

**The Impact of Health on Labor Market Outcomes:
Evidence from a Large-Scale Health Experiment**

Melvin Stephens Jr. and Desmond Toohey

Online Appendix

Table A.1: 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).

Table A.2: Earnings and Family Income Regressions

	Baseline		Year Six			Year Six Age \leq 48
	(1)	(2)	(3)	(4)	(5)	(6)
A. Earnings						
<i>SI</i>	-0.015	-0.010	0.020	0.027	0.028	0.023
[Heteroskedastic <i>p</i> -value]	[0.056]	[0.131]	[0.058]	[0.001]	[0.001]	[0.011]
{Clustered <i>p</i> -value}	{0.025}	{0.074}	{0.025}	{0.011}	{0.011}	{0.019}
N	12,326	12,321	9,508	9,215	9,212	5,982
B. Family Income						
<i>SI</i>	-0.013	-0.009	0.035	0.038	0.040	0.030
[Heteroskedastic <i>p</i> -value]	[0.099]	[0.181]	[0.003]	[0.000]	[0.000]	[0.007]
{Clustered <i>p</i> -value}	{0.111}	{0.183}	{0.005}	{0.003}	{0.004}	{0.040}
N	12,399	12,395	10,845	10,524	10,521	6,425
Additional Controls:						
Baseline health & demographics		X			X	X
Baseline outcome				X	X	X

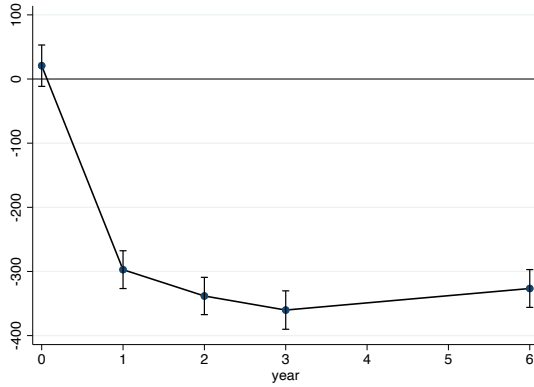
Notes: *p*-values generated using heteroskedasticity-consistent variances are shown in brackets. Kline-Santos wild cluster bootstrap *p*-values are reported in braces with clustering at the clinic level. Both sets of *p*-values test the null hypothesis that the given parameter is equal to 0. This table reports interval regression 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.

Table A.3: The Impact of CHD Risk on Earnings and Family Income

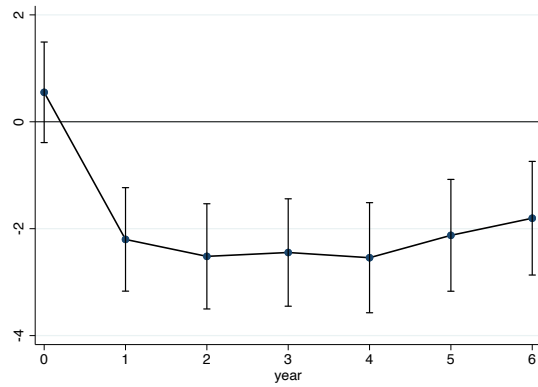
	Baseline	Year Six			
	(1)	(2)	(3)	(4)	(5)
A. Earnings					
<i>CHD Risk</i>	0.009 (0.002)	0.012 (0.005)	-0.058 (0.024)	0.005 (0.005)	-0.077 (0.021)
N	12,321	9,330	9,330	9,077	9,077
B. Family Income					
<i>CHD Risk</i>	0.013 (0.003)	0.003 (0.005)	-0.091 (0.025)	-0.010 (0.006)	-0.100 (0.023)
N	12,395	10,687	10,687	10,410	10,410
<i>CHD Risk</i> Endogenous:			X		X
Additional Controls:					
Baseline demographics	X	X	X	X	X
Baseline outcome				X	X

Notes: Heteroskedasticity-consistent standard errors are reported in parentheses. This table reports interval regression estimates of equation (2) where the binary treatment indicator *SI* is replaced with the measure of CHD risk, *CHD Risk*, and where the cutpoints are known and the unobserved latent outcome is assumed to be log normally distributed. The baseline demographic controls are 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)-(5) are a set of indicators for the corresponding outcome at baseline. The specifications in columns (3) and (5) treat *CHD Risk* as an endogenous regressor. That is, we first estimate (1) using *CHD Risk* as the dependent variable. We then take the residuals from this regression and insert them as control functions in equation (2) to account for the endogeneity of *CHD Risk* (e.g., Wooldridge 2015). The reported standard errors are bootstrapped when using this two-step procedure in columns (3) and (5). The earnings regressions are restricted to those who are employed for the relevant survey waves. 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.

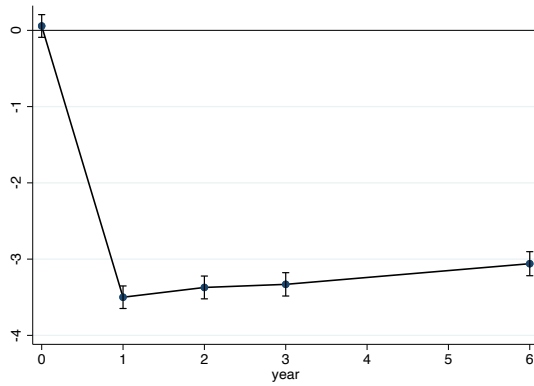
Figure A.1: Experimental Impact on Additional Cholesterol-Related Outcomes



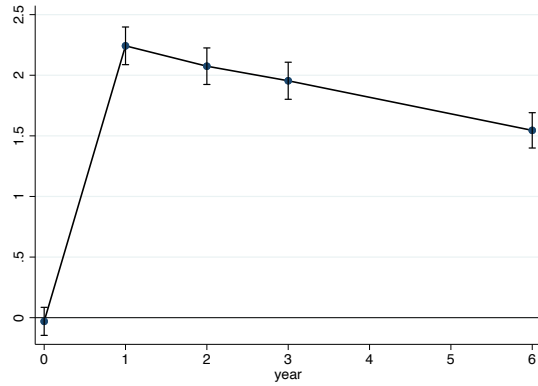
(a) Calories (Baseline=2,369)



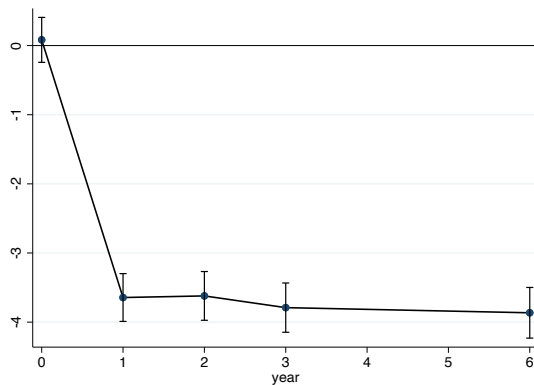
(b) Body Weight (Baseline=189 lbs)



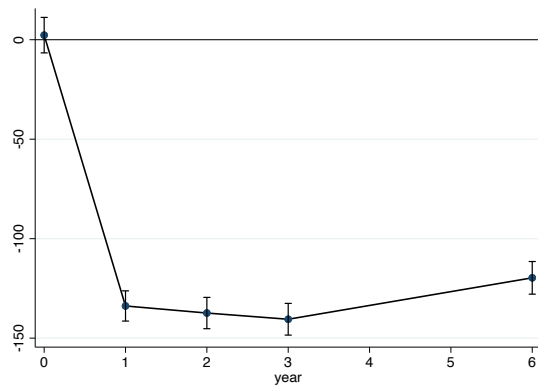
(c) Saturated Fat (Baseline=13.7%)



(d) Unaturated Fat (Baseline=6.3%)



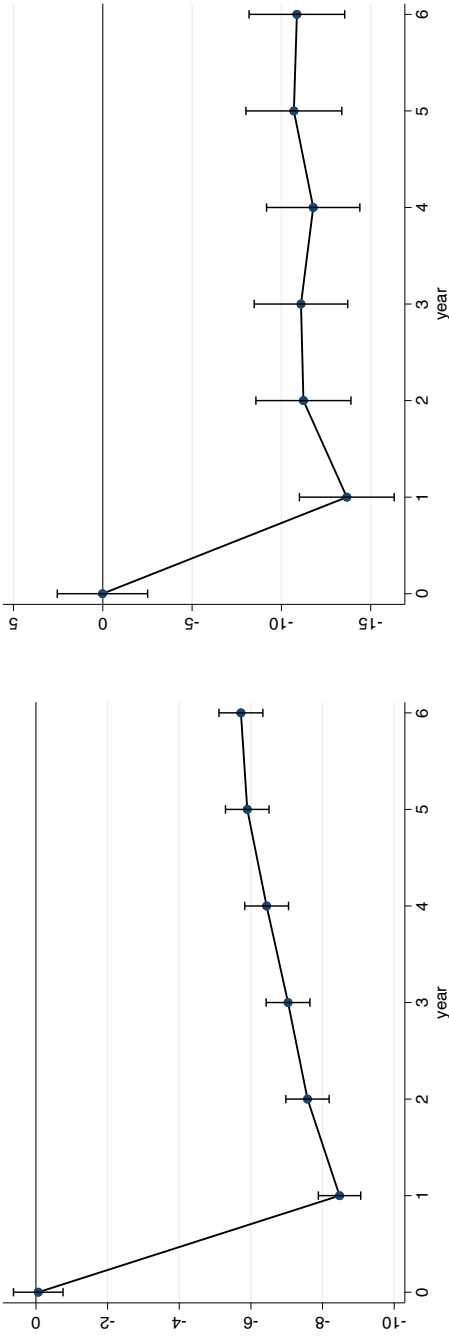
(e) Dietary Cholesterol (Baseline=384 mg/dl)



(f) Total Fat (Baseline=37.8%)

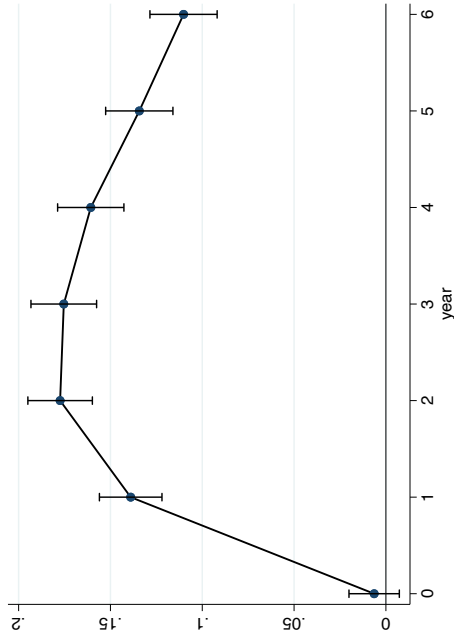
Notes: Each point is coefficient from a different regression of the form of equation (1). The 95% confidence interval bars are generated using heteroskedastic-consistent variances. The regression controls are baseline measures and include a full set of indicators for age, an indicator for being white, 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 A.2: Experimental Impact on Additional Health Outcomes



(a) Cigarettes/Day (Baseline=19)

(b) Serum Thiocyanate (Baseline=131)

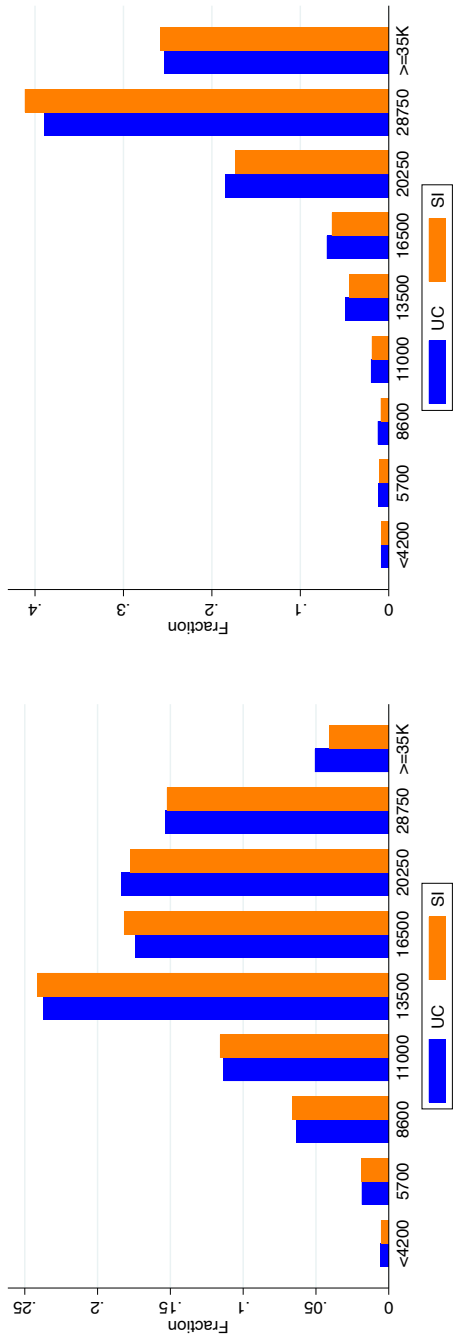


(c) Systolic Blood Pressure (Baseline=135 mm Hg)

(d) Taking Blood Pressure Medication (Baseline=19.3%)

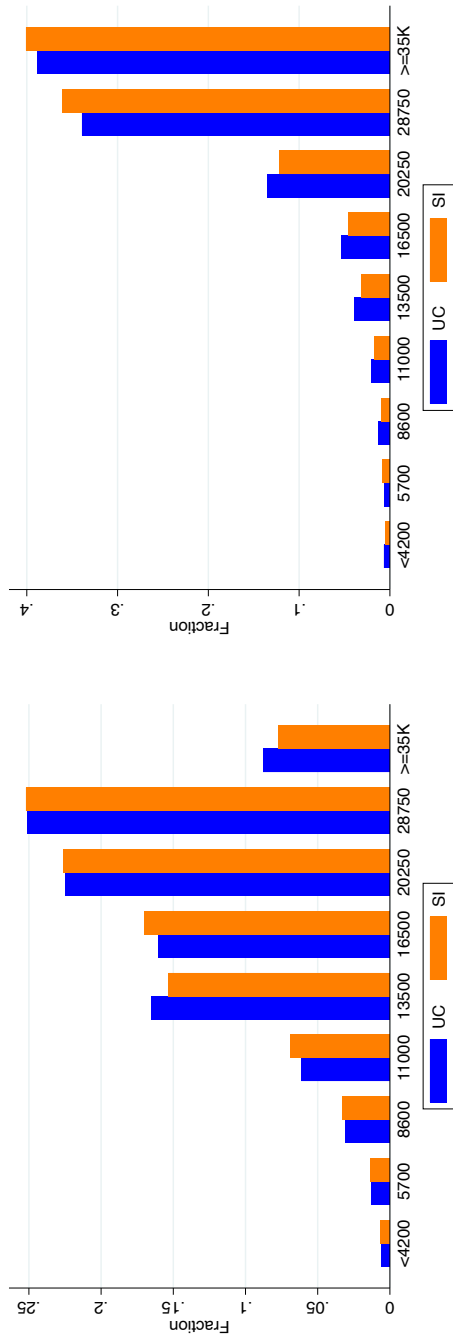
Notes: Each point is coefficient from a different regression of the form of equation (1). The 95% confidence interval bars are generated using heteroskedastic-consistent variances. The regression controls are baseline measures and include a full set of indicators for age, an indicator for being white, 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 A.3: 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.