Empir Econ
DOI 10.1007/s00181-010-0434-z

The role of self-reporting bias in health, mental health and labor force participation: a descriptive analysis

Justin Leroux · John A. Rizzo · Robin Sickles

Received: 9 October 2009 / Accepted: 17 September 2010

Abstract Previous research on male subjects has conjectured that subjective self-reports of health status may lead to an upward bias in the estimated effect of health on labor force participation 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. In the descriptive analysis conducted in this article, we compare the effects of mental and physical health status on labor force participation, employing propensity score methods to investigate whether these effects differ for self- and proxy respondents. The authors initially find some evidence that seems to suggest systematic differences between proxy and self-reporters in the effects of health on labor force participation, raising the possibility that self-reporters may be biased in their health assessments. After we control for objective baseline indices of mental and physical health status, however, differences between subjective health assessments and labor force participation become smaller and statistically insignificant. These results suggest that self-reports do not lead to overestimates of the importance of good physical or mental health on labor force participation, after one controls for objective health conditions in the models. Although we conclude that propensity score matching is a useful way

J. Leroux
Institute of Applied Economics, HEC Montréal,
CIRANO and CIRPEE, Montreal, QC, Canada

J. A. Rizzo
Department of Economics, Stony Brook University, Stony Brook, NY, USA

J. A. Rizzo
Department of Preventive Medicine, Stony Brook University, Stony Brook, NY, USA

R. Sickles (✉)
Department of Economics, Rice University, Houston, TX, USA
e-mail: rsickles@rice.edu

Published online: 29 December 2010

to align observations with covariates in estimating the effects of health on labor force participation, we find that the appropriate specification of matching variables—and in particular, the inclusion of objective health measures—is critical for understanding whether self-reporting bias matters in this context.

Keywords Health economics · Propensity scoring · Logistic regression · Self-reported health · Labor force participation · Mental health

JEL Classification C10 · C14 · I18 · J24

1 Introduction

The impact of health on labor force participation is a topic of considerable interest in the literature on labor and health economics. 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 force participation. 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 men than women.

Consistent with this view, research has found that men "perceive the state of unemployment as more stigmatic" (Kulik 2000a, p. 487).[1] However, the dearth of information relating mental health status to labor force participation[2] and the difficulties in quantifying the relative importance of health versus mental health in affecting labor force participation 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, to employer-based insurance in general, and to policymakers concerned with understanding the full social benefits of alternative health insurance plans.

In this article, we provide descriptive evidence on the relationships between labor force participation, health, and mental health status. The authors seek to understand whether self-reported health and mental health status measures lead to biased estimated associations with labor force participation because of misreporting by the respondents. The authors utilize a unique set of data, the Medical Expenditure Panel Survey (MEPS) (2006). The authors use the MEPS to relate health and mental health status to labor force participation, and these relationships among self-reporting and proxy respondents are compared. The MEPS is well suited to this purpose because it includes both measures of health and mental status as well as self-reports and proxy responses to questions regarding health and mental health. In the MEPS survey, members of a subject's household responded to questions regarding the subject's health and mental health status when the subject was unavailable to answer the surveyor's questions. Separate analyses are conducted for men and women.

[1] See also Kulik (2000b) for further evidence on gender differences in attitudes toward unemployment.

[2] Hereafter, we use "health" to refer to physical health and "mental health" otherwise.

The authors use propensity score methods to estimate the associations between labor force participation and, respectively, health and mental health. Robins (1999) suggested a new class of non-nested marginal structural models that may be used to estimate the effects of time-dependent variables on a binary outcome, while Hernan et al. (2001) applied this method in a model of joint outcomes. Since it was first introduced by Rosenbaum and Rubin (1983), propensity scoring 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 (2000) extended this methodology to allow for estimation of average effects with multi-valued treatments.

Frölich (2007) used propensity score matching to study gender differences in wages. He found that choice of degree among university graduates was a key factor explaining gender differences in wages, and that omitting this variable substantially increased the proportion of the gender earnings gap that was attributable to unobserved factors often thought to include discrimination. In a similar vein, the authors find that including objective measures of health and mental health status in our propensity score matching routine has important effects on interpretation of the results. In the absence of such objective measures, the results seem to suggest the presence of systematic differences between proxy and self-reporters in the effects of health on labor force participation, raising the possibility that self-reporters may be biased in their health assessments. After the authors control for objective baseline indices of health and mental health status, however, differences between subjective health assessments and labor force participation become small and statistically insignificant.

The article is structured as follows. In the next section, the authors discuss in more detail the empirical evidence that links labor force participation and health. Section 3 discusses the propensity scoring procedures used in this study and the rationale for propensity scoring methods and propensity score binning in particular. Section 4 discusses the descriptive results of our propensity scoring analyses. Section 5 concludes.

2 Evidence from the Medical Expenditure Panel Survey

The authors base their examination of propensity scoring methods in analyzing relationships between labor force participation, health, and mental health status on the Household Component Consolidated Data file from the MEPS for the years 2000–2005. This database, cosponsored by the Agency for Healthcare Research and Quality (AHRQ) and the National Center for Health Statistics (NCHS), provides nationally representative estimates of medical care services, health care expenditures, health status, labor force participation, and sociodemographic characteristics for the civilian, non-institutionalized population in the United States. The sample that is used includes subjects aged from 18 to 65. Students were excluded from the sample.[3] A total of

[3] The MEPS sample was chosen as a nationally representative subsample of the ongoing National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics, and may be linked to the NHIS database as well. The MEPS survey respondents were interviewed in person. The survey achieved a response rate of 77.7% (see Cohen et al. 1996 for further details). MEPS has a complex survey design involving stratification and clustering. The authors have used the weights in MEPS to deal with these issues.

Table 1 Subjective health and mental health, by respondent and labor force participation status

Health measure	In labor force proxy $N = 22,943$	In labor force self $N = 13,647$	Not in labor force proxy $N = 3,814$	Not in labor force self $N = 3,127$
Males				
Good health[a]	0.93	0.92	0.71	0.61
Good mental health[a]	0.97	0.96	0.82	0.80
Health measure	In labor force proxy $N = 8,104$	In labor force self $N = 26,409$	Not in labor force proxy $N = 3,713$	Not in labor force self $N = 12,570$
Females				
Good health[a]	0.93	0.91	0.79	0.74
Good mental health[a]	0.97	0.96	0.87	0.86

[a] "Good" health is defined as a binary variable equal to 1 if the subject indicates that he or she is in good, very good, or excellent health and 0 if in poor or fair health. The same procedure is used to define good mental health

26,757 men and 11,817 female subjects had health and mental health status reported by a proxy (someone in that person's household). Another 16,774 men and 38,978 women self-reported their health and mental health status.

Table 1 shows respondents' answers to questions about their health and mental health according to whether they are in or out of the work force. Separate summaries of responses are reported by gender and respondent status (e.g., self-reporter or proxy). Turning first to the men, Table 1 indicates that among subjects in the labor force, there is close agreement between proxies and self-reporters in terms of health status. Thus, 92 (respectively 91) % of male (respectively female) self-reporters in the work force indicate that they are in good health. "Good" health is defined as a binary variable equal to 1 if the subject indicates that he or she is in good, very good, or excellent health and 0 if in poor or fair health. The same procedure is used to define good mental health. Proxy respondents indicate that 93% of both men and women in the work force are in good health. Proxy and self-reporters' assessments of mental health status are nearly identical for men and women in the work force.

In contrast, among subjects out of the work force, substantial differences emerge between proxy and self-reporters for health. In particular, while 61% of out-of-work men indicate that they are in good health, the corresponding figure for proxies is 71%. A similar though less pronounced pattern occurs for women. Proxy and self reports of mental health are similar for out-of-work subjects.

Collectively, this pattern seems to suggest that out-of-work male and female self-reporters may be understating their health status and that this tendency is stronger among the men. These discrepancies are all the more striking since proxies and self-reporters are found to be remarkably similar among subjects who were in the labor force.

While the patterns observed in Table 1 are suggestive, differences in the characteristics of subjects whose health status is self or proxy reported may confound these results. In the next section, we attempt to isolate the linkages between health conditions and labor force participation utilizing robust nonparametric procedures. The methods are described next.

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 were found in the biometrics literature to analyze medical interventions. Propensity scoring techniques have been found to be an efficient alternative to the 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 intervention on a population (e.g., the effect of smoking on mortality). A randomized experiment with human subjects would be not only impractical but also unethical. Observational data are typically used in such cases to study causal effects.

The main problem in observational studies is that study and control groups are not randomized. Therefore, these groups may differ in characteristics other than treatment intake. In the present case, subjects are not randomly assigned to be in good or bad health. These cohorts may differ from each other in a variety of respects besides subjective health status. Hence, to isolate the effects of self-reported health status on labor force participation, it is necessary to condition on these factors.

The propensity scoring approach requires one to estimate the probability that each subject receives the intervention, whether or not the subject actually does. At first blush, this may not seem reasonable since we already know which subjects were assigned to a particular intervention. However, if we use the probability that a subject would have received the intervention given his values of the covariates (the propensity score), then we can create a "quasi-randomized" experiment. In other words, two subjects (regardless of their having received the intervention) who have similar propensity scores can be thought of as having the same *ex ante* probability of receiving the intervention or among subjects with similar propensity scores. One can imagine that subjects were "randomly" assigned the intervention in the sense of being equally likely to be the intervention group or control.

The authors used several alternative matching methods in our analyses. The first method, nearest neighbor matching with replacement, chooses one individual from the comparison group as a matching partner for a treated individual who is the closest in terms of propensity score. The untreated individual can be used more than once as a match. If T is the set of treated units and C the set of control units, with Y_i^T and Y_j^C their corresponding observed outcomes, then $C(i) = \min_j ||p_i - p_j||$ is the set of control units matched to the treated unit i with an estimated value of propensity score of p_i. Let N_i^C be the number of controls matched with $i \in T$, and set weights $w_{ij} = \frac{1}{N_i^C}$ if $j \in C(i)$ and $w_{ij} = 0$ otherwise, with $w_j = \sum_i w_{ij}$. Then the average intervention effect on those receiving the intervention is calculated as

$$\tau^N = \frac{1}{N^T} \sum_{i \in T} \left[Y_i^T - \sum_{j \in C(i)} w_{ij} Y_j^C \right].$$

Radius matching is similar to the nearest neighbor matching, except that it imposes a tolerance level on the propensity score distance to avoid bad matches. The set of controls matched with the treated is defined as $C(i) = \{p_j || |p_i - p_j|| < r\}$, where r is the tolerance level of propensity score distance and the average intervention effect is

$$\tau^R = \frac{1}{N^T} \sum_{i \in T} \left[Y_i^T - \sum_{j \in C(i)} w_{ij} Y_j^C \right].$$

Another matching method, kernel matching, uses weighted averages of all the individuals in the control group to construct the counterfactual outcomes. It assigns a higher weight to observations close to a treated individual (in terms of their propensity score) and a lower weight to observations that are not close. Using kernel matching, the average intervention effect on those receiving the intervention is calculated as

$$\tau^K = \frac{1}{N^T} \sum_{i \in T} \left\{ Y_i^T - \frac{\sum_{j \in C} Y_j^C G\left(\frac{p_i - p_j}{h_n}\right)}{\sum_{k \in C} G\left(\frac{p_i - p_j}{h_n}\right)} \right\},$$

where $G(.)$ is the kernel function, and h_n is the bandwidth parameter.

The last method, stratification matching, ranks the treated and controls and groups them into intervals (strata) on the basis of their propensity score. After calculating the impact for each stratum, we take mean difference in outcomes between the treated and control observations with weights proportional to the number of treated units in each stratum. This gives us an ATT given by

$$\tau_q^S = \frac{1}{N_q^T} \sum_{i \in I(q)} Y_i^T - \frac{1}{N_q^C} \sum_{j \in I(q)} Y_j^C,$$

where q refers to the q-th interval, and $I(q)$ is the set of units in stratum q. The terms N_q^T and N_q^C are the numbers of treated and control units in stratum q, and this can be rewritten as

$$\tau^S = \sum_{q=1}^{Q} \tau_q^S \frac{\sum_{i \in I(q)} D_i}{\sum_{\forall i} D_i},$$

where Q is the number of intervals.[4]

To implement the propensity score, we matched subjects on the following characteristics: age, education, race, marital status, health insurance status, and geographic

[4] Standard errors for the nearest neighbor matching have been corrected using an algorithm in Abadie and Imbens (2006). For the radius and kernel matching method, we used STATA code "psmatch2", see Leuven and Sianesi (2003) for further details. For the stratification matching method, we used STATA code "pscore" and "atts". See Becker and Ichino (2002), for further details.

location. The authors also conditioned on two indices of health and mental health provided in MEPS. The health index is known as the Physical Component Summary Scale Score (PCS), and the mental health index is the Mental Component Summary Scale Score (MCS). These indices are designed to provide comprehensive summary measures of mental and health status based on the ability to perform a number of specific daily activities as well as feelings and vitality. Because they are based on the ability to perform specific activities and measure specific mental health-related attitudes, these provide more objective measures of health and mental health than the subjective health assessments. These scales have been described in more detail elsewhere (Jenkinson et al. 1997; Ware et al. 1994). To gauge the importance of conditioning on objective health and mental health status, we use propensity score models with and without these health indices.

4 Propensity scoring and the relationship between health and labor force participation

Table 2 provides descriptive statistics for the variables used in this study, by respondent status (proxy or self-reporter) and gender. Self- and proxy respondents are quite similar in terms of most observable characteristics. The main exception is marital status. For both men and women, substantially lower shares of self-reporters are married. This likely reflects that fact that unmarried subjects may have more difficulty obtaining a proxy respondent from their household. Self-reporters have slightly lower health and mental health status measures.

In this section, we utilize propensity scoring techniques to analyze the relationship between health and labor force participation. The nearest neighbor, radius, kernel, and stratification matching methods are employed here.

4.1 No matching on health indices

The authors first analyze the effect of good health and mental health on labor force participation *without matching on the health and mental health indices described above*.

The main results are presented in Table 3. Turning first to the subgroup of male 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 31%. In other words, being in good health increases one's chances of being in the labor force by 31%. This marginal increment is greater for self-reporting men—35.6%. Moreover, this difference is significant at the 1% level. The authors assume that the proxy responses are unbiased because they do not have an incentive to misrepresent. This is of course an estimate but one that we believe is less subject to bias. Thus, we use proxy results as the benchmark. Under this assumption, the results appear to suggest that self-reporting men who are out of the labor force may be exaggerating their poor health status to justify why they are not working. The association between mental health status and labor force participation is also higher for self-reporting men than for proxies, but this differences is not statistically significant.

Table 2 Descriptive statistics for matching variables, by reporting status and gender

Variable name	Male self $N = 16,774$	Male proxy $N = 26,757$	Female self $N = 38,979$	Female proxy $N = 11,817$
Dependent variables				
Labor force participation rate	0.84 (0.36)	0.87 (0.33)	0.74 (0.44)	0.73 (0.44)
Subject is in good health	0.87 (0.34)	0.90 (0.30)	0.86 (0.34)	0.89 (0.31)
Subject is in good mental health	0.94 (0.24)	0.95 (0.21)	0.93 (0.25)	0.94 (0.23)
Matching variables				
Age in years	42.13 (12.08)	41.81 (11.96)	42.34 (11.65)	41.90 (12.83)
Education in years	13.45 (2.79)	12.82 (2.93)	13.24 (2.73)	12.95 (3.02)
Race				
Hispanic	0.12 (0.32)	0.15 (0.35)	0.11 (0.32)	0.14 (0.35)
Black	0.11 (0.32)	0.09 (0.29)	0.13 (0.33)	0.11 (0.31)
Other non white	0.06 (0.24)	0.05 (90.21)	0.05 (0.21)	0.10 (0.30)
Caucasian	0.71 (N/A)	0.70 (N/A)	0.71 (N/A)	0.65 (N/A)
Subject is married	0.41 (0.49)	0.77 (0.42)	0.58 (0.49)	0.72 (0.45)
Subject lacks health insurance	0.18 (0.39)	0.18 (0.38)	0.13 (0.33)	0.15 (0.35)
PCS index: health	50.99 (9.62)	51.39 (8.69)	50.01 (10.10)	50.29 (9.38)
MCS index: mental health	51.22 (9.32)	52.19 (8.80)	49.60 (10.05)	50.48 (9.66)
Census region				
Northeast	0.18 (0.38)	0.19 (0.39)	0.19 (0.39)	0.19 (0.39)
Midwest	0.23 (0.42)	0.23 (0.42)	0.24 (0.42)	0.21 (0.40)
South	0.35 (0.48)	0.36 (0.48)	0.36 (0.48)	0.37 (0.48)
West	0.25 (0.43)	0.22 (0.41)	0.22 (0.41)	0.24 (0.43)
Subject lives in urban location	0.85 (0.36)	0.80 (0.40)	0.82 (0.39)	0.84 (0.36)

For women, we find no difference in the relationship between health status and labor force participation for self-reporters and proxy respondents. In contrast, we find a larger effect of mental health on labor force participation among proxy responders (32%) than among self-reporters (23%), a difference that is statistically significant at the 1% level.

4.2 Matching on health indices

The authors repeat the propensity score approach, now adding the health and mental health indices described above to our matching procedure. The results are summarized in Table 4. As the table indicates, the magnitudes of the treatment effects decline and in most case the differences between self-reporters and proxies decline as well. Most importantly, we now find that none of the differences between self-reporters and proxy respondents is statistically significant.

The results suggest that any differences between proxies and respondents in the effects of perceived health and mental health on labor force participation become

Table 3 Average treatment effect of health and mental health on labor force participation MEPS years 2000–2005[a]

NOT using health indices for PS matching	Female proxy	Female self	Male proxy	Male self
Mentally healthy as treatment				
Nearest neighbor	0.2715 (0.025461)	0.21917 (0.011162)	0.28773 (0.014286)	0.33373 (0.017722)
Radius	0.3091163 (0.0246366)	0.2359612 (0.0111091)	0.3098336 (0.0156345)	0.3586642 (0.0195694)
Kernel (Gaussian)	0.3446849 (0.0192412)	0.2874018 (0.0093878)	0.3595106 (0.0137895)	0.3877317 (0.0161532)
Stratification	0.32 (0.025)	0.23 (0.01)	0.31 (0.015)	0.356 (0.019)
Healthy as treatment				
Nearest neighbor	0.20431 (0.016)	0.20863 (0.008)	0.21728 (0.009)	0.3046 (0.011)
Radius	0.220769 (0.017)	0.218928 (0.008)	0.23255 (0.009)	0.307276 (0.012)
Kernel (Gaussian)	0.265207(0.013)	0.254268 (0.007)	0.260729 (0.008)	0.349654 (0.009)
Stratification	0.218 (0.017)	0.216 (0.007)	.232 (0.010)	0.311 (0.014)

[a] Standard errors are in parentheses. The standard error estimations for nearest neighbour matching have been corrected using algorithm in Abadie and Imbens (2006)

Table 4 Average treatment effect of health and mental health on labor force participation MEPS years 2000–2005[a]

Using health indices for p score matching with pcs42 and mcs42	Female proxy	Female self	Male proxy	Male self
Mentally healthy as treatment				
Nearest neighbor	0.1451 (0.046)	0.10008 (0.023)	0.16157 (0.026)	0.12312 (0.035)
Radius	0.2520168 (0.022)	0.1037337 (0.019)	0.1264216 (0.019)	0.1428332 (0.025)
Kernel (Gaussian)	0.2595518 (0.023)	0.1603171 (0.013)	0.223597 (0.017)	0.2035948 (0.020)
Stratification	0.163 (0.033)	0.108 (0.020)	0.129 (0.015)	0.140 (0.036)
Healthy as treatment				
Nearest neighbor	0.060165 (0.032)	0.075328 (0.019)	0.0778 (0.015)	0.078986 (0.023)
Radius	0.042057 (0.020)	0.053391 (0.017)	0.066348 (0.010)	0.081493 (0.018)
Kernel (Gaussian)	0.111988 (0.018)	0.097212 (0.009)	0.084742 (0.008)	0.099639 (0.012)
Stratification	0.045 (0.023)	0.058 (0.015)	0.067 (0.015)	0.086 (0.022)

[a] Standard errors are in parentheses

insignificant once subjects are matched on more objective health and mental health indices. This, in turn, implies that the apparent reporting biases in the estimates that did not control for these health indices in fact reflect actual differences in health.

4.3 Limitations

Our empirical results are not meant to be given a structural interpretation. They are meant to provide researchers with descriptive information about the propensity scoring methods used in assessing associations between labor force participation and, respectively, health, and mental health status. Clearly, the subjective health and mental health measures may be endogenous, and our estimates which did not control for objective measures of health and mental health status may suffer from endogeneity bias. One may be more confident in the estimates that adjust for the objective health and mental health indices, but it is possible that this matching strategy, too, has missed an important yet unobservable factor affecting perceived health. The results must be viewed with these limitations in mind.

5 Conclusion

This article relates subjective measures of health and mental health status to labor force participation and compares these relationships among self-reporters and proxy respondents. The authors used propensity scoring methods to help control for confounding factors.

The authors do not find significant differences in the associations between these measures of health and labor force participation, after treatments and controls have been matched on more objective measures of health and mental health. The results of this study suggest that the labor force participation costs of poor health may not be seriously overstated when based on estimates using self-reports of health. Such measures, which are much more readily available in health economic databases than detailed objective health indices, thus may be useful in understanding the labor market consequences of health and mental health. From a policy perspective, estimates of health and labor force participation based on subjective health measures do not appear to significantly overstate employer labor force participation costs associated with poor health in the work force due to response bias.

As Frölich (2007) and DiNardo et al. (1996) have demonstrated, propensity score methods are quite useful in ascertaining relationships when confounding factors are present. Nevertheless, this technique is only as reliable as the specification of the matching variables. As Frölich (2007) demonstrated, failure to include a single explanatory variable in his matching model greatly increased the percentage of gender earnings attributable to unobserved factors. This omission would lead to the conclusion that discrimination—a popular unobservable thought to affect gender differences in wages—was more important in accounting for wages differences than may in fact be the case. In this study, failure to condition on objective measures of health leads to differences in the effects of health status on labor force participation between self- and proxy responders, and the suspicion that self-reporting bias may be at work. However, after one properly conditions on objective health status measures, these differences vanish. Nevertheless, these results must be viewed with caution, because, although we have estimated associations between health status measures and labor force participation, our analysis does not demonstrate causal relationships.

Acknowledgments The authors would like to thank Ronghua Guo, Rice University, for her crucial and substantive research assistance. The usual caveat applies.

References

Abadie A, Imbens G (2006) Large sample properties of matching estimators for average treatment effects. Econometrica 74:235–267

Becker SO, Ichino A (2002) Estimation of average treatment effects based on propensity scores. Stata J, StataCorp LP 2(4):358–377

Blinder A (1973) Wage discrimination, reduced form, and structural estimates. J Hum Resour 8:436–455

Bound J (1991) Self-reported versus objective measures of health in retirement models. J Hum Resour 26(1):106–138

Cohen J, Monheit A, Beauregard K et al (1996) The Medical Expenditure Panel Survey: a national health information resource. Inquiry 33:373–389

DiNardo J, Fortin N, Lemieux T (1996) Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach. Econometrica 64:1001–1044

Frölich M (2007) Propensity score matching without conditional independence assumption—with an application to the gender wage gap in the United Kingdom. Econom J 10:359–407

Hernan MA, Brumback B, Robins J (2001) Marginal structural models to estimate the joint causal effect of nonrandomized treatments. J Am Stat Assoc 96:440–447

Imbens GW (2000) The role of propensity score in estimating dose response functions. Biometrika 87: 706–710

Jenkinson C, Layte R, Jenkinson D et al (1997) A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies? J Public Health Med 19:179–286

Kulik L (2000a) Jobless men and women: a comparative analysis of job search intensity, attitudes toward unemployment, and related responses. J Occup Organ Psychol 73:487–500

Kulik L (2000b) The impact of gender and age on attitudes and reactions to unemployment: the Israeli case. Sex Roles 43:85–104

Leuven E, Sianesi B (2003) PSMATCH2: Stata module to perform mahalanobis and propensity score matching, commons support graphing, and covariate imbalance testing. Statistical Software Components S432001, Boston College Department of Economics. Accessed at http://ideas.repec.org/c/boc/bocode/s432001.html

Medical Expenditure Panel Survey (MEPS) (2006) Household Component Full Year Files (2000–2005). Agency for Healthcare Research and Quality, Rockville

Robins JM (1999) Marginal structural models versus structural nested models as tools for causal inference. In: Halloran ME, Berry D (eds) Statistical models in epidemiology: the environment and clinical trials. Springer-Verlag, New York, pp 95–134

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55

Ware J, Kosinski M, Keller SD (1994) SF-36 physical and mental health summary scales: a user manual. New England Medical Center, Boston