



Published in final edited form as:

Commun Stat Theory Methods. 2016 ; 45(9): 2538–2555. doi:10.1080/03610926.2014.887105.

SYSTEMATICALLY MISCLASSIFIED BINARY DEPENDENT VARIABLES

VIDHURA TENNEKOON^a and ROBERT ROSENMAN^b

^aDepartment of Economics, 308 Cate Center Drive, University of Oklahoma, Norman, OK 73019, USA. vtennekoon@ou.edu. 1-405-325-3614. Fax 1-405-325-5842. Corresponding author

^bSchool of Economic Sciences, Washington State University, Pullman, WA 99164, USA. yamaka@wsu.edu. 1-509-335-1193. Fax 1-509-335-1173

Abstract

When a binary dependent variable is misclassified, that is, recorded in the category other than where it really belongs, probit and logit estimates are biased and inconsistent. In some cases the probability of misclassification may vary systematically with covariates, and thus be endogenous. In this paper we develop an estimation approach that corrects for endogenous misclassification, validate our approach using a simulation study, and apply it to the analysis of a treatment program designed to improve family dynamics. Our results show that endogenous misclassification could lead to potentially incorrect conclusions unless corrected using an appropriate technique.

Keywords

binary choice model; misclassification; measurement error; response shift bias; Likert scales

1. INTRODUCTION

Misclassification of a dichotomous categorical variable means that an observation with a true value of '0' is observed as '1' or an observation that is truly a '1' is observed as a '0'. When the misclassified variable is the dependent variable, probit or logit estimates may lead to biased and inconsistent estimates if the misclassification is ignored or modeled incorrectly (Hausman, 2001).

Misclassification of a variable can happen for various reasons, although one can categorize them broadly into two groups; response errors that are random in nature, and those that vary systematically with some respondent characteristic. The case we explore here is the latter, when the probability of misclassification is observation specific and dependent on covariates.

The source of the error sometimes offers some insight into whether possible misclassification is systematic or not. In labor market data, for example, some respondents may misreport their employment status or a correctly reported labor status may be

mistranscribed (Chua and Fuller, 1987; Poterba and Summers, 1995). If the data is misreported because of language difficulties or a lack of understanding, the probability of misclassification could vary systematically with education or primary language. Moreover, while we might think mistranscribed data is a random event, if the mistranscriptions are due to transcriber quality and transcribers are correlated with location, then misclassification probabilities could vary systematically with location.

Previous work on misclassified dependent variables has taken two paths. The first approach uses supplemental data to verify the accuracy of responses. Chua and Fuller (1987) develop a parametric model that incorporated all $J(J-1)$ misclassification (possibilities of an outcome variable with J categories, but their approach requires a minimum of three independent sets of survey responses obtained by re-interviewing the original respondents, and has a limited practical use. The conditional logit procedure proposed in Poterba and Summers (1995) also incorporates all possibilities of misclassification and requires misclassification probabilities found by analyzing the discrepancies between interview and re-interview outcomes.

An alternative path builds the probability of misclassification into the estimation procedure, allowing for errors in the data, and using statistical methods to correct for it. Hausman et al. (1998) and Abrevaya and Hausman (1999) suggest both parametric and semi-parametric approaches for misclassification probabilities that cannot be independently verified and are independent of covariates. The focus of their parametric model is a dichotomous outcome variable with two types of misclassification, which they denote as α_0 (the probability that a true 0 is recorded as a 1) and α_1 (the probability that a true 1 being recorded as 0)¹. With their parametric approach the unknown misclassification probabilities are estimated simultaneously with the usual coefficients of the binary choice model. Their semi-parametric method provides consistent estimates of the model parameters, but not of the misclassification probabilities. Dustmann and van Soest (2004) extend the parametric model of Hausman et al. (1998) to a trichotomous case.

Lewbel (2000) allows the misclassification probabilities to be covariate-dependent functions and shows that (given some regularity) the binary choice model with covariate-dependent misclassification is completely identified even when the functional forms of α_0 , α_1 and the distribution of the error term are unknown. However, he also acknowledges that his estimator is ‘not likely to be very practical since they involve up to third order derivatives and repeated applications of nonparametric regression’ (pp. 607-608). The lack of any empirical work exploiting his estimator indicates the need for a more practical estimator in the case of covariate dependent misclassification, even at the cost of some additional assumptions.

Our paper extends the parametric approach of Hausman et al. (1998) to the case, where the misclassification probabilities are functions of one or more covariates². The parametric estimator that we propose is a more tractable way to identify a model similar to Lewbel

¹Throughout this paper we use the same notation.

²Although Hausman, et al. (1998) briefly discusses a limited extension of systematic misclassification in section 5.5, they do not fully characterize or implement the approach. A semi-parametric approach to deal with covariate-dependent misclassification of the

(2000), but is conditional on functional form assumptions. The paper proceeds as follows. In section 2, we present our structural approach to deal with covariate-dependent misclassification of the dependent variable and the identification requirements. Section 3 has a Monte Carlo experiment that compares our approach with the ordinary probit model and the basic model presented in Hausman et al. (1998). In section 4, we present an empirical application to demonstrate the applicability of the model. Finally, in section 5 we discuss implications and conclusions from our generalization.

2. THE GENERALIZED MODEL TO CORRECT FOR COVARIATE-DEPENDENT MISCLASSIFICATION

Assume, y_i^* is an unobserved latent variable such that

$$y_i^* = X_i \beta + \varepsilon_i \quad (1)$$

where, X_i is a vector of observed independent variables, β is a vector of coefficients to be estimated and ε_i is an iid error term with a known common distribution. We observe

$$y_i = 1 (y_i^* \geq 0). \quad (2)$$

If no misclassification is present, we always observe the dichotomous outcome variable, y_i , correctly. However, if there is misclassification, the outcome variable that we observe, y_i^o , includes some true '1's classified as '0's and some true '0's classified as '1's. As a result, in general, $y_i^o \neq y_i$. Accordingly, the binary variable we observe (y_i^o) also includes an additional measurement error ζ_i such that

$$y_i^o = y_i + \zeta_i \quad \text{where} \quad \zeta_i = \begin{cases} -1 & \text{with probability } \alpha_{1i} \text{ if } y_i = 1 \\ 1 & \text{with probability } \alpha_{0i} \text{ if } y_i = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

In other words,

$$\alpha_{0i} = Pr (y_i^o = 1 | y_i = 0) \quad (4)$$

and

$$\alpha_{1i} = Pr (y_i^o = 0 | y_i = 1). \quad (5)$$

dependent variable is discussed in detail in Abrevaya and Hausman (1999). Our interest is in the parametric model and in methods that provide misclassification probabilities.

The addition of ζ_j not only increases the variance of the econometric error term, but also adds heteroskedastity in a specific way. The overall stochastic mechanism that determines the values ultimately observed with random misclassification is a conditional Bernoulli process that can be characterized via the following data generating process.

$$y_i^0 = 1 (y_i^* \geq 0) 1 (u_i \leq 1 - \alpha_{1i}) + 1 (y_i^* < 0) 1 (u_i \leq \alpha_{0i}), \text{ where } u_i \sim \text{Uniform}(0, 1), \forall i \quad (6)$$

We assume u_j and ε_j in (1) are independent. If the values of α_{0j} and α_{1j} are dependent on γ_j as in (4) and (5), but independent of X_j , and the probability distribution function of ε_j is $F(\cdot)$, then, as Hausman, et al. (1998) show, we can express the expected value of the observed dependent variable as

$$E(y_i^0 | X_i) = Pr(y_i^0 = 1 | X_i) = \alpha_{0i} + (1 - \alpha_{0i} - \alpha_{1i}) F(X_i \beta). \quad (7)$$

When α_{0j} and α_{1j} are constants (α_0 and α_1) and $\alpha_0 + \alpha_1 < 1^3$, the parameters of the above model can be consistently estimated either by MLE or NLLS⁴.

Suppose instead that the misclassification probabilities α_{0j} and α_{1j} are functions of a set of variables, Z_i^0 and Z_i^1 respectively as in Lewbel (2000). In particular, the probabilities in (4) and (5) are now given by

$$\alpha_{0i}(Z_i^0) = Pr(y_i^0 = 1 | y_i = 0, Z_i^0) = F_0(Z_i^0 \gamma_0) \quad (8)$$

$$\alpha_{1i}(Z_i^1) = Pr(y_i^0 = 0 | y_i = 1, Z_i^1) = F_1(Z_i^1 \gamma_1) \quad (9)$$

where Z_i^0 and Z_i^1 may be but are not necessarily subsets⁵ of X_i , and F_0 and F_1 are the cumulative distribution functions of stochastic components that determine each type of misclassification.⁶ Inserting the preceding generalized representation of the misclassification probabilities into (7), the expected value of the observed dependent variable with a covariate-dependent misclassification can be expressed as

$$E(y_i^0 | X_i, Z_i^0, Z_i^1) = Pr(y_i^0 = 1 | X_i, Z_i^0, Z_i^1) \\ = F_0(Z_i^0 \gamma_0) + (1 - F_0(Z_i^0 \gamma_0) - F_1(Z_i^1 \gamma_1)) F(X_i \beta) \quad (10)$$

³This condition, termed the “monotonicity condition” in Hausman et al. (1998) must be satisfied to identify $(\beta, \alpha_0, \alpha_1)$, separately from $(-\beta, -\alpha_0, -\alpha_1)$.

⁴The relevant objective functions are given by equations (6) and (7) in Hausman et al. (1998).

⁵This allows one or both misclassification probabilities to depend on variables that do not affect the outcome.

⁶A generalization of the model could include a correlated error structure between the error terms of the latent variable equations.

If each of the vectors Z_i^0 and Z_i^1 include only a constant term, we have

$\alpha_{0i} = F_0(Z_i^0 \gamma_0) = F_0(\gamma_0) = \alpha_0$ and $\alpha_{1i} = F_1(Z_i^1 \gamma_1) = F_1(\gamma_1) = \alpha_1$ equivalent to α_0 and α_1 in Hausman et al. (1998). Accordingly, equation (10) nests the basic parametric model presented in Hausman et al. (1998), hereafter referred to as HAS1, allowing a statistically testable proposition.⁷

Assuming the functional forms of F_0 , F_1 and F are known the parameters of the model can be estimated with NLLS by minimizing

$$f(\beta, \gamma_0, \gamma_1) = \sum_{i=1}^n \left(y_i^0 - F_0(Z_i^0 \gamma_0) - (1 - F_0(Z_i^0 \gamma_0) - F_1(Z_i^1 \gamma_1)) F(X_i \beta) \right)^2 \quad (11)$$

over $(\beta, \gamma_0, \gamma_1)$. Alternatively, MLE can be applied to the following log likelihood function:

$$l(\beta, \gamma_0, \gamma_1) = n^{-1} \sum_{i=1}^n \left(\begin{array}{l} y_i^0 \ln [F_0(Z_i^0 \gamma_0) + (1 - F_0(Z_i^0 \gamma_0) - F_1(Z_i^1 \gamma_1)) F(X_i \beta)] + \\ (1 - y_i^0) \ln [1 - F_0(Z_i^0 \gamma_0) - (1 - F_0(Z_i^0 \gamma_0) - F_1(Z_i^1 \gamma_1)) F(X_i \beta)] \end{array} \right) \quad (12)$$

In the Monte-Carlo simulations and the application to real data that we present in subsequent sections, all the parameter estimations are based on MLE using equation (12) and also approximate all three functions F_0 , F_1 and F above by a normal CDF.

As we explained earlier, HAS1 is a special case of our generalization, which we refer to hereafter as GHAS, without any covariates affecting each type of misclassification probabilities. The generalization of the Hausman et al. (1998) data generating process in (5) that applies to the GHAS specification is given by

$$y_i^0 = 1 (y_i^* \geq 0) 1(u_i \leq 1 - F_1(Z_i^1 \gamma_1)) + 1(y_i^* < 0) 1(u_i \leq F_0(Z_i^0 \gamma_0)), \quad (13)$$

where $u_i \sim \text{Uniform}(0, 1), \forall i$

and again ε_i in (1) and u_i are independent. The nesting of HAS1 in GHAS and of the standard binary choice model in HAS1 facilitates statistical testing for the most suitable model in a given application. The significance tests for parameters in Z_i^0 and Z_i^1 other than the constant terms serve as tests for the suitability of GHAS over HAS1. Given that no elements of Z_i^0 and Z_i^1 pass this threshold, one may estimate HAS1 and the significant tests of the terms α_0 and α_1 serve as tests for the suitability of HAS1 model over the standard binary choice model⁸.

⁷If $F_0(Z_i^0 \gamma_0) = F_1(Z_i^1 \gamma_1) \equiv 0$ (10) further collapses to a standard binary choice specification. However, as discussed in footnote 9, it is not possible to directly test for this condition.

Identification of the parameters of (12) stems from the non-linearity of F . The first order necessary conditions and the Fisher information matrix of (12) can be expressed as below.

$$\mathbf{g} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \beta} \\ \frac{\partial \mathcal{L}}{\partial \gamma_0} \\ \frac{\partial \mathcal{L}}{\partial \gamma_1} \end{bmatrix} = n^{-1} \sum_{i=1}^n \left(y_i^0 / P_i^{-1-y_i^0} / (1 - P_i) \right) C_i = 0 \quad (14)$$

$$\mathbf{I} = -E \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \beta'} & \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \gamma_0'} & \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \gamma_1'} \\ \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \gamma_0'} & \frac{\partial^2 \mathcal{L}}{\partial \gamma_0 \partial \gamma_0'} & \frac{\partial^2 \mathcal{L}}{\partial \gamma_0 \partial \gamma_1'} \\ \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \gamma_1'} & \frac{\partial^2 \mathcal{L}}{\partial \gamma_0 \partial \gamma_1'} & \frac{\partial^2 \mathcal{L}}{\partial \gamma_1 \partial \gamma_1'} \end{bmatrix} = n^{-1} \sum_{i=1}^n (1/P_i (1 - P_i)) C_i C_i' \quad (15)$$

where $P_i = F_0(Z_i^0 \gamma_0) + (1 - F_0(Z_i^0 \gamma_0) - F_1(Z_i^1 \gamma_1)) F(X_i \beta)$,

$$C_i = \begin{bmatrix} (1 - F_0(Z_i^0 \gamma_0) - F_1(Z_i^1 \gamma_1)) F'(X_i \beta) X_i \\ (1 - F(X_i \beta)) F_0'(Z_i^0 \gamma_0) Z_i^0 \\ -F(X_i \beta) F_1'(Z_i^1 \gamma_1) Z_i^1 \end{bmatrix} \text{ and } F_0' \text{ and } F_1' \text{ are the first derivatives of } F_0, F_1 \text{ and } F_1 \text{ respectively.}$$

When F_0 and F_1 are symmetric and $F_0 = F_1$ identification requires that Z_i^0 be different from Z_i^1 . To demonstrate this point consider the case where $Z_i^0 = Z_i^1 = Z_i$, $F_1(v) = F_0(v) = 1 - F_0(-v)$ and $F(v) = 1 - F(-v)$. Then the log-likelihoods, $l(\beta, \gamma_0, \gamma_1) = l(-\beta, -\gamma_0, -\gamma_1)$. Hence, to identify $(\beta, \gamma_0, \gamma_1)$ from $(-\beta, -\gamma_0, -\gamma_1)$ we need $Z_i^0 \neq Z_i^1$. However, if F is asymmetric or $F_1 \neq F_0$, we do not necessarily require this exclusion restriction to identify the model parameters.

A merit of our estimator, however, is that the identification does not require

$X_i \neq Z_i^0$ or $X_i \neq Z_i^1$ when $Z_i^0 \neq Z_i^1$ and $F_0(\bullet), F_1(\bullet)$ and $F(\bullet)$ are non-linear transformations. Additional exclusive restrictions will help strong identification of parameters but are not necessary. Moreover, if $Z_i^0 \neq Z_i^1$ we no longer need $\alpha_0 + \alpha_1 < 0$ as Hausman et al. (1998) requires. In spite of these advantages our estimator has certain limitations too. The Hausman et al. estimator allows the misclassification probabilities to be zero (but not 1, since that would violate the monotonicity condition). If $F_0'(F_0^{-1}(0)) = F_0'(F_0^{-1}(1)) = 0$ and/or $F_0'(F_0^{-1}(0)) = F_0'(F_0^{-1}(1)) = 0$ as is the case with most functional forms⁹, ours require each type of misclassification probability to be bounded between 0 and 1, and not at the possible extremes, because if either of the two types of misclassifications takes an extreme value, the matrix $C_i C_i'$ becomes singular. A

⁸As noted above, the standard probit model, in general, is not nested in GHAS in a directly testable manner and thus we propose this sequential procedure. As the misclassification probability, α_k for $k = 0, 1$ reaches 0, $Z_i^k \gamma_k$ approaches the lower bound of $F(\bullet)$, which is $-\infty$ in case of a normal distribution, potentially leading to convergence issues. As such, convergence issues of GHAS may indicate a misspecified model and that HAS1 could be a more appropriate choice.

⁹A notable exception is the uniform distribution function.

related consequence would be large standard errors when misclassification probabilities are too small or too large. In contrast to HAS1, our estimator performs best when the misclassification probabilities are large in both directions. If the misclassification is known to be one-sided we can use a restricted version of our model by imposing $F_0(\bullet) \equiv 0$ or $F_1(\bullet) \equiv 0$ as appropriate, circumventing this identification issue while improving the efficiency of the estimator.

3. MONTE CARLO EXPERIMENT

In order to assess the impact of covariate-dependent misclassification on estimates with and without an appropriate correction mechanism we set up a Monte Carlo experiment which mimics the experiment used in Hausman et al. (1998). We first generated the X matrix in equation (1) including three random variables and a constant as covariates. Our X matrix is identical to the one they use in section 4 of their paper and comprises of X_1 , drawn from a lognormal distribution X_2 , a dummy variable equal to one, with probability 1/3, X_3 , a uniform (0,1) random variable, and a constant. The ε econometric error term, was drawn from a standard normal distribution. The parameter vector β also is identical to theirs. Based on this data generation process, the latent dependent variable is given by,

$$y^* = X\beta + \varepsilon \quad \text{where} \quad X = \begin{bmatrix} 1 & x_1 & x_2 & x_3 \end{bmatrix}, \beta = \begin{bmatrix} -1 & 0.2 & 1.5 & 0.6 \end{bmatrix}' \quad (16)$$

In our experiment the two types of misclassification probabilities are functions of subsets of X . More specifically, we have designed our experiment such that, the covariates in equations (7) and (8) are given by $Z_0 = \begin{bmatrix} 1 & x_3 \end{bmatrix}$ and $Z_1 = \begin{bmatrix} 1 & x_2 & x_3 \end{bmatrix}$. Denoting $\gamma_0 = \begin{bmatrix} \gamma_{00} & \gamma_{01} \end{bmatrix}$ and $\gamma_1 = \begin{bmatrix} \gamma_{10} & \gamma_{11} & \gamma_{12} \end{bmatrix}$ given the distribution of Z_0 and Z_1 , expected values of α_{0i} and α_{1i} in equations (8) and (9) are, respectively,

$$E(\alpha_{0i}) = E[\Phi(Z_0\gamma_0)] = \frac{1}{3\gamma_{01}} \int_{\gamma_{00}}^{\gamma_{00}+\gamma_{01}} \Phi(\theta) d\theta \quad (17)$$

$$E(\alpha_{1i}) = E[\Phi(Z_1\gamma_1)] = \frac{1}{3\gamma_{12}} \int_{\gamma_{10}+\gamma_{11}}^{\gamma_{10}+\gamma_{11}+\gamma_{12}} \Phi(\theta) d\theta + \frac{2}{3\gamma_{12}} \int_{\gamma_{10}}^{\gamma_{10}+\gamma_{12}} \Phi(\theta) d\theta \quad (18)$$

where $\Phi(\theta)$ denotes the normal distribution function. For consistency and comparison with the Hausman et al. (1998) experiment, we first chose the parameter vectors γ_0 and γ_1 by numerically integrating (16) and (17) using Gauss-Legendre quadrature such that $E(\alpha_{0i}) = E(\alpha_{1i}) \approx 0.05$. We also ran experiments with two additional symmetric expected values, 0.1 and 0.2, and two asymmetric and larger misclassification probabilities, $E(\alpha_{0i}, \alpha_{1i}) = (0.3, 0.75)$ and $(0.75, 0.3)$. The observed dependent variable, y^0 was generated by adding misclassification according to equation (13).

For each set of parameters, we generated a random sample, and used that sample to estimate the model parameters using, (i) the standard probit model (Probit); (ii) HAS1; and (iii) GHAS. The results are based on 200 Monte Carlo runs, each with a random sample of 5000 observations, for each of the sets of parameter values described in the preceding paragraph. The standard errors reported are the standard deviations of each set of 200 estimates.

Our findings with regard to probit estimates, shown in table 1, though based on a different data generating process, are broadly in line with the findings of Hausman et al., (section 4): (i) Even in the case of a small amount of misclassification, ordinary probit produces estimates are biased by 15-25%; (ii) The problem worsens as the amount of misclassification grows; (iii) Not only does probit yield inconsistent estimates, but it can also overstate the precision of the estimates. Our results show that the three observations are valid, not only for the case with random misclassification, but also for the more general case with covariate-dependent misclassification. The problems with the ordinary probit model in the presence of a misclassified dependent variable, whether random or covariate-dependent, are not small sample problems and thus cannot be overcome by increasing the sample size. As the sample size increases, $\Phi(Z_0Z_1)$ and $\Phi(Z_1Z_1)$ approaches their expected values $E(\alpha_{0j})$ and $E(\alpha_{1j})$. The consistency of the ordinary probit estimator requires

$$\hat{\beta}_{MLE} [E(\alpha_{0i}), E(\alpha_{1i})] = \hat{\beta}_{MLE} [0, 0] \text{ which is not the usual case.}$$

The overstated precision of estimates, together with a significant bias of estimates is a more severe issue than having the biased estimates alone. Even when the misclassification probabilities are 5%, ordinary probit estimates are at least two standard deviations away from the true values, and any statistically significant estimates are but a mere illusion due to the false precision, possibly leading a researcher towards incorrect conclusions. The problem worsens as the misclassification probabilities increase.

Despite not being the correct model, one may expect HAS1 to perform better than the ordinary probit model in the presence of covariate-dependent misclassification. As the result show, there is no guarantee that HAS1 will perform better, even though it may partially correct the bias under certain conditions. More specifically, when the misclassification probabilities are small and only depend on one or few covariates which are independent of the covariates of the main equation, HAS1 is a better alternative than the conventional probit model. In real applications, however, misclassification probabilities may be large and may depend on a large number of covariates; hence the random component of misclassification may be much smaller relative to the systematic component. Under such conditions HAS1 may increase the bias while also reducing the efficiency and thus may not be a better option than ordinary probit.

Mroz and Zayats (2008) argue that one must be cautious of the scale invariability when comparing the coefficient estimates of different specifications of a multilevel binary choice model. They note (page 409), "The coefficients might differ only because one is implicitly conditioning on information sets that differ by the inclusion of additional, independent information." and the 'relative effects' or the coefficient ratios could be a better measure of comparison. The idea is that the ratio cancels out the common scale factor. Although the magnitude differences are less severe, the superiority of GHAS prevails even when

evaluating coefficient ratios. For example, when the symmetric misclassification probabilities are 0.1, the average ratio of the estimated β_1/β_2 in the Monte Carlo with probit was 0.101, compared to 0.111 for HAS1 and 0.136 for GHAS, while the true value of the ratio is 0.133. This pattern persists over for all parameters and all misclassification probabilities¹⁰.

As our experimental results show, the superiority of GHAS over HAS1 and ordinary probit becomes more apparent both with the increased misclassification probabilities and with the increased heterogeneity of the distribution of misclassification probabilities. This holds when the probability of misclassification is symmetric or asymmetric. We intentionally used the two last sets of parameters to show a potential outcome of probit estimates when the misclassification probabilities (in expectation) are so high that the HAS1 monotonicity condition is not satisfied. Misclassification probabilities of these magnitudes are not always unrealistic. In fact, we may not abandon a project due to large covariate dependent misclassification probabilities, particularly when one or both the misclassification probabilities are very large with a specific sub group, but small with others. If we ignore misclassification and use probit estimates, as the results show, the co-efficient estimates are not only biased downward but also may show up with their signs toggled. As the probit estimates have no correspondence to the underlying data generation process when there are large misclassification probabilities it could lead to one or more of the following consequences.

- i. Downward biased estimates with the same sign and with reduced statistical significance;
- ii. The coefficient of an important variable may appear to be insignificant;
- iii. The bias could be sufficiently large to flip the sign of the estimate;
- iv. An insignificant variable may appear to be significant if it affects misclassification probabilities; and/or
- v. The estimates may show an impact larger than the true impact.

The HAS1 model should not be employed when the misclassification probabilities are large. When mean misclassification probabilities sum to a value greater than 1, violating HAS1 monotonicity requirement, as shown in tables 2 and 3, HAS1 in general predicts very low or zero misclassification. In addition, HAS1 coefficient estimates are not qualitatively different from the biased probit estimates. Typically the magnitudes of the misclassification probabilities are not known; using HAS1 when the means of the misclassification probabilities are large and systematic may mislead a researcher into believing misclassification is not a problem.

In addition to its superiority over other models in precisely estimating the coefficients of the main equation, GHAS also helps to correctly and precisely estimate the impact of each covariate on the two type of misclassification. As the results reported in tables 2 and 3 indicate, to precisely estimate the parameters of equations (8) and (9) when the

¹⁰These results are available from the authors.

misclassification probabilities are small, we need a sizable sample. However, when the misclassification probabilities are large, those parameters can be estimated with a high precision even with a relatively small sample.

As a final check, we tested what damage is done if we use GHAS when the probabilities of misclassification are not covariate dependent, so HAS1 would be more appropriate. Not surprisingly HAS1 is more efficient than GHAS. However, using GHAS when the misclassification is not covariate dependent does little harm. These results are reported in the appendix (Tables A1-A3).

4. APPLICATION TO ESTIMATE THE EFFECIVENESS OF A FAMILY IMPROVEMENT PROGRAM

We demonstrate the applicability of GHAS by using it to estimate the determinants of improvement in family functioning after participating in the Strengthening Families Program for Parents and Youth 10-14 (SFP) in Washington State and Oregon. For comparison we estimate the same model using HAS1 and ordinary probit. The Strengthening Families Program (SFP) is an internationally recognized parenting and family strengthening program for high-risk families. The program is designed to be delivered in local communities for groups of 7-12 families¹¹. Families attend SFP once a week for seven weeks and participate in educational activities that bring parents and their children together in learning environments designed to strengthen entire families through improved family communication, parenting practices, and parents' family management skills¹².

4.1 Applicability of the Model

The dependent variable of our application is a binary indicator equal to 1 if a participant's self-reported family functionality after the program is higher than the pre-program functionality. This indicator variable is derived using the pre-treatment and post-treatment scores measured on a Likert scale. One fundamental assumption that we make here is that there are true (latent) objective scores before and after the treatment, but neither the researcher nor the respondent observes these true values. Each participant makes a subjective assessment of her score and then translates it into an integer value within the range of the Likert scale used by the researcher. Response bias is the difference between the subjective measures of same objective outcome used by different individuals, while response *shift* bias comes from the response bias of the same individual changing at two measurement points (Sprangers and Hoogstraten, 1989; Hill and Betz, 2005).

Our study is essentially a "before-after" comparison at the surface. However, under certain assumptions the comparison is analogous to a true treatment effect. The family functionality of a household, the target of the intervention that we discuss here, in general is a slowly-changing variable and highly unlikely to change autonomously within a 7-week period, the duration of the intervention. This assumption leads two more results. First, any change in the

¹¹The two variables within-program average and within-program standard deviation we discuss later and use in our estimations are based these groups.

¹²<http://sfp.wsu.edu>

family functionality of a participant's household is due to the program effect since the impact of any other potential factors is negligible. Second, the family functionality of non-participants does not change during this short period. The two results together imply that the "before-after" comparison is a good practical measure of treatment effect in our case.

Our concern here is the misclassification of the indicator variable of improvement that we derive. Suppose both pre-treatment and post-treatment scores reported by each participant include response bias. If the magnitude of the bias remains unchanged after the program the reported scores show the true change. The issue we face here is that the intervention not only changes the family functionality, but also the knowledge about what good functionality is. As a result, participants may recalibrate their metrics used to measure and report family functionality after the program. As an example, suppose a participant reports her pre-treatment score is 2. After the program her family functionality has not changed but due to recalibrating her metric she realizes that her initial score should be 3, which she reports as her post-treatment score, seeing no improvement in her family functionality from the program. A researcher now observes an improvement while she really has not improved, contributing to misclassification probability α_{0j} . Suppose another participant reports her pre-treatment score as 4 but after recalibrating the metric she finds that her true score before the program should have only been 3 and now it has improved to 4. The researcher observed no improvement while she has really improved and we have misclassification type α_{1j} . Rosenman et al. (2011) has shown substantial response bias and response shift bias in SFP data.

In addition to the misclassification in our binary variable due to response shift bias we suspect there is also Likert imbalance bias (Tennekoon and Rosenman, 2012). Likert imbalance bias occurs when subjective measures are translated to a Likert scale value and may complement response shift bias.

By this nature the misclassification in our variable is probably not random. Any response shift change of a participant after the treatment likely depends on family and social background including her demographics and the characteristics of her SFP group, making HAS1, which assumes constant misclassification probabilities, a poor choice. The impact of Likert imbalance bias too is uneven across participants with different reported pre-treatment family functioning levels. In particular, the participants at one of the extremes of the Likert scale prior to the program are more likely to unintentionally misreport.

Available SFP data are limited and we only have some demographic information and reported pre-treatment and post-treatment scores of participants, which impacts not only the improvement in family functionality but also the misclassification probabilities. Accordingly, we have no way to proceed with the Lewbel (2000) approach, which requires at least one continuous variable affecting the improvement but not misclassification, even if we ignore the computational complexity of his approach.

In addition to the variables that we have at hand, unobserved individual effects are likely to affect the true improvement in family functionality as well as the bias hence a normal distribution appears to be the best functional form choice for these unobserved effects. This

motivates us to choose normal CDFs for F_0, F_1 and F . We have one variable, a dummy equal to one if the pre-score is near the upper bound, to differentiate Z_i^1 from Z_i^0 , which is unlikely to affect α_0 . Since we assume our F_0 and F_1 are the same function and also assume our F is symmetric, we need this exclusion restriction to distinguish $(\beta, \gamma_0, \gamma)$ from $(-\beta, -\gamma_0, -\gamma)$ ¹³.

4.2 Data

Our data consisted of 1,437 observations of parents who attended one of the 94 SFP programs in Washington and Oregon states through 2005-2009. Variables used in the analysis, including definitions, and summary statistics are presented in Table 4. The average family functioning, as measured by the change in self-assessed functioning from the pretest to the posttest increased from 3.98 to 4.27 after participation in SFP. Seventy-one percent of the participants showed an improvement in family functioning. The remaining 29% showed either a negative or no change in family functioning.

Twenty-five percent of the participants identified themselves as male, 72% as female, and 3% did not report their gender. Twenty-seven percent of the participants identified themselves as Hispanic/Latino, 60% as White, 2% as African-American; 4% as American Indian/Alaska Native, and 3% as other or multiple race/ethnicity, while 3% of the participants did not report their race/ethnicity. Seventy-four percent of the participants reported that they are living with a partner or a spouse, and 19% reported not having a spouse or partner. Almost 8% of participating parents did not report whether they are living with a partner or a spouse. The average of the within-program average pre-score was 3.99, not statistically different from the overall average pre-score of 3.98. The average of the within-program standard deviation of pre-score was 0.499, compared to the overall standard deviation of pre-score of 0.566. The implications of these statistics are that there does not seem to be much variation in the attendees of different cycles. Around 3% of the sample had reported pre-score values larger than 4.9.

We used the two gender related variables, the five variables related to race/ethnicity, the two variables related to partner/spouse, age, pre-score, within-program average and standard deviation of pre-score (despite the seeming consistency in those attracted to the program whenever and wherever it was offered) and a constant as the covariates of the main equation. Our covariates determining the propensity to record improvement as no-improvement (α_{0i}) were three race categories (native and other categories were combined with the category who did not report their race/ethnicity)¹⁴, age, pre-score, a dummy equal to 1 if the pre-score is larger than 4.9, and a constant. As the covariates determining the propensity to record no-improvement as improvement (α_1) we used the same three race related variables, age, pre-score and a constant. The choice of these variables was partly motivated by the findings of Rosenman et al. (2011). The dummy variable pre-score 4.9 was used as a covariate because people with very high initial functioning have little room to show improvement, even if they improve. This variable helps specifically to capture Likert scale

¹³As explained in section 2, our model allows Z_i^0 and Z_i^1 to be subsets of X_j and even one to be equal to be subsets of X_j . Accordingly, our use of GHAS is not constrained by the unavailability of additional exclusion restrictions in vector X_j .

¹⁴This combined category was not significantly different from whites. The result was robust when we used the three categories separately but the standard errors were very large.

bias, while serving as an exclusion restriction. A similar variable was not included among the covariates of equation (6) because only 3 participants had pre-scores below 1.5 and the lowest value of the scale, unlike the highest value, did not appear to be binding.

4.3 Analysis of Results

The results from GHAS, together with the results of HAS1 and traditional probit, are presented in tables 5 and 6. According to the traditional probit model, improvement after participating in SFP is a function of four covariates. Male participants are less likely to improve after the program than are females and those who did not report their gender; African Americans are less likely to improve than are other race categories; those who did not report whether they are living with a partner or a spouse are less likely to improve than the participants who reported that information; and, participants with higher pre-scores are less likely to improve than the participants with lower pre-scores.

HAS1 finds improvement covariates qualitatively similar to those found with the ordinary probit, but predicts misclassification probabilities as well. According to the results, the probability that a participant with no improvement reporting an improvement (α_0) takes the lowest possible value, zero. The model also predicts a 3.2% probability that participants who improved their family functioning after the program may report that they have not improved (α_1).

The GHAS estimates are noticeably different from those found with ordinary probit and HAS1, albeit not without some similarities. In contrast to HAS1, GHAS indicates that the misclassification probabilities in each direction are substantial (based on model predictions) and depends on several covariates. When considering α_{0j} , the coefficients of Hispanic dummy, age and pre-score are significant. However, the coefficient of the constant term is not significant at conventional levels. Older participants, participants with Hispanic origin and people with self-perceived low initial family functioning levels are more likely to show improvement even when they do not improve.

According to GHAS, the probability that true improvement would be reported as no-improvement (α_{1j}) also depends on several covariates. Among the statistically significant determinants of α_{1j} are the constant term, age, pre-score, and pre-score being close to the upper bound. The results suggest that older people and people with high initial family functioning levels are more likely to misclassify improvement as not happening. Consistent with Likert Scale Bias, people with initial functioning levels closer to the upper bound of the scale have very little or no room to show any improvement and therefore are also likely to be misclassified.

Our most important result, especially in light of the Monte Carlo analysis, is that the predictors of improvement found with GHAS model are not the same as those found consistently using HAS1 and probit. The male and African American dummies, which were significant in HAS1 and probit, are not significant in GHAS. Pre-score and the constant term continue to be significant, but with opposite signs. In addition, several variables that were indicated not important by HAS1 and probit are significant at conventional levels using GHAS. GHAS indicates that Hispanics are more likely to improve than Whites, that the

participants from two-parent families are more likely to improve than single parents, as are the group that did not report the details of their partner/spouse. Participants who do not report their gender or race, however, are less likely to improve than the participants who report their information. Finally, programs with participants from initially better functioning families and programs with more heterogeneous participants in terms of their pre-scores are more successful than other programs.

Of the differences, the most important is that GHAS indicates that better functioning families are more likely to improve than poor functioning families, a finding that contrasts with what was found with ordinary probit and HAS1. However, when the initial functioning increases, it increases not only the propensity to improve, but also the propensity to be misclassified and not to show the improvement. This explains why ordinary probit, which does not account for this misclassification, and HAS1, which does not account for the dependence of misclassification on initial functioning, show the opposite.

The expected values of misclassification probabilities predicted by GHAS, $E(a_{0j})=.755$ and $E(a_{1j})=.339$, are very large and sharply contrast with the HAS1 estimates ($a_0 = 0$ and $a_1 = 0.032$). The results, however, are in conformity with the findings of the Monte Carlo study, which showed a severe underestimation of misclassification probabilities by HAS1 when they are systematic and of these magnitudes.

Given the difference in results, one must wonder which model is the most appropriate. Overall, GHAS has the best fit among the three models in terms of the log-likelihood, adjusted pseudo R-squared (McFadden) and the number of successful predictions (Table 7). The model, successfully predicts 1,079 of 1,437 outcomes as reported by the data (75.1%), and estimates that 1,264 participants (88.0%) really improve after the SFP program compared to the reported 70.8%. The probit estimate of the number of people improved, for comparison, is 990 (68.9%) which, perhaps not surprisingly, is very close to the observed number. HAS1 lags significantly in the number of correct predictions of the data as a whole and reports, by far, the smallest number of participants who actually improved¹⁵. Since HAS1 reports there is no probability of someone who improved recording themselves as not improved and a positive probability someone who improved reporting that they did not, this indicates that the main equation seriously underreports the predicted improvement, calling into question the validity of its results. Accordingly, the ultimate effect of misclassification in our observed data could well be a serious underestimation of SFP's efficacy, unless corrected appropriately, with systematic misclassification.

¹⁵As noted in the text, our application is provided as an illustration of GHAS rather than a comprehensive analysis of the SFP. When GHAS is used in a purposeful evaluation one needs to weigh all evidence about the appropriateness of the specification. Although the measures discussed in the text favored GHAS over HAS1 and ordinary probit, a cross-validation log likelihood test did not favor GHAS. However, GHAS asks significantly more of the data than does HAS1 or ordinary probit, meaning one is trading off potential specification bias against efficiency. As illustrated in our Monte Carlo analysis, the bias costs of using HAS1 when GHAS is appropriate can be large while, as we show in the Appendix, the costs (in terms of the expected value of the misclassification probability) of using GHAS when HAS1 is the correct model are low.

5. CONCLUSIONS

When the dependent variable is misclassified, parameter estimates of the binary choice model are biased and inconsistent, a condition exacerbated if the misclassification is systematic rather than random. Although nonparametric methods can provide consistent estimates of model parameters, those that also provide estimates of misclassification (which may be of significant interest to policy makers) are cumbersome and often impossible to implement because of additional data needs. We provide a straightforward method to properly account for endogenous misclassification that provides both consistent estimates of the model parameters and yields estimates of misclassification probabilities for the sample and for each individual. Our experimental results document the importance of controlling for endogenous misclassification, and demonstrate that little harm is done if our approach is used for random misclassification. Moreover, our results indicate that possible systematic misclassification is not a factor that a researcher can simply ignore. The presence of systematic misclassification can toggle overall conclusions and lead analysts to substantially underestimate program benefits. Our application to real data from the Strengthening Families Program shows how large misclassification can be with subjective self-reported data, and how it can radically affect parameter estimates.

The ultimate goal of evaluating the efficacy of a treatment is identifying its costs and benefits, whether the treatment is preventive, curative or educational. If the results produced are spurious, the researchers and any other users of such results may easily end up with wrong conclusions, which may have severe policy implications. The model presented here provides an effective and easily implemented way to deal with the issue and estimate treatment effects more accurately.

The applicability of GHAS to the research problem we explained does not prove its superiority under all situations. Since MLE consistency is an asymptotic property, the relative merits of GHAS and HAS1 are not clearly visible when either the sample size is small or the misclassification probabilities are small.

In our application we ignored the impact of a potential selection bias that could arise if the participants of SFP are systematically different from the non-participants. We can easily correct for selection bias by combining a selection probit equation with equation (12) and estimating a modified bivariate probit with selection. Limitations of our data did not allow us to pursue this extension, although it is straightforward. If there is reason to believe that there are unobserved variables that affect the outcome as well as the misclassification probabilities, it may be appropriate to allow the error terms to be correlated, which is also straight forward. Finally, a misclassified polychotomous variable can be dealt with by enhancing the models presented in Abrevaya and Hausman (1999) and Dustmann and van Soest (2004) in a manner similar to ours.

ACKNOWLEDGEMENT

This research was supported in part by a grant from the National Institute of Drug Abuse (R21-DA 025139-01A1). We thank the parents and facilitators who participated in the program evaluation of SFP. This paper was improved by comments from Ron Mittelhammer, Tom Mroz, Dan Friesner, Laura Hill, Sean Murphy, Bidisha Mandal,

participants of the 3rd Annual Health Econometrics Workshop, the Editor and an anonymous referee. We thank Jason Abrevaya for sharing his Gauss code. Any remaining errors are our own.

REFERENCES

- Abrevaya J, Hausman JA. Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Panel Data on Employment Spells. *Annales D'Economie et de Statistique*. 1999; 55-56:243–75.
- Chua TC, Fuller WA. A Model for Multinomial Response Error Applied to Labor Flows. *Journal of the American Statistical Association*. 1987; 82:46–51.
- Dustmann C, van Soest A. An analysis of speaking fluency of immigrants using ordered response models with classification errors. *Journal of Business and Economic Statistics*. 2004; 22(3):312–321.
- Hausman J. Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *The Journal of Economic Perspectives*. 2001; 15(4):57–67.
- Hausman JA, Abrevaya J, Scott-Morton FM. Misclassification of the Dependent Variable in a Discrete-Response Setting. *Journal of Econometrics*. 1998; 87:239–269.
- Hill LG, Betz D. Revisiting the Retrospective Pretest. *American Journal of Evaluation*. 2005; 26:501–517.
- Lewbel A. Identification of the Binary Choice Model with Misclassification. *Econometric Theory*. 2000; 16(4):603–609.
- Mroz TM, Zayats YV. Arbitrarily Normalized Coefficients, Information Sets, and False Reports of “Biases” in Binary Outcome Models. *The Review of Economics and Statistics*. 2008; 90(3):406–413.
- Poterba JM, Summers LH. Unemployment Benefits and Labor Market Transitions: A Multinomial Logit Model with Errors in Classification. *The Review of Economics and Statistics*. 1995; 77:207–216.
- Rosenman R, Tennekoon V, Hill LG. Bias in Self Reported Data. *International Journal of Behavioural and Healthcare Research*. 2011; 2(4):320–332. [PubMed: 25383095]
- Sprangers M, Hoogstraten J. Pretesting Effects in Retrospective Pretest-Posttest Designs. *Journal of Applied Psychology*. 1989; 74(2):265–272.
- Tennekoon, V.; Rosenman, R. Likert Imbalance Bias, manuscript, School of Economic Sciences. Washington State University; Pullman, WA: 2012.

Table 1
Determinants of $\Pr(y=1)$ with covariate dependent misclassification (coefficients)

Variable	True Value	Probit		HAS1		GHAS	
		Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
$E(a_{0j}) = E(a_{1j}) = 0.05$							
Intercept	-1.0	-0.745	0.049	-0.725	0.069	-1.006	0.170
beta1	0.2	0.168	0.014	0.187	0.016	0.206	0.020
beta2	1.5	1.419	0.043	1.504	0.083	1.504	0.092
beta3	-0.6	-0.835	0.073	-0.906	0.102	-0.605	0.172
$E(a_{0j}) = E(a_{1j}) = 0.1$							
Intercept	-1.0	-0.489	0.049	-0.448	0.070	-0.970	0.148
beta1	0.2	0.142	0.015	0.170	0.019	0.205	0.018
beta2	1.5	1.414	0.043	1.554	0.111	1.488	0.074
beta3	-0.6	-1.138	0.074	-1.283	0.145	-0.631	0.162
$E(a_{0j}) = E(a_{1j}) = 0.2$							
Intercept	-1.0	-0.135	0.045	0.037	0.141	-1.012	0.262
beta1	0.2	0.096	0.013	0.142	0.023	0.208	0.026
beta2	1.5	1.267	0.043	1.622	0.215	1.504	0.113
beta3	-0.6	-1.356	0.074	-1.720	0.321	-0.595	0.256
$E(a_{0j}) = 0.3, E(a_{1j}) = 0.75$							
Intercept	-1.0	0.606	0.046	0.632	0.054	-1.104	0.474
beta1	0.2	-0.020	0.010	-0.021	0.011	0.222	0.085
beta2	1.5	1.265	0.047	1.290	0.052	1.582	0.795
beta3	-0.6	-3.112	0.091	-3.154	0.086	-0.645	0.694
$E(a_{0j}) = 0.75, E(a_{1j}) = 0.3$							
Intercept	-1.0	1.385	0.056	2.204	0.473	-0.995	0.094
beta1	0.2	-0.016	0.010	0.032	0.034	0.201	0.021
beta2	1.5	0.615	0.054	0.984	0.071	1.521	0.162
beta3	-0.6	-1.592	0.087	-2.626	0.630	-0.634	0.252

Table 2
Determinants of $\Pr(y_0=1|y=0)$ with covariate dependent misclassification

Variable	True Value	HAS1		GHAS	
		Est.	Std. Err.	Est.	Std. Err.
$E(a_{0j}) = E(a_{1j}) = 0.05$					
Intercept	-0.50			-0.443	0.210
gamma01	-3.96			-4.779	2.611
$E(\hat{\alpha}_0)$	0.05	0.004	0.009	0.052	0.018
$E(a_{0j}) = E(a_{1j}) = 0.1$					
Intercept	0.25			0.306	0.168
gamma01	-5.36			-5.833	1.151
$E(\hat{\alpha}_0)$	0.10	0.008	0.014	0.100	0.012
$E(a_{0j}) = E(a_{1j}) = 0.2$					
Intercept	0.50			0.528	0.141
gamma01	-3.49			-3.638	0.596
$E(\hat{\alpha}_0)$	0.20	0.012	0.020	0.199	0.016
$E(a_{0j}) = 0.3, E(a_{1j}) = 0.75$					
Intercept	1.00			1.027	0.264
gamma01	-3.60			-3.614	0.304
$E(\hat{\alpha}_0)$	0.30	0.001	0.002	0.305	0.039
$E(a_{0j}) = 0.75, E(a_{1j}) = 0.3$					
Intercept	5.00			5.048	0.419
gamma01	-6.64			-6.710	0.606
$E(\hat{\alpha}_0)$	0.75	0.000	0.000	0.750	0.010

Table 3
Determinants of $\Pr(y_0=0|y=1)$ with covariate dependent misclassification

Variable	True Value	HAS1		GHAS	
		Est.	Std. Err.	Est.	Std. Err.
$E(a_{0j}) = E(a_{1j}) = 0.05$					
Intercept	-0.50			-0.417	0.919
gamall	-1.00			-1.472	1.800
gama12	-2.83			-6.605	17.684
$E(\hat{\alpha}_1)$	0.05	0.041	0.033	0.060	0.040
$E(a_{0j}) = E(a_{1j}) = 0.1$					
Intercept	0.25			0.410	0.589
gamall	-2.00			-2.580	1.295
gama12	-3.63			-4.828	4.695
$E(\hat{\alpha}_1)$	0.10	0.058	0.035	0.111	0.040
$E(a_{0j}) = E(a_{1j}) = 0.2$					
Intercept	0.50			0.601	0.443
gamall	-1.60			-1.767	0.529
gama12	-2.40			2.854	1.590
$E(\hat{\alpha}_1)$	0.20	0.121	0.049	0.204	0.050
$E(a_{0j}) = 0.3, E(a_{1j}) = 0.75$					
Intercept	1.00			1.571	1.169
gamall	-4.00			-4.774	1.298
gama12	4.15			4.336	0.908
$E(\hat{\alpha}_1)$	0.75	0.012	0.014	0.746	0.024
$E(a_{0j}) = 0.75, E(a_{1j}) = 0.3$					
Intercept	1.50			1.542	0.340
gamall	-2.00			-2.034	0.259
gama12	-3.60			-3.688	0.600
$E(\hat{\alpha}_1)$	0.30	0.087	0.016	0.301	0.030

Table 4
Variable Names, Descriptions and Summary Statistics

Name	Description	Mean	Std. Dev.
Improved	Observed (misclassified) binary dependent variable: Equal to 1 if post-test score > pre-test score	0.708	0.455
Male	Equal to 1 if the gender is reported as male; 0 otherwise	0.250	0.433
Gender Not Reported	Equal to 1 if the gender is not reported; 0 otherwise	0.030	0.170
Black/African American	Equal to 1 if the race is reported as African American; 0 otherwise	0.023	0.150
Hispanic/Latino	Equal to 1 if the race is reported as Hispanic; 0 otherwise	0.269	0.443
Native American	Equal to 1 if the race is reported as Native American; 0 otherwise	0.040	0.195
Other Races	Equal to 1 if the race is reported as other or of multiple ethnicity; 0 otherwise	0.034	0.182
Race Not Reported	Equal to 1 if the race is not reported; 0 otherwise	0.034	0.182
Age	Integer (17-73)	38.822	7.846
Living with Partner or spouse	Equal to 1 if reported living with partner or spouse; 0 otherwise	0.736	0.441
Partner/Spouse Details Not Reported	Equal to 1 if the partner/spouse details not reported; 0 otherwise	0.077	0.266
Program Average of Pre-score	Average of the pre-scores of the participants enrolled in the same program; Continuous variable between 1-5	3.987	0.237
Program Std. Dev. of Pre-score	Standard deviation of the pre-scores of the participants enrolled in the same program; Continuous variable	0.499	0.173
Pre-test Score	Self-reported pre-test score; Semi-continuous variable between 1-5	3.979	0.546
Pre-test Score > 4.9	Equal to 1 if the pre-score > 4.90; 0 otherwise	0.033	0.178

Table 5
Determinants of True Improvement in Family Functionality

Variable	Probit		HAS1		GHAS	
	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Improvement						
Male	-0.250 ***	0.090	-0.300 ***	0.103	0.572	0.517
Gender Not Reported	0.436	0.305	0.548	0.403	-2.032 ***	0.786
Excluded: Female						
Black or African-American	-0.525 **	0.239	-0.563 **	0.252	-0.368	0.744
Hispanic	-0.148	0.096	-0.120	0.111	1.139 *	0.680
Native American	-0.294	0.194	-0.319	0.213	-0.424	0.735
Other Races	-0.115	0.201	-0.052	0.241	3.166	6.290
Race Not Reported	0.375	0.290	0.357	0.326	-1.441 *	0.827
Excluded: White						
Age	-0.002	0.005	-0.003	0.006	-0.015	0.022
Living with Partner/Spouse	-0.158	0.105	-0.162	0.118	1.151 ***	0.402
Partner/Spouse Details Not						
Reported	-0.500 ***	0.166	-0.573 ***	0.187	2.367 ***	0.832
Excluded: Not Living with						
Partner/Spouse						
Program Average of Pre-score	0.165	0.199	0.307	0.241	1.568 **	0.762
Program Std. Dev. of Pre-score	-0.337	0.250	-0.225	0.296	2.029 **	1.023
Pre-score	-1.358 ***	0.096	-1.612 ***	0.173	1.454 ***	0.440
Intercept	5.951 ***	0.834	6.527 ***	1.060	-12.635 ***	3.387

p<0.01;

**
p<0.05;

*
p<0.10

Table 6
Determinants of Probabilities of Misclassification

Variable	HAS1		GHAS	
	Est.	Std. Err.	Est.	Std. Err.
Recording No Improvement as Improvement				
Black or African-American			-7.831	5.534
Hispanic			-15.139 *	9.007
Race Not Reported and other races			-3.317	4.032
Excluded: White				
Age			0.629 *	0.358
Pre-score			-3.922 *	2.088
Intercept			2.849	5.747
$E(\hat{\alpha}_0)$	0.000 ***	0.000	0.7549 ***	0.015
Recording Improvement as No Improvement				
Black or African-American			0.428	0.368
Hispanic			-0.143	0.137
Race Not Reported and other races			0.149	0.202
Excluded: White				
Age			0.013 *	0.007
Pre-score			1.139 ***	0.146
Pre-score > 4.9			1.086 ***	0.355
Intercept			-5.533 ***	0.702
$E(\hat{\alpha}_1)$	0.0320	0.020	0.3387 ***	0.025

p<0.01;

**
p<0.05;

*
p<0.10

Table 7
Overall Comparison of three Models

	Probit	HAS1	GHAS
Number of observations	1437	1437	1437
Number of free parameters	14	16	27
Log-likelihood	-709.091	-707.051	-687.508
Adjusted Pseudo-R2 (McFadden)	0.1683	0.1678	0.1781
Correct predictions	1048 (72.9%)	932 (64.9%)	1079 (75.1%)
Estimated number of participants improved their family functionality	990 (68.9%)	686 (47.7%)	1264 (88.0%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table A1
Determinants of $\Pr(y=1)$ with random misclassification (coefficients)

Variable	True Value	Probit		HAS1		GHAS	
		Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
$E(a_{0j}) = E(a_{1j}) = 0.05$							
Intercept	-1.0	-0.844	0.048	-1.035	0.170	-1.112	0.290
beta1	0.2	0.159	0.015	0.207	0.033	0.217	0.042
beta2	1.5	1.290	0.041	1.546	0.182	1.587	0.242
beta3	-0.6	-0.498	0.068	-0.644	0.133	-0.573	0.278
$E(a_{0j}) = E(a_{1j}) = 0.1$							
Intercept	-1.0	-0.716	0.050	-1.025	0.207	-1.101	0.376
beta1	0.2	0.128	0.014	0.208	0.046	0.214	0.054
beta2	1.5	1.107	0.040	1.548	0.241	1.580	0.317
beta3	-0.6	-0.411	0.066	-0.652	0.180	-0.553	0.370
$E(a_{0j}) = E(a_{1j}) = 0.2$							
Intercept	-1.0	-0.507	0.046	-0.998	0.319	-1.094	0.596
beta1	0.2	0.085	0.012	0.204	0.065	0.217	0.092
beta2	1.5	0.790	0.037	1.509	0.383	1.563	0.549
beta3	-0.6	-0.278	0.066	-0.638	0.263	-0.561	0.525

Table A2
Determinants of $\Pr(y_0=1|y=0)$ with random misclassification

Variable	True Value	HAS1		GHAS	
		Est.	Std. Err.	Est.	Std. Err.
$E(a_{0j}) = E(a_{1j}) = 0.05$					
Intercept		-1.645		-2.165	3.383
gamma01		0.000		-0.024	4.125
$E(\hat{\alpha}_0)$	0.05	0.055	0.039	0.062	0.044
$E(a_{0j}) = E(a_{1j}) = 0.1$					
Intercept		-1.282		-1.534	0.872
gamma01		0.000		-0.071	0.858
$E(\hat{\alpha}_0)$	0.10	0.099	0.050	0.100	0.056
$E(a_{0j}) = E(a_{1j}) = 0.2$					
Intercept		-0.841		-1.513	2.326
gamma01		0.000		0.412	2.626
$E(\hat{\alpha}_0)$	0.20	0.181	0.070	0.175	0.090

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table A3
Determinants of $\Pr(y_0=0|y=1)$ with random misclassification

Variable	True Value	HAS1		GHAS	
		Est.	Std. Err.	Est.	Std. Err.
$E(\alpha_0) = E(\alpha_1) = 0.05$					
Intercept	-1.645			-2.443	1.539
gamall	0.000			-0.524	1.758
gama12	0.000			-0.822	5.328
$E(\hat{\alpha}_1)$	0.05	0.049	0.033	0.060	0.051
$E(\alpha_0) = E(\alpha_1) = 0.1$					
Intercept	-1.282			-1.725	1.016
gamall	0.000			0.250	1.102
gama12	0.000			-0.125	1.795
$E(\hat{\alpha}_1)$	0.10	0.098	0.048	0.100	0.059
$E(\alpha_0) = E(\alpha_1) = 0.2$					
Intercept	-0.841			-1.378	1.518
gamall	0.000			0.183	1.256
gama12	0.000			0.081	1.465
$E(\hat{\alpha}_1)$	0.20	0.183	0.069	0.178	0.091