Health Serv Outcomes Res Method (2004) 5: 175–191
DOI 10.1007/s10742-006-6827-y

Sociodemographic correlates of psychiatric diseases: accounting for misclassification in survey diagnoses of major depression, alcohol and drug use disorders

Elizabeth Savoca

Received: 13 January 2005 / Revised: 1 September 2005 / Accepted: 12 October 2005

Abstract This paper illustrates how validation data can be used to correct for errors in survey indicators of psychiatric disorders in models where the outcome of interest is the probability of a positive diagnosis. Nonlinear models of the risks associated with a broad range of sociodemographic factors for three disorders (major depression, alcohol and drug use disorders) are estimated with adjustments for classification errors in the survey diagnoses. Estimates show that inferences drawn from the unadjusted models may seriously understate gender and regional differences in the prevalence rates of all three disorders, the effects of education and ethnicity on the development of alcohol use disorders, and the relationship between marital status and the risk of major depression.

Keywords Misclassification bias · Risk factors · Depression · Substance use disorders · National Comorbidity Survey

1. Introduction

Statisticians from the medical and social sciences have developed a variety of techniques for correcting for biases resulting from misclassification in categorical variables in survey data (Kuha et al. 1998; Hausman et al. 1998; Hausman et al. 2001). The most straightforward approach uses validation studies, which match the survey data to more accurate records, to infer the probabilities of errors in the survey responses. For example, Espeland and Hui (1987) and Chen (1989) examine the effect of seat-belt use on the occurrence of car accident injuries by using hospital records to correct for errors in data taken from police reports. Greenland (1988) uses medical records to correct survey data on women's antibiotic use during pregnancy in a study of the association between drug use and sudden-infant-death syndrome. Labor economists have frequently followed this approach using administrative and employer records to account for errors in household survey responses to questions about union

E. Savoca (✉)
Department of Economics, Smith College,
Northampton, MA 01063
e-mail: esavoca@email.smith.edu

status, industry of employment, and participation in transfer income programs (Freeman 1984; Card 1996; Krueger and Summers 1988; Bollinger and David 1997).

For the most part, health services researchers have ignored the issue of misclassification, even though at least one large-scale ongoing population-based survey, the National Health Interview Survey, periodically provides data from medical records to validate household responses to questions about the presence or absence of a wide range of physical diseases, data which imply fairly high false positive and false negative rates in the survey responses (United States Department of Health and Human Services 1994). Scholars who study questions related to mental health have largely avoided the issue of misclassification by relying on specialized population based surveys that are well known for the sophisticated design of their diagnostic survey instruments, and by assuming that the simulated psychiatric diagnoses recorded in the surveys are reasonably accurate. Validation studies involving direct clinical appraisals of the survey respondents' mental health, however, have found large discrepancies between the survey outcomes and the psychiatrists' assessments (Anthony et al. 1985; Helzer et al. 1985; Kessler et al. 1998).

Mental health researchers, who have been sensitive to this issue, have focused exclusively on correcting for measurement error bias in linear models where the survey indicators of health conditions serve as control variables rather than as outcome variables (Frank and Gertler 1991; Savoca 2000). A well-known finding in the statistics and econometrics literature is that random errors in regressors, whether continuous or categorical, always lead to biased coefficients. Only recently have social scientists recognized that the theoretical implications of measuring a continuous outcome variable with error do not extend to the case where the mismeasured dependent variable is dichotomous (Poterba and Summers 1995; Bollinger and David 1997; Hausman et al. 1998; Hausman et al. 2001; Dustmann and van Soest 2001). The classical case for continuous dependent variables argues that measurement error introduces a loss of efficiency but not bias in regression coefficients. In contrast, even purely random measurement error in a categorical outcome variable biases the coefficients on regressors in both linear and nonlinear models.

This paper illustrates how validation data can be used to correct for errors in survey indicators of psychiatric disorders in models where the outcome of interest is the probability of a positive diagnosis. Models of the risks associated with a broad range of sociodemographic factors for three psychiatric disorders are estimated with adjustments for classification errors in the survey diagnoses. The primary data for this study is the National Comorbidity Survey (NCS), a survey carried out in the early 1990s, which randomly sampled adults in the general U.S. population and administered to each survey participant the Composite International Diagnostic Interview (CIDI) (World Health Organization 1990). Responses to the CIDI allowed the survey to determine clinical diagnoses of several psychiatric disorders according to professional diagnostic criteria (American Psychiatric Association 1987; World Health Organization 1991a, b). In the first stage of estimation, summary statistics from a validation study of the CIDI are used to estimate its false positive and false negative rates. These estimated error rates are then used to transform nonlinear least squares estimates into consistent estimates of the effects of the risk factors on the probability of developing three debilitating diseases: major depression, alcohol and drug abuse or dependence.

The paper is organized into five sections. Section 2 reviews the theoretical results for probability models when the binary outcome variable is measured with error. This section shows that misclassification in a binary outcome variable will bias the predicted differences in the probability of a positive outcome between two groups towards zero and the odds ratio towards one. Section 3 describes the primary data set and derives the relevant error rates in

its survey diagnoses. Section 4 presents estimates of the logit probability model with and without adjustments for classification errors. The estimates are relatively easy to compute and demonstrate a significant quantitative difference between the adjusted and unadjusted estimated risk factor effects on the probabilities of developing clinical diagnoses. The paper concludes with a discussion of the implications of these results and suggestions for future work in this area.

2. A probability model with misclassified outcomes

It is well understood in the social sciences that the regression coefficient of an imperfectly measured binary covariate can be biased toward zero in a simple linear regression framework. Aigner (1973), one of the first econometricians to publish on this topic, is widely cited in the economics literature, even in some introductory econometrics textbooks (See Judge et al. 1985). A lesser-known theoretical result is that error rates in binary outcome variables may also lead to a systematic downward bias in the estimated impact of an error-free covariate on the true underlying binary outcome. This section demonstrates that misclassification in a binary outcome variable will bias the predicted differences in the probability of a positive outcome between two groups towards zero and the odds ratio towards one. It draws on the epidemiology literature in particular, on the work of Diamond and Lilienfeld (1962) and Gullen et al. (1968).

Consider the bivariate framework where y, the true diagnosis is a function of a risk factor x, and Y is the observed diagnosis. Suppose that the errors in Y's classification of the population are nondifferential. That is, given the true diagnosis, the level of x provides no additional information about the observed disease classification: $\Pr(Y = 1|y, x) = \Pr(Y = 1|y)$. Also, suppose that the false positive and false negative rates in Y sum to less than one. This essentially amounts to the assumption that the number of persons correctly classified exceeds the number of persons incorrectly classified.[1]

The survey diagnosis, Y misclassifies each survey participant with these probabilities, $\Pr(Y = 1|y = 0) = r_0$ and $\Pr(Y = 0|y = 1) = r_1$. Therefore,

$$\begin{aligned}\Pr(Y = 1|x) &= \Pr(Y = 1|y = 1) \cdot \Pr(y = 1|x) + \Pr(Y = 1|y = 0) \cdot \Pr(y = 0|x) \quad (1)\\ &= (1 - r_1) \cdot \Pr(y = 1|x) + r_0 \cdot \Pr(y = 1|x)\\ &= r_0 + (1 - r_0 - r_1) \cdot \Pr(y = 1|x).\end{aligned}$$

Hence, the marginal effect of x on the probability of a positive survey diagnosis is a fraction of its effect on the probability of truly developing the disorder:

$$\Delta\Pr(Y = 1|x)/\Delta x = (1 - r_0 - r_1) \cdot (\Delta\Pr(y = 1|x)/\Delta x). \quad (2)$$

Misclassification in a response variable can be shown to bias odds ratios toward one. Suppose that we wish to compute the relative odds of receiving a positive diagnosis for two

[1] $\Pr(Y = 1|y = 0) + \Pr(Y = 0|y = 1) < 1$ implies that $\Pr(Y = 1|y = 0) < \Pr(Y = 1|y = 1)$ and that $\Pr(Y = 0|y = 1) < \Pr(Y = 0|y = 0)$. That is, the survey diagnosis must predict a higher (lower) probability of developing the disorder among true positives (negatives) than among true negatives (positives). If not then the observed marginal effect of x is opposite in sign from the true effect and the observed odds ratio flips to the opposite side of one.

groups that differ in their level of x (x_1 versus x_2). Let OR denote the true odds ratio based on the relationship between the y and x and OR^* denote the odds ratio based on the relationship between the survey diagnosis Y and x:

$$OR = \frac{\dfrac{\Pr(y=1|x=x_1)}{1-\Pr(y=1|x=x_1)}}{\dfrac{\Pr(y=1|x=x_2)}{1-\Pr(y=1|x=x_2)}} = \frac{\Pr(y=1|x=x_1)}{\Pr(y=1|x=x_2)} \cdot \frac{1-\Pr(y=1|x=x_2)}{1-\Pr(y=1|x=x_1)} \tag{3}$$

$$OR^* = \frac{\dfrac{\Pr(Y=1|x=x_1)}{1-\Pr(Y=1|x=x_1)}}{\dfrac{\Pr(Y=1|x=x_2)}{1-\Pr(Y=1|x=x_2)}} = \frac{\Pr(Y=1|x=x_1)}{\Pr(Y=1|x=x_2)} \cdot \frac{1-\Pr(Y=1|x=x_2)}{1-\Pr(Y=1|x=x_1)} \tag{4}$$

Recall from equation (1) that $\Pr(Y=1|x) = r_0 + (1-r_0-r_1)\cdot\Pr(y=1|x)$ and multiply the numerator and denominator in (3) by $(1-r_0-r_1)/(1-r_0-r_1)$. Then we can express OR, the true odds ratio, in terms of the observed probabilities:

$$\begin{aligned} OR &= \frac{\dfrac{(1-r_0-r_1)\Pr(y=1|x=x_1)}{(1-r_0-r_1)-(1-r_0-r_1)\Pr(y=1|x=x_1)}}{\dfrac{(1-r_0-r_1)\Pr(y=1|x=x_2)}{(1-r_0-r_1)-(1-r_0-r_1)\Pr(y=1|x=x_2)}} = \frac{\dfrac{\Pr(Y=1|x=x_1)-r_0}{1-\Pr(Y=1|x=x_1)-r_1}}{\dfrac{\Pr(Y=1|x=x_2)-r_0}{1-\Pr(Y=1|x=x_2)-r_1}} \\ &= \frac{\Pr(Y=1|x=x_1)-r_0}{\Pr(Y=1|x=x_2)-r_0} \cdot \frac{1-\Pr(Y=1|x=x_2)-r_1}{1-\Pr(Y=1|x=x_1)-r_1} \end{aligned} \tag{5}$$

The expression for OR, the true odds ratio, in (5) exceeds the expression for the observed odds ratio, OR^* in (4) whenever $\dfrac{\Pr(Y=1|x=x_1)-r_0}{\Pr(Y=1|x=x_2)-r_0} > \dfrac{\Pr(Y=1|x=x_1)}{\Pr(Y=1|x=x_2)}$ and $\dfrac{1-\Pr(Y=1|x=x_2)-r_1}{1-\Pr(Y=1|x=x_1)-r_1} > \dfrac{1-\Pr(Y=1|x=x_2)}{1-\Pr(Y=1|x=x_1)}$.

Simple algebraic manipulation shows that these two conditions hold only when $\Pr(Y=1|x=x_1) > \Pr(Y=1|x=x_2)$, that is, only when the observed odds ratio, OR^*, is greater than one. Similarly, when OR^* is less than one, the true odds ratio (5) will be less than the observed ratio. Hence, we can conclude that misclassification in the survey diagnosis will bias the odds ratio toward one.[2]

3. The national comorbidity survey

Over the past two decades psychiatric epidemiology has made great strides in advancing our knowledge of the prevalence and correlates of psychiatric disorders in the general U. S.

[2] This conclusion relies on the conditions that $\Pr(Y=1|x) > r_o$ and that $1-\Pr(Y=1|x) > r_1$. Both conditions follow from the assumption of nondifferentiability and the assumption that $r_1 + r_0 < 1$. The first condition, $\Pr(Y=1|x) > \Pr(Y=1|y=0) = \Pr(Y=1|y=0,x)$, implies that, given x, the survey screening test has a higher probability of detecting the disorder in the total population than in the sub-population of healthy persons. The second condition, $\Pr(Y=0|x) > \Pr(Y=0|y=1) = \Pr(Y=0|y=1,x)$, implies that, given x, the survey screening instrument would record a higher disease-free prevalence rate in the total population than in the sub-population of persons who are genuinely sick.

population (Robins and Regier 1991; Kessler et al. 1994; Kessler et al. 2003). These advances are due to the availability of large-scale population-based surveys, the most influential of which is the National Comorbidity Survey (NCS), sponsored by the U. S. National Institute of Mental Health. This survey is well known for the design of its diagnostic survey instrument, the Composite Diagnostic Interview Schedule (CIDI), which enabled lay-interviewers to simulate clinical diagnoses of a wide range of mental disorders. The CIDI has been adapted for use in many countries with different diagnostic customs and methods. Indeed, it is the primary epidemiologic instrument upon which cross-national comparisons of disease prevalence and correlates are made (World Health Organization 2000, 2004).

Kessler et al. (1994) provide full details of the survey's design. In short, the NCS is a nationally representative sample of the noninstitutionalized U.S. civilian population who were between the ages of 15 and 54 in 1990-1992. The survey contains a household sample of 8098 respondents, living in the 48 contiguous states. This paper focuses on three of the disorders assessed in the study: alcohol use and dependency, drug use and dependency, and major depression. The survey records 6-month, 12-month and lifetime diagnoses for each of these disorders. This analysis is restricted to lifetime diagnoses, that is, a diagnosis for whether the respondent ever experienced the disorder, since they were the only diagnoses subject to a validation study.

3.1. Evaluating errors in the CIDI diagnoses

A validation study of the CIDI, involving direct clinical reappraisals of several diagnosis-specific subsamples of the NCS, was carried out between 13 and 33 months after the initial NCS interview. Published reports of the findings present summary statistics on the positive and negative predictive values of the CIDI for each of the 9 diseases included in the validation study (Kessler et al. 1998).

The clinicians involved in the validation study administered the Structured Clinical Interview (SCID), a standardized interview specifically designed for experienced professionals. Both the CIDI and SCID were structured to produce a wide range of psychiatric diagnoses consistent with widely accepted professional criteria established by the American Psychiatric Association (1997). The CIDI, however, was tailored for lay interviewers. The symptom questions rely solely on the subject's accounts. The interviewers are given no discretion in either the phrasing or interpretation of the questions and responses. Professionals using the SCID, however, are expected to use their best clinical judgment as to the clearest way to pose questions and to resolve inconsistencies in the subjects' responses.

Generally speaking, a straightforward application of Bayes Theorem can derive false positive and false negative rates from these predictive values. Using the notation in Section 2, let y represent the binary variable denoting the clinician's diagnosis (0 = negative, 1 = positive), and Y the binary variable denoting the CIDI diagnosis. Let PPV and NPV denote the positive and negative predictive values, respectively: $PPV = \Pr(y = 1|Y = 1)$ and $NPV = \Pr(y = 0|Y = 0)$.

The CIDI's false positive rate, r_0 can be written as:

$$r_0 = \Pr(Y = 1|y = 0) = \frac{\Pr(y = 0|Y = 1) \cdot \Pr(Y = 1)}{\Pr(y = 0|Y = 1) \cdot \Pr(Y = 1) + \Pr(y = 0|Y = 0) \cdot \Pr(Y = 0)}$$

$$= \frac{(1 - PPV) \cdot \Pr(Y = 1)}{(1 - PPV) \cdot \Pr(Y = 1) + NPV \cdot (1 - \Pr(Y = 1))}.$$

Table 1 Estimates of false positive and false negative rates, lifetime CIDI diagnoses (standard errors in parentheses)

	Major depression	Alcohol use disorders	Drug use disorders
False Positive Rate[1] (r_0)	0.078[a] (0.024)	0.034[c] (0.025)	0.027[b] (0.015)
False Negative Rate[1] (r_1)	0.387[b] (0.241)	0.454[a] (0.140)	0.726[a] (0.086)
True Positive Rate ($1 - r_1$)	0.613	0.546	0.274
κ value[2]	0.53[a] (0.14)	0.54[a] (0.09)	0.39[a] (0.12)
χ^2 statistic[2]	0.00	1.30	2.40

[1]Standard errors, in parentheses, are computed using the delta method (Green 1997).

[2]Reported in Kessler et al. (1998).

[a]p-value $\leq .01$ [b].01 < p-value $\leq .05$ [c].05 < p-value $\leq .10$.

Similarly, the CIDI's false negative rate, r_1 can be written as:

$$r_1 = \Pr(Y = 0|y = 1) = \frac{(1 - NPV) \cdot (1 - \Pr(Y = 1))}{(1 - NPV) \cdot (1 - \Pr(Y = 1)) + PPV \cdot \Pr(Y = 1)}.$$

Certain features of the validation study must be considered before calculating estimates of r_0 and r_1 that one could reasonably assume applied to the primary study population.[3] First, to insure a sufficient number of positive CIDI cases, the validation study oversampled persons who received positive CIDI diagnoses in the initial NCS interview. Disease prevalence according to the CIDI diagnoses in the validation study, therefore, would overestimate the probability of a positive CIDI diagnosis, $\Pr(Y = 1)$, in the general population. I use the CIDI disease prevalence rates in the primary data, the NCS, instead. Second, all of the negative cases selected for the validation study had some symptoms but not enough to meet the threshold for a positive CIDI diagnosis. Therefore, the *NPV* in the reported study may understate the *NPV* in the general population, since presumably fewer negative cases that reported no symptoms would be reclassified in the clinical reappraisal. Indeed, Kessler et al. (1998) report that in test-retest reliability studies it was rare for NCS respondents who initially reported no symptoms to receive a positive diagnosis in the retest. Hence, this analysis uses an adjusted estimate of the *NPV*, provided by Kessler et al. (1998), which assumes that none of the respondents in the NCS who presented no symptoms would have received a positive diagnosis in the clinical reappraisal had they been included in the validation study. The *PPV* of the CIDI from the validation study was assumed to be a reasonable estimate of the *PPV* in the primary study population.

Table 1 summarizes the error rate estimates. Assuming independence in the estimates of the *NPV*, *PPV*, and the CIDI disease prevalence in the primary study population, standard errors of the error rates were obtained via the delta method (Green 1997). The first row of numbers shows false-positive rates that are quite low. The CIDI rarely misclassifies a health person. On the other hand, the estimated false-negative rates, reported in the second row, are quite high. The CIDI fails to recognize drug use disorders in nearly three-fourths of the population afflicted with this disease. Alcohol use disorders go undetected in nearly half of the afflicted population. Major depression is missed in over one-third of true cases.

[3] See Wittchen et al. (1995) for a discussion of the rationale for the design of the validation study.

Despite the high false-negative rates the CIDI, nonetheless, satisfies some professional standards for a valid screening test.[4] A comparison of the third with the first row of numbers shows that for each of the disorders, the probability that the CIDI detects the disorder among the ill population $(1 - r_1)$ is much higher than the probability of a positive CIDI diagnosis in the healthy population (r_0). In the fourth row we see that for all three disorders, the κ values are positive and statistically significant at less than the 1% significance level, allowing us to reject the hypothesis that the observed positive associations between the CIDI diagnoses and the clinical diagnoses were due purely to chance.[5]

Furthermore, the presence of large error rates in a survey diagnosis does not necessarily imply a systematic bias in its estimate of true disease prevalence. The relationship in equation (1) also applies to the unconditional probabilities: $\Pr(Y = 1) = r_0 + (1 - r_0 - r_1) \cdot \Pr(y = 1)$. With some algebraic manipulation of terms we can derive the condition that $\Pr(Y = 1) = \Pr(y = 1)$ whenever $r_1 \cdot \Pr(y = 1) = r_0 \cdot (1 - \Pr(y = 1))$. For diseases like drug abuse, for which the probability of a positive diagnosis is believed to be relatively low, a high false negative rate (r_1) and a low false positive rate (r_0) could completely offset each other when computing the population prevalence estimate. Reported in the fifth row of Table 1 are the χ^2 statistics for testing the hypothesis that the prevalence estimates based on the CIDI are identical to the rates implied by the clinical appraisals. Indeed, we are unable to reject this null for all three disorders.

And, finally, these error rates are comparable to errors in surveys on other sensitive topics. For instance, recipients of public relief may be reluctant to admit accepting assistance, perhaps because of the social stigma attached to public dependence, or because they may have fraudulently established eligibility and wish to avoid detection. Marquis and Moore (1990) estimated false negative rates, in the *Survey of Income and Program Participation* as high as 49% for recipients of welfare and 44% for unemployment compensation beneficiaries. Here, too, the probability that a survey participant incorrectly reports receiving assistance is low, less than 5% (Bound et al. 2001).

4. Estimation of regression coefficients

Although statistical analysis of the validation data supports the conclusion that errors in the CIDI diagnoses do not lead to biased estimates of the prevalence rates for most of the diseases assessed in the NCS (Kessler et al. 1998), theory suggests that the high error rates reported in Table 1 may still lead to serious biases in measures of association. I assess the consequences of these errors for regression coefficients relating risk factors to psychiatric health outcomes in the following framework.

4.1. Population model

Consider y_i , the diagnosis that a randomly chosen person i, in a random sample of n, would receive from a medical professional. It takes on a value of 0 if negative and 1 if positive and is related to an observed vector of k sociodemographic characteristics $\mathbf{x_i} = (x_{1i}, x_{2i}, \ldots, x_{ki})$:

$$y_i = F(\mathbf{x_i}\boldsymbol{\beta}) + \varepsilon_i \qquad i = 1, \ldots, \mathrm{n} \tag{6}$$

[4] See Rogan and Gladen (1978) for a discussion of the various ways to judge the 'goodness' of a screening test.

[5] The κ values and χ^2 statistics reported in Table 1 are taken from Kessler et al. (1998). For an explanation of the statistical principles behind these evaluative measures, see Bishop et al. (1975).

where we assume that ε_i has a zero mean, is independent of $\mathbf{x}_i$ with $VAR(\varepsilon_i|\mathbf{x_i}) = F(\mathbf{x_i}\boldsymbol{\beta})/(1 - F(\mathbf{x_i}\boldsymbol{\beta}))$, and that $\boldsymbol{\beta}$ is a k dimensional parameter vector, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)$. Hence, $E(y_i|\mathbf{x_i}) = \Pr(y_i = 1|\mathbf{x_i}) = F(\mathbf{x_i}\boldsymbol{\beta})$.

4.2. Measurement model

The survey diagnosis, Y_i, classifies person i with these probabilities: $\Pr(Y_i = 1|y_i = 0) = r_0$; $\Pr(Y_i = 0|y_i = 1) = r_1$; $\Pr(Y_i = 1|y_i = 1) = 1 - r_1$; $\Pr(Y_i = 0|y_i = 0) = 1 - r_0$. It follows from equation (1) that $E(Y_i|\mathbf{x_i}) = \Pr(Y_i = 1|\mathbf{x_i}) = r_0 + (1 - r_0 - r_1)F(\mathbf{x_i}\boldsymbol{\beta})$. Consequently, Y_i can be expressed as a linear function of y_i:

$$Y_i = r_0 + (1 - r_0 - r_1)y_i + v_i \qquad i = 1, ..., \mathrm{n} \tag{7}$$

where $v_i = Y_i - E(Y_i|y_i)$, the expectation error, has zero mean and is independent of y_i.[6] The assumption of nondifferential error implies that v_i is also independent of $\mathbf{x_i}$.

4.3. Regression model

Substituting equation (6) into (7) and replacing the error rates in Y with the estimates, $\hat{r}_0$ and $\hat{r}_1$, derived in Section 3, provides us with an estimable regression model:

$$Y_i = \hat{r}_0 + (1 - \hat{r}_0 - \hat{r}_1)F(\mathrm{x_i}\boldsymbol{\beta}) + u_i \tag{8}$$

where $u_i = (1 - r_0 - r_1)\varepsilon_i + v_i + (r_0 - \hat{r}_0) + F(\mathbf{x_i}\boldsymbol{\beta})(\hat{r}_0 - r_0 + \hat{r}_1 - r_1)$.

Depending on the specification of $F(\mathbf{x_i}\boldsymbol{\beta})$, either linear or nonlinear least squares estimation provides a consistent estimate of $\boldsymbol{\beta}$. Since $VAR(u_i|\mathbf{x_i})$ is heteroscedastic, weighted least squares yields a more efficient estimator (Green 1997). Estimates of $VAR(u_i|\mathbf{x_i})$ are constructed from the unweighted least squares estimates of $\boldsymbol{\beta}$ along with $\hat{r}_0$ and $\hat{r}_1$ using the delta method. This variance estimator, derived in the Appendix, takes into account the sampling variability in the estimates of $r_0$and $r_1$as well as the dependency between the estimates of $r_0$and r_1 and the first stage estimator of $\boldsymbol{\beta}$. The function, $1/\sqrt{VAR(u_i|\mathbf{x_i})}$, provides the weights for the second and final set of regression estimates.

4.4. Logit specification

In the medical and social sciences, probit and logistic regressions, largely for ease of interpretation, are the most favored empirical specification for probabilities. In particular, a simple transformation of the logit coefficients allows one to interpret the estimates as odds ratios. The logit specifies $\Pr(y_i = 1|\mathbf{x_i}) = F(\mathbf{x_i}\boldsymbol{\beta}) = \left[1 + \exp(\mathbf{x_i}\boldsymbol{\beta})\right]^{-1}$. Suppose that one of the observable characteristics, say x_2, is a 0-1 dummy variable. Then the odds of receiving a positive diagnosis for a person for whom $x_2 = 1$ relative to a person for whom $x_2 = 0$ is: $\exp(\beta_2)$, everything else the same. For a continuous covariate, say x_3, the relative odds for a person whose level of $x_3 = x_3'$ versus a person whose level is x_3'' is: $\exp(\beta_3(x_3' - x_3''))$, everything else the same.

Tables 2 through 4 report the estimated odds ratios based on the logistic regression coefficient estimates, both unadjusted and adjusted for classification errors.[7] Ninety-five percent

[6] See Savoca (2000) for an explicit demonstration of this result.

[7] The logit coefficient estimates are available from the author, on request.

Table 2 Sociodemographic correlates of major depression, lifetime CIDI diagnosis

Variable	Odds ratio, W/O error correction (95% C. I.)		Odds ratio, with error correction (95% C. I.)	
Education				
0–11 years	0.8	(0.7, 1.0)	0.7	(0.5, 1.1)
12 years	0.9	(0.7, 1.0)	0.8	(0.6, 1.1)
13–15 years	1.1	(0.9, 1.3)	1.2	(0.8, 1.6)
16 years and above	1.0		1.0	
Employment Status				
Worker	0.7*	(0.5, 0.8)	0.4*	(0.3, 0.7)
Homemaker	0.7*	(0.5, 0.9)	0.5*	(0.3, 0.9)
Student	0.5*	(0.4, 0.7)	0.3*	(0.2, 0.5)
Other	1.0		1.0	
Marital Status				
Married	0.5*	(0.4, 0.5)	0.2*	(0.2, 0.3)
Never Married	0.7*	(0.5, 0.8)	0.5*	(0.3, 0.7)
Divorced/Separated/Widowed	1.0		1.0	
Family Income				
Below the poverty level	1.2	(0.9, 1.5)	1.3	(0.8, 2.1)
Between 1 and 3 times the poverty level	1.2	(1.0, 1.4)	1.3	(0.9, 1.8)
Between 3 and 6 times the poverty level	1.0	(0.8, 1.2)	0.9	(0.7, 1.3)
Over 6 times the poverty level	1.0		1.0	
Ethnicity				
Hispanic	1.0	(0.7, 1.7)	1.3	(0.6, 2.9)
Black	0.7	(0.5, 1.0)	0.5	(0.2, 1.1)
White	1.4	(1.0, 2.0)	1.9	(0.9, 3.9)
Other	1.0		1.0	
Region				
Midwest	0.9	(0.7, 1.0)	0.7	(0.5, 1.0)
Northeast	0.8*	(0.7, 1.0)	0.6*	(0.5, 0.9)
South	0.8*	(0.7, 0.9)	0.6*	(0.5, 0.9)
West	1.0		1.0	
Sex				
Female	1.8*	(1.6, 2.1)	3.1*	(2.4, 4.1)
Male	1.0		1.0	
Urbanicity				
Major metropolitan county	1.2*	(1.0, 1.4)	1.4	(1.0, 1.9)
Medium to small metropolitan county	1.2	(1.0, 1.4)	1.3	(1.0, 1.9)
Rural	1.0		1.0	
Age (in years)				
One Standard Deviation Above the Mean	1.0	(0.9, 1.2)	1.1	(0.8, 1.4)
One Standard Deviation Below the Mean	1.0		1.0	

Notes: Based on estimates from a logit model using nonlinear weighted least squares and heteroscedastic consistent standard errors using a method developed by White (1980).

*Significant at the 5% level.

Table 3 Sociodemographic correlates of drug use disorders, lifetime CIDI diagnosis

Variable	Odds ratio, W/O error correction (95% C. I.)		Odds ratio, with error correction (95% C. I.)	
Education				
0–11 years	1.4*	(1.0, 1.8)	1.0	(0.5, 1.9)
12 years	1.2	(1.0, 1.5)	1.5	(0.9, 2.4)
13–15 years	1.2	(1.0, 1.5)	1.5	(0.9, 2.5)
16 years and above	1.0		1.0	
Employment Status				
Worker	0.7*	(0.6, 1.0)	0.3*	(0.2, 0.7)
Homemaker	0.9	(0.6, 1.3)	0.6	(0.2, 1.4)
Student	0.2*	(0.1, 0.3)	0.0*	(0.0, 0.1)
Other	1.0		1.0	
Marital Status				
Married	0.7*	(0.5, 0.8)	0.4*	(0.2, 0.6)
Never Married	0.6*	(0.5, 0.8)	0.4*	(0.2, 0.8)
Divorced/Separated/Widowed	1.0		1.0	
Family Income				
Below the poverty level	1.0	(0.8, 1.4)	1.4	(0.7, 3.0)
Between 1 and 3 times the poverty level	1.0	(0.8, 1.3)	1.1	(0.7, 1.9)
Between 3 and 6 times the poverty level	0.8*	(0.6, 1.0)	0.6	(0.4, 1.0)
Over 6 times the poverty level	1.0		1.0	
Ethnicity				
Hispanic	1.2	(0.7, 2.1)	0.8	(0.3, 2.7)
Black	1.1	(0.6, 1.9)	0.4*	(0.1, 1.4)
White	2.3*	(1.4, 3.9)	2.8*	(1.0, 7.9)
Other	1.0		1.0	
Region				
Midwest	0.6*	(0.5, 0.7)	0.3*	(0.2, 0.5)
Northeast	0.7*	(0.6, 0.9)	0.4*	(0.2, 0.6)
South	0.6*	(0.5, 0.8)	0.3*	(0.2, 0.6)
West			1.0	
Sex				
Female	0.5*	(0.4, 0.6)	0.2*	(0.1, 0.3)
Male	1.0		1.0	
Urbanicity				
Major metropolitan county	1.8*	(1.5, 2.3)	3.7*	(2.2, 6.2)
Medium to small metropolitan county	1.6*	(1.2, 1.9)	2.5*	(1.5, 4.0)
Rural	1.0		1.0	
Age (in years)				
One Standard Deviation Above the Mean	0.5*	(0.2, 1.0)	0.1*	(0.1, 0.2)
One Standard Deviation Below the Mean	1.0		1.0	

Notes: See Table 2.

*Significant at the 5% level.

Table 4 Sociodemographic correlates of alcohol use disorders, lifetime CIDI diagnosis

Variable	Odds ratio, W/O error correction (95% C. I.)		Odds ratio, with error correction (95% C. I.)	
Education				
0–11 years	1.4*	(1.1, 1.7)	1.5*	(1.0, 2.1)
12 years	1.4*	(1.2, 1.7)	1.6*	(1.2, 2.1)
13–15 years	1.5*	(1.3, 1.8)	1.9*	(1.5, 2.6)
16 years and above	1.0		1.0	
Employment Status				
Worker	0.7*	(1.1, 1.7)	0.5*	(0.4, 0.8)
Homemaker	0.6*	(0.4, 0.8)	0.4*	(0.2, 0.6)
Student	0.2*	(0.1, 0.3)	0.1*	(0.0, 0.1)
Other	1.0		1.0	
Marital Status				
Married	0.6*	(0.5, 0.7)	0.5*	(0.4, 0.7)
Never Married	0.6*	(0.5, 0.8)	0.5*	(0.4, 0.7)
Divorced/Separated/Widowed	1.0		1.0	
Family Income				
Below the poverty level	1.1	(0.9, 1.4)	1.2	(0.8, 1.8)
Between 1 and 3 times the poverty level	1.0	(0.8, 1.2)	1.0	(0.8, 1.4)
Between 3 and 6 times the poverty level	0.9	(0.8, 1.1)	0.9	(0.7, 1.2)
Over 6 times the poverty level	1.0		1.0	
Ethnicity				
Hispanic	0.8*	(0.7, 0.9)	1.2	(0.6, 2.2)
Black	0.7*	(0.6, 0.9)	0.5*	(0.2, 0.9)
White	0.6*	(0.5, 0.7)	2.8*	(1.6, 4.9)
Other	1.0		1.0	
Region				
Midwest	1.0	(0.9, 1.2)	0.6*	(0.4, 0.8)
Northeast	1.1	(0.9, 1.3)	0.5*	(0.4, 0.7)
South	1.0		0.4*	(0.3, 0.5)
West			1.0	
Sex				
Female	0.4*	(0.3, 0.4)	0.2*	(0.1, 0.2)
Male	1.0		1.0	
Urbanicity				
Major metropolitan county	1.0	(0.9, 1.2)	1.1	(0.9, 1.5)
Medium to small metropolitan county	1.1	(0.9, 1.3)	1.3	(1.0, 1.7)
Rural			1.0	
Age (in years)				
One Standard Deviation Above the Mean	0.6*	(0.5, 0.7)	0.4*	(0.3, 0.5)
One Standard Deviation Below the Mean	1.0		1.0	

Notes: See Table 2.

*Significant at the 5% level.

confidence intervals for the odds ratios are calculated according to: $\exp(\hat{\beta} \pm 1.96 \cdot SE(\hat{\beta}))$ for the coefficients on dummy explanatory variables and $\exp(\hat{\beta}(x' - x'') \pm 1.96 \cdot (x' - x'') \cdot SE(\hat{\beta}))$ for the coefficient on a continuous variable where x' and x'' represent two different comparative values of the continuous covariate.

For both of the substance use disorders, the estimated odds ratios for females versus males are cut roughly in half when the logit estimates are corrected for classification errors, implying that these disorders are, relatively speaking, much less prevalent in the female population than the naïve estimates would indicate. On the other hand, the naïve estimates significantly understate women's risk of depression; the estimated odds ratio rises from 1.8 to 3.1 after adjusting for misclassification in the CIDI depression diagnosis. For each of the three disorders striking differences are present in the unadjusted versus adjusted estimates of regional differences in prevalence rates. The estimated relationships between urbanicity, employment status and the risk of drug use disorders, between ethnicity and the risk of alcohol use disorders and between marital status and the risk of major depression are also seriously understated by the uncorrected estimates.

5. Summary and conclusions

General population-based surveys of mental health are critical for policy planning, particularly since psychiatric diseases often go unnoticed by physicians and surgeons in the general medical sector and since a relatively small fraction of the at-risk population seeks treatment in the specialty mental health sector (Saravay et al 1991; Robins and Regier 1991). Recognizing this, in the 1980s and 1990s, the U. S. National Institute of Mental Health (NIMH) sponsored three large-scale surveys of the general population in order to estimate community-based prevalence rates and treatment needs and to identify high-risk groups. Because the cost of conducting clinical evaluations of each survey participant would be prohibitive, NIMH also funded the development of structured diagnostic interviews that could be administered by lay interviewers with no more than one week of training. A computer algorithm would then process the responses to the interview questions to arrive at psychiatric diagnoses that followed professional criteria. To assess the reliability of these diagnoses, for each of these surveys, validation studies, which subjected subsamples of the survey participants to direct psychiatric appraisals by experienced clinicians, were undertaken.

This paper is the first to analyze the consequences of errors in the survey diagnoses for logit estimators of the sociodemographic correlates of psychiatric diagnoses. This work focuses on the structured diagnostic interview included in the National Comorbidity Survey, the second of the NIMH sponsored surveys and the first nationally representative survey of the mental health of general U. S. population. Estimates of the sociodemographic correlates of positive lifetime diagnoses of depression and substance use disorders show that inferences drawn from logit models that do not account for misclassification in the diagnosis may be seriously biased.

The relevance of this paper's findings for health services research extends beyond studies of mental health outcomes. In a comprehensive review of the measurement error literature, Bound et al. (2001) discuss several analyses of misclassification in other categorical health-related variables that are of keen interest to policy makers. They note studies which have found that respondents' knowledge of detailed features of their health insurance coverage, including whether or not the plan covers mental health treatment, is extremely limited, and studies of the agreement between household reports of the presence of chronic medical

conditions and medical records which have found false positive and false negative rates as high as 40 percent.

An important limitation of this study is one that secondary analysts often face. Presumably, for reasons of confidentiality, individual level data from the validation study are not publicly available. Consequently, the published summary statistics from the study dictate the structure of the measurement model. In particular, the measurement error is assumed to be nondifferential; it is identical for all individuals regardless of differences in health-related characteristics like gender, ethnicity, and severity of illness. Relaxation of these assumptions may change the misclassification bias in regression coefficients in an unpredictable fashion (Carroll et al. 1995).

For this application, however, even if the validation data were publicly available, the small sample size of the study would most likely limit the complexity of any alternative measurement models one may wish to consider. The results of this paper clearly call for further exploration into the consequences of survey errors in psychiatric diagnoses. The development of larger scale validation studies would be a good place to start.

Appendix. Variance formulas

A1. False positive (r_0) and false negative (r_1) rates

Consider equations (A1) and (A2), the formulas for the estimators of r_0 and r_1.

$$\hat{r}_0 = \frac{(1 - P\hat{P}V) \cdot \hat{P}}{(1 - P\hat{P}V) \cdot \hat{P} + N\hat{P}V \cdot (1 - \hat{P})} \tag{A1}$$

$$\hat{r}_1 = \frac{(1 - N\hat{P}V) \cdot (1 - \hat{P})}{(1 - N\hat{P}V) \cdot (1 - \hat{P}) + P\hat{P}V \cdot \hat{P}} \tag{A2}$$

where $P\hat{P}V$ and $N\hat{P}V$ are the estimators for the positive and negative predictive values of the CIDI, taken from the validation study, and $\hat{P}$ is the estimator of the population disease prevalence rate implied by the CIDI (P), calculated from the primary data set (NCS).

The variances and covariance formulas for $\hat{r}_0$ and $\hat{r}_1$ are derived from the delta method. The estimators $P\hat{P}V$ and $N\hat{P}V$ are independent since they were derived from two independent subsamples of the validity study, participants with positive CIDI diagnoses and negative CIDI diagnoses, respectively. Since the validation sample is a subsample of the primary survey, $P\hat{P}V$ and $N\hat{P}V$ should each be correlated with $\hat{P}$. Sample estimation of these correlations is not possible because the individuals who participated in the reassessment were not matched to the primary data file. The validation samples, however, are small relative to the primary sample, so that the covariance of the estimators from the two samples should be negligible in size. Because these covariances terms enter the variance formulas additively, they are assumed to contribute little to the variance calculations and are, consequently set to zero.

$$VAR(\hat{r}_0) = VAR(P\hat{P}V) \cdot \left(\frac{\partial \hat{r}_0}{\partial P\hat{P}V}\right)^2 + VAR(N\hat{P}V) \cdot \left(\frac{\partial \hat{r}_0}{\partial N\hat{P}V}\right)^2 + VAR(\hat{P}) \cdot \left(\frac{\partial \hat{r}_0}{\partial \hat{P}}\right)^2 \tag{A3}$$

$$VAR(\hat{r}_1) = VAR(P\hat{P}V) \cdot \left(\frac{\partial \hat{r}_1}{\partial P\hat{P}V}\right)^2 + VAR(N\hat{P}V) \cdot \left(\frac{\partial \hat{r}_1}{\partial N\hat{P}V}\right)^2 + VAR(\hat{P}) \cdot \left(\frac{\partial \hat{r}_1}{\partial \hat{P}}\right)^2 \tag{A4}$$

$$\begin{aligned} COV(\hat{r}_0, \hat{r}_1) = {} & VAR(P\hat{P}V) \cdot \frac{\partial \hat{r}_0}{\partial P\hat{P}V} \cdot \frac{\partial \hat{r}_1}{\partial P\hat{P}V} + VAR(N\hat{P}V) \cdot \frac{\partial \hat{r}_0}{\partial N\hat{P}V} \cdot \frac{\partial \hat{r}_1}{\partial N\hat{P}V} \\ & + VAR(\hat{P}) \cdot \frac{\partial \hat{r}_0}{\partial \hat{P}} \cdot \frac{\partial \hat{r}_1}{\partial \hat{P}} \end{aligned} \tag{A5}$$

where $VAR(P\hat{P}V) = PPV \cdot (1 - PPV)/\mathrm{N}_{PPV}$, $VAR(N\hat{P}V) = NPV \cdot (1 - NPV)/\mathrm{N}_{NPV}$, $VAR(\hat{P}) = P \cdot (1 - P)/\mathrm{N}_P$, and where N_{PPV} is the number of CIDI cases reassessed by professional experts (20 each for major depression and alcohol use disorders, 25 for drug use disorders), N_{NPV} is the number of noncases reassessed (10 cases each for major depression and alcohol use disorders, 6 for drug use disorders), and N_P is the number of respondents in the NCS (8098).

Evaluated at the estimated values of $P\hat{P}V$, $N\hat{P}V$, and $\hat{P}$, equations (A3), (A4), (A5) provide consistent estimates of the asymptotic covariance matrix of $\hat{r}_0$ and $\hat{r}_1$.

A2. Variance of the regression error: $VAR(u_i|\mathbf{x_i})$

Recall the expression for the regression error in equation (8):

$$u_i = (1 - r_0 - r_1)\varepsilon_i + v_i + (r_0 - \hat{r}_0) + F(\mathbf{x_i}\boldsymbol{\beta})(r_0 - \hat{r}_0 + r_1 - \hat{r}_1) \tag{A6}$$

Since the expectation error in the measurement model, v_i, is independent of y_i, it follows that $COV(\varepsilon_i, v_i) = 0$. The variance of the regression error, then, is:

$$\begin{aligned} VAR(u_i|\mathbf{x_i}) = {} & (1 - r_0 - r_1)^2 VAR(\varepsilon_i) + 2(1 - r_0 - r_1)(1 + F(\mathbf{x_i}\boldsymbol{\beta}))COV(\varepsilon_i, \hat{r}_0) \\ & + 2(1 - r_0 - r_1)F(\mathbf{x_i}\boldsymbol{\beta})COV(\varepsilon_i, \hat{r}_1|\mathbf{x_i}) + VAR(v_i|\mathbf{x_i}) \\ & + 2(1 + F(\mathbf{x_i}\boldsymbol{\beta}))COV(v_i, \hat{r}_0|\mathbf{x_i}) + 2F(\mathbf{x_i}\boldsymbol{\beta})COV(v_i, \hat{r}_1|\mathbf{x_i}) \\ & + (1 + F(\mathbf{x_i}\boldsymbol{\beta}))^2 VAR(\hat{r}_0) + 2F(\mathbf{x_i}\boldsymbol{\beta})(1 + F(\mathbf{x_i}\boldsymbol{\beta}))COV(\hat{r}_0, \hat{r}_1) \\ & + (F(\mathbf{x_i}\boldsymbol{\beta}))^2 VAR(\hat{r}_1) \end{aligned} \tag{A7}$$

The equations for $VAR(\hat{r}_0)$, $VAR(\hat{r}_1)$, and $COV(\hat{r}_0, \hat{r}_1)$are given in (A3), (A4), and (A5), $VAR(\varepsilon_i|\mathbf{x_i}) = F(\mathbf{x_i}\boldsymbol{\beta})(1 - F(\mathbf{x_i}\boldsymbol{\beta}))$, and $VAR(v_i|\mathbf{x_i}) = r_0(1 - r_0)(1 - F(\mathbf{x_i}\boldsymbol{\beta})) + r_1(1 - r_1)F(\mathbf{x_i}\boldsymbol{\beta})$.

The estimators, $\hat{r}_0$ and $\hat{r}_1$ are nonlinear functions of $P\hat{P}V$ and $N\hat{P}V$, statistics taken from the validation study, which are treated as independent of the primary data source, the NCS; and of $\hat{P} = \sum_i Y_i/\mathrm{N}_P$, the estimator for the population disease prevalence rate P, which is based on the disease prevalence in the NCS. This latter relationship introduces a correlation between $\hat{r}_0$ and $\hat{r}_1$ the population regression error ε_i and the measurement error, v_i.

A first order Taylor series expansion of equations (A1) and (A2) around *PPV*, *NPV*, and P, allows us to approximate the expressions for $\hat{r}_0$ and $\hat{r}_1$ with

$$\hat{r}_0' = r_0 + \left(\frac{\partial \hat{r}_0}{\partial PPV}\right)^2 (P\hat{P}V - PPV) + \left(\frac{\partial \hat{r}_0}{\partial NPV}\right)^2 (N\hat{P}V - NPV) + \left(\frac{\partial \hat{r}_0}{\partial P}\right)^2 (\hat{P} - P) \tag{A8}$$

$$\hat{r}_1' = r_1 + \left(\frac{\partial \hat{r}_1}{\partial PPV}\right)^2 (P\hat{P}V - PPV) + \left(\frac{\partial \hat{r}_1}{\partial NPV}\right)^2 (N\hat{P}V - NPV) + \left(\frac{\partial \hat{r}_1}{\partial P}\right)^2 (\hat{P} - P) \tag{A9}$$

Substituting (A8) and (A9) for $\hat{r}_0$ and $\hat{r}_1$, and setting the covariances of ε_i and v_i with $P\hat{P}V$ and $N\hat{P}V$ to zero, gives us these approximation formulas for the remaining covariance terms in the expression for $VAR(u_i|\mathbf{x_i})$in (A7).

$$\begin{aligned} COV(\varepsilon_i, \hat{r}_0|\mathbf{x_i}) \approx COV(\varepsilon_i, \hat{r}_0'|\mathbf{x_i}) &= \left(\frac{\partial r_0}{\partial P}\right)^2 COV(\varepsilon_i, \hat{P}_i) \\ &= \left(\frac{\partial r_0}{\partial P}\right)^2 COV(\varepsilon_i, (1 - r_0 - r_1)(\sum_i \varepsilon_i/\mathrm{N}_P)) \\ &= \left(\frac{\partial r_0}{\partial P}\right)^2 ((1 - r_0 - r_1)VAR(\varepsilon_i)/\mathrm{N}_P)) \end{aligned} \tag{A10}$$

Similarly,

$$COV(\varepsilon_i, \hat{r}_1|\mathbf{x_i}) \approx COV(\varepsilon_i, \hat{r}_1'|\mathbf{x_i}) = \left(\frac{\partial r_1}{\partial P}\right)^2 ((1 - r_0 - r_1)VAR(\varepsilon_i)/\mathrm{N}_P)) \tag{A11}$$

$$\begin{aligned} COV(v_i, \hat{r}_0|\mathbf{x_i}) \approx COV(v_i, \hat{r}_0'|\mathbf{x_i}) &= \left(\frac{\partial r_0}{\partial P}\right)^2 COV(v_i, \hat{P}_i) \\ &= \left(\frac{\partial r_0}{\partial P}\right)^2 COV(v_i, (\sum_i v_i/\mathrm{N}_P)) \\ &= \left(\frac{\partial r_0}{\partial P}\right)^2 (VAR(v_i)/\mathrm{N}_P) \end{aligned} \tag{A12}$$

and, by analogy,

$$COV(v_i, \hat{r}_1|\mathbf{x_i}) \approx COV(v_i, \hat{r}_1'|\mathbf{x_i}) = \left(\frac{\partial r_1}{\partial P}\right)^2 (VAR(v_i)/\mathrm{N}_P) \tag{A13}$$

Equations (A7) through (A13) can be evaluated at $\hat{r}_0$, $\hat{r}_1$, and $\hat{\boldsymbol{\beta}}$, the unweighted regression estimate, to form a consistent estimate of $VAR(u_i|\mathbf{x_i})$.

Acknowledgements Financial support for this research came from the U.S. National Institute of Mental Health Grant MH067598-01.

References

Aigner, D.: Regression with a binary independent variable subject to errors of observation. Journal of Econometrics **1**, 49–60 (1973).

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd edition, Revised. American Psychiatric Association, Washington D.C. (1987).

Anthony, J.C., Folstein, M., Romanoski, A.J., Von Korff, M.R., Nestadt, G. R., Chahal, R., Merchant, A., Brown, H., Shapiro, S., Kramer, M., Gruenberg, E.M.: Comparisons of the DIS and a standardized psychiatric diagnosis. Archives of General Psychiatry **42**, 667–675 (1985).

Bishop, Y.M.M., Fienberg, S.E., Holland, P.W.: Discrete multivariate analysis: Theory and practice. The MIT Press, Cambridge, MA (1975).

Bollinger, C.R., David, M.H.: Modeling discrete choice with response error: food stamp participation. Journal of the American Statistical Association **92**(439), 827–835 (1997).

Bound, J., Brown, C., Mathiowetz, N.: Measurement error in survey data. In Heckman, J.J., and Leamer, E., (eds.), Handbook of econometrics, vol. 5, Elsevier Science B.V., Amsterdam (2001).

Card, D.: The effects of unions on the structure of wages: a longitudinal analysis. Econometrica. **64**(4), 957–979 (1996).

Carroll, R.J, Ruppert, D., Stefanski, L.A.: Measurement errors in nonlinear models. Chapman and Hall, London, U.K. (1995).

Chen, T.T.: A review of methods for misclassified categorical data in epidemiology. Statistics in Medicine **8**, 1095–1106 (1989).

Diamond, E.L., Lilienfeld, A.M.: Effects of errors in classification and diagnosis in various types of epidemiological studies. American Journal of Public Health **52**, 1137–1144 (1962).

Dustmann, C., van Soest, A.: Language fluency and earnings: estimation with misclassified language indicators. The Review of Economics and Statistics **83**(4), 663–674 (2001).

Espeland, M., Hui, S.L.: A general approach to analyzing epidemiologic data that contain misclassification errors. Biometrics. **43**(4), 1001–1012 (1987).

Freeman, R.B.: Longitudinal analysis of the effects of trade unions. Journal of Labor Economics **2**, 1–26 (1984).

Frank, R., Gertler, P.: Assessment of measurement error bias for estimating the effect of mental distress on income. Journal of Human Resources **26**(1), 154–164 (1991).

Green, W.H.: Econometric analysis. Prentice Hall, Saddle River, NJ (1997).

Greenland, S.: Variance estimation for epidemiologic effect estimates under misclassification. Statistics in Medicine **7**, 745–757 (1988).

Gullen, W.H., Bearman, J.E., Johnson, E.A.: Effects of misclassification in epidemiologic studies. Public Health Reports **83**, 914–918 (1968).

Hausman, J.A.: Abrevaya, J., Scott-Morton, F.M.: Misclassification of the dependent variable in a discrete response setting. Journal of Econometrics **87**, 239–269 (1998).

Hausman, J.A.: Mismeasured variables in econometric analysis: problems from the right and problems from the left. Journal of Economic Perspectives **15**(4), 57–68 (2001).

Helzer, J., Robin, L.N., McEvoy, L.T., Spitnagel, E.L., Stoltzman, R.K., Farmer A., Brockington, I.F.: A comparison of clinical and diagnostic interview schedule diagnoses. Archives of General Psychiatry **42**, 657–666 (1985).

Judge, G.G., Griffiths, W.E., Hill, R.C., Lürkepohl, H., Lee, T.: The theory and practice of econometrics. 2nd edition. John Wiley and Sons, New York (1985).

Kessler, R.C., McGonagle, K.A., Zhao, S., Nelson, C.B., Hughes, M., Eshleman, S., Wittchen, H., Kendler, K.S.: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the U.S. Archives of General Psychiatry **51**, 8–19 (1994).

Kessler, R.C., Berglund, P., Demler, O., Jim, R., Koretz, D., Merikangas, K.R., Rush, A.J., Walters, E.E., Wang, P.S.: The epidemiology of major depressive disorder. The Journal of the American Medical Association **289**(33), 3095–3105 (2003).

Kessler, R.C., Wittchen, H., Abelson, McGonagle, J.M., Schwarz, N., Kendler, K.S., Knauper, B., Zhao, S.: Methodological studies of the composite international diagnostic interview (CIDI) in the U.S. National Comorbidity Survey (NCS). International Journal of Methods in Psychiatric Research **7**(1), 33–55 (1998).

Krueger, A., Summers, L.: Efficiency wages and inter-industry wage structure. Econometrica **56**(2), 259–293 (1988).

Kuha, J., Skinner, C., Palmgren, J.: Misclassification error. In Armitage, P. and Colton, T. (eds.), Encyclopedia of biostatistics, Volume 4. Wiley, New York (1998).

Marquis, K., Moore, J.: Measurement errors in sipp program reports. In: Proceedings of the bureau of the census 1990 annual research conference. Bureau of the Census, Washington D.C. (1990).
Poterba, J.M., Summers, L.: Unemployment benefits and labor market transitions: A multinomial logit model with errors in classification. The Review of Economics and Statistics **77**(2), 207–216 (1995).
Robins, L., Regier, D.: Psychiatric disorders in america. Free Press, New York (1991).
Rogan, W.J., Gladen, B.: Estimating prevalence from the results of a screening test. American Journal of Epidemiology **107**(1), 71–76 (1978).
Saravey, S.M., Steinberg, M.D., Weinschel, B., Pollack, S., Alovis, N. Psychological comorbidity and length of stay in the general hospital. American Journal of Psychiatry **148**, 324–329 (1991).
Savoca, E.: Measurement errors in binary regressors: An application to measuring the effects of specific psychiatric diseases on earnings. Health Services and Outcomes Research Methodology **1**(2), 149–164 (2000).
United states department of health and human services. Vital health and statistics: Evaluation of national health interview survey diagnostic reporting: Series 2: Data Evaluation and Methods Research, No. 120. DHHS, Hyattesville, MD (1994).
White, H.: A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica **48**, 817–838 (1980).
Wittchen, H., Kessler, R.C., Zhao, S., Abelson, J.: Reliability and clinical validity of UM-CIDI DSM III-R generalized anxiety disorder. Journal of Psychiatric Research **29**, 95–110 (1995).
World health organization. Composite international diagnostic interview (CIDI). version 1.0. World Health Organization, Geneva, Switzerland (1990).
World health organization. ICD-10, chapter V, mental and behavioral disorders. Diagnostic criteria for research, draft for field trials. World Health Organization, Geneva, Switzerland (1991a).
World health organization. Tenth revision of the international classification of diseases chapter V(F): Mental and behavioral disorders. clinical descriptions and diagnostic guidelines. World Health Organization, Division of Mental Health, Geneva, Switzerland (1991b).
World health organization international consortium in psychiatric epidemiology. Cross-national comparisons of the prevalences and correlates of mental disorders. Bulletin of the World Health Organization **78**, 413–426 (2000).
World health organization world mental health survey consortium. Prevalence, severity, and unmet need for treatment of mental disorders in the world health organization world mental health surveys. The Journal of the American Medical Association **291**(21), 2581–2590 (2004).