Health, Mental Health and Labor Productivity: The Role of Self-Reporting Bias

Justin Leroux[1], John Rizzo[2] and Robin Sickles[1]
[1]Department of Economics, Rice University
[2]College of Medicine and Public Health and Department of Economics, The Ohio State University

July 15, 2003

Abstract

This paper relates physical and mental health status to labor force participation and compares these relationships among self-report and proxy respondents. Previous research has conjectured that self-reports of health status may lead to an upward bias in the estimated effect of health on labor productivity because subjects who are out of the labor force may be more likely to understate their health status so as to justify their lack of employment. Also, we integrate mental health into our study by using two alternative approaches-logistic regression analysis and propensity scoring methods. We find that among the cohort of self-reporters, physical health has a substantially stronger impact on labor productivity than mental health; precisely the opposite patterns were obtained when physical and mental health status were reported by proxy respondents. These results suggest the self-reports may lead to a bias in estimating labor productivity costs of physical versus mental health on labor force participation by overestimating the importance of good physical health and underestimating the role of good mental health. This in turn suggests that the benefits of more generous mental health insurance benefits may have been underappreciated in the medical policy debates.

Keywords: health economics, propensity scoring, logistic regression, self-reported health, labor force participation, mental health.

JEL classitfication numbers:

1 Introduction

The impact of health on labor productivity is a topic of considerable interest in the labor and health economics literatures. Growing recognition of the indirect labor productivity costs associated with poor health and chronic disease has raised the interest of policymakers as well. Previous research (e.g., Bound, 1991) has conjectured that self-reports of health status may lead to an upward bias in the estimated effect of health on labor productivity. The reason is that subjects who are out of the labor force may be more likely to understate their health status so as to justify their lack of employment. This effect is believed to be more prevalent among males than females. However, the dearth of information relating mental health status to labor force participation and the difficulties in quantifying the relative importance of physical health versus mental health in effecting labor productivity suggest that linkages between health and labor force participation may not be firmly established. Empirical evidence on these issues would be quite useful to self-insuring employers who are designing health insurance benefits packages, and in policymakers concerned with understanding the full social benefits of alternative health insurance plans.

In this paper we apply two alternative methodological approaches, logistic regression analysis and propensity scoring methods, to analyze treatment effects among labor force participation, physical, and mental health status. We utilize a unique set of data, the Medical Expenditure Panel Survey (MEPS). We use the MEPS to relate physical and mental health status to labor force participation and we compare these relationships among self-report and proxy respondents. The MEPS is well-suited to this purpose because it includes both measures of physical and mental status as well as self-reports and proxy responses to questions regarding physical and mental health. In the MEPS survey, members of a subject's household responded to questions regarding the subject's physical and mental health status when the subject was unavailable to answer the surveyor's questions. Both logistic regression analysis and propensity scoring approaches condition on respondents' age, education and race. Separate analyses are conducted for males and females.

Methodologically, treatment effects have been a topic of interests in the literature for some time although recent interest appears to center on issues of robustness. Robins (1999) has suggested a class of new class of non-nested marginal structural models that may be used to estimate the causal effects of time dependent treatment on a binary outcome while Hernan et al. (2001) recently has applied this method in model of joint outcomes. Since it was first introduced by Rosenbaum and Rubin (1983) prosensity scoring, which examines the conditional probability of receiving treatment given pre-treatment variables, has been motivated by its non-parametric methodology. It eliminates bias due to an assumed (and incorrect) parametric relationship between the outcome and the observed covariates whose parametric relationship with the outcome variable is unknown. Imbens (1999) extended this methodology to allow for estimation of average causal effects with multi-valued treatments.

The paper is structured as follows. In the next section we discuss in more detail the empirical evidence that links labor force participation and health. We also examine whether such linkages appear in the MEPS using multivariate logistic regression methods. Section 3 provides theoretical treatments and the rationale behind propensity scoring methods and propensity score binning in particular. Section 4 discusses the results of our propensity scoring analyses and their implications. Section 5 concludes.

2 Evidence from the Medical Expenditure Panel Survey

The data in our study is a national representative sample of males whose ages varied between 18 and 64. A total of 3,374 subjects reported physical and mental health status by a proxy (someone in that male's household). Another 2,449 self-reported their physical and mental health status. Multivariate logistic regression models, adjusting for age, race and education, were estimated to establish a relationship among physical and mental health status, and the probability of being in the labor force. The results indicate that physical health has a much stronger impact on labor force participation for those who self-report than for those whose reports are by proxy. In contrast, mental health has a weaker effect when self-reports are used. This pattern appears to reflect the tendency for subjects who are out of the labor force to self-report their physical and mental health differently than proxy-respondents. Among unemployed subjects, self-reporters are more likely to indicate that their physical health status is poor or fair (the reference cohort). On the other hand, self-reporters are more likely to indicate that their mental health status is good or better. the information can be summarized in the following table where a subject is considered healthy when his physical condition is reported to be "good', "very good" or "excellent".

Table 1: Frequencies of Larob Force Participation by Mode of Representation

Mode of representation	Labor force participation	% of healthy subjects
Self:	In	92
	Out	59
Proxy	In	93
	Out	75

These discrepancies are all the more striking as proxy and self reports were found to be remarkably similar among subjects who were in the labor force. An interpretation of these results is that subjects who are out of the labor force tend to understate their physical health status and overstate their mental health status. This would lead to a stronger estimate of the impact of physical health on labor force participation among self-reporters, and a correspondingly weaker effect of mental health status; this is precisely what has been observed. An other interpretation is that subjects out of the labor force find it easier to blame a physical ailment —real or imagined— on their (negavite) labor force status than to admit they are suffering from emotional difficulties that limit their productivity.

Causality clearly may go both ways. Being out of the labor force can greatly affect one's mental health status. Suffering from emotional and psychological difficulties also has an impact on labor force partisipation. The combination of these factors may have a role in biasing self-reports of subjects out of the labor force. These respondents tend to downplay their emotional affection and may blame their employment status on a physical ailment, real or imagined.

Since most studies of health and labor force participation have relied only on self-reports (c.f., Sickles and Taubman, 1996), then if the bias discussed above is true, it would suggest that existing studies may be overstating the importance of physical health status on labor productivity, and understating that of mental health status. Thus when allocating health care dollars to promote worker productivity, relatively more dollars should be spent on mental health benefits and relatively less on physical health benefits. These findings could thus have important implications for the optimal design of employer-based health insurance plans. However, if the causality is reversed, social structures and health care for the unemployed could be made more effective by focusing on addressing emotional and psychological deficits.

In the next section we attempt to isolate the casual linkages between health conditions and labor force participation utilizing robust nonparametric procedures. Since there is no evidence that proxy and self-reports have been randomly assigned, the problem of sample selection of subjects may be an issue. Propensity scoring methods were employed to deal with the selection bias that arises in such a situation. The methods are introduced below.

3 Propensity Scoring

Propensity scoring methods were first introduced by Rosenbaum and Rubin (1983) as a way to significantly reduce bias in observational studies. Early applications found in the biometrics literature analyze medical treatment effects. Propensity scoring techniques have been found to be an efficient alternative to most common econometric bias-reducing techniques (e.g. Heckman's two-step estimation procedure). The objective is to statistically evaluate the effect of a particular treatment on a population (e.g. the effect of smoking on mortality). A randomized experiment with human subjects would be not only unethical but also impractical. Observational data are typically used in such cases to study causal effects.

The main problem in observational studies is that selection for treatment is not randomized. Therefore, the treated and the non-treated may differ in characteristics other than treatment intake. In an example provided by Cochran (1986), the yearly death rates of the cigarette-smoking population in the U.S. are identical to that of the non-smoking population: 13.5 per thousand in 1968. This would suggest that cigarette-smoking has little to no effect on mortality. This conclusion is, of course, incorrect. Cigarette smokers differ from non-smokers in characteristics other than smoking habits; one of the most important of these characteristics is their age. Indeed, in Cochran's studies, the average age for a non-smoker in 1968 was 57 years-old and 53 years of age for a cigarette smoker. In other words, non-smokers were found significantly older than cigarette-smokers. Failure to take into account these differences in age would not result in meaningful conclusions. When controlling for differences in age, the mortality rate of cigarette-smokers climbed to 17.7 per thousand while that of non-smokers remained at 13.5 per thousand. These figures are in accordance with other studies about the subject. In this smoking example, in order to control for age, one can compare different smoking-habit groups that have the same age distribution. One practical solution is to reweigh the sample so that the distribution of age for different smoking-habit groups is approximately the same. In the smoking example, the weights that have been chosen were simply the sample frequencies of the age subgroups for non- smokers —here, three subgroups of approximately equal size; this explains why the mortality rate for non-smokers remained unchanged in the process. The main concept behind propensity scoring methods is very similar to what has been done in this example and is developed in the next subsection.

The reweighing method becomes intractable when we want to control for more than one characteristic, or covariate. We need a one-dimensional variable that would summarize background characteristics accurately in the sense that subjects with similar values of this proxy would have similar background characteristics. The *propensity score* is precisely that.

We need to introduce some notation in order to better explain how propensity scoring can be used in our context. Let y_i and x_i denote the values of the observed outcome and covariate variables for subject i. Let z^* denote the observed treatment assignment that takes on binary values; i.e. $z_i^* = 1$ if the treatment has been assigned to subject i, $z_i^* = 0$ otherwise. The propensity score is defined as the probability of receiving the treatment given the value of the covariates.

$$p(x_i) = \Pr\{z_i^* = 1|x_i\} = \text{propensity score for subject } i.$$

The joint distribution of the covariates and the observed treatment given the propensity score is given by:

$$\begin{aligned}
\Pr\{x, z^*|p(x)\} &= \Pr\{x|p(x)\} \times \Pr\{z^*|x, p(x)\} \\
&= \Pr\{x|p(x)\} \times \Pr\{z^*|x\} \\
&= \Pr\{x|p(x)\} \text{ times } p(x) \text{ or } 1 - p(x) \\
&= \Pr\{x|p(x)\} \times \Pr\{z^*|p(x)\}.
\end{aligned}$$

where it is assumed that x and z^* are conditionally independent given the propensity score $p(x)$.

Since the objective of propensity scoring methods is to match subjects with similar background characteristics regardless of the treatment they receive one can achieve this matching by classifying subjects into a finite number of categories (*bins*) according to their propensity score. This procedure is called propensity score binning and can be decomposed in the following five steps[1].

Step 1 Model the probability of receiving the treatment ($z^* = 1$) given the covariates (x). This is usually done through a logistic regression with linear functional $x^\top\theta$.

Step 2 Sort all observations on the estimated propensity score, $p(x) = p\{z^* = 1|x\}$.

Step 3 Form five (or more) relatively homogenous subgroups (bins) of subjects with similar estimated propensity scores by dividing fitted probabilities into quintiles (or finer divisions).

Step 4 Calculate a mean outcome difference, $\bar{y}_{k,z^*=1} - \bar{y}_{k,z^*=0}$, within each subgroup k.

Step 5 Form an overall average difference, $\bar{y}_{z^*=1} - \bar{y}_{z^*=0}$, using appropriate weights.

The appropriate weights indicated in step 5 are usually proportional to the number of subjects within each bin or to the inverse of the variance of the propensity score within each bin. The overall difference from the latter option will always appear to be more precise, but this weighing typically downweighs results from the outer (first and last) bins. The overall difference using weights proportional to the sample frequencies of each bin (usually nearly equal across bins) may be much less biased as it is practically insensitive to outliers. Indeed, outliers can greatly inflate within-bin variances because within-bin sample sizes are reduced by a factor of five or more (depending on the number of subgroups).

It may not seem reasonable at first that this method requires estimating the probability that each subject receives the treatment since we already know which subjects were assigned a particular treatment. However, if we use the probability that a subject would have been treated given his values of the covariates (the propensity score, precisely) in order to adjust our estimates of the treatment effect, we can create a "quasi-randomized" experiment. In other words, two subjects (regardless of their treatment intake) who have similar propensity scores can be thought of having the same *ex ante* probability of being assigned the treatment or among subjects with similar propensity scores. One can imagine that subjects were "randomly" assigned the treatment in the sense of being equally likely to be treated or control.

[1] See Obenchain, [9] and [10].

4 Propensiting Scoring and the Relationship between Health and Labor Force Participation

In this section we utilize propensity scoring techniques to analyze the relationship between health and labor force participation. Propensity score binning is applied to subjects whose health conditions have been proxy-reported and to self-reporters, separately. In the former, however, it may be reasonable to divide subjects in two groups according to their treatment assignment, and then apply propensity score binning where the "pseudo-treatment" is now whether subjects self-report their health status. When evaluating the extent of the reporting bias itself, propensity score binning will be applied to subjects out of the labor force only, taking mode of representation as the "pseudo-treatment"; this will measure the effect of self-reporting (the treatment) on reported health (the outcome variable) for subjects out of the labor force; i.e. the reporting bias we are after.

4.1 Impact of Health on Labor Force Participation

We first analyze the effect of health (the treatment) on labor force participation. We apply propensity score binning to the proxy-reported subjects and to the self-reporters separately. For each estimation, we obtain two estimates of the effect of health on labor force participation: one using sample frequency within each bin as a weighing scheme when averaging across bins, the other using the inverse of the within-bin variance as weights. The following numerical results were chiefly obtained thanks to the S-Plus propensity scoring functions designed by Bob Obenchain[2]. The main results are presented in Table 2.

Table 2: Impact of Health on Labor Force Participation Estimation Results

	Proxy report	Self report
Raw Average Treatment Difference	0.232017	0.4294142
Standard Deviation of Raw Difference	0.03305922	0.02505341
Average within Bin Treatment Difference	0.1558927	0.2848299
Standard Deviation of Average within Bin Difference	0.09007506	0.0656173
Inverse Variance Weighted Difference	0.1910878	0.350969
Standard Deviation of Inverse Variance Difference	0.0323359	0.02486205

For the subgroup of subjects whose health status was reported by a proxy, the marginal increment in the probability of being in the labor force due to good health condition is of 16% (resp. 19%) when using within-bin sample frequency (resp. the inverse of the within-bin variance) as a weighing scheme. In other words, being in good health increases one's chances of being in the labor force by 16 to 19%. When selection bias is not taken into account, the "raw" effect is estimated at 23%.

[2] At this time, these functions are available online at the following URL: http://www.math.iupui.edu/~indyasa/bobodown.htm

As for the subgroup of subjects who self-reported their health status, the marginal increment in the probability of being in the labor force due to good health condition is of 28% (resp. 35%) when using within-bin sample frequency (resp. the inverse of the within-bin variance) as a weighing scheme. In other words, all other things being equal, a subject who considered himself in good health had a probability of being in the labor force greater than that of someone considering himself in poor health by 28 to 35%. The estimated raw impact is 43%.

These results are remarkable in several ways. First of all, these numbers corroborate the fact that the impact of physical health reports on labor force participation is greater for self-reporters, which was expected. What is most striking, however, is that even though proxy reports significantly down-play the effect of health on labor force participation, this effect is still quite important; this suggests that the proxy reports provide us with somewhat of a lower bound for our estimates. Finally, we observe a reduction in the difference between the effects on proxy- and self-reported when a bias-reducing method is employed.

4.2 Impact of Labor Force Participation on Health

It is quite reasonable to expect labor force participation to have an impact on health. Subjects out of the labor force may suffer emotionally and feel discarded and/or stigmatized (especially being males); also, as emotional and psychological difficulties often translate into physical symptoms, their physical health may be affected. This could help explain why males out of the labor force are statistically in poorer health condition than employed subjects. Moreover, the discrepancies between proxy and self-reports of physical health status may also be understood thanks to this approach. Indeed, the stigma associated with being unemployed along with the stigma associated with depression (especially for males, once again) may lead these subjects to misreport (consciously or subconsciously) their physical health status.

We apply propensity score binning estimation separately to the proxy-reported subjects and on the self-reporters. For each estimation, two values of the estimate of the effect of labor force participation on health are given; one uses sample frequency within each bin as a weighing scheme when averaging across bins, the other uses the inverse of the within-bin variance as weights. The results are presented in Table 3.

Table 3: Impact of Labor Force Participation on Health Estimation Results

	Proxy report	Self report
Raw Average Treatment Difference	0.1814896	0.3378763
Standard Deviation of Raw Difference	0.02651639	0.02132619
Average within Bin Treatment Difference	0.2325843	0.2652681
Standard Deviation of Average within Bin Difference	0.09727791	0.05495876
Inverse Variance Weighted Difference	0.1309656	0.2911695
Standard Deviation of Inverse Variance Difference	0.02531794	0.02191314

For the subgroup of subjects whose health status was reported by a proxy, the marginal increment in the probability of being in good health due to being in the labor force is of 23% (resp. 13%) when using within-bin sample frequency (resp. the inverse of the within-bin variance) as a weighing scheme. In other words, under the hypothesis made in this segment of the paper, being in the labor force increases one's chances of being in good health by 13 to 23%. The estimated raw effect is 18%.

As for the subgroup of subjects who self-reported their health status, the marginal increment in the probability of being in good health due to being in the labor force is of 27% (resp. 29%) when using within-bin sample frequency (resp. the inverse of the within-bin variance) as a weighing scheme. In other words, all other things being equal, a subject who is in the labor force has 27 to 29% greater chance of considering himself in good health than someone out of the labor force. The estimated raw effect is 34%.

Again, estimates are greater for self-reporters. These last two averages are very similar, this suggests that there are few outliers in the sample. However, the results obtained with the proxy-reported cohort are quite different, which suggest the presence of outliers; thus, the weighted average using sample frequency as weights (23%) should be more reliable. Again, the effects obtained are quite important, even when selection bias is taken care of, and are lower than, yet comparable to, the raw effect.

4.3 Measuring the Reporting Bias

??, shows that health reports for self-reporters out of the labor force were significantly lower than that for proxy-reported subjects. We suspected that this discrepancy is due to a direct relationship between labor force participation and health status. We applied a bias-reducing method in order to isolate the impact of labor force participation on health status —and *vice versa*— on self-reporters and on the proxy-reported group separately. The result is that these effects are different although more comparable than **??** suggested. One possible reason for this is that the way the data has been gathered is itself biased. We applied propensity score binning to the cohort out of the labor force (regardless of who reported their health status) in order to isolate the "effect" of self-reporting on health. It turns out that this effect is quite small compared to what could have been expected. The results are presented in Table 4.

Table 4: Reporting Bias Estimation Results

Raw Average Treatment Difference	-0.1597055
Standard Deviation of Raw Difference	0.03327768
Average within Bin Treatment Difference	-0.05576054
Standard Deviation of Average within Bin Difference	0.08102068
Inverse Variance Weighted Difference	-0.05230291
Standard Deviation of Inverse Variance Difference	0.03170674

The estimated raw bias is 20% but the "bias-less" effect of self-reporting is only 5-6%. One interpretation of this result is that there is a selection problem in "choosing" who would be the self-reporters and who would be the proxy-reported subjects. The self-reporters may self-reported precisely because their poor health constrained them to stay at home and they were thus available for the survey.

Since the "effective" impact of self-reporting is actually much smaller than what **??** suggests, it is not surprising that the difference between estimates for proxy- and self-reporters is not as great as what could have been expected. Nevertheless, this impact exists and must be taken into consideration.

These results are striking. Among the cohort of self-reporters, physical health has a substantially stronger impact on labor productivity than mental health; precisely the opposite patterns were obtained when physical and mental health status were reported by proxy respondents. Specifically, among self-reporting males, results from the propensity score estimates reveal that labor force participation is 27 % greater when subjects are in good or excellent health. The corresponding figure for proxy respondent is 17 %. In contrast, labor force participation is 24 % higher among male self-reporters in good or excellent mental health, but 33 % larger when proxy respondents are used. Similar results are obtained with the logistic regression analysis. the results for females following a similar, though less pronounced, pattern.

The reasons for this difference appear due to different responding patterns between self-reporters and proxies regarding physical and mental health status. Among males, for example, 54 % of self-reporters who were out of the labor force reported themselves to be in good or excellent health, while proxy respondents reported that 73 % of subjects out of the labor force were in good or excellent physical health. Precisely the opposite pattern occurs in reporting with respect to mental health status. It in the In this case, 82% of self-reporters indicate that they are in good or excellent mental health status, while only 79% of subjects out of labor force are rated as being in good or excellent mental health status.

These results suggest the self-reports may lead to serious bias in estimating labor productivity costs of physical versus mental health on labor force participation, by overestimating the importance of good physical health an underestimating the role of good mental health. This in turn suggests that there may be a role for relatively more generous mental health insurance benefits.

5 Conclusion

The paper relates physical and mental health status to labor force participation and compares these relationships among self-reporter and for proxy respondents. We used two alternative approaches-logistic regression analysis and propensity scoring methods, each approach condition on respondents' age, education and race.

The results suggest that the self-reports may lead to a bias in estimating labor productivity costs of physical versus mental health on labor force participation, by overestimating the importance of good physical health an underestimating the role of good mental health. This in turn suggests that there may be a role for relatively more generous mental health insurance benefits. More specifically, existing studies on which the design of health insurance plans is based have mostly relied on self-reports. Since there is evidence that these self reports are biased, there is a great possibility that existing employer-based insurance plans are not optimal. Here, we tried to isolate the effective impact of health on labor force participation. This approach will be a useful tool in the design of more efficient employer-based insurance plans. Results also suggest that this impact is substantial even when stripped of the reporting bias but that it is significantly lower than that inferred from self-reports.

In this paper, we also examined the relationship between physical health and labor force participation using a somewhat different approach. Our hypothesis is labor force participation has a strong impact on health and therefore, we can have a psychosomatic interpretation: there is a chain reaction from "being unemployed" to "suffering from physical symptoms". This approach appears to be useful when we consider the allocation of spending in social structures for the unemployed. The effects are smaller after reducing the bias but are still quite significant.

Both free-of-bias impacts measured for self-reporters turned out to be somewhat close to what could have been inferred from Table 1. This raises the question of whether there exists a significant reporting bias. In order to answer this question, we applied the same bias-reducing method (propensity score binning) on the unemployed cohort (regardless of reporting procedure) to isolate the impact of "self-reporting" on "reported health". Numerical results suggest that this impact is significant though much lower than expected— 6% instead of 16%— and, as a consequence, that the "choice" of whether a subject is proxy- or self-reported is extremely non-random and greatly biased towards unhealthy self-reporters. One possible interpretation is that unemployed subjects who are in poor health condition tend to be more available at home than healthier subjects.

6 Appendix I-The Medical Expenditure Panel Survey (MEPS) Data

The data set used for this paper is a full year 1996 Household Component consolidated Data file (HC-012) from the Medical Expenditure Panel Survey (MEPS). The variable extracted from this data set are described below.

LABFOR Binary Variable (BV) = 1 if subject is in the labor force; else =0.

HEALTHY BV =1 if reported health of the subject is "good", "very good" or "excellent"; else =0.

AGE1834[3] BV =1 if subject aged 18-34; else =0.

AGE4554 BV =1 if subject aged 45-54; else =0.

AGE5565 BV =1 if subject aged 55-65; else =0.

NOHS[4] BV =1 if subject did not attend high school; else =0.

SOMEHS BV =1 if subject attended but did not graduate from high school; else =0.

SOMECOLL BV =1 if subject attended but did not graduate from college; else =0.

COLLGRAD BV =1 if subject is a college graduate; else =0.

GRADSCHL BV =1 if subject attended graduate school; else =0.

HISPANIC[5] BV =1 if subject is Hispanic; else =0.

BLACK BV =1 if subject is African American; else =0.

OTHRACE BV =1 if subject is of other non-white race; else =0.

SELFRESP BV =1 if subject self-represented his health condition; else =0.

[3]Reference age cohort is 35-44.

[4]Reference education cohort is high school education.

[5]Reference race group is white.

7 Appendix II-Logistic Regression Results

Table 5-Results from Proxy Reports

Parameter	Estimate	Standard Error	Chi-square	Pr>ChiSq
Intercept	2.1813	0.4563	22.8539	<.0001
hlthexc	1.2349	0.4057	9.2675	0.0023
hlthvg	1.0824	0.3570	9.1944	0.0024
hlthgood	0.4798	0.3074	2.4366	0.1185
hlthpoor	-0.7711	0.3765	4.1940	0.0406
mlthexc	1,7791	0.4154	18.3554	<.0001
mlthvg	1.5771	0.4145	15.2000	<.0001
mlthgood	1.1839	0.3657	10.4807	0.0012
mlthpoor	-1.2143	0.5659	4.6046	0.0319
age1834	-0.9714	0.3779	6.6075	0.0102
age4554	-1.0537	0.3588	8.6263	0.0033
age5564	-2.7290	0.3349	66.4042	<.0001
nohs	-0.5127	0.3342	2.3537	0.1250
somehs	-0.1465	0.3003	0.2380	0.6256
somecoll	0.0183	0.3240	0.0032	0.9550
collgrad	0.1835	0.3857	0.2265	0.6341
gradschl	0.3320	0.4478	0.5494	0.4585
hispanic	0.5547	0.3118	3.1642	0.0753
black	-0.2629	0.3063	0.7369	0.3906
othrace	0.1815	0.6642	0.0746	0.7847

Table 6-Results from Self Report

Parameter	Estimate	Standard Error	Chi-square	Pr>ChiSq
Intercept	0.7329	0.3439	4.5406	0.0331
hlthexc	2.0461	0.2950	48.1147	<0.0001
hlthvg	1.5650	0.2643	35.0750	<0.0001
hlthgood	1.3810	0.2441	32.0099	<0.0001
hlthpoor	-1.1769	0.3062	14.7690	0.0001
mlthexc	1.2156	0.3104	15.3343	<0.0001
mlthvg	1.4438	0.3196	20.4038	<0.0001
mlthgood	0.9414	0.3020	9.7193	0.0018
mlthpoor	-0.5967	0.6370	0.8777	0.3488
age1834	-0.0332	0.2805	0.0140	0.9057
age4554	-0.3730	0.2610	2.0430	0.1529
age5564	-2.4688	0.2415	104.4716	<0.0001
nohs	-0.4087	0.3084	1.7559	0.1851
somehs	-0.6523	0.2456	7.0528	0.0079
somecoll	0.0685	0.2399	0.0816	0.7751
collgrad	0.1599	0.2762	0.3352	0.5626
gradschl	0.00176	0.3041	0.0000	0.9954
hispanic	0.8127	0.2807	8.382	0.0038
black	-0.3862	0.2319	2.7726	0.0959
othrace	0.6365	0.5460	1.3587	0.2438

Estimation of the effect of health on labor force participation

Table 7-Estimation of the direct effect on proxy-reporters

Coefficients	Value	Std. Error	t value
Intercept	2.6777175	0.2046813	13.0823774
AGE1834	0.2823608	0.2197123	1.2851389
AGE4554	-0.8776531	0.2231059	-3.9337957
AGE5565	-1.3319584	0.2379992	-5.5964822
NOHS	-0.6006229	0.2755286	-2.1798935
SOMEHS	-0.6853505	0.2294515	-2.9869084
SOMECOLL	0.7840215	0.2255925	3.4753876
COLLGRAD	0.9163796	0.2613299	3.5066012
GRADSCHL	1.7332922	0.3727030	4.6505993
HISPANIC	-0.5398733	0.1913124	-2.8219465
BLACK	-0.9078931	0.2214222	-4.1002805
OTHRACE	-0.3659532	0.4520527	-0.8095367

Table 8-Correlation of Coefficients

	(Intercept)	AGE1834	AGE4554	AGE5565	NOHS	SOMEHS
AGE1834	-0.6184003					
AGE4554	-0.6605102	0.5853671				
AGE5565	-0.6101941	0.5490948	0.5511841			
NOHS	-0.1388347	-0.0022162	0.0036449	-0.0206877		
SOMEHS	-0.1921420	-0.1029693	-0.0005667	-0.0303273	0.2762547	
SOMECOLL	-0.3149545	0.0339930	0.0031446	0.0133492	0.1797715	0.2326071
COLLGRAD	-0.2712112	0.0338795	0.0024102	-0.0379662	0.1469217	0.1936892
GRADSCHL	-0.1970960	0.0333021	-0.0126118	-0.0301042	0.1002941	0.1310080
HISPANIC	-0.3310053	-0.0441280	0.0761193	0.0685336	-0.3479526	-0.1717387
BLACK	-0.2464559	-0.0635043	0.0167087	0.0333071	-0.0083066	-0.0151766
OTHRACE	-0.1001362	-0.0436780	-0.0251642	0.0180434	-0.0372268	-0.0173057

Table 8 (continued)

	SOMECOLL	COLLGRAD	GRADSCHL	HISPANIC	BLACK
AGE1834					
AGE4554					
AGE5565					
NOHS					
SOMEHS					
SOMECOLL					
COLLGRAD	0.2321381				
GRADSCHL	0.1639708	0.1482688			
HISPANIC	0.0739044	0.0920318	0.0749595		
BLACK	-0.0174633	0.0147794	0.0335910	0.2811911	
OTHRACE	0.0093811	-0.0422876	0.0110207	0.1496658	0.1174859

Table 9-Estimation of the direct effect on self-reporters

Coefficients	Value	Std. Error	t value
(Intercept)	1.89977772	0.1363685	13.9312059
AGE1834	0.78676980	0.1639703	4.7982456
AGE4554	-0.30127060	0.1576668	-1.9108050
AGE5565	-0.90047152	0.1507847	-5.9719042
NOHS	-1.03551816	0.1962409	-5.2767713
SOMEHS	-0.84565413	0.1829074	-4.6234015
SOMECOLL	0.53672480	0.1496793	3.5858311
COLLGRAD	1.22933918	0.2185395	5.6252492
GRADSCHL	1.08334098	0.2084939	5.1960329
HISPANIC	-0.03413954	0.1543593	-0.2211693
BLACK	-0.66087222	0.1515056	-4.3620325
OTHRACE	-0.13999465	0.3017320	-0.4639701

Table 10-Correlation of Coefficients

	(Intercept)	AGE1834	AGE4554	AGE5565	NOHS	SOMEHS
AGE1834	-0.5098055					
AGE4554	-0.5646113	0.4623342				
AGE5565	-0.6189303	0.4745455	0.5090807			
NOHS	-0.2206865	0.0206923	0.0077221	-0.0371620		
SOMEHS	-0.2519280	-0.0559883	-0.0293530	-0.0151666	0.2505598	
SOMECOLL	-0.4168726	0.0343134	0.0240404	0.0474792	0.2516465	0.2816133
COLLGRAD	-0.2612937	0.0003965	-0.0110890	-0.0105320	0.1707920	0.1893999
GRADSCHL	-0.2749218	0.0362348	-0.0349459	-0.0227533	0.1791437	0.1949544
HISPANIC	-0.2450677	-0.0928869	0.0240479	0.0995133	-0.2909566	-0.1161175
BLACK	-0.2335040	-0.0468231	0.0270920	0.0580948	-0.0457615	-0.0725550
OTHRACE	-0.0952808	-0.0596441	-0.0171938	0.0207383	-0.0829279	-0.0068297

Table 10 (continued)

	SOMECOLL	COLLGRAD	GRADSCHL	HISPANIC	BLACK
AGE1834					
AGE4554					
AGE5565					
NOHS					
SOMEHS					
SOMECOLL					
COLLGRAD	0.2409391				
GRADSCHL	0.2522734	0.1781445			
HISPANIC	0.0445727	0.0473390	0.0455268		
BLACK	-0.0367938	-0.0027055	0.0096205	0.2111868	
OTHRACE	0.0115776	-0.0495118	-0.0431278	0.1379878	0.1133385

Estimation of the effect of labor force participation on health

Table 11-Estimation of the direct effect on proxy-reporters

Coefficients	Value	Std. Error	t value
(Intercept)	3.36054562	0.2448693	13.7238320
AGE1834	-1.63806980	0.2384672	-6.8691608
AGE4554	-0.75720165	0.2871384	-2.6370615
AGE5565	-1.96938668	0.2726974	-7.2218760
NOHS	-0.66928463	0.2669143	-2.5074887
SOMEHS	-0.56964307	0.2125012	-2.6806587
SOMECOLL	0.39892901	0.1887624	2.1133919
COLLGRAD	0.65624220	0.2260319	2.9033173
GRADSCHL	0.69844895	0.2553392	2.7353774
HISPANIC	-0.05201517	0.1786209	-0.2912043
BLACK	-0.98517917	0.1860746	-5.2945391
OTHRACE	-0.76996806	0.3341771	-2.3040719

Table 12-Correlation of Coefficients

	(Intercept)	AGE1834	AGE4554	AGE5565	NOHS	SOMEHS
AGE1834	-0.8387522					
AGE4554	-0.7054372	0.7156976				
AGE5565	-0.7399210	0.7525245	0.6273870			
NOHS	-0.1175768	0.0111593	-0.0060176	-0.0165806		
SOMEHS	-0.1512310	-0.0440214	-0.0043997	-0.0256582	0.2668396	
SOMECOLL	-0.2723961	0.0172927	0.0088490	0.0157612	0.1946537	0.2665163
COLLGRAD	-0.2306278	0.0176350	0.0099243	-0.0261018	0.1538496	0.2158331
GRADSCHL	-0.2234835	0.0370786	-0.0013017	-0.0259535	0.1357679	0.1859404
HISPANIC	-0.2063337	-0.0258571	0.0379493	0.0361517	-0.3455750	-0.1852521
BLACK	-0.1864687	-0.0155731	0.0040317	0.0350431	-0.0002709	-0.0256784
OTHRACE	-0.0814936	-0.0249522	-0.0193137	0.0238536	-0.0442892	-0.0346177

Table 12 (continued)

	SOMECOLL	COLLGRAD	GRADSCHL	HISPANIC	BLACK
AGE1834					
AGE4554					
AGE5565					
NOHS					
SOMEHS					
SOMECOLL					
COLLGRAD	0.2757122				
GRADSCHL	0.2445181	0.2147110			
HISPANIC	0.0804601	0.1027945	0.0999059		
BLACK	-0.0149013	0.0257541	0.0492945	0.2587503	
OTHRACE	0.0068188	-0.0634760	0.0127076	0.1601539	0.1373109

Table 13-Estimation of the direct effect on self-reporters

Coefficients	Value	Std. Error	t value
(Intercept)	2.059150967	0.1343294	15.3291113
AGE1834	0.006406614	0.1420324	0.0451067
AGE4554	-0.050658722	0.1636326	-0.3095883
AGE5565	-1.713336817	0.1419293	-12.0717651
NOHS	-0.745710927	0.1968959	-3.7873352
SOMEHS	-0.798535425	0.1762538	-4.5305991
SOMECOLL	0.377363351	0.1336331	2.8238763
COLLGRAD	0.878178527	0.1737453	5.0544006
GRADSCHL	0.650039880	0.1692235	3.8413095
HISPANIC	-0.028386529	0.1425542	-0.1991280
BLACK	-0.853420564	0.1361675	-6.2674316
OTHRACE	-0.783704537	0.2254438	-3.4762750

Table 14-Correlation of Coefficients

	(Intercept)	AGE1834	AGE4554	AGE5565	NOHS	SOMEHS
AGE1834	-0.6064877					
AGE4554	-0.5478817	0.5118485				
AGE5565	-0.6647572	0.5774316	0.5165996			
NOHS	-0.2070514	0.0374324	-0.0015806	-0.0141380		
SOMEHS	-0.2443213	-0.0350757	-0.0322345	0.0104730	0.2270804	
SOMECOLL	-0.4106835	0.0272511	0.0262928	0.0362218	0.2458328	0.2849296
COLLGRAD	-0.2829906	-0.0051334	-0.0011865	-0.0299167	0.1862119	0.2130457
GRADSCHL	-0.3033743	0.0401537	-0.0246918	-0.0368089	0.1933922	0.2149910
HISPANIC	-0.2264972	-0.0849216	0.0212521	0.0869445	-0.2739518	-0.1173173
BLACK	-0.2411076	-0.0322696	0.0238348	0.0988645	-0.0362526	-0.0530523
OTHRACE	-0.1104667	-0.0755547	-0.0129289	0.0539466	-0.0646487	0.0048099

Table 14 (continued)

	SOMECOLL	COLLGRAD	GRADSCHL	HISPANIC	BLACK
AGE1834					
AGE4554					
AGE5565					
NOHS					
SOMEHS					
SOMECOLL					
COLLGRAD	0.2967395				
GRADSCHL	0.3037343	0.2432402			
HISPANIC	0.0479907	0.0534850	0.0514001		
BLACK	-0.0265625	-0.0110436	0.0134207	0.2141088	
OTHRACE	0.0006136	-0.0737926	-0.0692231	0.1547503	0.1468046

Table 15-Estimation of the reporting bias

Coefficients	Value	Std. Error	t value
(Intercept)	1.354332773	0.2648608	5.11337516
AGE1834	-1.698627982	0.2684028	-6.32865307
AGE4554	-0.632042016	0.3177594	-1.98905826
AGE5565	-0.029482088	0.2865840	-0.10287417
NOHS	0.204485812	0.2834031	0.72153693
SOMEHS	0.229043567	0.2464133	0.92950982
SOMECOLL	0.395526682	0.2245955	1.76106236
COLLGRAD	0.008283243	0.2875954	0.02880173
GRADSCHL	0.338133070	0.2988733	1.13135923
HISPANIC	-0.122712343	0.2089599	-0.58725293
BLACK	0.131540746	0.2102827	0.62554227
OTHRACE	0.563540501	0.3710736	1.51867602

Table 16-Correlation of Coefficients

	(Intercept)	AGE1834	AGE4554	AGE5565	NOHS	SOMEHS
AGE1834	-0.7619490					
AGE4554	-0.6873993	0.6743890				
AGE5565	-0.7725232	0.7426144	0.6303560			
NOHS	-0.1710152	0.0080728	0.0062792	-0.0320976		
SOMEHS	-0.1930244	-0.0669203	0.0157417	0.0111114	0.2572353	
SOMECOLL	-0.2353789	-0.0795095	-0.0063167	-0.0173412	0.2524999	0.3085231
COLLGRAD	-0.1914057	-0.0693520	0.0275512	-0.0562542	0.1996790	0.2244001
GRADSCHL	-0.2005237	-0.0254146	0.0055910	-0.0271502	0.1786811	0.2144099
HISPANIC	-0.1788754	-0.1021378	-0.0332495	0.0537711	-0.1956676	-0.1406460
BLACK	-0.2336323	-0.0182373	0.0225324	0.0889131	-0.0344587	-0.0958557
OTHRACE	-0.0939054	-0.0741737	0.0147944	0.0403233	-0.1012095	0.0061735

Table 16 (continued)

	SOMECOLL	COLLGRAD	GRADSCHL	HISPANIC	BLACK
AGE1834					
AGE4554					
AGE5565					
NOHS					
SOMEHS					
SOMECOLL					
COLLGRAD	0.2648404				
GRADSCHL	0.2460713	0.1952586			
HISPANIC	0.0287082	0.0609277	0.0655299		
BLACK	-0.0181689	0.0757528	0.0006213	0.3011063	
OTHRACE	-0.0193396	-0.0765121	0.0356291	0.1961641	0.1647513

References

[1] Abadie A., Observational studies I: Matching and Regression, Lecture notes of Spring 2001, handout #5. (Unpublished).

[2] Behrman, J., Sickles, R. C., Taubman, P. (1998), *The Causes, Correlates and Consequences of Death Among Older Adults: Some Methodological Approaches and Substantive Analysis*, Boston: Kluwer Academic, 1-186.

[3] Behrman, J., Sickles, R. C., Taubman, P., and Yazbeck, A. (1992), "Black-White Differentials in Mortality: Results from the Retirement History Survey,"*Journal of Econometrics*, 50, 183-204.

[4] Bound, J. (1991), "Self-reported versus Objective Measures of Health in Retirement Models," *Journal of Human Resources* 26(1), 106-138.

[5] Cochran, W. G. (1968), " The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies," *Biometrics*, 26, 295-313.

[6] D.Agostino, R. B. Jr (1998), "Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group," *Statistics in Medicine*, 17, 2265-2281.

[7] Hernan, M. A., Brumback, B., and Robins, J. (2001), "Margianl Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments," *Journal of American Statistical Association,* 96, 440-447.

[8] Imbens, G. W. (1999), " The Role of Propensity Score in Estimating Dose Response Functions," NBER Working Paper.

[9] Obenchain R. L., Propensity score binning and smoothing in S-Plus, online http://www.math.iupui.edu/~indyasa/bobodown.htm

[10] Obenchain R. L. andMelÞ C., Adjustment for treatment selection bias in retrospective studies, http://www.math.iupui.edu/~indyasa/bobodown.htm

[11] Medical Expenditure Panel Survey (MEPS); Household Component Full Year Files (year 1996). October 2001. Agency for Healthcare Research andQuality, Rockville, MD.

[12] Robins, J.M. (1999), "Marginal Structural Models Versus Structural Nested Models as Tools for Causal Inference," *Statistical Models in Epidemiology: The Environment and Clinical Trials,* eds. M. E. Halloran and D. Berry, New York: Springer-Verlag, pp. 95-134.

[13] Rosenbaum, P. R., and Rubin, D. B. (1983), " The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika,* 70, 41-55.

[14] Sickles, R. C., and Taubman, P. (1986), "An Analysis of the Health and Retirement Status of the Elderly," *Econometrica,* 54, 1339-1356.

[15] Sickles, R. C., and Taubman, P. (1997), "Mortality and Morbidity Among Adults and the Elderly," Chapter 16 of the *Handbook of Population and Family Economics*, edited by Mark R. Rosenzweig and Oded Stark, Amsterdam: North-Holland, 559-643.