Natalie Lewis-Vass, Jan Osea, Yiqi Wang, Zipeng Yang
Multiple Regression, 001-2
May 12, 2022

More Degrees, Less Disease? The Effects of Education and Income on Self-Reported Health for Individuals Over Age 50

Introduction

Longer life expectancies and fewer births have begun to rapidly transform United States demographics, resulting in a population surge of older Americans. By 2035, adults over the age of 65 are expected to outnumber those under the age of 18.[1] Government programs that support older adults, such as Social Security and Medicare, will need to respond to a surge in beneficiaries and increased healthcare costs. A better understanding of older adults and trends in aging can help to inform the critical policy required to respond to such a significant change in our country's composition. 88% of the respondents used for our analysis has at some point or continues to receive Social Security, and 74% are covered by a federal government health insurance program. Self-reported health status influences personal decisions related to retirement, health insurance, and finances, all of which have broad policy implications. Our study aims to take a closer look at the relationship between socioeconomic characteristics and self-reported health status. With a linear probability model, looking at the likelihood of self-identifying as healthy, we hope to shed light on non-health related factors associated with better health, such as education level, income, gender, and race.

Data

Our analysis assesses the impact of income and education on self-reported health for individuals over the age of 50 using longitudinal panel data from the RAND Health and

[1] Vespa J. The U.S. Joins Other Countries With Large Aging Populations. United States Census Bureau. 2018 https://www.census.gov/library/stories/2018/03/graying-america.html

Retirement Study (HRS). Table 1 presents a descriptive overview of the sample. The sample analyzed contains survey answers from 9,962 respondents over five years, including 2008, 2010, 2012, 2014 and 2016. The dataset contains survey responses from people over the age of 50 and their spouses, including 600 observations from spouses between the ages of 26 and 49. For the purposes of our analysis, we excluded observations for anyone under the age of 50, creating a range of 50 years old to 107, with a mean age of 70. 88 individuals were removed from our analysis entirely because they were under the age of 50 for the entirety of the survey. However, 120 from the total 49,370 observations were included for respondents who reached the age of 50 at some point after 2008. It is important to note that all respondents have lived past the age of 50, meaning they are likely healthier than those who passed away at an earlier age due to health issues. Our sample was also 61.3% women, possibly due to the longer life expectancies of women.[3] The original dataset contained a Hispanic indicator variable and a categorical race variable. Because we were looking closely at demographic factors, we combined these to create a more thorough description of race. The resulting racial makeup of our study was 73.5% White (non-Hispanic) and 13.8% Black (non-Hispanic), approximately reflecting the racial makeup of the United States.[3] People categorized as "Other" made up 2.6% of respondents and 10.1% identified as Hispanic.

The average BMI observed was 28.4, which the CDC classifies as overweight.[2] BMI values ranged from 8.9 to 92.8, which could be due to self-reporting error or miscommunication about the unit of measurement. We conducted a robustness analysis by dropping respondents with BMIs above the 99th percentile (BMI>47.2). Because we saw no change in our results, we opted to include all values for *BMI*. The most frequently observed health condition of our study

[2] Center for Disease Control. Healthy Weight, Nutrition, and Physical Activity. https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/english_bmi_calculator/bmi_calculator.html

sample was heart disease and/or heart problems at 26%, followed by 24% with diabetes, 16.5% with or having had cancer, and 7.2% that had experienced a stroke.

Model and Empirical Strategy

Our analysis assesses the impact of education and income on self-reported health, controlling for demographic factors and common health conditions. Our dependent variable, self-reported health status, is commonly used as an indicator of a person's overall well-being in terms of social, biological, and psychological health,[3, 4] and has been found to be an accurate predictor of mortality.[5, 6] The original HRS data set categorized self-reported health on a scale from 1 to 5, including 1.Excellent, 2.Very good, 3.Good, 4.Fair, and 5.Poor. To analyze the probability of identifying as *Healthy*, we divided the categories into two groups. To determine which statuses would be considered *Healthy,* we conducted a robustness analysis. Individuals who identified themselves as 1.Excellent and 2.Very Good were redefined as *Healthy* with a value of 1, and those identified as 3.Good. 4.Fair and 5.Poor were redefined as not *Healthy,* with a value of 0. Our sample consisted of 29,175 observations (59.09%) that were not *Healthy*, and 20,195 observations (40.91%) of *Healthy*.

---

[3] Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA 1995;273:59–65. https://pubmed.ncbi.nlm.nih.gov/7996652/

[4] Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. J Health Soc Behav. 1997 Mar;38(1):21-37. PMID: 9097506. https://pubmed.ncbi.nlm.nih.gov/9097506/

[5] Wuorela, M., Lavonius, S., Salminen, M. *et al.* Self-rated health and objective health status as predictors of all-cause mortality among older people: a prospective study with a 5-, 10-, and 27-year follow-up. *BMC Geriatr* 20, 120 (2020). https://doi.org/10.1186/s12877-020-01516-9

[6] S. Miilunpalo, I. Vuori, P. Oja, M. Pasanen, H. Urponen. *Self-rated health status as a health measure: The predictive value of self-reported health status on the use of physician services and on mortality in the working-age population.* Journal of Clinical Epidemiology, 50 (5) (1997) https://www.jclinepi.com/article/S0895-4356(97)00045-0/pdf

To assess the income of individuals in the survey, we wanted to capture the variation in earning sources for people over the age of 50, including the 32% of our sample working for pay, and the 88% receiving Social Security. To create our variable *Income* we combined household income, Social Security, Disability Insurance, Unemployment Insurance, and worker's compensation from the last calendar year. We found that *Income* ranged from $0 to $4.58 million, with a mean of $28,387. Higher values for *Income* could allow individuals to afford better medical care or help fund healthier lifestyle activities, leading us to hypothesize that on average, higher incomes will be associated with a higher probability of identifying as *Healthy*.

To analyze the relationship between education and self-reported health, we created a dummy variable from survey responses of the highest education level obtained. The survey asked respondents whether they attended some high school, received their GED, graduated high school, attended some college, or obtained a college degree or more. We defined our variable *College* to have a value of 1 if the respondent had attended some college or more, and a value of 0 for some high school, GED, or high school graduation. We decided not to include years of education as a continuous variable because exactly quantifying educational years will diffuse our intended variable and will result in an ambiguous takeaway. For example, examining the effect of an additional year of education could mean the individual repeated a grade, or it could mean they started law school. Instead, we thought it would be more fitting to use a dichotomous measure. Of 49,370 observations, 52.52% (25,928) were categorized as having never attended any college, and 47.48% (23,442) were categorized as attending at least some college. We predict that on average, having attended college will increase the likelihood of identifying as *Healthy*. A positive association could be attributed to knowledge gained in higher education that can increase health awareness and promote positive lifestyles choices. For our base model, we

estimate an OLS linear probability regression including our main explanatory variables (*College* and *Income*), controlling for the year of the survey.

In our second model we controlled for demographic factors such as age, gender, race (White, Hispanic, Black, Other), and marital status. We also controlled for BMI and included four dummy variables for common health conditions to capture whether respondents had diabetes, heart disease, cancer, or reported a stroke. Although incident rates for these health conditions are likely correlated with factors such as race, income, and education level, we wanted to analyze the impact of *College* and *Income* on *Healthy,* even when controlling for these common conditions. Including health conditions will likely unbias the coefficients for our independent variables of interest and give us a more accurate estimation of their effect on *Healthy*. We chose to exclude behavioral factors that could be influenced by socioeconomic and demographic factors such as smoking and exercising. Because of gender disparities in life expectancy, we hypothesize that, on average, identifying as a woman will increase the likelihood of reporting *Healthy*. Because of social support and companionship, we also hypothesize that marriage will be positively associated with reporting as *Healthy*. We believe the indicators for *Black*, *Hispanic*, and *Other* will decrease the probability of identifying as *Healthy* because of racial disparities in healthcare.

In our third model we included interactions between *College* and the variable representing income, and the indicators for female and Black, represented by *CollegexIncome*, *CollegexFemale*, and *CollegexBlac*k. We included these interaction terms because we believe that the predicted positive effect of having attended college will depend on an individual's income, gender, and race. We suspect that the positive effect of college will be intensified for those with higher values for *Income* and moderated for Black respondents because of pervasive

racism in medical care and racial disparities in healthcare access. Our analysis includes an interaction between *Female* and *College* to take a closer look at shifting college enrollment trends in the U.S. While women now are more likely to enroll in college, universities were primarily male-dominated institutions within the lifetime of all of the respondents included in our model.[7]

To achieve a more unbiased regression estimate, we introduced fixed effects in our fourth model. We first tried using individual fixed effects to control for time-invariant factors for each respondent. However, this dropped most of our socioeconomic and demographic variables of interest such as race and gender. Because our sample is made up of respondents over the age of 50, a demographic unlikely to return to school, *College* was also dropped. *Income* and *Married* had very little variance for individuals over time, resulting in statistical insignificance. To still control for any time-invariant omitted variables, we proceeded with a regional level fixed effect regression:

$$\text{Healthy}_{it} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\,\text{Income}_{it} + \boldsymbol{\beta}_2\,\text{College}_i + \boldsymbol{\beta}_3\,\text{Age}_{it} + \boldsymbol{\beta}_4\,\text{CollegexIncome}_{it} + \boldsymbol{\beta}_5\,\text{CollegexFemale}_i + \boldsymbol{\beta}_6\,\text{CollegexBlack}_i + \boldsymbol{\beta}_7\,\text{Female}_i + \boldsymbol{\beta}_8\,\text{Hispanic}_i + \boldsymbol{\beta}_9\,\text{Black}_i + \boldsymbol{\beta}_{10}\,\text{Other}_i + \boldsymbol{\beta}_{11}\text{Married}_{it} + \boldsymbol{\beta}_{12}\text{BMI}_{it} + \boldsymbol{\beta}_{13}\text{Diabetes}_{it} + \boldsymbol{\beta}_{14}\text{Heart}_{it} + \boldsymbol{\beta}_{15}\text{Cancer}_{it} + \boldsymbol{\beta}_{16}\text{Stroke}_{it} + \boldsymbol{\beta}_{17}\text{Year10}_t + \boldsymbol{\beta}_{18}\text{Year12}_t + \boldsymbol{\beta}_{19}\text{Year14}_t + \boldsymbol{\beta}_{20}\text{Year16}_t + \alpha_i + \varepsilon_{it}$$

where $Healthy_{it}$ is the probability of an individual i categorizing themselves as "healthy" in year t, and $Income_{it}$ and $College_i$ are our main explanatory variables, conditional on census region, controlling for demographic factors and health conditions.

[7] Parker K. What's behind the growing gap between men and women in college completion? 2021 https://www.pewresearch.org/fact-tank/2021/11/08/whats-behind-the-growing-gap-between-men-and-women-in-college-completion/#:~:text=Young%20women%20are%20more%20likely,adults%20ages%2025%20to%2034.

We looked at the possibility of using a non-linear model, such as logit or probit, to estimate the impact of *Income* and *College* on the probability of self-reporting as *Healthy*. However, we were unable to use xtreg for a fixed effects regression using either of these models. To achieve the most unbiased estimate and for continuity across our models, we moved forward with the linear probability model to interpret our results. However, this introduced underlying problems in our model, including the inaccuracy of the R^2 value, the possibility of *Healthy* being incorrectly valued at greater than 1 or less than 0, and heteroskedasticity. Additionally, because we used multiple year panel data, we assumed autocorrelation. We attempted to alleviate these issues by reducing omitted variable bias and using robust standard errors. After conducting a VIF test following our third regression resulting in a value of 2.2, we determined there was no need to address multicollinearity.

Results

The results of the OLS regression analyzing the probability of self-reporting as *Healthy* are presented in Table 2. For our base model specification in column 1, we observe statistically significant positive coefficients for both *College* and *Income*. Adding controls in Column 2, we observed statistically significant negative coefficients for *Female*, *Age*, *Black*, *Hispanic*, and *Other*, and a positive coefficient for the *Married* indicator. All of the health-related controls were estimated to have a statistically significant, negative association with the probability of self-reporting as *Healthy*. For our third model, we found the coefficients for all 3 of our interaction terms, *CollegexIncome, CollegexFemale* and *CollegexBlack* were negative, meaning the positive effects of having attended college were moderated by having a higher *Income*, being *Female* or being *Black*.

In column 4, we included the results of our individual fixed effects regression to display the dropped time-invariant variables. In our preferred model in column 5, we included the interaction terms and controlled for all time-invariant regional level factors, demographic characteristics, and common health conditions. We found that these estimates confirmed our hypotheses for our independent variables of interest. The discrete effects of having attended some college (*College*) and having a higher *Income* are both positively associated with *Healthy* and statistically significant. However, the coefficient for the interaction term *CollegexIncome* is negative, indicating that the positive impact of college is weakened by a higher income. At an income of $374,285, for a non-Black man, the positive effect of *College* on the probability of reporting *Healthy* is completely nullified, and becomes negative with higher incomes or for people who are *Female* or *Black*, on average. It is unclear to us why *Income* would minimize the positive effect of *College*, but one hypothesis is reckless and unhealthy behavior that could be associated with more disposable income. As we theorized, the positive effect of *College* on *Healthy* was reduced when interacted with both *Black* and *Female*. On average, non-Black men with no income experienced the largest positive effect from having attended college (13.1 percentage points), whereas a Black woman with no income was only 8.96 percentage points more likely to identify as *Healthy* when having attended college. We have presented a more detailed interpretation of the interaction terms in Table 3 of the Appendix, looking at the changes in the effect of *College* on *Healthy* at 3 different values for *Income*.

Our results indicated that, contrary to our hypothesis, identifying as *Female* decreased an individual's probability of being *Healthy.* However, the coefficients for *Female* and *CollegexFemale* in our fifth model were not statistically significant. As expected, because of their well-known detriment to well-being, all our included health factors (*BMI, Diabetes, Heart,*

Cancer, Stroke) were statistically significant at the 1% level and negatively associated with *Healthy*. We also found that, on average, married individuals were 4.4 percentage points more likely to self-report as *Healthy*, all else being equal. We suspect individuals receive care and companionship from their spouses, resulting in better health outcomes. All the race indicators we included had a statistically significant, negative impact on *Healthy*, likely due to racial disparities in health care and access. On average, White people (our omitted variable) were more likely to identify as healthy, all else being equal.

Conclusion:

After analyzing our results, we have found that when controlling for health conditions and socioeconomic characteristics, higher incomes and college attendance increases one's probability of reporting better health outcomes. This suggests that increasing the accessibility and affordability of higher education and increasing monetary benefits for older Americans could result in higher rates of individuals reporting *Healthy*. While most seniors are unlikely to return to college, expanding programs like Social Security could be a practical and effective method for improving health in older populations. However, further research is needed to better understand why the positive influence of *College* is the moderated by higher values of *Income*.

We suspect that the positive effect of *Married* on the probability of reporting as *Healthy* is due to the associated companionship, intimacy and expanded social network of partnership.[8] This could have implications for senior living accommodations and retirement age. Individuals who are married could also benefit from a greater household income through spousal earnings and tax deductions.[10] Additional analyses should investigate the interaction between *Married* and

[8] Amato PR. Marriage, cohabitation and mental health. Family Matters 96 (2015) 5-13

Income on self-reported health results, as well as if non-marital unions and partnerships have similar effects.

Because our variable for *College* included both college graduates and those who only attended some college, further research could be done to look at what aspects of higher education influence one's health. One possible avenue for exploration is the relationships, habits and identities individuals form during this time and their possible influence on healthier choices. Better understanding the positive impact of higher education could help policymakers recreate these conditions elsewhere.

Appendix

TABLE 1: *Characteristics of Study Sample*

VARIABLES	(1) mean	(2) sd	(3) min	(4) max
Healthy	0.41	0.49	0.00	1.00
Income	28,386.52	84,559.01	0.00	4,058,000.00
College	0.47	0.50	0.00	1.00
Female	0.61	0.49	0.00	1.00
Age	70.45	9.25	50.00	107.00
White	0.73	0.44	0.00	1.00
Black	0.14	0.35	0.00	1.00
Hispanic	0.10	0.30	0.00	1.00
Other	0.03	0.16	0.00	1.00
Married	0.60	0.49	0.00	1.00
BMI	28.38	5.83	8.90	92.80
Diabetes	0.24	0.43	0.00	1.00
Heart	0.26	0.44	0.00	1.00
Cancer	0.16	0.37	0.00	1.00
Stroke	0.07	0.26	0.00	1.00
Year08	0.20	0.40	0.00	1.00
Year10	0.20	0.40	0.00	1.00
Year12	0.20	0.40	0.00	1.00
Year14	0.20	0.40	0.00	1.00
Year16	0.20	0.40	0.00	1.00
Number of Region	5	5	5	5

N=49370

TABLE 2: *Effects of College Attendance and Income on Probability of Self-Reporting as Healthy*

VARIABLES	(1) **Model 1**: Base	(2) **Model 2**: Controls	(3) **Model 3**: Interactions	(4) **Model 4:** Individual Fixed Effects	(5) **Model 5:** Regional Fixed Effects
Income	5.224e-07***	3.26e-07***	6.35e-07***	-1.66e-08	6.27e-07***
	(5.336e-08)	(3.90e-08)	(1.69e-07)	(6.08e-08)	(1.07e-07)
College	0.170***	0.116***	0.131***		0.131***
	(0.005)	(0.004)	(0.008)		(0.013)
Female		-0.009**	-0.007		-0.007
		(0.004)	(0.006)		(0.006)
CollegexIncome			-3.60e-07**	4.60e-08	-3.54e-07**
			(1.72e-07)	(6.90e-08)	(1.20e-07)
CollegexFemale			-0.004		-0.005
			(0.009)		(0.011)
CollegexBlack			-0.035***		-0.037*
			(0.012)		(0.016)
Black		-0.122***	-0.107***		-0.103***
		(0.006)	(0.007)		(0.011)
Hispanic		-0.190***	-0.187***		-0.184***
		(0.007)	(0.007)		(0.007)
Other		-0.122***	-0.122***		-0.121***
		(0.013)	(0.013)		(0.018)
Age		-7.82e-04***	-7.05e-04***	-2.61e-04	-7.04e-04
		(2.56e-04)	(2.63e-04)	(0.004)	(5.05e-04)
Married		0.044***	0.043***	-0.006	0.044***
		(0.004)	(0.004)	(0.010)	(0.008)
BMI		-0.009***	-0.009***	-0.004***	-0.009***
		(3.50e-04)	(3.50e-04)	(0.001)	(0.001)
Diabetes		-0.141***	-0.140***	-0.039***	-0.140***
		(0.005)	(0.005)	(0.010)	(0.012)
Heart		-0.181***	-0.181***	-0.094***	-0.181***
		(0.005)	(0.005)	(0.009)	(0.006)
Cancer		-0.080***	-0.080***	-0.097***	-0.080***
		(0.006)	(0.006)	(0.013)	(0.010)
Stroke		-0.130***	-0.130***	-0.089***	-0.129***
		(0.007)	(0.007)	(0.014)	(0.012)
Year10	0.012*	0.027***	0.027***	0.016	0.027**
	(0.007)	(0.007)	(0.007)	(0.012)	(0.006)
Year12	-0.013*	0.012*	0.013*	-0.004	0.012**
	(0.007)	(0.007)	(0.007)	(0.021)	(0.004)

Year14	-0.054***	-0.017***	-0.017**	-0.041	-0.017*
	(0.007)	(0.007)	(0.007)	(0.029)	(0.006)
Year16	-0.083***	-0.035***	-0.034***	-0.065	-0.034***
	(0.007)	(0.007)	(0.007)	(0.039)	(0.004)
Constant	0.341***	0.768***	0.753***	0.610*	0.753***
	(0.005)	(0.023)	(0.024)	(0.319)	(0.064)
Observations	49,370	49,370	49,370	49,370	49,370
R-squared	0.049	0.149	0.150	0.023	0.147
Year FE	YES	YES	YES	YES	YES
Individual FE	NO	NO	NO	YES	NO
Regional FE	NO	NO	NO	NO	YES
Number of Region					5
Number of Respondents				9,962	
F-test: College CollegexIncome CollegexFemale CollegexBlack			179.9		585.9
Prob >F			<0.01		<0.01

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

TABLE 3: *Interpretation of Interactions between College and Income, Female and Black and their effect on Healthy*

		Black	Non-Black
No income	Male	9.45 percentage points	13.10 percentage points
	Female	8.96 percentage points	12.61 percentage points
Average income*	Male	8.46 percentage points	12.11 percentage points

	Female	7.96 percentage points	11.61 percentage points
$400,000 income	Male	-4.55 percentage points	-0.9 percentage points
	Female	-5.04 percentage points	-1.39 percentage points

*Average income for our sample: $28,386.52

**All calculations based on the coefficients from Model 5 of Table 2

Interpretation: On average, for a Black/Non-Black woman/man with an income of 0/$28,386.52/$400,000, having attended at least some college is associated with a _____ percentage point increase/decrease in the likelihood of self-reporting as healthy, all else being equal.

```
      name:  <unnamed>
       log:  /Users/nugget/Desktop/NYU/Econometrics/Project/Health Project.smcl
  log type:  smcl
 opened on:  12 May 2022, 13:29:28

 1 . use "$datadir/hrsdata_mr.dta", clear

 2 .
 3 . /*Variable name changes*/
 4 . rename agey Age

 5 . rename cenreg Region

 6 . rename hhidpn Respondents

 7 .
 8 . /*excluding those under the age of 50*/
 9 . drop if Age<50
   (600 observations deleted)

10 .
11 . /*Dichotomous self reported health for dependent variable*/
12 .
13 . generate Healthy = 1

14 . replace Healthy = 1 if shlt==1
   (0 real changes made)

15 . replace Healthy = 1 if shlt==2
   (0 real changes made)

16 . replace Healthy = 0 if shlt==3
   (16,710 real changes made)

17 . replace Healthy = 0 if shlt==4
   (9,533 real changes made)

18 . replace Healthy = 0 if shlt==5
   (2,932 real changes made)

19 . label define Healthy 0 "0.not healthy" 1 "1.healthy", modify
```

```
20 . label values Healthy Healthy

21 .
22 . /*creating race variable with hispanic*/
23 .
24 . replace race = 4 if hispan==1
   (5,001 real changes made)

25 . label define race  1 "1.white non-hispanic" 2 "2.black non-hispanic" 3 "3.other" 4 "4.hispanic", modify

26 . label values race race

27 .
28 . gen White = (race==1)

29 . gen Black = (race==2)

30 . gen Other = (race==3)

31 . gen Hispanic = (race==4)

32 .
33 . /*Year dummies*/
34 . gen Year08 = (year==2008)

35 . gen Year10 = (year==2010)

36 . gen Year12 = (year==2012)

37 . gen Year14 = (year==2014)

38 . gen Year16 = (year==2016)

39 .
40 . /*gender*/
41 . gen Female = (gender==2)

42 .
43 . /*replace missing value with mean*/
44 . gen BMI_im=bmi
   (583 missing values generated)
```

```
45 . egen mean_bmi= mean(bmi)

46 . replace BMI_im= mean_bmi if missing(BMI_im)
   (583 real changes made)

47 . gen bmi_mis = (bmi==.)

48 .
49 . /*marital status*/
50 . replace mstat = 0 if mstat>1
   (19,993 real changes made)

51 . label define mstat 0 "0.not married" 1 "1.married", modify

52 . label values mstat mstat

53 . gen Married = (mstat==1)

54 .
55 . /*education level*/
56 . gen College = 1

57 . replace College = 0 if educ<4
   (25,928 real changes made)

58 . replace College = 0 if missing(educ)
   (15 real changes made)

59 . replace College = 1 if educ>3
   (15 real changes made)

60 . label define College 0 "no college" 1 "Atleast some college"

61 . label values College College

62 .
63 . /*Income*/
64 . gen Income = iearn + issdi + iunwc + icap

65 .
66 . /*Interactions*/
67 . gen CollegexIncome = Income*College
```

```
68 . gen CollegexFemale= Female*College

69 . gen CollegexBlack= Black*College

70 .
71 . /*Health, replacing with 0 if no data*/
72 . gen Diabetes = diab
   (35 missing values generated)

73 . replace Diabetes = 0 if missing(diab)
   (35 real changes made)

74 . label define Diabetes 0 "0.no" 1 "1.yes", modify

75 . label values Diabetes Diabetes

76 .
77 . gen Heart = heart
   (44 missing values generated)

78 . replace Heart = 0 if missing(heart)
   (44 real changes made)

79 . label define Heart 0 "0.no" 1 "1.yes", modify

80 . label values Heart Heart

81 .
82 . gen Cancer = cancr
   (43 missing values generated)

83 . replace Cancer = 0 if missing(cancr)
   (43 real changes made)

84 . label define Cancer 0 "0.no" 1 "1.yes", modify

85 . label values Cancer Cancer

86 .
87 . gen Stroke = strok
   (29 missing values generated)
```

```
88 . replace Stroke = 1 if strok>0
   (287 real changes made)

89 . replace Stroke = 0 if missing(strok)
   (29 real changes made)

90 . label define Stroke 0 "0.no" 1 "1.yes", modify

91 . label values Stroke Stroke

92 .
93 . /*Replacing missing regional observations*/
94 . replace Region = 5 if missing(Region)
   (33 real changes made)

95 .
96 .
97 . /*test regression and tests before outreg*/
98 .
99 . reg Healthy Income College Age CollegexIncome CollegexFemale CollegexBlack Year10 Year12 Year14 Year16 His
   > panic Black Other Female Married BMI Diabetes Heart Cancer Stroke, robust

   Linear regression                               Number of obs     =     49,370
                                                   F(20, 49349)      =     555.89
                                                   Prob > F          =     0.0000
                                                   R-squared         =     0.1498
                                                   Root MSE          =     .45343

------------------------------------------------------------------------------
             |               Robust
     Healthy |  Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      Income |   6.35e-07   1.69e-07     3.75   0.000     3.04e-07    9.67e-07
     College |    .1306039   .0078105    16.72   0.000     .1152952    .1459126
         Age |   -.0007054   .0002627    -2.68   0.007    -.0012203   -.0001904
CollegexIn~e |   -3.60e-07   1.72e-07    -2.09   0.036    -6.96e-07   -2.31e-08
CollegexFe~e |   -.0042901   .0085226    -0.50   0.615    -.0209945    .0124143
CollegexBl~k |   -.0349218   .0120196    -2.91   0.004    -.0584804   -.0113633
      Year10 |    .0272629   .0065461     4.16   0.000     .0144324    .0400934
      Year12 |    .0124596   .0065885     1.89   0.059     -.000454    .0253733
      Year14 |   -.0169176   .0066165    -2.56   0.011     -.029886   -.0039492
      Year16 |   -.0342544   .0067164    -5.10   0.000    -.0474186   -.0210902
    Hispanic |   -.1869544     .00659   -28.37   0.000    -.1998709   -.1740379
       Black |   -.1072543   .0074223   -14.45   0.000    -.1218021   -.0927065
       Other |   -.1221269   .0126954    -9.62   0.000      -.14701   -.0972438
      Female |   -.0065622    .005974    -1.10   0.272    -.0182714    .0051469
     Married |    .0433455    .004497     9.64   0.000     .0345314    .0521596
      BMI_im |   -.0085467   .0003502   -24.41   0.000    -.0092331   -.0078604
    Diabetes |   -.1403687   .0047951   -29.27   0.000    -.1497672   -.1309702
       Heart |   -.1810574   .0046564   -38.88   0.000     -.190184   -.1719309
      Cancer |   -.0803017   .0055259   -14.53   0.000    -.0911326   -.0694709
      Stroke |   -.1298911   .0071603   -18.14   0.000    -.1439253   -.1158569
       _cons |    .7534631   .0244432    30.83   0.000     .7055541    .8013721
------------------------------------------------------------------------------
```

```
100 .
101 . estat vif
```

Variable	VIF	1/VIF
CollegexIn~e	**7.81**	**0.128116**
Income	**7.56**	**0.132342**
CollegexFe~e	**3.42**	**0.292533**
College	**2.99**	**0.334230**
Female	**2.05**	**0.487584**
CollegexBl~k	**1.78**	**0.562954**
Black	**1.77**	**0.564606**
Year16	**1.75**	**0.570890**
Year14	**1.69**	**0.593099**
Year12	**1.65**	**0.607537**
Year10	**1.62**	**0.616795**
Age	**1.31**	**0.765961**
Married	**1.15**	**0.866014**
BMI_im	**1.12**	**0.893301**
Diabetes	**1.11**	**0.902647**
Hispanic	**1.09**	**0.913545**
Heart	**1.09**	**0.921409**
Stroke	**1.04**	**0.961458**
Cancer	**1.03**	**0.972212**
Other	**1.02**	**0.982280**
Mean VIF	**2.20**	

```
102 .
103 . test College CollegexIncome CollegexFemale CollegexBlack

     ( 1)  College = 0
     ( 2)  CollegexIncome = 0
     ( 3)  CollegexFemale = 0
     ( 4)  CollegexBlack = 0

           F(  4, 49349) =  179.91
                Prob > F =    0.0000

104 .
105 .
```

```
106 . /*Table output*/
107 .
108 . ssc install outreg2
    checking outreg2 consistency and verifying not already installed...
    all files already exist and are up to date.

109 .
110 . outreg2 using HealthProject.doc, replace sum(log) keep(Healthy Income College Age Year08 Year10 Year12 Yea
    > r14 Year16 Hispanic Black Other White Female Married BMI Diabetes Heart Cancer Stroke) eqkeep (mean sd min
    >  max) sortvar(Healthy Income College Female Age White Black Hispanic Other Married BMI Diabetes Heart Canc
    > er Stroke) title(Table 1: Characteristics of Study Population) dec (2) addnote(N=49370)label
```

Variable	Obs	Mean	Std. dev.	Min	Max
Respondents	49,370	1.77e+08	1.78e+08	1.00e+07	5.03e+08
year	49,370	2012.022	2.825994	2008	2016
hispan	49,365	.1013066	.3017373	0	1
race	49,370	1.493842	.9536083	1	4
byear	49,370	1941.17	8.838528	1909	1966
bmonth	49,370	6.546709	3.434572	1	12
gender	49,370	1.612943	.4870818	1	2
edyrs	49,291	12.78158	3.173436	0	17
edegrm	49,370	2.740267	1.919887	0	8
educ	49,355	3.320616	1.355877	1	5
Age	49,370	70.45161	9.245036	50	107
meduc	45,243	9.895011	3.690986	0	17
feduc	42,604	9.580145	3.995823	0	17
mstat	49,370	.5950375	.4908898	0	1
Region	49,370	2.663399	.9564264	1	5
relig	49,187	1.570781	.9221554	1	5
vetrn	49,320	.1888483	.3913921	0	1
bplace	49,330	5.275329	2.925444	1	11
cohbyr	49,370	3.679846	1.188486	0	7
shlt	49,339	2.812907	1.037326	1	5
depres	47,532	.1118825	.3152252	0	1
effort	47,501	.2051325	.4038027	0	1
sleepr	47,509	.2891873	.4533898	0	1
whappy	47,443	.8778534	.3274585	0	1
flone	47,513	.1475175	.3546248	0	1
enlife	47,482	.9224759	.2674241	0	1
fsad	47,505	.167014	.3729923	0	1
going	47,428	.1871679	.3900504	0	1
cesd	47,554	1.305821	1.88777	0	8
bmi	48,787	28.37991	5.866131	8.9	92.8
height	49,324	1.6754	.1036046	.9144	2.3241
weight	48,822	79.92146	18.81644	22.6795	181.436
smokev	49,001	.5444583	.4980246	0	1
smoken	48,992	.1011594	.3015427	0	1

hibp	49,299	.6332988	.4819088	0	1
hibp_e	49,266	.6190273	.4856308	0	1
diab	49,335	.2421607	.4283954	0	1
diab_e	49,288	.236305	.424816	0	1
cancr	49,327	.1648387	.3710387	0	1
cancr_e	49,208	.1650748	.3712518	0	1
lung	49,329	.1017657	.302343	0	1
lung_e	49,290	.0946439	.2927255	0	1
heart	49,326	.2620727	.4397665	0	1
heart_e	49,270	.2550842	.435913	0	1
strok	49,341	.0777649	.2866654	0	2
strok_e	49,303	.085938	.2802753	0	1
psych	49,323	.185674	.3888473	0	1
psych_e	49,261	.1793102	.3836157	0	1
arthr	49,314	.6508497	.4767064	0	1
arthr_e	49,311	.6410132	.4797082	0	1
hosp	49,214	.2553745	.4360759	0	1
doctor	49,278	.9286497	.2574117	0	1
oopmd	49,370	3371.675	9805.034	0	634821
totn	49,370	364050.7	1004064	-1510000	3.03e+07
totb	49,370	542115.3	1208034	-1512500	3.71e+07
iearn	49,370	12287.78	36190	0	1425000
issdi	49,370	456.8931	2434.207	0	74400
iunwc	49,370	113.7951	1147.386	0	36000
icap	49,370	15528.06	73240.36	0	4052000
ssrecv	49,370	.8775775	.3277765	0	1
ssdi	49,370	.0446628	.2065644	0	1
higov	49,143	.742181	.4374383	0	1
prpcnt	48,964	.5841639	.6150852	0	21
lifein	48,793	.5980366	.4902997	0	1
sayret	48,599	1.004403	.7463791	0	3
work	49,314	.31997	.4664693	0	1
lbrf	49,370	4.131112	1.676773	1	7
jjobs	49,370	2.152785	1.294527	0	9
jyears	49,370	35.45345	15.15815	0	74
inlbrf	48,594	.3403712	.4738389	0	1
hhres	49,370	2.140065	1.131641	1	15
hchild	48,316	3.234477	2.116324	0	20
livsib	49,351	2.831635	2.45302	0	20
evbrn	49,335	2.719428	1.835352	0	11
peninc	48,659	.2930804	.4551797	0	1
Healthy	49,370	.4090541	.4916643	0	1
White	49,370	.7345554	.4415741	0	1
Black	49,370	.1383431	.3452633	0	1
Other	49,370	.0258051	.1585552	0	1

Hispanic	49,370	.1012963	.3017238	0	1
Year08	49,370	.197225	.3979077	0	1
Year10	49,370	.1994329	.3995781	0	1
Year12	49,370	.2003848	.4002924	0	1
Year14	49,370	.2011748	.4008825	0	1
Year16	49,370	.2017825	.4013347	0	1
Female	49,370	.6129431	.4870818	0	1
BMI_im	49,370	28.37991	5.831391	8.9	92.8
mean_bmi	49,370	28.37992	0	28.37992	28.37992
bmi_mis	49,370	.0118088	.1080258	0	1
Married	49,370	.5950375	.4908898	0	1
College	49,370	.4748228	.4993708	0	1
Income	49,370	28386.52	84559.01	0	4058000
CollegexIn~e	49,370	21236.04	80197.89	0	4058000
CollegexFe~e	49,370	.2762001	.4471215	0	1
CollegexBl~k	49,370	.0523395	.222713	0	1
Diabetes	49,370	.2419891	.428292	0	1
Heart	49,370	.2618392	.43964	0	1
Cancer	49,370	.1646952	.370909	0	1
Stroke	49,370	.0724934	.2593058	0	1

HealthProject.doc
dir : seeout

```
111 .
112 . reg Healthy Income College Year10 Year12 Year14 Year16, robust
```

Linear regression

Number of obs	=	49,370
F(6, 49363)	=	374.87
Prob > F	=	0.0000
R-squared	=	0.0490
Root MSE	=	.47949

Healthy	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
Income	5.22e-07	5.34e-08	9.79	0.000	4.18e-07	6.27e-07
College	.1696056	.004606	36.82	0.000	.1605778	.1786334
Year10	.0121808	.0069434	1.75	0.079	-.0014283	.0257899
Year12	-.0129311	.0069333	-1.87	0.062	-.0265203	.0006582
Year14	-.0537807	.0068826	-7.81	0.000	-.0672708	-.0402907
Year16	-.0825498	.0068219	-12.10	0.000	-.0959207	-.0691788
_cons	.341332	.005411	63.08	0.000	.3307263	.3519377

```
113 . outreg2 using Table2.doc, replace ctitle (Model 1) addtext(Year FE, YES, Individual FE, NO, Regional FE, N
    > O) sortvar(Income College) label
    Table2.doc
    dir : seeout

114 .
115 . reg Healthy Income College Female Age Year10 Year12 Year14 Year16 Hispanic Black Other Married BMI Diabete
    > s Heart Cancer Stroke, robust

Linear regression                               Number of obs     =     49,370
                                                F(17, 49352)      =     652.91
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1493
                                                Root MSE          =     .45357

------------------------------------------------------------------------------
             |               Robust
     Healthy |  Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      Income |   3.26e-07   3.90e-08     8.37   0.000     2.50e-07    4.03e-07
     College |   .1162555   .0044111    26.35   0.000     .1076096    .1249014
      Female |  -.009421   .0043787    -2.15   0.031    -.0180034   -.0008386
         Age |  -.000782   .0002559    -3.06   0.002    -.0012836   -.0002804
      Year10 |   .0268275   .0065501     4.10   0.000     .0139891    .0396658
      Year12 |   .0119539   .0065939     1.81   0.070    -.0009703     .024878
      Year14 |  -.0174698   .0066205    -2.64   0.008     -.030446   -.0044936
      Year16 |  -.0345928    .006718    -5.15   0.000      -.04776   -.0214255
    Hispanic |  -.1899253   .0065021   -29.21   0.000    -.2026696   -.1771811
       Black |  -.1222253   .0060081   -20.34   0.000    -.1340012   -.1104494
       Other |  -.1223989   .0126943    -9.64   0.000    -.1472799    -.097518
     Married |   .0436671   .0044789     9.75   0.000     .0348884    .0524457
      BMI_im |   -.008551     .00035   -24.43   0.000     -.009237    -.007865
    Diabetes |  -.1408259   .0047835   -29.44   0.000    -.1502016   -.1314502
       Heart |  -.1813385   .0046563   -38.94   0.000    -.1904649   -.1722121
      Cancer |  -.0800264   .0055252   -14.48   0.000    -.0908558   -.0691969
      Stroke |  -.1300958    .007156   -18.18   0.000    -.1441217   -.1160698
       _cons |   .7683759   .0233263    32.94   0.000      .722656    .8140958
------------------------------------------------------------------------------

116 . outreg2 using Table2.doc, append ctitle (Model 2) addtext(Year FE, YES, Individual FE, NO, Regional FE, NO
    > ) sortvar(Income College Female Black Hispanic Other Age Married BMI Diabetes Heart Cancer Stroke) label
    Table2.doc
    dir : seeout
```

```
117 .
118 . reg Healthy Income College Age CollegexIncome CollegexFemale CollegexBlack Year10 Year12 Year14 Year16 His
    > panic Black Other Female Married BMI Diabetes Heart Cancer Stroke, robust

Linear regression                               Number of obs     =     49,370
                                                F(20, 49349)      =     555.89
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1498
                                                Root MSE          =     .45343

------------------------------------------------------------------------------
             |               Robust
     Healthy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Income |   6.35e-07   1.69e-07     3.75   0.000     3.04e-07    9.67e-07
     College |   .1306039   .0078105    16.72   0.000     .1152952    .1459126
         Age |  -.0007054   .0002627    -2.68   0.007    -.0012203   -.0001904
CollegexIn~e |  -3.60e-07   1.72e-07    -2.09   0.036    -6.96e-07   -2.31e-08
CollegexFe~e |  -.0042901   .0085226    -0.50   0.615    -.0209945    .0124143
CollegexBl~k |  -.0349218   .0120196    -2.91   0.004    -.0584804   -.0113633
      Year10 |   .0272629   .0065461     4.16   0.000     .0144324    .0400934
      Year12 |   .0124596   .0065885     1.89   0.059     -.000454    .0253733
      Year14 |  -.0169176   .0066165    -2.56   0.011     -.029886   -.0039492
      Year16 |  -.0342544   .0067164    -5.10   0.000    -.0474186   -.0210902
    Hispanic |  -.1869544     .00659   -28.37   0.000    -.1998709   -.1740379
       Black |  -.1072543   .0074223   -14.45   0.000    -.1218021   -.0927065
       Other |  -.1221269   .0126954    -9.62   0.000      -.14701   -.0972438
      Female |  -.0065622    .005974    -1.10   0.272    -.0182714    .0051469
     Married |   .0433455    .004497     9.64   0.000     .0345314    .0521596
      BMI_im |  -.0085467   .0003502   -24.41   0.000    -.0092331   -.0078604
    Diabetes |  -.1403687   .0047951   -29.27   0.000    -.1497672   -.1309702
       Heart |  -.1810574   .0046564   -38.88   0.000     -.190184   -.1719309
      Cancer |  -.0803017   .0055259   -14.53   0.000    -.0911326   -.0694709
      Stroke |  -.1298911   .0071603   -18.14   0.000    -.1439253   -.1158569
       _cons |   .7534631   .0244432    30.83   0.000     .7055541    .8013721
------------------------------------------------------------------------------

119 . test College CollegexIncome CollegexFemale CollegexBlack

 ( 1)  College = 0
 ( 2)  CollegexIncome = 0
 ( 3)  CollegexFemale = 0
 ( 4)  CollegexBlack = 0

       F(  4, 49349) =  179.91
            Prob > F =    0.0000
```

```
120 . outreg2 using Table2.doc, append ctitle (Model 3) addtext(Year FE, YES, Individual FE, NO, Regional FE, NO
    > ) sortvar(Income College Female CollegexIncome CollegexFemale CollegexBlack Black Hispanic Other Age Marri
    > ed BMI Diabetes Heart Cancer Stroke) addstat(F-test: College CollegexIncome CollegexFemale CollegexBlack,
    > r(F), Prob >F, (r(p))) label
    Table2.doc
    dir : seeout

121 .
122 . xtreg Healthy Income College Age CollegexIncome CollegexFemale CollegexBlack Year10 Year12 Year14 Year16 H
    > ispanic Black Other Female Married BMI Diabetes Heart Cancer Stroke, robust fe i(Respondents)
    note: College omitted because of collinearity.
    note: CollegexFemale omitted because of collinearity.
    note: CollegexBlack omitted because of collinearity.
    note: Hispanic omitted because of collinearity.
    note: Black omitted because of collinearity.
    note: Other omitted because of collinearity.
    note: Female omitted because of collinearity.

    Fixed-effects (within) regression               Number of obs     =     49,370
    Group variable: Respondents                     Number of groups  =      9,962

    R-squared:                                      Obs per group:
         Within  = 0.0230                                         min =          1
         Between = 0.1204                                         avg =        5.0
         Overall = 0.0779                                         max =          5

                                                    F(13,9961)        =      55.70
    corr(u_i, Xb) = 0.1509                          Prob > F          =     0.0000

                            (Std. err. adjusted for 9,962 clusters in Respondents)
------------------------------------------------------------------------------
             |               Robust
     Healthy |  Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      Income |  -1.66e-08   6.08e-08    -0.27   0.785    -1.36e-07    1.03e-07
     College |          0  (omitted)
         Age |  -.0002612   .0047906    -0.05   0.957    -.0096517    .0091293
CollegexIncome|   4.60e-08   6.90e-08     0.67   0.505    -8.92e-08    1.81e-07
CollegexFemale|          0  (omitted)
 CollegexBlack|          0  (omitted)
      Year10 |    .015878   .0119892     1.32   0.185    -.0076233    .0393793
      Year12 |  -.0042595   .0201775    -0.21   0.833    -.0438116    .0352925
      Year14 |  -.0405541   .0289523    -1.40   0.161    -.0973064    .0161982
      Year16 |  -.0646278   .0394509    -1.64   0.101    -.1419596     .012704
    Hispanic |          0  (omitted)
       Black |          0  (omitted)
       Other |          0  (omitted)
      Female |          0  (omitted)
     Married |  -.0055211   .0095251    -0.58   0.562    -.0241922      .01315
      BMI_im |   -.003664   .0008034    -4.56   0.000    -.0052388   -.0020891
    Diabetes |  -.0385432   .0102262    -3.77   0.000    -.0585887   -.0184977
       Heart |  -.0942208   .0089519   -10.53   0.000    -.1117682   -.0766733
      Cancer |  -.0967396   .0127694    -7.58   0.000    -.1217702    -.071709
      Stroke |  -.0893987    .014129    -6.33   0.000    -.1170943    -.061703
------------------------------------------------------------------------------
```

_cons	.6095179	.3191419	1.91	0.056	-.0160648	1.235101
sigma_u	.37134345					
sigma_e	.33313828					
rho	.55407257	(fraction of variance due to u_i)				

```
123 . outreg2 using Table2.doc, append ctitle (Model 4: Individual Fixed Effects) addtext(Year FE, YES, Individu
    > al FE, YES, Regional FE, NO) sortvar(Income College Female CollegexIncome CollegexFemale CollegexBlack Bla
    > ck Hispanic Other Age Married BMI Diabetes Heart Cancer Stroke) label
    Table2.doc
    dir : seeout

124 .
125 . xtreg Healthy Income College Age CollegexIncome CollegexFemale CollegexBlack Year10 Year12 Year14 Year16 H
    > ispanic Black Other Female Married BMI Diabetes Heart Cancer Stroke, robust fe i(Region)
    warning: existing panel variable is not Region

    Fixed-effects (within) regression               Number of obs     =     49,370
    Group variable: Region                          Number of groups  =          5

    R-squared:                                      Obs per group:
         Within  = 0.1469                                         min =        144
         Between = 0.9028                                         avg =    9,874.0
         Overall = 0.1498                                         max =     20,078

                                                    F(4,4)            =          .
    corr(u_i, Xb) = 0.1001                          Prob > F          =          .

                                        (Std. err. adjusted for 5 clusters in Region)
```

Healthy	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
Income	6.27e-07	1.07e-07	5.85	0.004	3.30e-07	9.25e-07
College	.1308738	.0131078	9.98	0.001	.0944807	.167267
Age	-.0007013	.0005045	-1.39	0.237	-.002102	.0006993
CollegexIncome	-3.54e-07	1.20e-07	-2.94	0.042	-6.88e-07	-1.97e-08
CollegexFemale	-.0051343	.0111437	-0.46	0.669	-.036074	.0258055
CollegexBlack	-.0365725	.0155605	-2.35	0.078	-.0797755	.0066304
Year10	.0271803	.0063943	4.25	0.013	.0094269	.0449336
Year12	.0123744	.0035957	3.44	0.026	.0023912	.0223577
Year14	-.0169812	.0063758	-2.66	0.056	-.0346834	.0007209
Year16	-.0343624	.0037766	-9.10	0.001	-.044848	-.0238768
Hispanic	-.183576	.00676	-27.16	0.000	-.2023447	-.1648073
Black	-.1028873	.0110997	-9.27	0.001	-.1337051	-.0720695
Other	-.1211381	.0179627	-6.74	0.003	-.1710106	-.0712657
Female	-.0057841	.0063217	-0.91	0.412	-.0233361	.0117678
Married	.0434515	.0078879	5.51	0.005	.0215511	.0653519
BMI_im	-.0085971	.0009661	-8.90	0.001	-.0112794	-.0059149
Diabetes	-.1399043	.0121837	-11.48	0.000	-.1737318	-.1060768
Heart	-.1807635	.0062349	-28.99	0.000	-.1980745	-.1634526
Cancer	-.0800774	.0095965	-8.34	0.001	-.1067214	-.0534333
Stroke	-.1290757	.0121399	-10.63	0.000	-.1627816	-.0953699

```
       _cons |   .7531733   .0637321    11.82   0.000     .5762246    .9301221
-------------+----------------------------------------------------------------
     sigma_u |   .03454788
     sigma_e |   .45334628
         rho |   .00577388   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
126 . outreg2 using Table2.doc, append ctitle (Model 5: Regional Fixed Effects) addtext(Year FE, YES, Individual
    >  FE, NO, Regional FE, YES) sortvar(Income College Female CollegexIncome CollegexFemale CollegexBlack Black
    >  Hispanic Other Age Married BMI Diabetes Heart Cancer Stroke) label
    Table2.doc
    dir : seeout

127 .
128 . log close
          name:  <unnamed>
           log:  /Users/nugget/Desktop/NYU/Econometrics/Project/Health Project.smcl
      log type:  smcl
     closed on:  12 May 2022, 13:29:36
-------------------------------------------------------------------------------------------------------------------
```