57 in the upper end of the CHD risk distribution as determined by a model predicting CHD mortality risk using Framingham Study data.[4] The upper age limit was chosen so as to avoid participants moving at retirement since, as discussed below, participants in the SI group needed to make regular visits to the clinics. There was not a unified method for sampling possible participants, so clinics were free to enroll participants in different ways, provided that they met the age and CHD risk requirements. Initially, the intention was to exclude men with diastolic blood pressure readings over 110 mm Hg or who were taking antihypertensive medication, but both of these requirements were ultimately relaxed.

Potential participants were enrolled through three separate screening visits separated by three to four weeks. At each screening, measures needed to compute CHD risk were taken and the potential respondents were analyzed for the likelihood of responding to intervention. Only those with elevated CHD risk, initially targeted as the upper 15 percent of the distribution, and a willingness to change risk factors were contacted again after each screening. The vast majority of the initially-screened men were excluded after the first screening for having estimated CHD risk that was below the study's threshold (Sherwin et al., 1981). Despite targeting the upper 15 percent of the distribution based on Framingham data, this restriction actually eliminated some 90 percent of the 360,000 men attending the first screening. In total, 22,088 men returned for the second screening. Further respondents were excluded because they expected to move from the area, had previously been hospitalized for more than two weeks due to a heart attack, had been prescribed diabetes medication, or had very high levels of serum cholesterol (greater than 350 mg/dl) or diastolic blood pressure (greater than 115 mm Hg). Following the third screening, participants were randomized into the SI and UC groups.

[4] We discuss the details of the CHD risk prediction model below.

2.3 The Intervention

MRFIT was a non-blind randomized trial. The SI group was subject to the intervention, which is described in detail throughout this section, while the UC group was advised to seek their usual avenues of care in the community. Information on their medical conditions and risk factors were disclosed to the UC group and their medical providers. As the MRFIT organizers were aware, the UC group was not a control group *per se* and, hence, avoided the standard "treatment" and "control" terminology by instead using the SI and UC labels. UC participants were informed of their elevated risk along with a number of medical measures and were followed-up with throughout the study. However, the interventions described below induced differences in the two groups, which were effectively identical at baseline. The SI and UC means for key baseline variables are displayed in Table 1.[5] There are no statistically significant differences in the means of these variables between the two groups.[6]

The intervention for the SI group had multiple arms and aspects of implementation that were used throughout the six years of the study. After initial meetings and screenings at baseline, SI group members participated in a series of 10 meetings with MRFIT staff and other SI participants over the first two or three of months of the study (Benfari, 1981). These sessions were aimed at communicating to the participants the specific risks associated with various risk factors, giving information on changing behaviors, and providing support for doing so. Participants were encouraged to bring their wives or "homemakers," which was intended to be especially helpful in affecting changes in diet and smoking behavior. Following this intensive period of meetings, depending on their progress in modifying risk factors, the participants entered a period of extended intervention or maintenance for each risk factor. Extended intervention involved continued efforts at changing behavior, while maintenance provided support for changes already observed.

[5]Our analysis sample, which is used in Table 1, contains the 12,562 of the original 12,866 MRFIT participants that have non-missing data for age, race, education, marital status, and employment status measured at baseline. Additional information regarding the available data is discussed below in the Data and Measurement sub-section.

[6]The p-values for the differences reported in Table 1 are based on a wild cluster bootstrap which clusters at the level of the 22 clinics (Cameron, Gelbach, and Miller 2008). All the differences shown in Table 1 remain statistically insignificant, with little movement in the corresponding p-values, if we do not apply clustering methods.

2.3.1 Cholesterol Intervention

MRFIT organizers sought to decrease serum cholesterol levels for all men in the SI group, but a particular focus was placed on achieving a ten percent decrease for participants with baseline levels greater than 220 mg/dl (Caggiula et al., 1981). Based on evidence from prior studies, it was thought that such a reduction could be achieved by recommending a diet with ten percent of calories each from saturated and polyunsaturated fats and a limit on dietary cholesterol of 300 mg per day.[7] The diet also set a goal for total fat intake as 35 percent of calories. Weight loss was also a target for the subset of men in the SI group who were over 1.2 times "ideal body weight," which is solely determined as a function of height.[8]

When MRFIT participants turned out to have, on average, healthier pre-intervention diets than those seen in the prior studies, these targets were made more ambitious. Compared to existing evidence, MRFIT participants in both experimental groups consumed fewer calories, a lower percentage of calories from total fat and saturated fat, and a higher percentage of calories from polyunsaturated fat at baseline. Diets were particularly better-than-expected among participants with high baseline cholesterol levels. Caggiula et al. (1981) hypothesize that in response to screenings showing high cholesterol levels, participating men had already begun to adjust their diets. In 1976, some two to three years into the study, the saturated fat and dietary cholesterol limits were lowered to eight percent of calories and 250 mg per day, respectively. The weight loss targets were extended to include men over 1.15 times ideal body weight, and a goal of reducing bodyweight by at least 10 pounds was added for most of the men in the SI group.

The specifics of the MRFIT intervention for modifying diet amounted to targeting particular levels of intake for various food groups. A written manual was distributed to SI participants, categorizing food types according to whether they were "OK" or should be avoided. SI sessions were

[7] It should be noted that more recent evidence shows no impact of dietary cholesterol on serum cholesterol, which has resulted in the USDA dropping dietary cholesterol recommendations from its guidelines. See p. 17 of the USDA's "Scientific Report of the 2015 Dietary Guidelines Advisory Committee." Regardless, as shown in results reported in this paper as well as prior research, MRFIT did achieve serum cholesterol reductions, presumably due to the fat intake recommendations and changes to participant weight.

[8] "Ideal weight" in this context is 0.9 times the average height-specific weight for men aged 18–34 in the National Health Survey, 1960-1962. This makes ideal weight for a six-foot man approximately 162 pounds, and 1.2 times ideal weight is just over 194 pounds. For most heights, 1.2 times ideal body weight amounts to a BMI in the range of 26 to 28.

aimed at providing this kind of relevant information to participants and their wives or homemakers. In addition to basic information on what foods to target and avoid, MRFIT sessions highlighted shopping skills, label-reading, food demonstrations, and tastings. Over the long term, the intensity of follow-up with each participant was a function of their cholesterol response. Throughout the study, the serum blood cholesterol levels of the SI group was tracked at least every four months.

2.3.2 Smoking Intervention

Smoking cessation for the 59 percent of participants that smoked was a key design goal of MRFIT. Given the relatively high CHD risk among MRFIT participants due to other factors, smoking was viewed as a particularly high risk to this population (Hughes et al., 1981).[9] MRFIT clinics were staffed with smoking specialists, psychological consultants, and health counselors to aid participants in quitting. As in the dietary aspects of the intervention, participants' wives were also engaged to support smoking cessation. An initial intervention at baseline involved participants receiving a "strong antismoking message from a physician" (Hughes et al., p.482) who provided tailored information based on the participant's health measures. This was followed by a meeting with a smoking specialist, from which the specialist could identify preferred cessation techniques and whether the participant appeared prepared to attempt cessation.

The smoking intervention from this point forward was structured around documentation created specifically for MRFIT participants. The intervention group meetings highlighted the risks of smoking and the benefits of cessation and encouraged participants to examine details of their smoking behaviors. Meetings involved behavioral modification techniques and group discussions. The smoking specialists followed up regularly to offer support to participants who reported they had stopped smoking. Participants received feedback on a number of objective medical measures including serum thiocyanate and expressed carbon monoxide. Over the extended intervention period, those who had not quit or who had relapsed were considered for additional intervention. These interventions were tailored more specifically to each individual and potentially involved more information from physicians, additional types of group meetings, or further cessation therapies.

[9] However, because the sample participants were selected to meet a CHD risk threshold, there is a negative within-sample correlation between smoking intensity and the other risk factors.

2.3.3 Blood Pressure Intervention

The blood pressure intervention in MRFIT targeted reducing diastolic blood pressure for hypertensive participants (Cohen et al., 1981). Blood pressure readings were taken throughout the screening process, as part of sample selection, but categorization as hypertensive was initially based on the third screening, when systolic and diastolic blood pressure readings were both taken. Participants were categorized as hypertensive if they had a diastolic blood pressure reading of 90 mm Hg at third screening and again at a confirmation follow-up. If a participant's readings later exceeded these thresholds at a regular visit and a confirmation follow-up, they were categorized as hypertensive at that point. Any participant on antihypertensive medication was considered hypertensive throughout. Those who were taking such medication were given a diastolic blood pressure target of 80 mm Hg. For participants not initially on medication, the specific target was 89 mm Hg or a 10 mm Hg reduction, whichever was lower.

Hypertensive participants were treated with a "stepped care" approach. This involved steady increases in the level of hypertensive medication in a way that was standardized across clinics. During a period of close monitoring, participants were put on increasingly potent blood pressure medications if the desired blood pressure reductions were not observed. The medications used in MRFIT were from a centralized source, prepackaged for participants, and provided free of charge (Cohen et al., 1981). If blood pressure readings fell consistently below 80 mm Hg or weight loss was achieved, participants' medications were eligible to be stepped down. In addition to the medication, dietary advice on weight reduction and reduced sodium intake were counseled for hypertensive treatment.

2.4 Data and Measurement

Men in the UC group were invited back for annual visits and examinations after randomization. Men in the SI group were invited for these visits as well as interim follow-ups approximately every four months (Sherwin et al., 1981). The annual visits allowed MRFIT to record new information on medical history, 24-hour dietary recall, leisure activity, smoking history, and other behaviors. The annual examinations of all participants included, in addition to a physical examination, the

recording of a number of biomarkers. For tracking changes to health, this paper uses data from all these annual visits. However, certain labor force information was only recorded at baseline and the six-year follow-up. This includes reported earnings from participants' main jobs as well as family income, both of which are reported in categories. Other labor force information, including layoff, firing, and disability over the prior year, are reported at annual visits.[10]

Participation throughout the experimental period was relatively high. Sherwin et al. (1981) note that a "large majority" of men participated in the group sessions and that those who did not were "usually" willing to participate in individual sessions. Nearly 75% of wives of the SI group participated in at least some of the group sessions. Through the fourth year of the experiment, over 91% were either attending their annual visits or known to be deceased, with participation nearly equal for the SI and UC groups. We discuss sample attrition in more detail in section 4.3.

2.5 Coronary Heart Disease (CHD) Risk

We follow previous MRFIT researchers in using calculated measures of CHD risk to summarize the overall health of the participants and the intervention's effects. These risk scores are generated using estimates from the Framingham Study as reported in Neaton et al. (1981).[11] Risk is estimated with a logit model with the outcome being mortality due to CHD within a six year period and using serum cholesterol, diastolic blood pressure, and cigarettes smoked per day as predictors.[12] While other participant characteristics (e.g., age) could be used in calculating risk scores, as in many publicly-available risk calculators based on the Framingham sample, we follow previous researchers and the MRFIT organizers in using these risk scores that are based on key indicators for MRFIT's targeted dimensions of health. We control for additional characteristics of participants, including age, in our empirical analysis as described in the following sections. Using serum cholesterol, diastolic blood pressure, and cigarettes per day with the logit coefficients from Neaton et al. (1981), we calculate

[10]The original survey forms for MRFIT can be found at `https://biolincc.nhlbi.nih.gov/studies/mrfit/Forms/`.

[11]We deviate slightly from Neaton et al. (1981) in that we do not adjust self-reported smoking levels for the respondent's measured serum thiocyanate, which is a biomarker for smoking. To the extent that individuals in the SI might be more likely to underreport their smoking behavior (perhaps because they are expected to reduce their smoking levels), we will tend to overstate the CHD risk differences between the SI and UC groups. However, we show below that serum thiocyanate levels are lower in the SI group than the UC group.

[12]The logit coefficients in this model, found in Neaton et al. (1981, Table 18), are 0.0088 on serum cholesterol, 0.0464 on diastolic blood pressure, and 0.0286 on cigarettes per day, with an intercept of 11.0336.

risk scores at screening and at each follow-up year.[13]

Kernel-smoothed densities of baseline CHD risk by experimental status are displayed in the top left panel of Figure 1. The magnitudes of the risk scores map directly into CHD mortality risk from the original Framingham logit model. That is, a risk score of two percent indicates a two percent probability of mortality due to CHD within six years. In addition to being statistically equivalent at the mean (see Table 1), the densities for the SI and UC groups are quite similar across the entire distribution (Figure 1a).[14] Further, while the mean risk scores are just above two percent, both distributions exhibit a noticeable positive skew.

In order to perform power calculations and determine the necessary sample size, MRFIT organizers anticipated the effects of the intervention on the SI and UC groups. They made these predictions based on evidence from prior interventions. The predicted effects are shown in Appendix Table A1. While the organizers anticipated cholesterol and blood pressure impacts to be concentrated among participants with high baseline levels of those risk factors, they anticipated larger percentage reductions in cigarettes smoked for those at the low end of the smoking distribution. In examining possible heterogeneous effects of the intervention in section 4.5, we use these subgroups defined by the MRFIT organizers based on baseline risk factors.

3 The Impact of MRFIT on Health Outcomes

Before turning to employment-related outcomes, we briefly document the impact of the MRFIT intervention on health outcomes that were directly impacted by the three interventions: serum cholesterol reduction, smoking cessation, and lowering blood pressure. The longitudinal impacts of MRFIT on serum cholesterol (Caggiula et al. 1981), smoking behavior (Hughes et al. 1981), and blood pressure (Cohen et al. 1981) have been previously shown for the first four years of the experiment. We extend these findings to cover the first six years of the experiment and also present findings for the longitudinal impact on CHD risk scores.

[13] All three variables were not recorded at the same screening visits, the estimates reported here use first-screening serum cholesterol and third-screening blood pressure and cigarette smoking levels.

[14] We can compare the risk scores of the MRFIT sample to a nationally-representative sample from the National Health and Nutrition Examination Survey (NHANES I), which was in the field from 1971 to 1975. We find that the average CHD risk score for men ages 35 to 57 in NHANES I was 1.26, consistent with MRFIT participants being drawn from the upper end of the risk distribution.

3.1 The Impact on CHD Risk Scores

The reductions in CHD risk were large and sustained as shown in Figure 1. After the first year of the intervention, Figure 1b shows that the average CHD risk in the SI group is over 40 percent lower than that found in the UC group. Consistent with the prediction that the cholesterol and blood pressure interventions will affect those with the highest initial levels, the long right tail of the SI distribution is particularly diminished. By year six of the experimental period, as displayed in Figure 1c, the average risk scores for both groups had fallen further. Decreases in the UC group, however, outpaced those in the SI group, leading to a smaller effect for the intervention on CHD mortality risk by year six.[15]

To examine the longitudinal effects of the MRFIT intervention, we estimate the equation

$$y_{it} = \alpha_t + \beta_t SI_i + \boldsymbol{\gamma_t} \mathbf{X_i} + \varepsilon_{it}, \tag{1}$$

where y_{it} is an outcome for participant i in year t, SI_i is an indicator equal to 1 for those in the SI group, and $\mathbf{X_i}$ is a vector of controls measured at baseline. These baseline controls include a full set of indicators for age, an indicator for being white, indicators for four education groups, and an indicator for being married. Inclusion of the controls is not expected to affect the estimated experimental impacts as SI_i is randomly assigned. Indeed, the intervention is balanced as shown in Table 1 and, in practice, dropping the controls has a negligible impact on the experimental effect point estimates. Rather, we include the controls to increase the precision of the estimates.

To produce confidence intervals displayed with these results, we cluster standard errors at the level of the 22 clinics involved in the experiment. Since we have relatively few clusters, we use the wild cluster bootstrap to construct confidence intervals and test statistics (Cameron, Gelbach, and Miller 2008). Our subsequent analysis of the primary labor market outcomes of interest employ alternate variance estimation methods that are robust to small numbers of clusters.[16]

[15] It should be noted that the MRFIT participants had aged six years beyond the sample used to estimate the coefficients for CHD risk score. The CHD risk scores may no longer reflect actual six-year CHD mortality group for the sample as it ages, however, the measure remains a time-consistent summary of the effects of the intervention on key variables.

[16] Abadie et al. (2017) provide formal conditions under which researchers should report clustered standard errors. Their Corollary 2 states that clustering is correct if one of the following three conditions is correct: i) no treatment effect heterogeneity, ii) the sample contains a relatively small number of clusters from the total number of clusters in

The evolution of the average experimental CHD mortality risk differences is shown in Figure 1d. Each point in this figure represents the coefficient on the intervention indicator from a regression based on equation (1) where CHD risk for different years is the outcome. The effects of the intervention on risk scores are largest after one year and decay over time. The coefficients indicate a relative initial decrease in CHD risk on the order of one-third for the SI group. The experimental differences fade slightly over the subsequent years, but remain precisely estimated and meaningfully different across the groups.

3.2 The Impact on Cholesterol

The effect of the intervention on serum cholesterol is displayed in the top two panels of Figure 2. As shown in Figure 2a, we find that the intervention reduced serum cholesterol by approximately 8 mg/dl with the impact falling slightly in the later years of the experimental period. Figure 2b displays the estimated intervention effects on serum cholesterol split by the baseline levels used for the predicted impacts shown in Appendix Table A1. Among participants with baseline serum cholesterol levels of 220 mg/dl, the intervention reduced serum cholesterol by 10 mg/dl, which is smaller than the 10% reduction predicted for this group. This effect fades slightly during the last two experimental years. The intervention significantly also lowered serum cholesterol levels for those under 220 mg/dl at baseline, although the magnitude of the response is smaller for this group relative to those with higher serum cholesterol levels at baseline.

The SI group self-reports a large initial reduction in total daily caloric intake of approximately 300 calories which grows to approximately 350 calories by year three of the experimental period (Figure 2c). Given the baseline average caloric intake of 2,369 calories, these differentials represent intake reductions of 12 to 15 percent.[17] The calorie reductions remain in this range over the entire sample period, suggesting long-term food-intake changes for the SI group relative to the UC group. However, in spite of the large self-reported reduction in calories, the SI group only experiences a

the population of interest, and iii) there is, at most, one unit per cluster. As we view the population of interest as being the entire United States as opposed to only the areas covered by the 22 clinics, the number of clusters included in the sample is relatively small when compared to the number of potential clusters in the population. As such, due to the design of the MRFIT sample, we present clustered standard errors with our analysis.

[17] Additional results for the food intake categories explicitly targeted by the intervention (saturated fat, polyunsaturated fat, dietary cholesterol, and total fat intake) are are displayed in Appendix Figure A1.

small, but significant, decline in weight of two pounds (Figure 2d). Although this last result suggests that the SI group may be overstating the extent to which they adhere to their food pattern, the serum cholesterol results clearly indicate that the intervention improves cholesterol levels.

3.3 The Impact on Smoking

The estimated impact of the MRFIT intervention on smoking outcomes is displayed in Figure 3. The effects of the intervention can be seen most pointedly at year one, as shown in the top left panel of the Figure, where the probability of smoking falls by nearly 20 percentage points or roughly one third of the baseline smoking population. The experimental effect shrinks over subsequent years which is due to more-rapidly decreasing smoking rates in the UC group (not shown here). A similar narrowing of the gap in daily cigarettes smoked also occurs following an initial reduction of more than 40 percent. The bottom left panel of Figure 3 displays the experimental effect on serum thiocyanate which is a biomarker for smoking levels. Although thiocyanate is present even among those who never smoke, its concentrations are dramatically higher among current smokers, increasing with the intensity of consumption (Hughes et al. 1981). The experimental effect on serum thiocyanate is negative and immediately apparent in the first year of the experiment and only exhibits modest amounts of decay over time.

The bottom right panel of Figure 3 shows the experimental impact on smoking cessation across the distribution of baseline smoking intensity. The groupings of baseline smoking intensity match those used for predicting experimental responses shown in Appendix Table A1. The relative decrease of over 40 percentage points among the lightest smokers (fewer than 20 cigarettes per day) is much larger is than the roughly 25 percentage point decline for the heaviest smokers (at least 40 cigarettes per day). This general pattern is in line with the predicted smoking effects described in Appendix Table A1. While the experimental effect fades for all baseline intensity groups, this happens most dramatically for the light-smoking group and becomes indistinguishable from the moderate-smoking effect in later years.

3.4 The Impact on Blood Pressure

Figure 4 shows the impact of the intervention on blood pressure. Baseline diastolic blood pressure is 91 mm Hg, indicating that the average participant is at the low end of stage 1 hypertension. As shown in the top left panel, the intervention lowers the SI group's diastolic blood pressure by nearly -4.5 mm Hg in year two before decreasing to near -3 mm Hg by year six. Consistent with the experimental predictions, we find heterogeneous effects when splitting participants by baseline blood pressure based on Appendix Table A1 (Figure 4b). The intervention successfully lowers blood pressure levels for those with the highest starting blood pressure although the decrease does not quite reach the ten percent reduction predicted for this group. Both groups have effects that peak in the middle years of the study and fade slightly in the later years.

A similar overall pattern appears in the systolic blood pressure effects displayed in Figure 4c. The treatment effect peaks in size at year two, with a decrease of approximately 7 mm Hg on a base of 135 mm Hg. Much of these effects is likely driven by variation in the likelihood of taking hypertension medication over time. At baseline, just under one-fifth of the sample is taking hypertension medication. Blood pressure medication use is higher for the SI group in all experimental years, with the relative difference peaking in year two as shown in Figure 4d. At that point, the differential effectively doubles the medication rate in the SI group relative to baseline levels.

4 The Impact of MRFIT on Earnings and Family Income

4.1 The Impact of MRFIT on Earnings and Income CDFs

Given the large and sustained impact of the MRFIT intervention on health, we next turn to investigating the impact on labor market outcomes. We first examine the impact of the intervention on the discrete earnings and income distributions. We use specifications similar to equation (1) to measure the effects of the experiment on the discrete CDFs of observed earnings and income. This approach provides insight into which parts of the earnings and income distributions are affected by the intervention.

The top left panel of Figure 5 displays differences at baseline between the SI and UC groups in 1-CDF of earnings.[18] That is, we examine the differences between the two experimental groups in the fraction of observations at or above a given earnings category. These estimates are generated from a series of linear probability models similar to equation (1) using an indicator for having earnings at or above the given category as the outcome and using baseline demographic and health controls that we include in equation (2). Across most of the distribution, the point estimates are negative and, in some cases, significant, which indicates that the SI group has slightly lower baseline earnings than the UC group.

The impact of the intervention on the 1-CDF of year six earnings is shown in the remaining two panels of Figure 5. We find a significant impact of the MRFIT intervention on the upper end of the earnings distribution in year six (Figure 5b). As the inflation rate was high throughout much of the sample period but the nominal thresholds for the earnings and income categories were the same in year six as at baseline, much of the distribution shifted into the upper earnings categories in year six.[19] Given that we find that the SI group has slightly lower earnings at baseline, we also present additional results in which we account for these initial differences by including a set of indicators for the categorical baseline earnings outcomes.[20] As displayed in Figure 5c, the intervention increased the fraction of individuals in the top three earnings categories after accounting for baseline earnings.

Figure 6 presents the experimental impact on 1-CDF of family income. The 1-CDF differences at baseline are not statistically different for the SI and UC groups although all of the point estimates, except for one in the top left panel, are negative. The top right panel of the Figure shows that at year six, there is a significant difference between the two experimental groups in three income categories and is even more pronounced than what we found for earnings.[21] The bottom panel of the Figure finds slightly larger effects on year-six income when including indicators for baseline family income as controls.

[18]The results shown in Figure 5 condition on being employed. However, as we show below in section 4.3, the rates of non-employment at year six in the SI and UC groups are identical. The PDFs for the earnings and income variables, by experimental group, are shown in Appendix Figure A2.

[19]See the PDF for year six earnings in the upper right panel of Appendix Figure A2.

[20]These estimates restrict the sample to participants employed both at baseline and year six. We provide a detailed discussion of sample selection issues in section 4.3. As we discuss below, we believe sample selection has a negligible impact on our earnings and family income results.

[21]We discuss possible explanations for differences in the estimated impact between earnings and family income in section 4.4.

4.2 Quantifying the Impact of MRFIT on Earnings and Income

We next quantify the impact of MRFIT on earnings and family income. To do so, we estimate a variant of an ordered probit to account for the categorical reporting of these outcomes. In place of actual earnings or income, inc^*_{it}, we observe inc_{it} which contains J categories where

$$
\begin{aligned}
inc_{it} = 1 &\iff inc^*_{it} \leq \mu_1 \\
inc_{it} = j &\iff \mu_{j-1} < inc^*_{it} \leq \mu_j \ \forall \ j \in \{2, \ldots, J-1\} \\
inc_{it} = J &\iff \mu_{J-1} < inc^*_{it},
\end{aligned}
$$

Assuming that log earnings or log income is normally distributed, we could estimate a standard ordered probit where the $\mu_j, \ j = 1, \ldots, J-1$ are unobserved parameters to be recovered. However, since we know the actual thresholds for the categorical variables, we can modify the ordered probit likelihood function to make explicit use of these thresholds rather than estimate the cutpoints.[22] Specifically, we assume that the log of earnings (income) is normally distributed and use the corresponding log cutpoints in estimation.

The equation we estimate using the modified ordered probit are

$$
log(inc^*_{it}) = \lambda_t + \delta_t SI_i + \boldsymbol{\theta_t} \mathbf{X_i} + \nu_{\mathbf{it}}, \tag{2}
$$

where inc^*_{it} is (unobserved) earnings or income of participant i in period t, SI_i is a binary indicator which equals one for those in the SI group, $\mathbf{X_i}$ are baseline demographic and health controls, and ν_{it} is a normally distributed error term. The baseline demographic controls are the same as we use to estimate equation (1) while the health controls include continuous baseline measures of serum cholesterol, diastolic blood pressure, and number of cigarettes smoked as well as an indicator for being a smoker. The parameter δ_t is the impact of the intervention on log earnings or log income. To conduct inference, we cluster at the level of the clinic. As we only have 22 clinics in our sample, we use the procedure developed by Kline and Santos (2012), analogous to a wild cluster bootstrap,

[22] In MRFIT, for both earnings and income, J=9 and and the thresholds are \$4200, \$7200, \$10,000, \$12,000, \$15,000, \$18,000, \$22,500, and \$35,000. These thresholds are the same at baseline and at year six.

for M-estimators.

4.3 Sample Selection

Issues involving sample selection arise in three possible ways in our analysis: attrition between baseline and year six, non-employment, and missing data. Table 2 shows the share of observations that are dropped from the analysis for each reason, both at baseline and at year six, and does so separately for the SI and UC groups. We show these rates separately for the earnings and family income analysis both because family income is non-zero even when the participant is not employed and because there are slightly different rates of missing data for earnings and family income.

The first two panels of Table 2 show the share of observations dropped from the baseline earnings and family income regressions. As shown in Panel A, roughly four percent of observations are dropped from the baseline earnings regressions in both the SI and UC groups. The rate of missing earnings data due to non-employment at baseline is extremely low, although this can be explained, at least in part, by the age restrictions on the MRFIT sample which limited participants to ages with the highest employment rates. Information on labor force status at baseline is rather limited which may also contribute to the high reported employment rates.[23] The fact that the rates of sample selection at baseline are almost identical across the experimental groups is not surprising as participants learn the results of the assignment to the SI or UC group after the baseline interview that collects labor force information. As shown in Panel B, the only source of sample selection for the baseline family income analysis is missing data as the non-employed reside in households with positive family income. Given these results, we do not account for sample selection in our analysis of baseline earnings and family income.

As shown in the bottom two panels of Table 2, a much higher share of observations are dropped from the year six regressions. However, the overall rates of dropped observations from the earnings regressions, at roughly 29.5%, is essentially the same for the SI and UC groups (Panel C). The

[23]Labor force status at baseline is determined from a question in which participants are asked whether they have two or more jobs to which they can provide one of three answers: yes, no, or retired. Thus, the non-employed are those who state that they are retired. In fact, the majority of those classified as having missing data in Panel A are those who do not respond to this labor force question. If we were to treat all of these observations as non-employed, the rates of non-employment are still quite low at baseline.

rate of attrition by year six is 9.8% in the SI group and is 11.6% in the UC group.[24] These attrition rates are substantially lower than what is found in a standard longitudinal dataset used in economic research. For example, Zabel (1998) finds that roughly 25% of participants in the nationally-representative portion of the Panel Study of Income Dynamics leave the sample by the sixth year of the study. The non-employment rate at year six of those who remain in MRFIT is 10.2% and is identical for the SI and UC groups.[25] The share of observations dropped due to missing data of 9.4% is slightly higher for the SI group than the 7.7% rate for the UC group. The total share of observations dropped from the year six family income regressions (Panel D) are very close for the SI and UC groups with slight differences in the rates being dropped for attrition and missing data.

Applying the standard approach to modeling sample selection in economics (e.g., Heckman 1979), the selection equation is

$$s_i^* = \pi_t SI_i + \boldsymbol{\omega_t} \mathbf{W_i} + \epsilon_{it}, \tag{3}$$

where the observed selection indicator, s_i, equals one if $s_i^* > 0$ and equals zero otherwise. Since we have three different mechanisms for selection, in general using a single selection equation is not appropriate. Moreover, finding one, much less three, valid exclusion restrictions in order to estimate the corresponding system of equations is quite challenging.

However, if we are willing to assume that two of the three sources of selection are (conditionally) random, then we can account for the remaining form of selection with a single equation. For example, many studies using longitudinal data do not account explicitly for sample attrition, even among studies with dramatically higher rates of attrition than are found in MRFIT. Similarly, the rates of missing data in MRFIT are dramatically lower than in the Current Population Survey where in recent years nearly one-third of earnings observations for employed individuals are missing

[24] The differential attrition rates by year six are not due to differential mortality for the SI and UI groups. Seven years after the beginning of the intervention, mortality was slightly higher for the SI group although this difference was statistically insignificant (MRFIT Research Group 1982). Mortality was lower for the SI group when measured ten years and again sixteen years after the intervention although both these differences were only marginally significant (MRFIT Research Group 1990, 1996).

[25] The question eliciting employment status at year six provides far more detail than the corresponding question at baseline. Respondents are asked "What is your present job status?" to which they can respond either "working full-time," "working part-time," or "unemployed." Those giving the final option as a response are given a follow-up question to determine whether they are laid off, disabled, retired, or other.

and subsequently imputed (Stephens and Unayama 2015). In these instances, researchers routinely treat sample selection as (conditionally) random by not modeling sample attrition, by using imputed values, or by dropping observations with imputed data.[26]

We can account for non-employment in multiple ways. Lee (2009) develops a method for bounding the impact of a binary treatment when the outcome of interest is subject to sample selection. If sample selection is based on a single index as in equation (3) and if whether a participant works, as a function of treatment assignment, is consistent with a monotonicity assumption, then bounds on the treatment effect can be constructed by trimming the highest and lowest outcomes in the treatment group with the larger share of positively selected observations.[27] If we treat attrition and missing data as random and then apply Lee's method, the fact that the rates of non-employment at year six are essentially identical for the SI and UC groups means that there are very few observations to trim when estimating the proposed bounds.[28] Thus, the resulting bounds on the treatment effect would differ very little from estimated treatment effect.

Non-employment at year six occurs for many reasons: just over half of the non-employed are retired, nearly 30% are either temporary or permanently disabled, and the remainder give other reasons including being laid off. The different factors that drive these multiple dimensions of non-employment may invalidate the single index function approach to sample selection required for applying Lee's bounding method. Alternatively, as shown in Figure 7, there is an important age component to the reason for non-employment in MRFIT.[29] While rates of attrition and missing data are fairly constant by baseline age, the rate of non-employment rises sharply for older MRFIT

[26]Results from estimating a logit to predict who remains in the sample in year six show that those in better baseline health, who are better educated, and who are married are more likely to appear in year six. However, we do not find that experimental status is not quite marginally significant (p-value=0.101). Thus, below we present earnings and income estimates which include controls for these baseline health and demographic characteristics to account for attrition along these dimensions.

[27]As Lee notes, the treatment effect that is bounded is for the specific sub-group that is always positively selected regardless of treatment status. In our context, this effect is the impact of MRFIT on earnings for the sub-population that would be employed regardless of being assigned to either the SI or UC group.

[28]In the simple case where we ignore covariates and treat attrition and missing year six earnings as random, the trimming proportion, which is the fraction of positively selected observations that need to be trimmed 0.0011. That amounts to trimming 0.1%, or five, observations before computing the upper and lower bounds.

[29]Although MRFIT participants were initially screened to fall between ages 35 and 57, inclusive, the baseline ages used in Figure 7 are from the third screening visit at which time baseline earnings, income, and employment information was collected. Hence, some participants may have had a birthday in the intervening period which is why the oldest age displayed in Figure 7 is 58.

participants.[30] The year six non-employment rates are highest for those who are ages 56 and older at baseline as these participants reached the Social Security early retirement age by year six. Thus, we also present results in which we limit the sample to those that are age 48 and under at baseline, for whom the rates of non-employment do not exceed seven percent, in order focus on an exogenous subset of the sample for which non-employment is a minor issue.

Before turning to the earnings and income results, we briefly note the impact of the intervention on basic labor force outcomes at year six which we estimate using equation (1). Since MRFIT participants were ages 35 to 57 at baseline, the vast majority of participants are still in their working years at year six with an 88 percent employment rate for the UC group. The intervention has no effect on employment at this horizon as the estimated impact is statistically and substantively zero. A small minority (5.9 percent) of UC participants are retired by year six and there is no significant difference in retirement status due to the intervention. Of those who are employed, 97 percent in both groups are working full-time and there is no significant difference in this outcome. Moreover, while 60 percent of working UC participants report a change in job title or kind of work over the study period, this rate of job change is not significantly different than what we find for the SI group.[31]

4.4 Earnings and Family Income Estimates

Estimates of the impact of MRFIT on earnings and family income using equation (2) are shown in Table 3. Since we are modelling the impact on the log of the outcome of interest, we can interpret the coefficient as the percentage effect on the outcome due to treatment assignment. The estimated effects on earnings are shown in Panel A of Table 3. When we include only the treatment indicator in estimating equation (2) for baseline earnings (column (1) of Table 3), we find that those in the

[30] The rate of missing data does decline slightly with age although the high rates of non-employment for these older individuals reduces the share of them that can be at risk for having missing earnings data.

[31] These results on year six labor force outcomes are based on linear regressions using equation (1) which include baseline health and demographics as controls. The estimated effects of being in the SI group are: employment $\beta = 0.003$ (p-value=0.645), retirement $\beta = 0.003$ (p-vaue = 0.603), full-time employment conditional on being employed $\beta = -0.0004$ (p-value=0.918), and job change $\beta = -0.001$ (p-value = 0.946). In addition, at baseline as well as during each of the first first annual interviews, respondents are asked whether in the last twelve months they were either fired or laid off. We find no difference in the likelihood of giving an affirmative response to this question between the SI and UC groups at any point in time.

SI group have slightly lower earnings, consistent with the results shown in Figure 5.[32] Including baseline health and demographic controls (column (2)) shrinks the estimated difference in baseline earnings between the experimental groups.

The next three columns of Table 3 show the impact of the intervention on earnings at year six. When only a treatment indicator is included (column (3)), we find that earnings are roughly two percent higher for the SI group, a difference which has a p-value of 0.025. However, as we have seen, there is a small difference in baseline earnings between the SI and UC groups. When we also include a complete set of indicators for the baseline earnings outcome (column (4)), the estimated impact of the intervention on earnings rises to 2.7 percent with a p-value of 0.011. Further including baseline health and demographic controls (column (5)) increases the estimated effect to 3.1 percent with an even lower p-value.[33] Thus, the MRFIT intervention has a significant impact on the earnings of the SI group.

The intervention effects on family income are shown in Panel B of Table 3. As shown in column (1), baseline family income is slightly lower for the SI group but this difference is not statistically significant, a finding which remains with the inclusion of baseline health and demographic characteristics. At year six (column (3)), family income is 3.5 percent larger for the SI group and this finding is highly statistically significant. Including controls for baseline family income categories raises this estimate to 3.9 percent and including baseline health and demography controls further raises it to 4.1 percent.

One possible reason why the impact on family income is larger than that on earnings is that, recalling that the earnings measure is for the participants' main job, the intervention induces effects on the participants' earnings beyond their main jobs. Using responses to the year six question "Do you presently work for 2 or more employers?" we find that 9.6% of workers in the SI group answer affirmatively while 8.6% of those in the UC group do so. However, this difference is not statistically significant (p-value=0.152) and the one percentage point gap in having a second employer between the groups is too small to explain the difference in the earnings and family income estimates.

[32] In Table 3 we report, in brackets, p-values for the test of the null hypothesis that the parameter equals zero based on the Kline-Santos procedure for conducting inference when using a limited number of clusters with a non-linear estimator.

[33] In year six earnings and income regressions where we condition on baseline health and demographics, we also require respondents to have non-missing year six health characteristics.

Another possibility is that these changes are the result of intrahousehold reallocation of time and resources, as in Thirumurthy, Graff Zivin, and Goldstein (2008), or they could reflect improved health and earnings of other household members. Given that wives were to be heavily involved in the intervention itself, this latter explanation is certainly possible although no information on the work effort of spouses was collected.

The final column of Table 3 limits the sample to those age 48 and under at baseline. As discussed in Section 4.3, non-employment at year six is a very limited concern for this subset of participants. As we see in Panel A, the estimated impact of the intervention on earnings for this subset is only slightly smaller than that of the full sample. As shown in Panel B, the estimated impact on family income falls from 4% to 3% when we limit the sample to this younger subset of participants. However, since our family income estimates are not affected by non-employment, the observed decline in this estimated parameter is due to the age restriction rather than sample selection issues. These results, along with our earlier discussion of Lee's bounding method, suggest that our main results are not substantively affected by selection due to non-employment.

4.5 Earnings Impacts by Baseline Health

We next examine the heterogeneity in the impact on earnings by baseline health characteristics.[34] We use the same categorizations that compose the predicted health impacts shown in Appendix Table A1 in order to delineate baseline health.[35] We use specifications analogous to those shown in column (5) of Table 3 which include baseline earnings, health, and demographics as controls. As we find little evidence that non-employment impacts our estimates in Table 3, we do not account for sample selection in estimating the heterogeneity in the earnings impacts by baseline health.

The results are presented in Figure 8. Each panel of the figure shows the results from a different regression.[36] As such, we do not estimate the impact of the intervention by jointly including each

[34]The estimated heterogeneity of the impacts on family income are qualitatively similar to those found for earnings and are not reported here. These results are available from the authors.

[35]The one exception is for serum cholesterol. Since only 16% of participants are below under 220 mg/dl at baseline, we raise the threshold to under 240 mg/dl for the earnings regressions at which point roughly one-third of participants are below the threshold.

[36]Following the discussion by Cameron and Miller (2015), each 95% wild cluster bootstrap confidence interval shown in Figure 8 is the set of candidate null hypotheses for which the p-value of the test of that null hypothesis equals or exceeds 0.05.

baseline health characteristic but rather do so separately for each health characteristic. The top left panel of Figure 8 finds a three percent earnings impact for the high cholesterol group that is statistically significant. The impact for the low cholesterol group is nearly identical to that of the high cholesterol group and we cannot reject the null hypothesis that the treatment effect is the same for those with low and high cholesterol at baseline (p-value=0.932). We also find that the impact by baseline blood pressure, shown in the center of the top row of Figure 8, is essentially the same for both blood pressure groups (p-value=0.654).

The impact by baseline smoking status is shown in the top right corner of Figure 8. The left two estimates show the earnings impacts by whether or not the participant was a smoker at baseline. While the estimated impact is over two percentage points greater for non-smokers than for smokers, the difference in the earnings impact between smokers and non-smokers is not statistically significant (p-value=0.241). The estimated impact for smokers hides some interesting differences across groups defined by the baseline daily number of cigarettes smoked. We also estimate a specification in which we allowed the impact on earnings to depend upon whether the individual is a non-smoker, smoked less than twenty cigarettes daily, smoked twenty to thirty-nine cigarettes daily, or smoked forty or more cigarettes per day. As shown in the top right corner of Figure 8, we find a large and significant effect of the intervention on the earnings of light smokers (less than twenty cigarettes per day), while the effect of the intervention diminishes as the number of cigarettes smoked at baseline increases. Moreover, the test of the null hypothesis that the impact of the intervention on earnings is the same for all three groups of smokers is marginally significant (p-value=0.053). We cannot reject, however, that the impact is the same for non-smokers and light smokers.

We also examine the earnings impact by baseline CHD mortality risk quartile. The estimated impacts are roughly three and a half percent for the first through third quartiles. The estimated earnings impact for the highest quartile is less than two percent and is statistically insignificant. However, despite these patterns, we cannot reject the null hypothesis that the earnings impact by is the same across all CHD mortality risk quantiles (p-value=0.605).[37] In addition, recall that a

[37]However, we can reject the null hypothesis that the treatment effects for the four quantiles all equal zero (p-value=0.045).

condition for being enrolled in MRFIT is to have high CHD risk relative to the broader population so that the results need to be interpreted in this context. That is, among individuals with relatively high CHD mortality risk, the point estimates suggest that the intervention raised the earnings for everyone except those with the highest CHD mortality risk.

Finally, we examine the earnings impact by baseline BMI across three groups: normal weight, overweight, and obese.[38] The intervention raises earnings by three percent for overweight and by five percent for obese individuals.[39] The impact on the earnings of normal weight individuals is less than one percent. However, we are unable to reject the hypothesis that the effects are the same for all three groups defined by BMI (p-value=0.223).

4.6 Earnings Impacts by Baseline Occupation Characteristics

We also examine whether the impact of the intervention on earnings varies systematically with occupational characteristics. Participant occupation information was collected both at baseline and at year six and was coded using the 1970 Census three-digit occupation classification system. Not surprisingly, we find that baseline earnings are higher for white collar workers relative to both blue collar workers and farm/service sector workers.[40] We estimate that the intervention increases year six earnings by over four percent for white collar workers, a finding which is statistically significant. The intervention increases earnings by less than one percent for blue collar workers. However, we cannot reject the null hypothesis that the treatment effects are equal across occupation groups, although when comparing just white collar and blue collar workers the difference is marginally significant.[41]

In addition, we examine whether the earnings impact differs by the routine and non-routine task

[38] We follow standard BMI based classifications in defining participants to be normal weight ($18.5 \leq \text{BMI} < 25$), overweight ($25 \leq \text{BMI} < 30$), and obese ($\text{BMI} \geq 30$). Roughly half of participants are classified as overweight at baseline with roughly a quarter of participants falling into each of the other two groups. Very few participants in MRFIT are classified as underweight ($\text{BMI} < 18.5$) and they are placed in the normal weight group for our analysis.

[39] Our specification includes main effects for the BMI categories in addition to the interactions of the BMI categories with experimental group assignment.

[40] We use standard classification schemes based on the 1970 Census three-digit occupation system to defined white collar workers (occupation code less than 400), blue collar (occupation codes greater than 400 but less than 800), and farm/service workers (occupation code greater than 800).

[41] The p-value for the test of the joint null hypothesis that the impact of the intervention is the same for all three occupation groups is 0.135. The p-value for the test of the joint null hypothesis that the impact is the same for blue and white collar workers is 0.054. Results by occupation group are not reported in tables but are available from the authors by request.

content of occupations using the methodology of Autor, Levy, and Murnane (2003) to define task content.[42] We find that baseline earnings increase with higher levels of an occupation's non-routine cognitive tasks (proxied by quantitative reasoning, and direction, control, and planning of tasks) and (insignificantly with) non-routine manual tasks (eye-hand-foot coordination) and with lower levels of an occupation's routine manual tasks (finger dexterity) and routine cognitive tasks (set limits, tolerances, or standards). However, we do not find any strong evidence that interactions between the intervention and the task content of occupations affect year six earnings as these interactions are jointly insignificant while only the finger dexterity dimension is marginal significant.[43]

4.7 The Impact of MRFIT on Work-Related Health Outcomes

Our results indicate that the MRFIT intervention not only improved health outcomes but also increased the earnings and family income of the SI group. However, while there is some suggestive evidence that the intervention has heterogeneous earnings impacts associated with baseline characteristics, none of the differences we examine are statistically significant. Thus, the exact mechanism(s) through which the improved health raises earnings and income is unclear.

The canonical framework for understanding the relationship between health and earnings is Grossman's (1972) model. In this framework, investments in health raise one's stock of health capital but do not increase worker productivity (i.e., human capital). However, more health capital increases the amount of healthy time that one is able to devote to various tasks, most notably market work, and thereby provides a link between health improvements and market earnings. While we lack exact measures of worker productivity (e.g., hourly wages) and non-health human capital (aside from educational attainment), MRFIT participants are asked about changes in their health and labor market outcomes over the past year at baseline as well as at each of the first five annual interviews. We examine two of these questions which shed some light on the mechanisms described by Grossman.

The first question is "Within the past 12 months, have you experienced a physical illness which

[42] We use male occupation specific task content based on the 1977 Dictionary of Occupational Titles linked to the 1970 Census codes from Autor, Levy, and Murnane found at http://economics.mit.edu/faculty/dautor/data/autlevmurn03. For twelve occupation codes used in MRFIT but not found in the Autor et al. data, we use values averaged over from the same occupation subgroup.

[43] Task content results are not reported in tables but are available from the authors.

kept you in bed for a week or more, or sent you to the hospital?" The top panel of Figure 9 shows the impact of the intervention responses to this question for the SI group using equation (1) in which baseline health and demographics are included as controls. While there is no difference in the response to this question at baseline, the SI group is significantly less likely to report an affirmative answer in four of the five experimental waves in which the question is asked. Cumulatively, 44 percent of the UC group report an affirmative answer to this question at least one time during the first five years of the experimental period while the SI group is 2.3 percentage points less likely to do so, a difference that is statistically significant.[44]

The second question is "Within the past 12 months, have you experienced not being able to work because of a disability?" The bottom panel of Figure 9 examines whether the respondent reported experiencing a work-limiting disability. We find that respondents in the SI group experience 15 to 30 percent lower rates of disabilities that prevent work during the experimental period with these differences being significant for years two through four. Cumulatively, 15.5 percent of the UC group report being unable to work at least one time during the first five years of the experimental period while the SI group is 1.7 percentage points less likely to do so (although this difference is marginally significant). Interestingly, over two-thirds of participants who give a positive response to this question only do so one time which suggests that many of these reported work preventing disabilities are transitory. Indeed, as we have noted above, there is no significant difference in employment rates at year six between the two treatment groups which further suggests that these periods of disability are mainly transitory. Overall, the findings from these two questions indicate decreased rates of illness and injury for the SI group.

5 Discussion

This paper examines the labor market effects of a randomized health intervention of working-age men which was focused on reducing mortality due to coronary heart disease, the leading cause of death in the United States. MRFIT succeeded in improving the health of the Special Intervention group along several dimensions. We find that the intervention also significantly increased earnings

[44]These results are conditional on responding to the fifth annual interview. The estimated cumulative difference is regression adjusted based on equation (1).

by three percent and family income by four percent with no concurrent effect on labor force participation. Although there is some sample attrition and labor force exit, we find that accounting for sample selection has no substantive impact on our results. Our findings suggest that there may be differential impacts on earnings for groups defined by baseline health and occupational characteristics although these differences are only statistically significant in the case of baseline smoking. We further show that a partial explanation for the effect of health on earnings is an increase in the availability of healthy time for market work. This finding is consistent with the basic prediction of the seminal model of Grossman (1972).

It is unlikely, given the magnitude of these results, that this increase in healthy time alone is enough to explain the three percent earnings increase due to the intervention. While the available data constrain the number of other potential mechanisms we can examine, results presented throughout this paper allow us to examine the roles of some mechanisms that appear to have limited effects. First, the weight loss of the SI group may represent an increase in their attractiveness, which previous research has shown to be rewarded in the labor market (Hamermesh and Biddle 1994). However, Figure 2 panel (d) shows that the relative weight loss is only two-to-three pounds (on a base of 189 pounds), which we consider to be too small to represent meaningful increases in attractiveness that would generate the observed higher earnings.[45] Second, other observed labor market outcomes do not appear to be driving the results because, as discussed at the end of Section 4.3, there are not meaningful differences in employment or retirement between the two study groups. Results in that section also show that the share of participants experiencing a change in job title or kind of work is the same across the groups, so there is little reason to believe that the effects are driven by promotions or career advancement for the SI group. Finally, while the analyses in Sections 4.5 and 4.6 suggest larger effects for blue collar workers, participants with higher levels of baseline BMI, and participants with lower levels of baseline smoking, we are not able to statistically distinguish the effects for these groups from the effects for their complements. Ultimately, there are surely additional mechanisms that we are unable to examine, but our results indicate that the flow of healthy time is at least one contributing factor to the earnings and family

[45]There was slightly greater weight loss among participants who were obese at baseline, a group that had larger estimated earnings effects as shown in panel (e) of Figure 8. Still, average relative weight loss for the SI group among the obese was only four to five pounds.

income differences that we observe.

If better health capital indeed improves labor market outcomes through a greater flow of healthy time, we might also expect to see effects over the entire course of MRFIT. Almost all measures of health capital improve by year one of the study and change over the subsequent five years. A fully dynamic analysis would relate these changes over time to health working time and earnings, which could differentiate between possible relationships between the interventions and labor market outcomes. It is possible that the health changes at year one represent improvements in health capital that immediately bring about a greater flow of available working time and, ultimately, earnings. Alternatively, the relevant dimensions of health capital could evolve slowly as healthier behaviors are maintained throughout the duration of the study. This would deliver gradual increases in healthy time and corresponding gradual increases in earnings. Finally, a dynamic analysis might reveal increases in earnings and income that are not associated with healthy working time at all. Such a finding would suggest other mechanisms and aspects of health lead to greater income and earnings. Unfortunately, the data limitations prevent us from further exploring the relationship between health, healthy time, and economic outcomes throughout the intervention period as earnings and family income are only available at baseline and at year six.

Much of the development literature on health and economic outcomes explores the role that health plays in changing other measures of human capital. In US data, the model of Restuccia and Vandenbroucke (2013) suggests that life expectancy gains over the last half-century were responsible for a quarter of the increase in education over the same period. Extending this logic to MRFIT, it is possible that the SI group is not more productive due to health improvements but simply accumulates more human capital in anticipation of lower future mortality. The endogenous relationship between longevity and human capital, including health, is further highlighted in the model of Ehrlich and Chuma (1990). If large adjustments to human capital or other behaviors are at play, the results presented in this paper would not indicate the effect of health on labor market outcomes. However, there are a number of reasons to think that this is unlikely in our context. First, men in the age range of MRFIT are relatively less likely to make human capital adjustments than, say, school-aged individuals. Second, if human capital is already optimized with respect to health, changes to health will only induce second-order effects on human capital and earnings, as

argued by Bleakley (2010). If a MRFIT participant has already optimized his level of human capital as a function of market forces and health, the envelope theorem would imply that innovations to health should not produce first-order effects on earnings through human capital re-optimization. On the other hand, direct effects of health on productivity, particularly for individuals who are constrained by the flow of healthy time, could induce first order impacts on earnings. Finally, prior research finds that the impact of MRFIT on mortality was generally small, with no detectable effects on mortality measured seven years after the intervention began and only marginally-significant effects at the ten and sixteen year marks (Multiple Risk Factor Intervention Trial Research Group, 1982, 1990, 1996). While it is possible that perceptions of increased longevity may have induced relatively large changes in the non-health behavior of the SI group, these perceptions would have not matched the realized gains in longevity.

Ultimately, we find that MRFIT raised earnings and family income for the SI group relative to the UC group. We identify reduced time lost to disability and illness as a possible mechanism underlying this effect. As heart disease remains the leading cause of death in the United States and the health dimensions addressed by MRFIT continue to be public health concerns today, our findings are highly relevant for the current relationship between health and the US labor market. While the available data preclude us from studying all other possible mechanisms, our findings and the context suggest that other mechanisms, like individuals' attractiveness or human capital adjustment, are less likely. As such, we interpret our findings as demonstrating positive direct effects of health on earnings and family income.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge (2017) "When Should You Adjust Standard Errors for Clustering?" *National Bureau of Economic Research Working Paper # 24003.*

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham (2016) "Management and Shocks to Worker Productivity" *Unpublished Manuscript.*

Alderman, Harold, John Hoddinott, and Bill Kinsey (2006) "Long Term Consequences of Early Childhood Malnutrition," *Oxford Economic Papers,* 58(3):450-474.

Almond, Douglas (2006) "Is the 1918 Influenza Pandemic Over? Long-Term Effects of *In Utero* Influenza Exposure in the Post-1940 U.S. Population," *Journal of Political Economy,* 114(4): 672-712.

Aragón, Fernando M., Juan Jose Miranda, and Paulina Oliva (2016) "Particulate Matter and Labor Supply: The Role of Caregiving and Non-Linearities" *Unpublished Manuscript.*

Basta, Samir S., Soekirman, Darwin Karyadi, and Nevin S. Scrimshaw (1979) "Iron deficiency anemia and the productivity of adult males in Indonesia," *The American Journal of Clinical Nutrition,* 32(4): 916-925.

Benfari, Robert C. (1981) "The Multiple Risk Factor Intervention Trial (MRFIT): III. The Model for Intervention," *Preventive Medicine,* 10(4):426-442.

Bleakley, Hoyt (2010) "Health, Human Capital, and Development," *Annual Review of Economics,* 2:283-310.

Caggiula, Arlene W., George Christakis, Marilyn Farrand, Stephen B. Hulley, Ruth Johnson, Norman L. Lasser, Jeremiah Stamler, Graham Widdowson (1981) "The Multiple Risk Factor Intervention Trial (MRFIT): IV. Intervention on Blood Lipids," *Preventive Medicine,* 10(4):443-475.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2008) "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics,* 90(3):414-427.

Cameron, A. Colin and Douglas L. Miller (2015) "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources,* 50(2):317-372.

Case, Anne, Darren Lubotsky, and Christina Paxson (2002) "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, 92(5):1308-34.

Cawley, John, and Christopher J. Ruhm (2011) "The Economics of Risky Health Behaviors," *Handbook of Health Economics,* 2:95-199.

Cohen, Jerome D., Richard H. Grimm, Jr., and W. McFate Smith (1981) "The Multiple Risk Factor Intervention Trial (MRFIT): VI. Intervention on Blood Pressure," *Preventive Medicine,* 10(4):501-518.

Currie, Janet and Brigette C. Madrian (1999) "Health, Health insurance and the Labor Market,"

in Ashenfelter, Orley C. and David Card (Eds.), Handbook of Labor Economics. Elsevier, Amsterdam, pp. 33093416.

Currie, Janet and Mark Stabile (2003) "Socioeconomic Status and Child Health: Why Is the Relationship Stronger for Older Children?" *American Economic Review*, 93(5):1813-23.

Dillon, Andrew, Jed Friedman, and Pieter Serneels (2014) "Health Information, Treatment, and Worker Productivity: Malaria Testing and Treatment among Nigerian Sugarcane Cutters," Discussion Paper no. 8074, IZA, Bonn, Germany.

Dow, Will, Paul Gertler, Robert F. Schoeni, John Strauss, and Duncan Thomas (1997) "Health Care Prices, Health and Labor Outcomes: Experimental Evidence," Working Paper no. 97-01, RAND, Santa Monica, CA.

Edgerton, V. R., G. W. Gardner, Y. Ohira, K. A. Gunawardena, and B. Senewiratne (1979) "Iron-deficiency anaemia and its effect on worker productivity and activity patterns," *British Medical Journal,* 2(6204):1546-1549.

Ehrlich, Isaac, and Hiroyuki Chuma (1990) "A Model of the Demand for Longevity and the Value of Life Extension," *Journal of Political Economy,* 98(4):761-782.

Field, Erica, Omar Robles, and Maximo Torero (2009) "Iodine Deficiency and Schooling Attainment in Tanzania," *American Economic Journal: Applied Economics*, 1(4):140-169.

Fink, Günther, and Felix Masiye (2015) "Health and agricultural productivity: Evidence from Zambia," *Journal of Health Economics*, 42:151-164.

Frankenberg, Elizabeth, Wayan Suriastini, and Duncan Thomas (2005) "Can Expanidng Access to Basic Healthcare Improve Children's Health Status? Lessons from Indonesia's 'Midwife in the Village' Programme," *Population Studies*, 59(1):5-19.

Frijters, Paul, John P. Haisken-DeNew, and Michael A. Shields (2005) "The causal effect of income on health: Evidence from German reunification," *Journal of Health Economics,* 24:997-1017.

Glewwe, Paul, and Edward A. Miguel (2008) "The Impact of Child Health and Nutrition on Education in Less Developed Countries," *Handbook of Development Economics,* 4:3561-13606.

Graff Zivin, Joshua, and Matthew Niedell (2012) "The Impact of Pollution on Worker Productivity" *American Economic Review,* 102(7):3652-3673.

Grossman, Michael (1972) "On the Concept of Health Capital and the Demand for Health" *Journal*

of Political Economy, 80(2):223-255.

Haas, Jere D., and Thomas Brownlie IV (2001) "Iron Deficiency and Reduced Work Capacity: A Critical Review of the Research to Determine a Causal Relationship " *The Journal of Nutrition,* 131(2):676S-690S.

Hamermesh, Daniel S. and Biddle, Jeff E. (1994) "Beauty and the Labor Market," *American Economic Review*, 84(5):1174-94.

Heckman, James J. (1979) "Sample Selection Bias as a Specification Error," *Econometrica,* 47(1): 153-62.

Hoddinott, John, John A. Maluccio, Jere R. Behrman, Rafael Flores, and Reynaldo Martorell (2008) "Effect of a nutrition intervention during early childhood on economic productivity of Guatemalan adults" *Lancet,* 371:411-416.

Hughes, Glenn H., Norman Hymowitz, Judith K. Ockene, Nathan Simon, and Thomas M. Vogt (1981) "The Multiple Risk Factor Intervention Trial (MRFIT): V. Intervention on Smoking," *Preventive Medicine,* 10(4):476-500.

Kline, Patrick and Andres Santos (2012) "A Score Based Approach to Wild Bootstrap Inference," *Journal of Econometric Methods,* 1(1): 23-41.

Kochanek Kenneth D., Sherry L. Murphy, Jiaquan Xu, and Betzaida Tejada-Vera. (2016) "Deaths: Final data for 2014," *National vital statistics reports* 65(4). Hyattsville, MD: National Center for Health Statistics.

Lee, David S. (2009) "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies,* 76(3)1071-1102.

Li, Ruowei, Xuecun Chen, Huaicheng Yan, Paul Deurenberg, Lars Garby, and Joseph G.A.J. Hautvast (1994) "Functional consequences of iron supplementation in iron-deficient female cotton mill workers in Beijing, China," *The American Journal of Clinical Nutrition,* 59(4): 908-913.

Lucas, Adrienne M. (2010) "Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka," *American Economic Journal: Applied Economics*, 2(2):46-71.

Maluccio, John A., John Hoddinott, Jere R. Behrman, Reynaldo Martorell, Agnes R. Quisumbing, and Aryeh D. Stein (2009) "The Impact of Improving Nutrition during Early Childhood on Education among Guatemalan Adults" *The Economic Journal,* 119(537):734-763.

Multiple Risk Factor Intervention Trial Research Group (1982) "Multiple Risk Factor Intervention Trial: Risk Factor Changes and Mortality Results," *Journal of the American Medical Association,* 248:1465-1477.

Multiple Risk Factor Intervention Trial Research Group (1990) "Mortality Rates After 10.5 Years for Participants in the Multiple Risk Factor Intervention Trial: Findings Related to A Priori Hypotheses of the Trial," *Journal of the American Medical Association,* 263:1795-1801.

Multiple Risk Factor Intervention Trial Research Group (1996) "Mortality After 16 Years for Participants Randomized to the Multiple Risk Factor Intervention Trial," *Circulation,* 94:946-951.

Neaton, James D., Steven Broste, Louis Cohen, Eugene L. Fishman, Marcus O. Kjelsberg, and James Schoenberger (1981) "The Multiple Risk Factor Intervention Trial (MRFIT): VII. A Comparison of Risk Factor Changes between the Two Study Groups," *Preventive Medicine,* 10(4):519-543.

Newhouse, Joseph P., and the Insurance Experiment Group (1993) *Free for All? Lessons from the RAND Health Insurance Experiment,* Harvard University Press, Cambridge.

ODonnell, Owen, Eddy Van Doorslaer, and Tom Van Ourti (2015) "Health and inequality," In Anthony B. Atkinson and Francois Bourguignon (eds.) Handbook of Income Distribution 2, 14911533.

Restuccia, Diego, and Guillaume Vandenbroucke (2013) "A Century of Human Capital and Hours," *Economic Inquiry,* 51(3):1849-1866.

Rivers, Douglas, and Quang H. Vuong (1988) "Limited Information Estimators and Exogeneity Tests for Simultaeneous Probit Models," *Journal of Econometrics,* 39:347-366.

Roseboom, Tessa J., Jan H.P. van der Meulen, Anita C. J. Ravelli, Clive Osmond, David J.P. Barker, and Otto P. Bleker (2001) "Effects of Prenatal Exposure to the Dutch Famine on Adult Disease in Later Life: An Overview," *Twin Research,* 4(5):293-298.

Sherwin, Roger, Charles T. Kaelber, Paul Kezdi, Marcus O. Kjelsberg, and H. Emerson Thomas, Jr. (1981) "The Multiple Risk Factor Intervention Trial (MRFIT): II. The Development of the Protocol," *Preventive Medicine,* 10(4):402-425.

Stephens Jr., Melvin and Takashi Unayama (2015) "Estimating the Impacts of Program Benefits: Using Instrumental Variables with Underreported and Imputed Data" *National Bureau of*

Economic Research Working Paper # 21248.

Terza, Joseph V., Anirban Basu, and Paul J. Rathouz (2008) "Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling," *Journal of Health Economics*, 27:531-543.

Thirumurthy, Harsha, Joshua Graff Zivin, and Markus Goldstein (2008) "The Economic Impact of AIDS Treatment: Labor Suply in Western Kenya," *Journal of Human Resources*, 43(3):511-552.

Thomas, Duncan (2009) "The causal effect of health on social and economic prosperity: Methods and findings" *Unpublished Manuscript.*

Thomas, Duncan, Elizabeth Frankenberg, Jed Friedman, Jean-Pierre Habicht, Mohammed Hakimi, Nicholas Ingwersen, Jaswadi, Nathan Jones, Christopher McKelvey, Gretel Pelto, Bondan Sikoki, Teresa Seeman, James P. Smith, Cecep Sumantri, Wayan Suriastini, and Siswanto Wilopo (2006) "Causal Effect of Health on Labor Market Outcomes: Experimental Evidence," Working Paper no. 070-06, California Center for Population Research, Los Angeles, CA.

Torpy Janet M., Alison E. Burke, and Richard M. Glass (2009) "Coronary Heart Disease Risk Factors," *Journal of the American Medical Association,* 302(21):2388.

Wooldridge, Jeffrey M. (2015) "Control Function Methods in Applied Econometrics," *Journal of Human Resources*, 50(2):420-445.

Zabel, Jeffrey E. (1998) "An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior," *The Journal of Human Resources,* 33(2):479-506.

Zukel, William J., Oglesby Paul, and Harold W. Schnaper (1981) "The Multiple Risk Factor Intervention Trial (MRFIT): I. Historical Perspectives," *Preventive Medicine,* 10(4):387-401.

Table 1: Balance of Baseline Characteristics

	SI	UC	Difference
Age	46.43	46.35	0.09
			[0.50]
White	0.898	0.905	-0.007
			[0.34]
HS Grad	0.211	0.208	0.003
			[0.66]
Some College	0.358	0.350	0.008
			[0.44]
College Grad	0.269	0.279	-0.010
			[0.19]
Married	0.887	0.889	-0.002
			[0.76]
Serum Cholesterol	254	254	0.22
			[0.75]
Smoker	0.593	0.590	0.004
			[0.65]
Cigs/Day (w/zeroes)	19.2	19.4	-0.13
			[0.73]
Diastolic Blood Pressure	90.7	90.7	0.02
			[0.85]
CHD Mortality Risk (%)	2.08	2.09	-0.016
			[0.50]

Notes: The p-values reported in brackets are from using a wild cluster bootstrap which clusters at the clinic level. See the text and Neaton et al. (1981) for the calculation of CHD mortality risk. The analysis sample contains 12,562 participants, 6,291 in the SI group and 6,271 in the UC group, who have non-missing data for age, race, education, marital status, and employment status measured at baseline.

Table 2: Sources of Sample Selection

	Share of Participants Dropped Due to:			
	Attrition	Non-Employment	Missing Data	Total
A. Baseline Earnings				
SI	-	0.5%	3.4%	3.9%
UC	-	0.6%	4.0%	4.5%
B. Baseline Family Income				
SI	-	-	3.5%	3.5%
UC	-	-	3.8%	3.8%
C. Year Six Earnings				
SI	9.8%	10.2%	9.4%	29.4%
UC	11.6%	10.2%	7.7%	29.5%
D. Year Six Family Income				
SI	9.8%	-	9.2%	19.0%
UC	11.6%	-	7.6%	19.2%

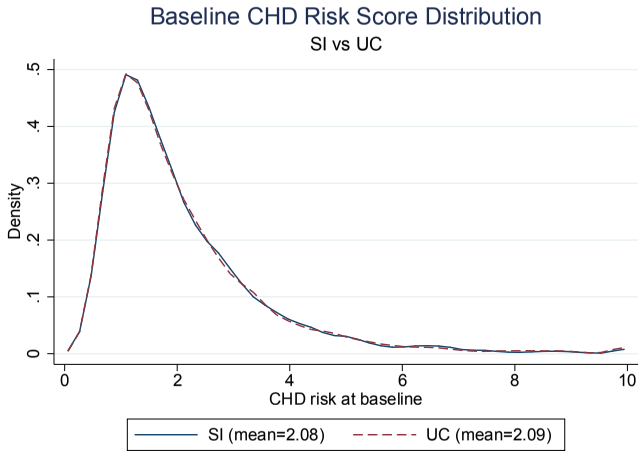
Notes: Authors' calculations. All percentages indicate the shares of the full sample of 12,866 respondents in MRFIT that are missing from baseline and year six analyses for the indicated reasons.

Table 3: Earnings and Family Income Regressions

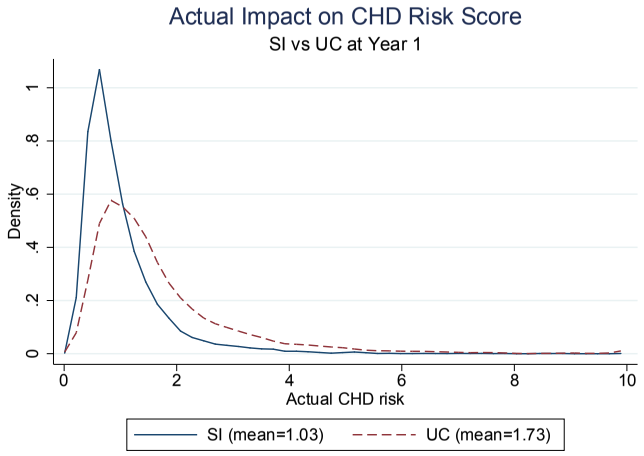
	Baseline		Year Six			Year Six Age≤ 48
	(1)	(2)	(3)	(4)	(5)	(6)
A. Earnings						
SI	-0.015	-0.010	0.020	0.027	0.031	0.027
	[0.025]	[0.074]	[0.025]	[0.011]	[0.004]	[0.010]
N	12,326	12,321	9,508	9,215	9,077	5,881
B. Family Income						
SI	-0.013	-0.010	0.035	0.039	0.041	0.031
	[0.111]	[0.183]	[0.006]	[0.004]	[0.004]	[0.043]
N	12,399	12,395	10,899	10,577	10,410	6,340
Additional Controls:						
Baseline health & demographics		X			X	X
Baseline outcome				X	X	X

Notes: This table reports ordered probit estimates in which the cutpoints are known and the unobserved latent outcome is assumed to be log normally distributed. The baseline health and demographic controls are serum cholesterol, diastolic blood pressure, number of cigarettes smoked, an indicator for being a smoker, a full set of indicators for age, an indicator for being white, indicators for four education groups, and a marital status indicator. The baseline outcome controls used for the year six outcomes in columns (4)-(6) are a set of indicators for the corresponding outcome at baseline. The earnings regressions are restricted to those who are employed for the relevant survey waves. Column (6) further restricts to participants who were 48 or younger at baseline. The outcomes are nine-group categorical earnings and income measures with cut points at \$4200, \$7200, \$10,000, \$12,000, \$15,000, \$18,000, \$22,500, and \$35,000. Kline-Santos wild cluster bootstrap p-values for the null that the parameter is equal to 0 are reported in brackets clustering at the clinic level.

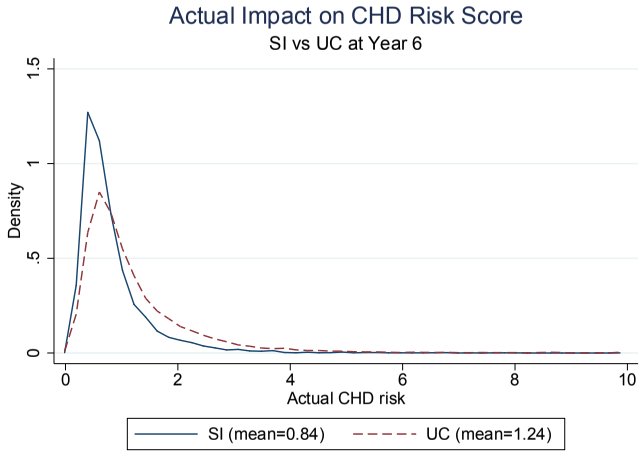
Figure 1: Experimental Impact on CHD Risk



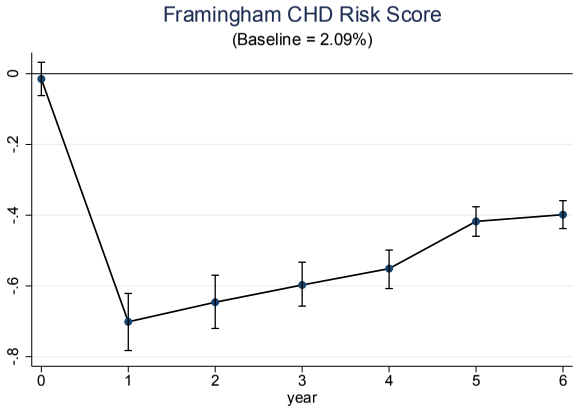
(a) Baseline CHD Risk



(b) Impact at Year One



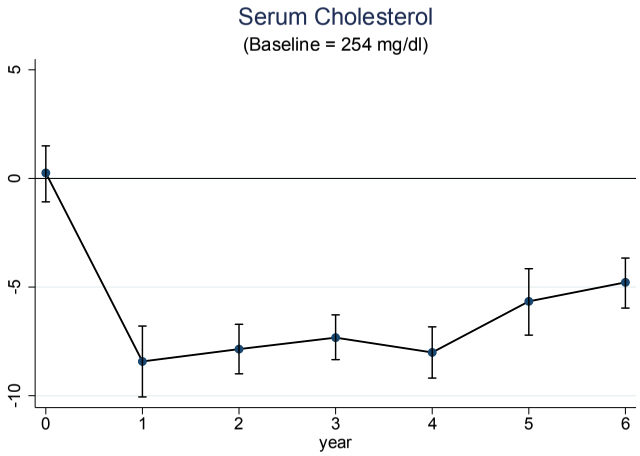
(c) Impact at Year Six



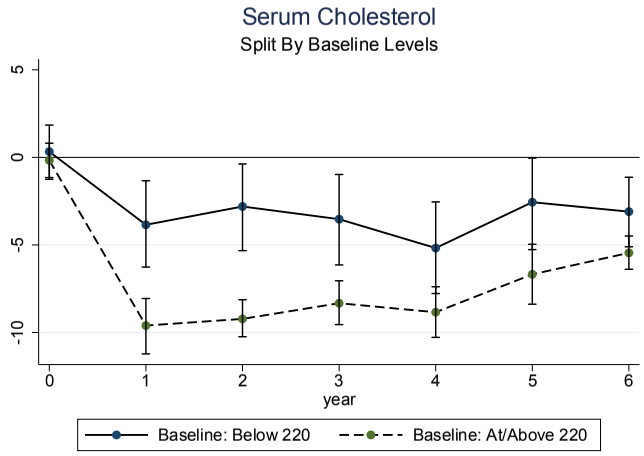
(d) Longitudinal Impact on CHD risk

Notes: See the text and Neaton et al. (1981) for the calculation of CHD mortality risk. Risk score levels are calculated using observed risk factors in each year. The densities are smoothed with an Epanechnikov kernel. Treatment effect estimates are the estimated coefficients on an indicator for being in the SI group from linear regressions of the form of equation(1). The regression controls include a full set of indicators for age, an indicator for being white, indicators for four education groups, and a marital status indicator. The 95% confidence interval bars shown in the bottom right panel are from using a wild cluster bootstrap which clusters at the clinic level. The sample is initially restricted to the 12,562 MRFIT respondents with nonmissing age, education, marital status, race, and employment status at baseline. Estimates for each year further restrict to observations with nonmissing outcomes and controls for that year.

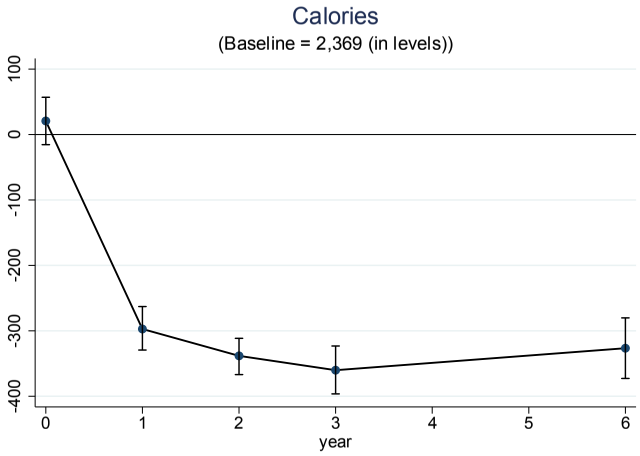
Figure 2: Experimental Impact on Serum Cholesterol, Calories, and Body Weight



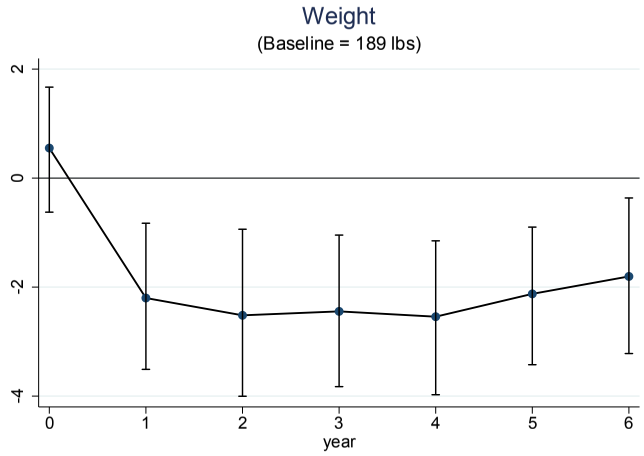
(a) Overall Serum Cholesterol



(b) Serum Cholesterol By Baseline Levels



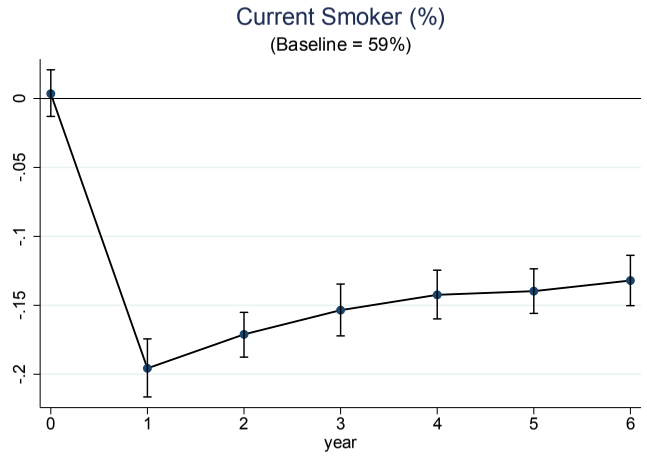
(c) Calories



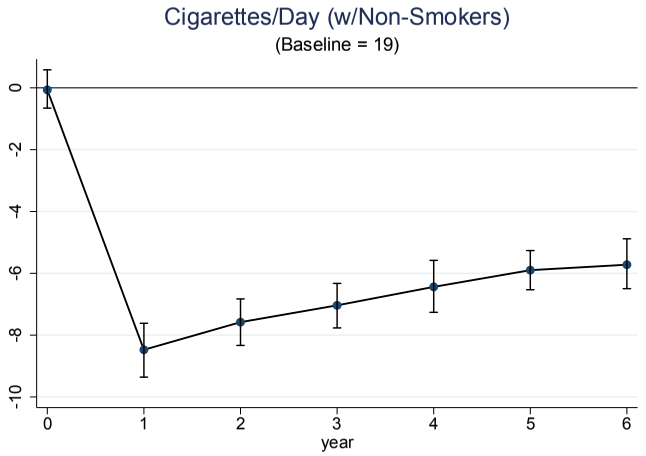
(d) Body Weight

Notes: Each point is coefficient from a different regression of the form of equation (1). The 95% confidence interval bars are from using a wild cluster bootstrap which clusters at the clinic level. The regression controls are baseline measures and include a full set of indicators for age, an indicator for being white, indicators for four education groups, and a marital status indicator. The sample is initially restricted to the 12,562 MRFIT respondents with nonmissing age, education, marital status, race, and employment status at baseline. Estimates for each year further restrict to observations with nonmissing outcomes and controls for that year.

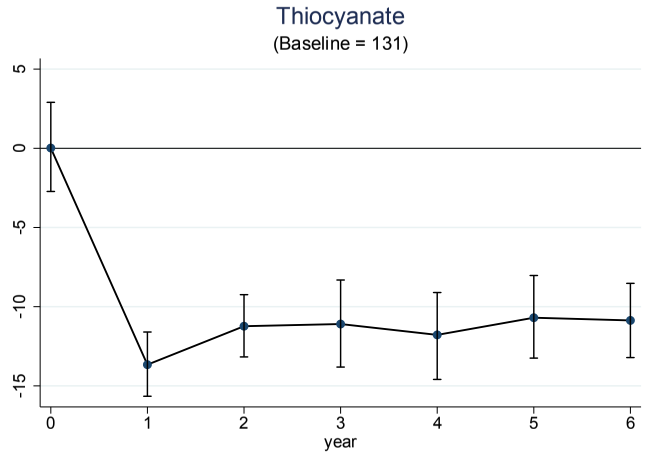
Figure 3: Experimental Impact on Smoking



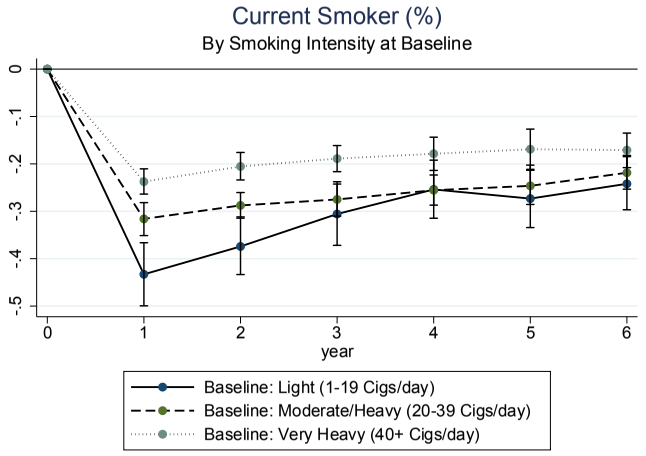
(a) Smoker



(b) Cigarettes/Day



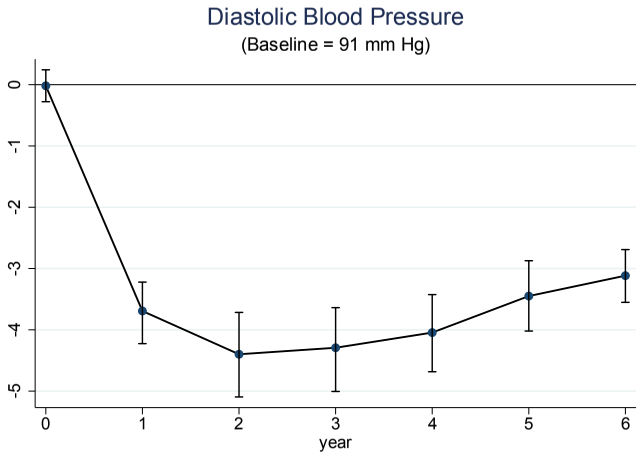
(c) Serum Thiocyanate



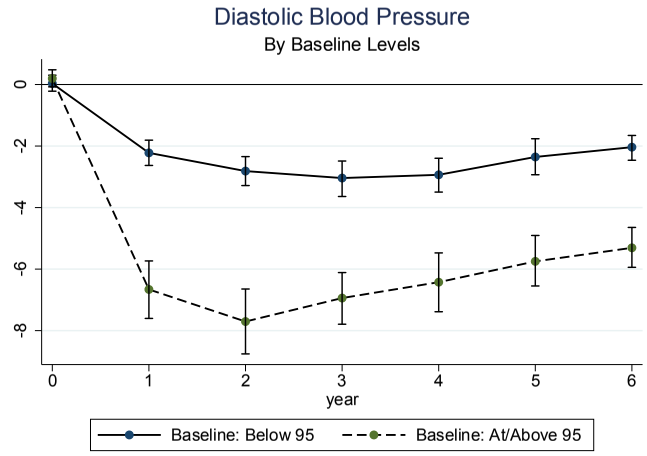
(d) Smoker

Notes: Each point is coefficient from a different regression of the form of equation (1). The 95% confidence interval bars are from using a wild cluster bootstrap which clusters at the clinic level. The regression controls are baseline measures and include a full set of indicators for age, an indicator for being white, indicators for four education groups, and a marital status indicator. The sample is initially restricted to the 12,562 MRFIT respondents with nonmissing age, education, marital status, race, and employment status at baseline. Estimates for each year further restrict to observations with nonmissing outcomes and controls for that year.

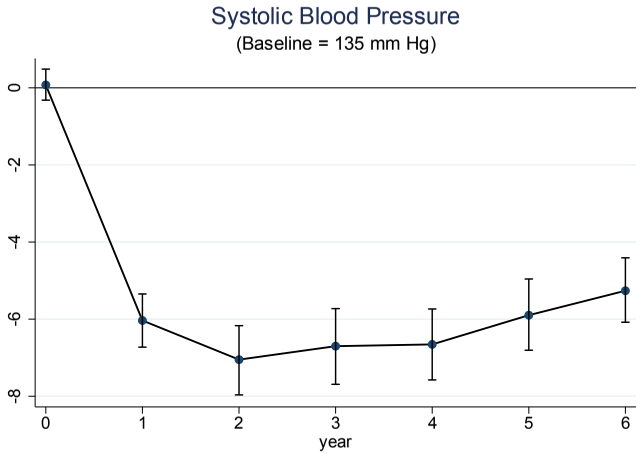
Figure 4: Experimental Impact on Blood Pressure



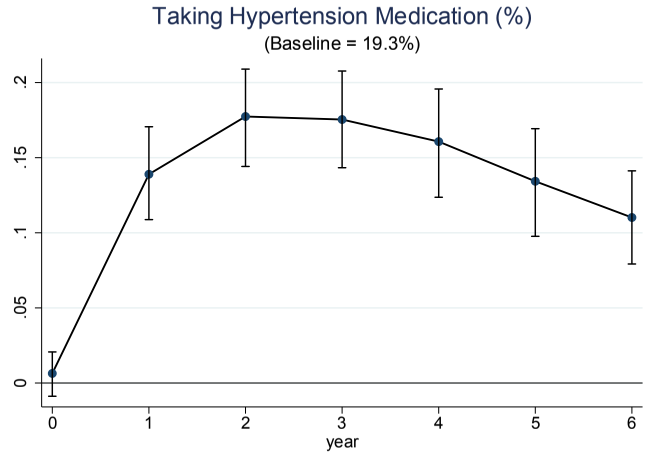
(a) Diastolic Blood Pressure



(b) Diastolic Blood Pressure by Baseline



(c) Systolic Blood Pressure



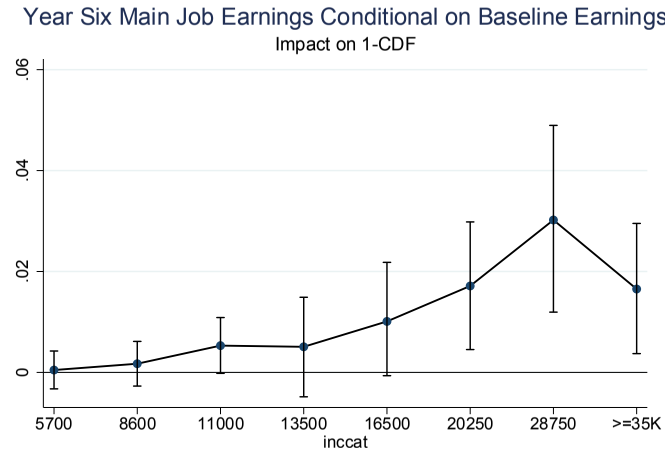
(d) Taking Blood Pressure Medication

Notes: Each point is coefficient from a different regression of the form of equation (1). The 95% confidence interval bars are from using a wild cluster bootstrap which clusters at the clinic level. The regression controls are baseline measures and include a full set of indicators for age, an indicator for being white, indicators for four education groups, and a marital status indicator. The sample is initially restricted to the 12,562 MRFIT respondents with nonmissing age, education, marital status, race, and employment status at baseline. Estimates for each year further restrict to observations with nonmissing outcomes and controls for that year.

Figure 5: Experimental Impact on Earnings

(a) (1-CDF) at Baseline

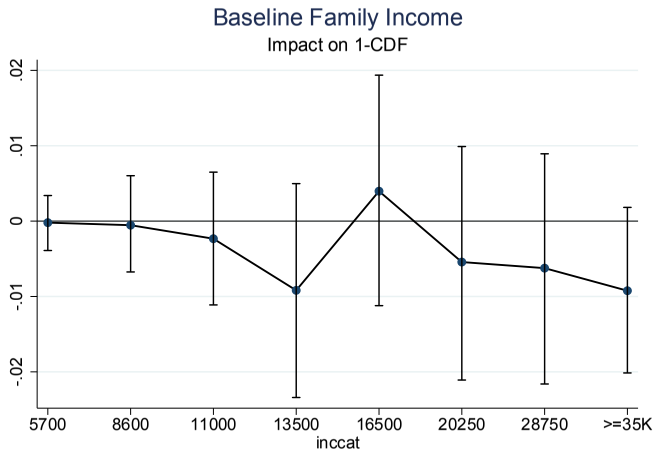
(b) (1-CDF) at Year Six



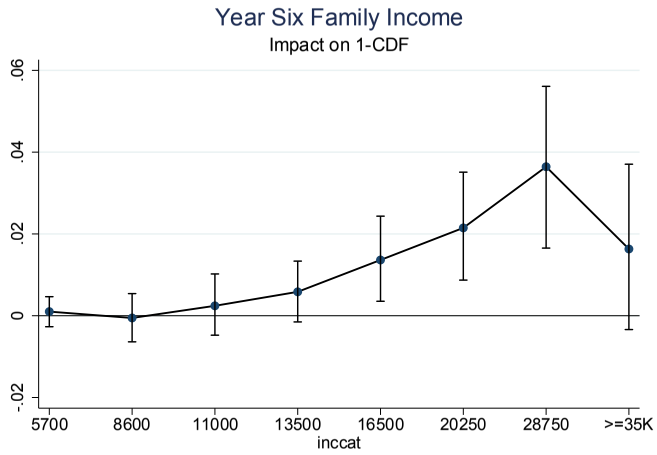
(c) (1-CDF) at Year Six Conditional on Baseline Earnings

Note: All three panels display estimated coefficients on an indicator for being in the SI group. Each point is a coefficient from a different regression of the form of equation (1), where the outcome is a binary variable for having earnings at or above the given earnings group at baseline (Panel (a)) or year six (Panels (b) and (c)). The 95% confidence interval bars are from using a wild cluster bootstrap which clusters at the clinic level. In both Panels (a) and (b), the regressions include baseline health and demographic controls: serum cholesterol, diastolic blood pressure, number of cigarettes smoked, an indicator for being a smoker, a full set of indicators for age, an indicator for being white, indicators for four education groups, and a marital status indicator. Panel (c) further includes a set of indicators for each of the baseline earnings categories. All three panels use a sample that is initially restricted to the 12,562 respondents with nonmissing baseline demographics. The results in panel (a) are estimated using the 12,321 baseline-employed respondents with nonmissing earnings. The results in panel (b) are estimated using the 9,179 respondents employed in year six with nonmissing earnings and all variables necessary to calculate CHD risk. The results in panel (c) are estimated using the 9,077 respondents meeting the sample requirements for both panels (a) and (b).

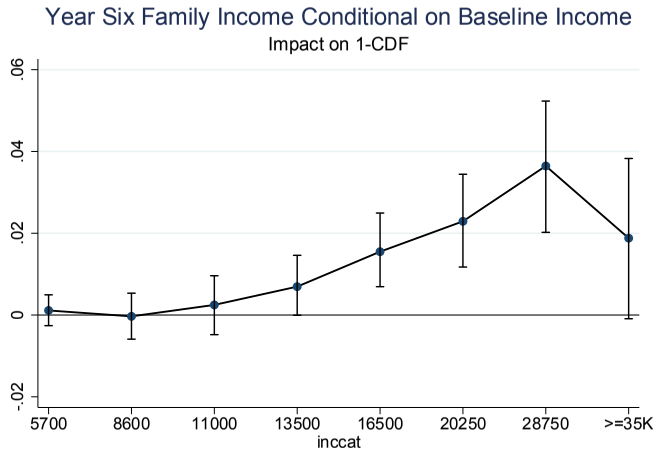
Figure 6: Experimental Impact on Income



(a) (1-CDF) at Baseline



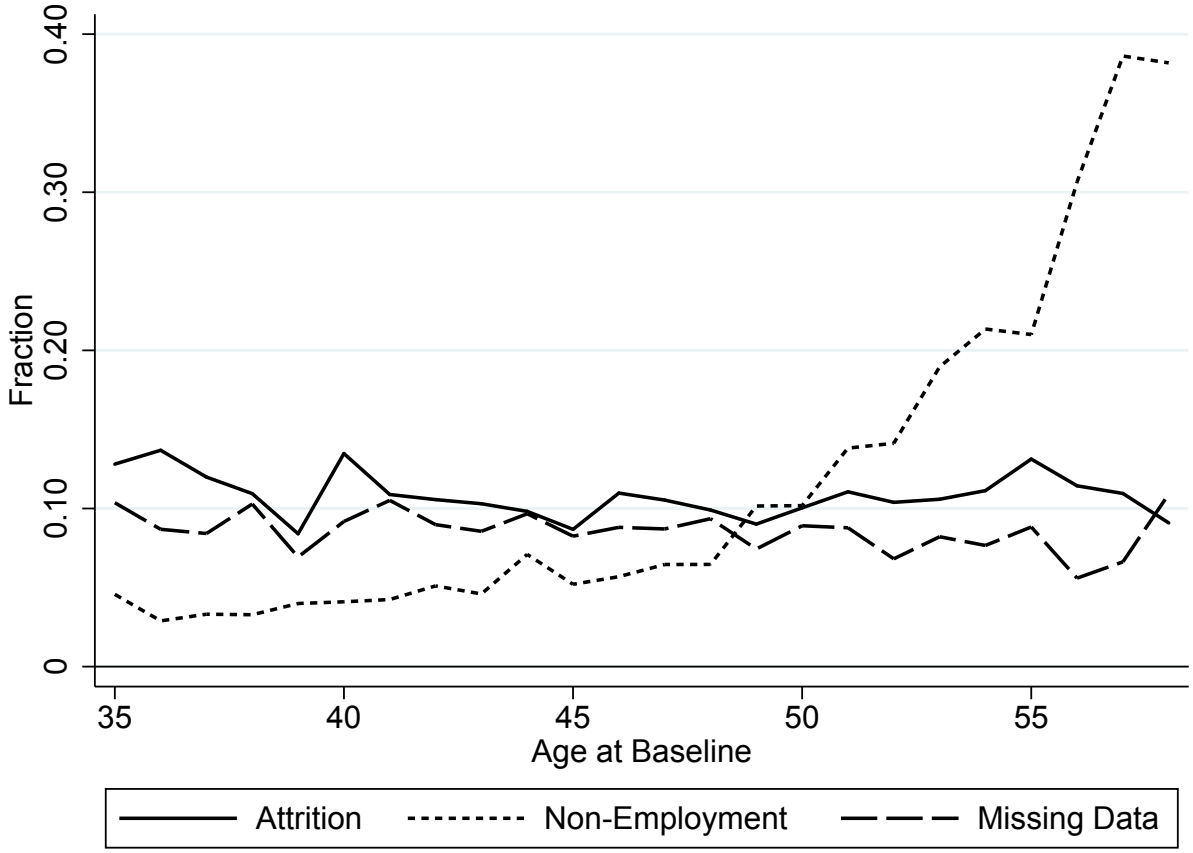
(b) (1-CDF) at Year Six



(c) (1-CDF) at Year Six Conditional on Baseline Income

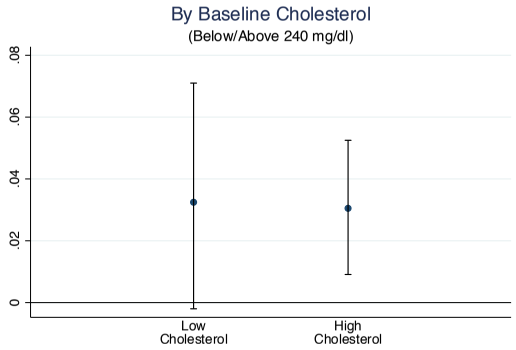
Note: All three panels display estimated coefficients on an indicator for being in the SI group. Each point is a coefficient from a different regression of the form of equation (1), where the outcome is a binary variable for having family income at or above the given income group at baseline (Panel (a)) or year six (Panels (b) and (c)). The 95% confidence interval bars are from using a wild cluster bootstrap which clusters at the clinic level. In both Panels (a) and (b), the regressions include baseline health and demographic controls: serum cholesterol, diastolic blood pressure, number of cigarettes smoked, an indicator for being a smoker, a full set of indicators for age, an indicator for being white, indicators for four education groups, and a marital status indicator. Panel (c) further includes a set of indicators for each of the baseline income categories. All three panels use a sample that is initially restricted to the 12,562 respondents with nonmissing baseline demographics. The results in panel (a) are estimated using the 12,395 nonmissing baseline earnings and income. The results in panel (b) are estimated using the 10,504 respondents with nonmissing year six employment and income and all variables necessary to calculate CHD risk. The results in panel (c) are estimated using the 10,410 respondents meeting the sample requirements for both panels (a) and (b).

Figure 7: Sources of Sample Selection at Year Six by Age at Baseline

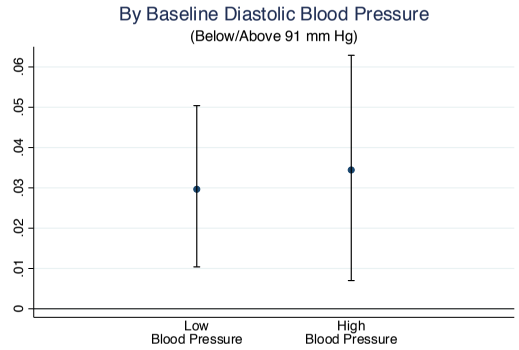


Notes: Missing data calculations are for year six earnings regressions. The three lines show the shares of the full 12,866 MRFIT respondents who are excluded from these regressions for three mutually exclusive reasons by baseline age. The attrition line indicates the share of respondents who do not appear in the year six data. The nonemployment line shows the share of respondents who appear in the year six data with nonmissing employment data but are not working. The missing data line shows the share of respondents who are labor force participants in the year six data but are not employed at baseline or are otherwise missing baseline or year six demographic or income data.

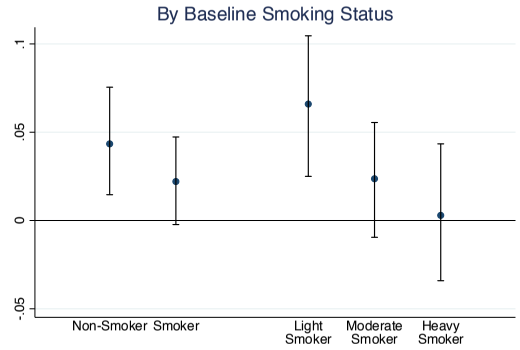
Figure 8: Heterogeneous Earnings Impacts By Baseline Health



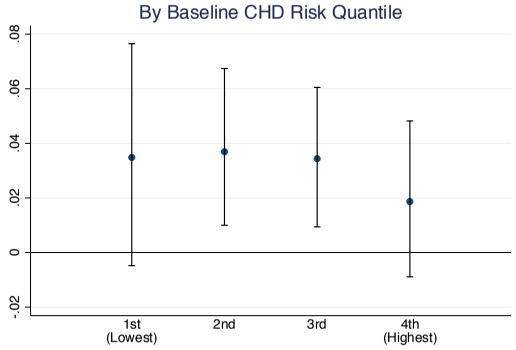
(a) By Baseline Cholesterol



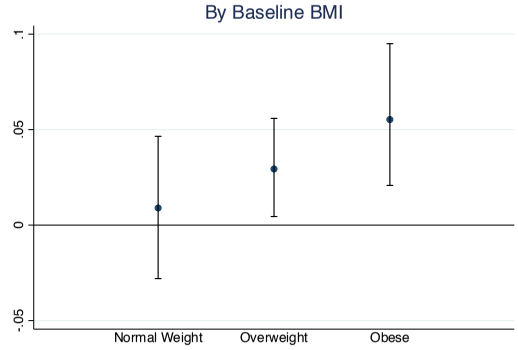
(b) By Baseline Diastolic Blood Pressure



(c) By Baseline Smoking Status



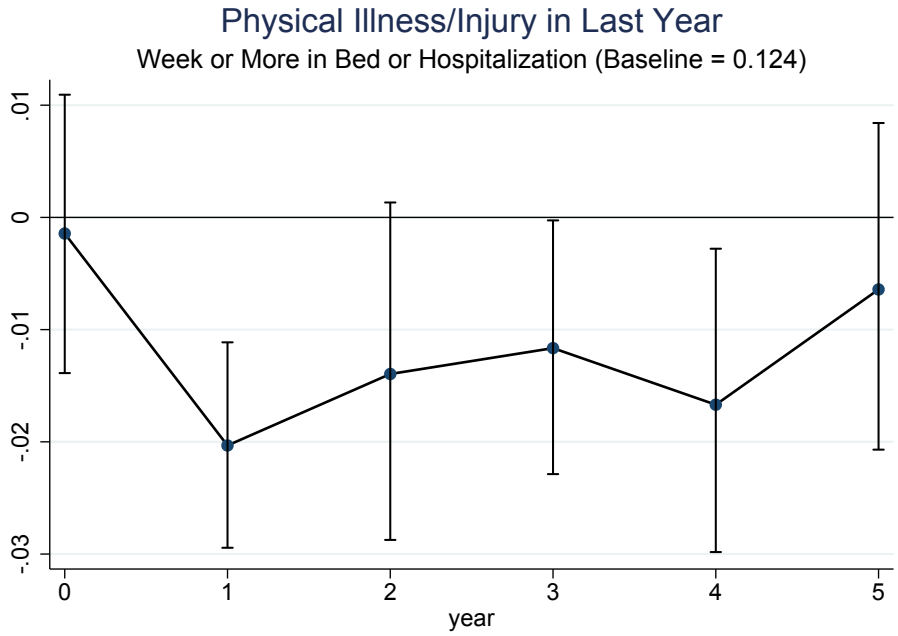
(d) By Baseline CHD Risk Quantile



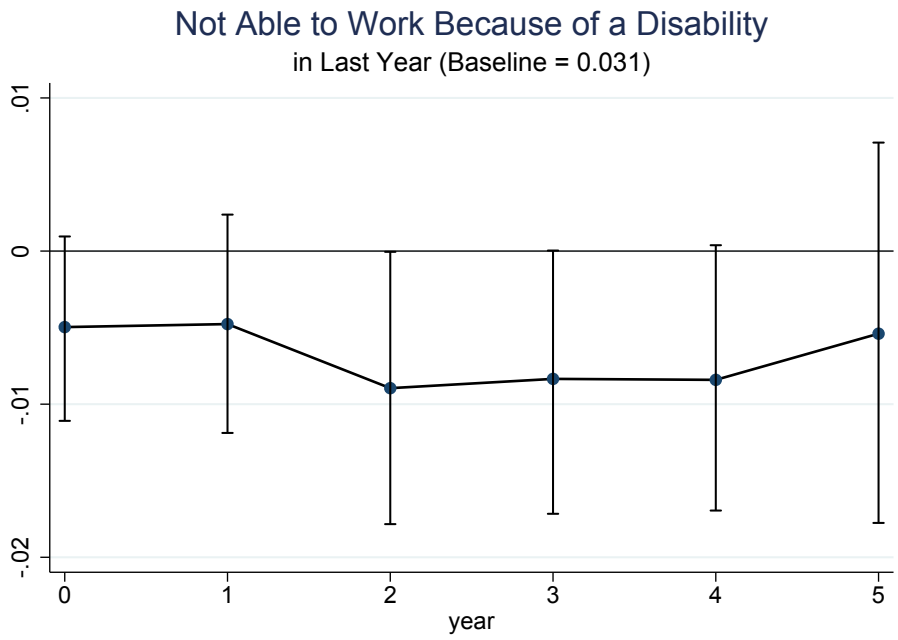
(e) By Baseline BMI

Notes: Each panel displays coefficients from a different regression of the form of equation (2), with the exception of Panel (c) in which the left two and right three coefficients are from different regressions. The displayed coefficients are SI-group indicators that are permitted to vary by the displayed baseline health groups. The group definitions for Panels (a) and (b) are defined in the panel subtitles. In Panel (c), light smokers are defined as smokers who self-report 19 or fewer cigarettes per day, while heavy smokers are those who self-report 40 or more cigarettes per day. In Panel (e), normal weight is defined as BMI < 25 and obese is defined as BMI ≥ 30. The 95 percent wild cluster bootstrap confidence interval bars contain the set of candidate null hypotheses for which the p-value of the test of that null hypothesis equals or exceeds 0.05.. The regressions include baseline health and demographic controls: linear serum cholesterol, linear diastolic blood pressure, number of cigarettes smoked, an indicator for being a smoker, a full set of indicators for age, an indicator for being white, indicators for four education groups, and a marital status indicator. The regressions further include a full set of indicators for baseline earnings categories. Because BMI does not appear in the general set of controls used throughout this paper, indicators for baseline BMI group are also included in the regression for Panel (e). The sample for all regressions is the 9,077 respondents who are not missing these measures and earnings at year six.

Figure 9: Experimental Impact on Health Outcomes in Last Year



(a) Physical Illness



(b) Disabled

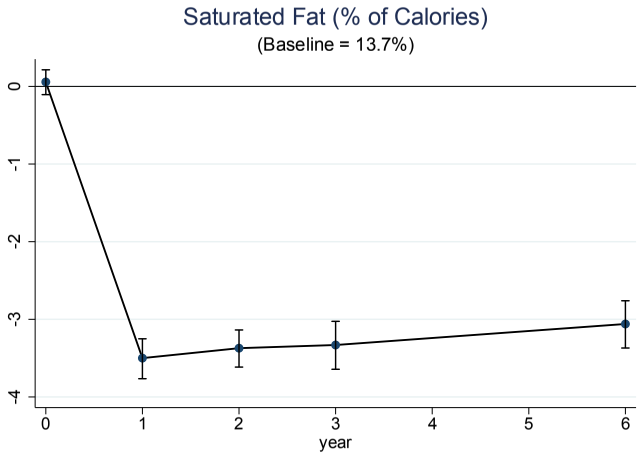
Notes: Each point is coefficient from a different regression of the form of equation (1). The 95% confidence interval bars are from using a wild cluster bootstrap which clusters at the clinic level. The regressions include baseline health and demographic controls: serum cholesterol, diastolic blood pressure, number of cigarettes smoked, an indicator for being a smoker, a full set of indicators for age, an indicator for being white, indicators for four education groups, and a marital status indicator. The sample is initially restricted to the 12,562 MRFIT respondents with nonmissing age, education, marital status, race, and employment status at baseline. Estimates for each year further restrict to observations with nonmissing outcomes and controls for that year.

Table A1: Predicted Percentage Change in Baseline Outcomes

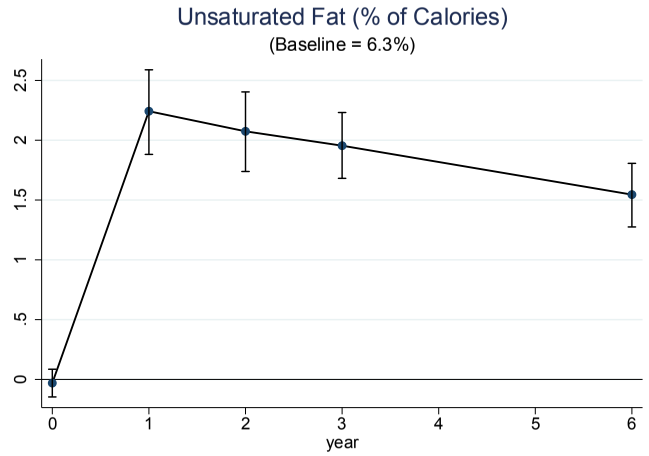
	SI	UC
Serum Cholesterol:		
≥220 mg/dl	10	0
<220 mg/dl	0	0
Diastolic Blood Pressure:		
≥95 mm HG	10	0
<95 mm HG	0	0
Cigarettes Smoked:		
1-19 Cigarettes/Day	55	15
20-39 Cigarettes/Day	40	10
40+ Cigarettes/Day	25	5

Notes: Sourced from Sherwin et al. (1981, Table 1). The Table presents the percentage changes in key CHD risk factors anticipated by MRFIT organizers as a function of baseline levels of the risk factors. The predicted serum cholesterol effects were informed by experimental results from the National Diet-Heart Study, the New York Anti-Coronary Club, and the Chicago Coronary Prevention Evaluation Program. The diastolic blood pressure predictions were informed by the Hypertension Detection and Follow-up Program. Anticipated effects of the anti-smoking intervention were less firm but were informed by prior studies suggesting that greater percentage reductions were possible among lighter smokers (Sherwin et al., 1981).

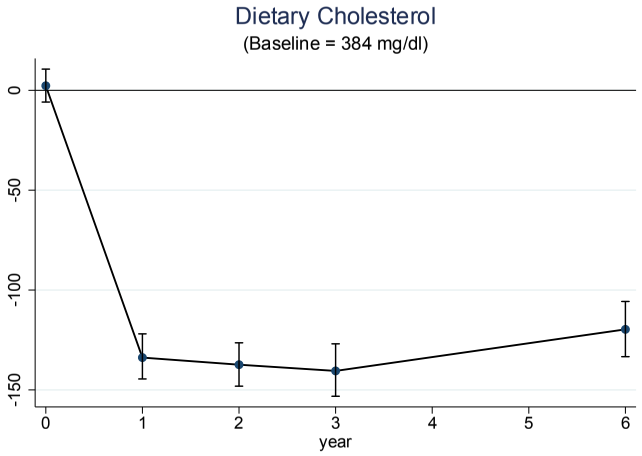
Figure A1: Experimental Impact on Cholesterol-Related Food Intake



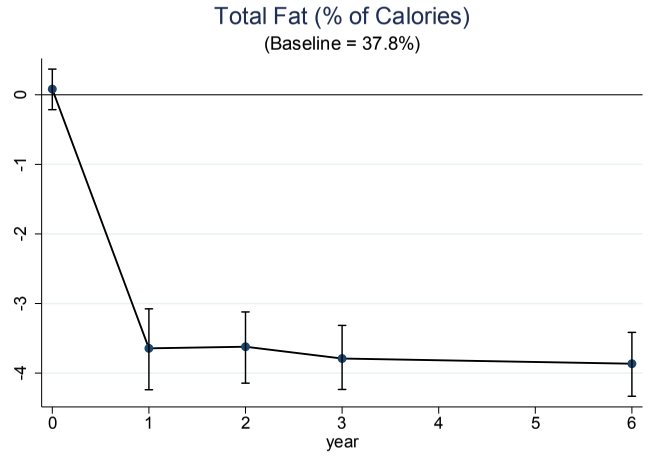
(a) Saturated Fat



(b) Unaturated Fat



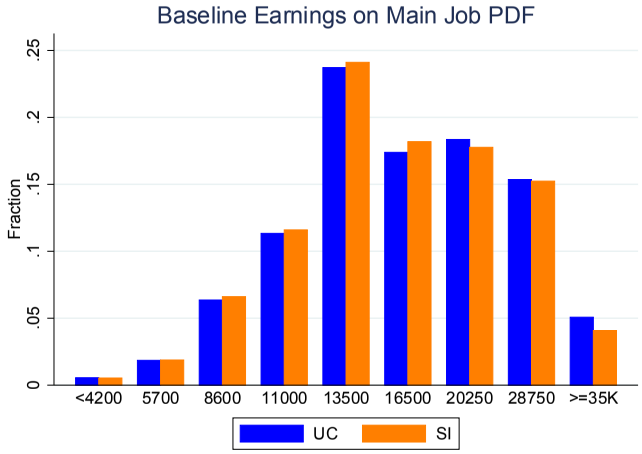
(c) Dietary Cholesterol



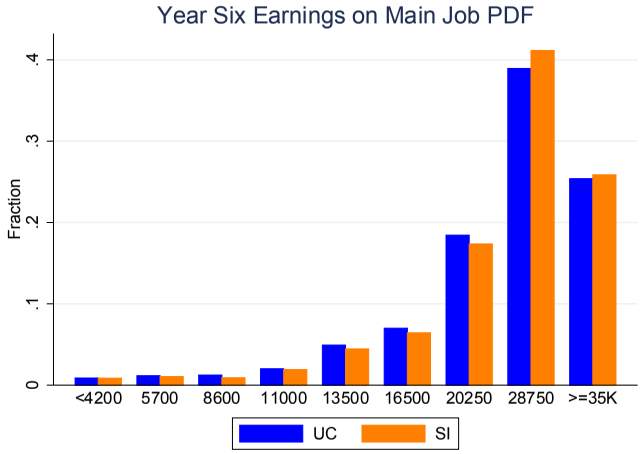
(d) Total Fat

Notes: Each point is coefficient from a different regression of the form of equation (1). The 95% confidence interval bars are from using a wild cluster bootstrap which clusters at the clinic level. The regression controls are baseline measures and include a full set of indicators for age, an indicator for being white, indicators for four education groups, and a marital status indicator. The sample is initially restricted to the 12,562 MRFIT respondents with nonmissing age, education, marital status, race, and employment status at baseline. Estimates for each year further restrict to observations with nonmissing outcomes and controls for that year.

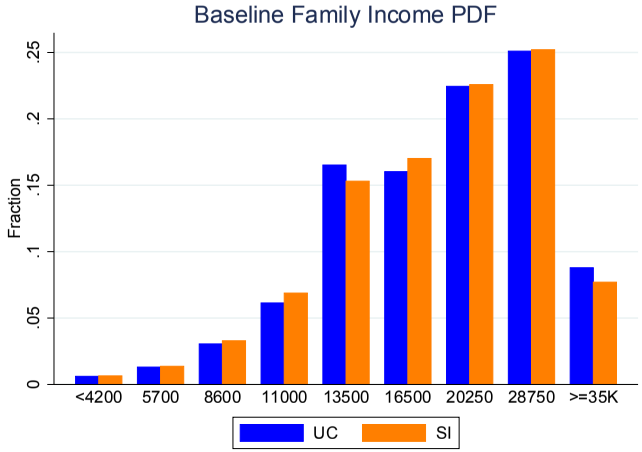
Figure A2: Distributions of Earnings and Family Income at Baseline and Year Six



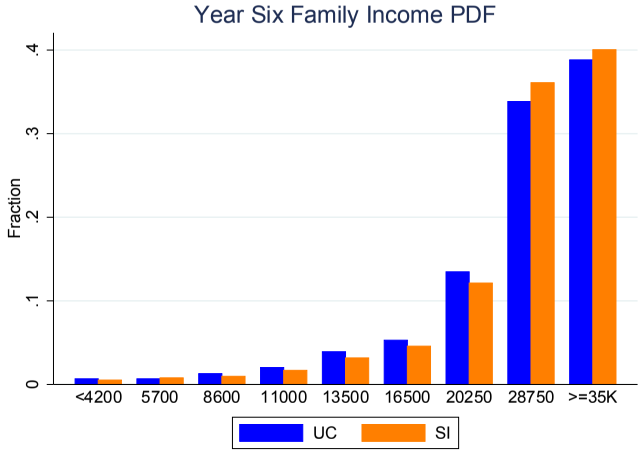
(a) Baseline Earnings PDF



(b) Year Six Earnings PDF



(c) Baseline Family Income PDF



(d) Year Six Family Income PDF

Notes: These figures present the discrete probability density functions, by experimental group, for the categorical earnings and family income measures at baseline and year six. All samples are restricted to respondents with nonmissing baseline demographic data. Panels (a) and (b) further restrict to individuals employed and with nonmissing earnings data at baseline and year six, respectively, resulting in samples of 12,322 and 9,316 observations. Panels (c) and (d) further restrict to individuals with nonmissing employment and income data at baseline and year six, respectively, resulting in samples of 12,396 and 10,670 observations.