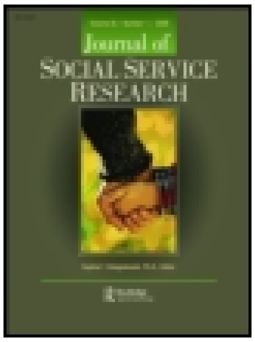

Routledge
Taylor & Francis Group

Journal of Social Service Research

ISSN: 0148-8376 (Print) 1540-7314 (Online) Journal homepage: http://www.tandfonline.com/loi/wssr20

Detecting and Statistically Correcting Sample Selection Bias

Gary Cuddeback MSW, MPH , Elizabeth Wilson MSW , John G. Orme PhD & Terri Combs-Orme PhD

To cite this article: Gary Cuddeback MSW, MPH , Elizabeth Wilson MSW , John G. Orme PhD & Terri Combs-Orme PhD (2004) Detecting and Statistically Correcting Sample Selection Bias, Journal of Social Service Research, 30:3, 19-33, DOI: 10.1300/J079v30n03_02

To link to this article: http://dx.doi.org/10.1300/J079v30n03_02

Published online: 10 Oct 2008.

Detecting and Statistically Correcting Sample Selection Bias

Gary Cuddeback
Elizabeth Wilson
John G. Orme
Terri Combs-Orme

ABSTRACT. Researchers seldom realize 100% participation for any research study. If participants and non-participants are systematically different, substantive results may be biased in unknown ways, and external or internal validity may be compromised. Typically social work re-

---

Gary Cuddeback, MSW, MPH, is Research Associate, Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, 101 Connor Drive, Suite 302, Chapel Hill, NC 27514.

Elizabeth Wilson, MSW, is a Doctoral Student, John G. Orme, PhD, is Professor, and Terri Combs-Orme, PhD, is Associate Professor, University of Tennessee, College of Social Work, Children's Mental Health Services Research Center, Henson Hall, Knoxville, TN 37996-3332.

Address correspondence to: Gary S. Cuddeback, Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, 101 Connor Drive, Suite 302, Chapel Hill, NC 27514 or John G. Orme, University of Tennessee, College of Social Work, Children's Mental Health Services Research Center, 128 Henson Hall, Knoxville, TN 37996-3332 (E-mail: Gary S. Cuddeback at *cuddeback@mail.schsr.unc.edu* or John G. Orme at *jorme@utk.edu*). The authors would like to express their appreciation to Shenyang Guo for his suggestions on an earlier version of this paper and to Niki Le Prohn at Casey Family Programs for her support.

This article is the result of a collaborative effort of the Casey Family Programs Foster Family Project at The University of Tennessee and Casey Family Programs, an operating foundation delivering foster care, adoption, and other permanency planning services in 14 states. You may review more information about Casey Family Programs at *http://www.casey.research.org/research* and more information about Casey Family Programs Foster Family Project at The University of Tennesee at *http://www.utcmhsrc.csw.utk.edu/caseyproject/*.

Journal of Social Service Research, Vol. 30(3) 2004
http://www.haworthpress.com/web/JSSR

Digital Object Identifier: 10.1300/J079v30n03_02

searchers use bivariate tests to detect selection bias (e.g., χ^2 to compare the race of participants and non-participants). Occasionally multiple regression methods are used (e.g., logistic regression with participation/non-participation as the dependent variable). Neither of these methods can be used to correct substantive results for selection bias. Sample selection models are a well-developed class of econometric models that can be used to detect *and* correct for selection bias, but these are rarely used in social work research. Sample selection models can help further social work research by providing researchers with methods of detecting and correcting sample selection bias.

KEYWORDS. Sample selection bias, statistical methods, social work research

Whether conducting a survey of life experiences and attitudes or examining relationships among predictor and response variables, social work researchers generally intend to make inferences beyond a study's participants to a population. It is the collection of a probability sample where each element has a known non-zero probability of selection that permits the use of statistics to make inferences about a population (Kish, 1965).

The selection of a probability sample is only the first step, however. Researchers must then ensure the participation of those selected. Researchers seldom realize 100% participation, usually for one of two reasons. First, selected individuals can, and frequently do, refuse to participate. This is problematic if collectively the individuals who do not participate are systematically different from those who do, and consequently the final sample may be biased. This is known as "sample selection bias." Second, selected individuals may agree to participate but then be "lost" over time due to transience, incarceration, death, or other reasons. The final sample might be biased if the individuals who are lost differ in some systematic way from the participants who remain. This is known as "attrition bias." Unlike sample selection bias, attrition bias mainly occurs in longitudinal studies. The remainder of this paper will concentrate on sample selection bias, however, some readings on attrition bias will be presented later.

Selection bias occurs because non-participation is rarely random (e.g., distributed equally across subgroups); instead, bias often is correlated with variables that also are related to the dependent variable of interest or that preclude using the sample to describe the target population (Goodfellow, Kiernan, Ahern, & Smyer, 1988). York (1998) defines selection bias as "any characteristic of a sample that is believed to make it different from the study population in some important way" (p. 239). Finally, Winship and Mare (1992) report that selection bias can occur when observations in social research are selected such that they are not independent of the outcome variables in the study, possibly leading to biased inferences about social processes.

Beginning in the late 1970s and early 1980s (Greene, 1981; Heckman, 1976, 1978, 1979), methods for detecting and statistically correcting selection bias were developed in economics and related areas. In the decades since, an extensive literature has evolved in the area of sample selection bias (Berk, 1983; Lee & Marsh, 2000; Miller & Wright, 1995; Stolzenberg & Relles, 1997; Vella & Verbeek, 1999; Winship & Mare, 1992). These methods are known as "sample selection" models.

Selection bias potentially threatens both internal and external validity (Berk, 1983; Miller & Wright, 1995). Selection bias is a threat to internal validity in that independent variables are correlated with a disturbance term (i.e., error) and analyses based on biased samples can lead to inaccurate estimates of the relationships between variables (e.g., regression coefficients). Thus, effects may be attributed to exogenous variables that actually are due to selection factors (Cook & Campbell, 1979). For example, consider the relationship between family income and approval to provide family foster care in a sample of foster family applicants. If data for income are missing systematically for applicants with higher incomes, the effect of income on approval might be underestimated as quantified using a regression coefficient, for example. Thus, the internal validity of the study might be compromised.

Selection bias also potentially threatens external validity because a final, biased sample might not be generalizable to the intended population. Using another example, consider the results of a study that evaluates a high school dropout prevention program based on an analysis of a random sample of students who completed the program. The final sample used in the analysis might underrepresent the high-risk students and overrepresent the students who are at low or medium risk if the students most at risk drop out of school prior to completing (or even starting) the intervention. And, any inferences (i.e., a conclusion that the program is successful for *all* students irrespective of their level of risk) drawn from the sample might not be generalizable to the students most in danger of dropping out of school (i.e., those that need the intervention the most).

These examples underscore the importance of attending to differences among participants and non-participants and participation rates. By establishing that no differences exist among participants and non-participants, or more importantly, detecting differences among participants and non-participants and correcting substantive results for these differences, these sample selection models are useful and important tools for social work researchers.

Moreover, selection models should be used whenever sufficient data for non-participants are available, and failing to do so can potentially lead to problems with the results of any research. However, failing to use these methods when appropriate is different than failing to use them when data for non-participants are unavailable, which is common. In this latter case, sample selection models are obviously of no use. This is an important distinction to note.

The importance of selection bias is known to social work researchers. However, with a few exceptions (Ards, Chung, & Myers, 1998; Brooks & Barth, 1999; Courtney, Piliavin, & Wright, 1997; Greenwell & Bengston, 1997; Grogan-Kaylor, 2001; McDonald, Moran, & Garfinkel, 1990; Vartanian, 1999), the available methods for detecting and statistically correcting selection bias have not been used in social work research, and only a limited number of the many available sample selection models have been used. Therefore, the purpose of this paper is to: (1) introduce sample selection models to social work researchers; (2) provide an overview of sample selection models; (3) illustrate the use of a sample selection model and compare the results with methods typically used in social work research; (4) note computer software for estimating sample selection models; and (5) direct readers to additional literature in this area.

SAMPLE SELECTION MODELS

The literature addressing the detection and correction of selection bias is extensive, and a complete review of this literature is beyond the scope of this paper. Overviews of sample selection models can be found in Moffitt (1991), Reynolds and Temple (1995), Shadish, Cook, and Campbell (2001), Stolzenberg and Relles (1997), Winship and Mare (1992), and Winship and Morgan (1999). Additional selected readings of potential use to social work researchers are suggested below.

One of the earliest sample selection methods is known as the "Heckman" two-step estimator (Heckman, 1976, 1978, 1979). However, there is some evidence that corrections using this method can sometimes worsen rather than improve estimates, even under ordinary circumstances (Stolzenberg & Relles, 1997; Winship & Mare, 1992). See Stolzenberg and Relles (1997) for a discussion of making reasonable judgments of when Heckman's (1976, 1979) two-step estimator is likely to improve or worsen regression coefficient estimates. Nevertheless, since Heckman (1976, 1979), numerous models for detecting and statistically correcting sample selection bias have been developed.

Current sample selection models typically involve the simultaneous estimation of two multiple regression models. One model (i.e., the substantive model) is used to examine the substantive question of interest (e.g., Is the probability of approval to foster different for African-American and European-American families?). In most respects this model is no different from any other multiple regression model, and continuous, binary, multi-categorical, or other types of dependent variables can be modeled using methods familiar to social work researchers (e.g., a linear "OLS" model for a continuous dependent variable, a binary probit or logit model for a binary dependent variable, a multinomial probit or logit model for a multi-categorical dependent variable) (Orme & Buehler, 2001). The other regression model (i.e., the selection model) is used to detect selection bias and to statistically correct the substantive model for selection bias. Binary probit regression typically is used for this purpose because the outcome modeled usually is binary (e.g., participation or not), but binary logit regression and other models can be used (Greene, 1995, 2000).

Illustration of a Sample Selection Model

The illustration we will use is based on a study of foster family applicants. In this study of a population of foster family applicants who completed pre-service training, 230 applicants were selected but only 161 participated (70%) (Orme, Buehler, McSurdy, Rhodes, Cox, & Patterson, in press; Orme, Buehler, McSurdy, Rhodes, Cox, & Cuddeback, in press). The substantive question of interest used for illustration here is whether African-American applicants are less likely than European-

American applicants to be approved to foster (i.e., the substantive model). First, it is important to consider whether there are systematic differences between participants and non-participants on selected variables (i.e., the selection model), in this case race, marital status, and education level. To make things more clear in the illustration, it is important to distinguish between the binary dependent variable in the substantive model (i.e., approval or disapproval to foster) and the dependent variable in the selection model (i.e., participation or non-participation), which is also a binary dependent variable. Descriptions of both the selection model and the substantive model follow.

Race, highest education, and marital status were determined for participants and non-participants. Race was coded 0 for "European-American" and 1 for "African-American/other"; marital status was coded 0 for "not married" and 1 for "married"; and education was coded 0 for "less than high school," 1 for "high school/GED," 2 for "some college, no degree," 3 for "associate/two-year degree," 4 for "bachelor's degree," and 5 for "advanced degree." Participation was determined for the sample of 230 applicants and coded 0 for "no" and 1 for "yes." Approval to foster was determined only for the 161 participants and coded as 0 for "no" and 1 for "yes."

Family-level education and race variables were created because approval and placement decisions are made at the family level, and so models were tested at the family level. Women's education was used for family-level education except in the four cases of unmarried men, and men's education was used in these cases. Family-level race for each unmarried applicant was the race of the individual (European-American = 0, African-American/other = 1). For same-race married couples, family-level race was the race shared by spouses, and for the four mixed-race married couples, family-level race was coded as African-American/other. Table 1 shows descriptive statistics for race, marital status, and education for both participants and non-participants.

First, to model the substantive question and to obtain results with which a selection model will be compared, binary probit regression was used with approval status as the dependent variable and race as the only independent variable entered into the regression equation. The relationship between race and approval status was not statistically significant ($b = -0.25$, χ^2 $(1) = 1.27$, $p = 0.26$) (see Table 2). It is important to examine whether there are systematic differences between participants and non-participants (i.e., detect possible selection bias), which may have led to biased results, however.

TABLE 1. Demographic Characteristics and Bivariate Comparisons

	Non-Participants	Participants	Test statistic	*p*
Characteristic	%	%		
Race			4.47*	.03***
European American	55.1	69.6		
African American/other	44.9	30.4		
Marital status			3.29*	.07
Not married	26.1	38.5		
Married	73.9	61.5		
Education			4740.50**	.07
< high school diploma	11.6	5.0		
High school/general equivalency diploma	34.8	34.8		
Some college/no degree	17.4	9.9		
Associate/two-year degree	15.9	18.0		
Bachelor's degree	11.6	23.0		
Advanced degree	8.7	9.3		

* Pearson chi square.
** Mann-Whitney test.
*** Indicates statistically significant test statistic.

TABLE 2. LIMDEP Output of Binomial Probit Model

Binomial Probit Model	
Maximum Likelihood Estimates	
Dependent variable	ASTATUS2
Weighting variable	ONE
Number of observations	161
Iterations completed	4
Log likelihood function	−95.68754
Restricted log likelihood	−96.32140
Chi-squared	1.267722
Degrees of freedom	1
Significance level	.2601945

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Index function for probability				
Constant	.6466520811	.12788011	5.057	.0000	
FRACE	−.2530536653	.22421983	−1.129	.2591	.30434783

One way to detect sample selection bias is to use participation status as the dependent variable, and then use bivariate statistical methods to compare participants and non-participants. This is the method typically used in social work research (e.g., Cohen, Mowbray, Bybee, Yeich, Risbisl, & Freddolino, 1993; Lauby, Kotranski, & Halbert, 1996). To illustrate this method with our example, a chi-square test was used to examine whether there was a systematic difference in terms of race and marital status between foster families who did and did not participate (see Table 1). A Mann-Whitney test was used to determine whether there was a systematic difference in highest education (see Table 1). Results indicated no differences between participants and non-participants in terms of marital status or highest education, but African-American/other families were less likely to participate than European-American families. This finding is particularly troubling given the substantive question of interest, and it raises questions about the validity of the results previously obtained from the above substantive regression model.

Bivariate comparisons can and oftentimes are used to detect selection bias but they cannot be used to estimate the combined or independent effects of the variables used to determine selection bias. This is especially important given that the demographic variables oftentimes used to examine selection bias typically are intercorrelated. Moreover, these methods often lack the necessary statistical power to detect sample selection bias and can give misleading results.

Multiple regression, specifically binary probit or logit regression, can be used to estimate the combined and independent effects of the variables used to determine selection bias (for social work examples see Brooks & Barth, 1999; Grogan-Kaylor, 2001; Littell, 1997), and these methods are better than bivariate comparisons for detecting selection bias.

To illustrate this method with our example, binary probit regression was used, participation status was the dependent variable, and race, marital status, and highest education were entered simultaneously in the regression equation. These results showed that African-American/other families ($b = -0.54$, $p = .007$) and two-parent families ($b = -0.54$, $p = 0.01$) ($\chi^2(3) = 13.88$, $p = .003$) were less likely to participate in the study. There was no difference between participants and non-participants in terms of highest education.

The bivariate and multiple regression methods described above can be used to detect selection bias. However, if selection bias is detected using these methods, a researcher's determination of if and how selection bias influences results concerning the substantive question of interest is usually relegated to the realm of speculation, typically in the discussion of results and limitations. Sample selection models provide a quantitative basis for examining the presence of selection bias and the nature of the effects of that bias on the substantive findings.

To illustrate a sample selection model with our example, approval to foster was the substantive dependent variable of interest, and a binary probit model was used to estimate the effect of race on approval. A second binary probit model was used to estimate the effects of marital status, highest education, and race on participation, and to correct the substantive model for selection bias (see Figure 1 and Table 3). These two regression models were estimated simultaneously, and overall this model is referred to as a "bivariate probit sample selection model" (Greene, 1995, 2000).

Results of the bivariate probit sample selection model show that race had no effect on the approval status of foster care applicants, and there was only a slight correction in the regression coefficient for race compared to the results initially obtained with the binary probit regression model without the correction for selection (i.e.,-0.25 vs. -0.23). The absence of selection bias is further confirmed as the correlation between the error terms of the two equations, as indicated by the fact that ρ (rho) in the LIMDEP output at the bottom of Table 3 is not statistically significant ($\rho(1,2) = -.13$, $p = .88$). Rho is the correlation between the error terms of the substantive and selection models. Rho has a potential range between -1 and $+1$ and can give some indication of the likely range of selection bias. A correlation with an absolute value of 1 would occur if the regression coefficients of the selection model and the regression coefficients of the substantive model were estimated by identical processes (i.e., potential selection bias). Conversely, a value of rho closer to zero would suggest that data are missing randomly or the regression coefficients of the selection model and the regression coefficients of the substantive model were estimated by unrelated processes (i.e., less evidence of selection bias).

Thus, though African-American/other and two-parent families were significantly less likely to participate in the study, the initial analyses of the sample produced estimates that appear to be generally accurate and free of selection bias. Thus, confidence in those estimates is improved with the modeling of the bivariate probit sample selection model.

Computer Software for Estimating Sample Selection Models

LIMDEP was used to model the bivariate probit sample selection model illustrated here. LIMDEP (Greene, 1995) can be used to estimate

FIGURE 1. Simultaneous Estimation of Two Multiple Regression Models

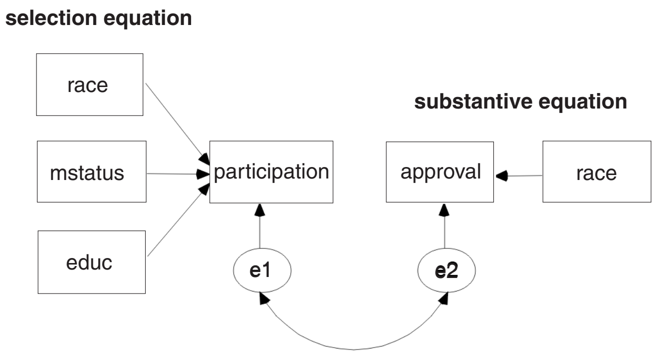

TABLE 3. LIMDEP Output of Bivariate Probit Selection Model

FIML Estimates of Bivariate Probit Model	
Maximum Likelihood Estimates	
Dependent variable	ASTPAR
Weighting variable	ONE
Number of observations	230
Iterations completed	12
Log likelihood function	−229.2320
Selection model based on PARTICIP	
Means for vars. 1-2 are after selection.	

Variable	Coefficient	Standard Error	b/St.Er.	P [|Z| >z]	Mean of X
	Index equation for ASTATUS2				
Constant	.6980012544	.33921139	2.058	.0396	
FRACE	−.2277555423	.27274375	−.835	.4037	.30434783
	Index equation for PARTICIP				
Constant	.8899888230	.25395851	3.504	.0005	
FRACE	−.5380531760	.20159996	−2.669	.0076	.34782609
MARRIED	−.5323110861	.21329976	−2.496	.0126	.65217391
FHIGHGRD	.8481873193E-01	.580605119E-01	1.461	.1441	2.3521739
	Disturbance correlation				
RHO (1,2)	−.1300628211	.82703100	−.157	.8750	

the widest variety of sample selection models. Information regarding LIMDEP can be found at *www.limdep.com*. Also, a student version, along with documentation, can be downloaded free from *www.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm*. The student version of this software and accompanying data sets are included with Greene's (2000) text.

For SAS users, Jaeger (1993) provides the code for performing Heckman's two-step estimation of sample selection bias. This program can be downloaded from the SAS Institute web page using the following link (*http://ftp.sas.com/techsup/download/stat/heckman.html*). Some adjustments to the code are necessary for the program to work (i.e., your own variable names must be inserted).

Finally, Stata 7 (2001) (*http://www.stata.com/site.html*) also can be used to estimate Heckman's (1976, 1979) two-step detection and correction of sample selection bias. Specific programming information can be found at the following Stata link (*http://www.stata.com/help.cgi?heckman)*.

Limitations

Although sample selection methods are useful for detecting and statistically correcting selection bias, they do have limitations. Any method for detecting and correcting selection bias is only as good as the selection model. Therefore, if the selection model is misspecified (e.g., important variables are missing from the model, only main effects are specified when interactions are present, or linearity is specified in the presence of non-linearity), methods of detecting and statistically correcting selection bias may be inaccurate or, unbeknownst to the researcher, may make estimates worse. As Shadish et al. (2002) note, in the context of selection in quasi-experimental designs: ". . . these models would probably work better if they used predictors that were selected to reflect theory and research about variables that affect selection into treatment, a procedure that requires studying the nature of selection bias as a phenomenon in its own right" (p. 168).

It can be helpful, though not always possible, for the researcher to have some general idea about the source and direction of the bias *before* applying any methods of correcting selection bias, and some idea about the validity of the corrections *after* they are made. In the example presented here, we did not have any sense of the source or direction of the bias and thus could not anticipate the direction in which corrections would occur. However, consider the example used earlier to illustrate how selection bias threatens internal validity. If data for those with higher incomes in a sample of foster care applicants were missing systematically, in general, it is reasonable to expect that the corrected estimate of the effect (e.g., regression coefficient) of income on approval to foster would be greater than the original, uncorrected estimate. Conversely, if the corrected coefficient is less than the original, uncorrected estimate, there may be reason for concern and some cause to question the validity of the correction.

Moreover, as many of these methods have their own subtleties and assumptions and no one method has emerged as superior, it may be appropriate to use and report the results of more than one technique, if possible (Shadish et al., 2002).

Also, Heckman has built upon his earlier works and has more recently introduced the propensity score as a way to address selection bias (e.g., see Shadish, Cook, & Campbell [2002] for a brief overview of propensity scores and Heckman & Smith [1995] and Heckman, Ichimura, & Todd [1997] for more detailed descriptions and applications of these methods). Briefly, propensity scores take into account all variables that might play a role in the selection process and create a predicted probability (i.e., propensity score) of participation vs. non-participation from a logistic regression equation (Shadish, Cook, & Campbell, 2001). These scores then can be used to match participants and non-participants or as a covariate in a regression model, for example. A more thorough discussion of these methods is beyond the scope of this paper, and it is important to note that propensity scores have limitations and their application to social work research have yet to be explored.

Sample Selection Readings

Much of the literature regarding selection bias originated in economics (Greene, 1981; Heckman, 1976, 1978, 1979; Lee & Marsh, 2000; Vella & Verbeek, 1999). However, articles addressing selection bias are now being applied to other areas as well, such as sociology (Berk, 1983; Hughes, 1997; Stolzenberg & Relles, 1997; Winship & Mare, 1992), substance abuse (Treno, Gruenewald, & Johnson, 1998), marriage and family studies (Miller & Wright, 1995), and evaluation and program planning (Devine, Brody, & Wright, 1997; Reynolds & Temple, 1995).

For examples of the application of selection bias correction methods to models with continuous dependent variables, see Leonard and Jiang (1999) and Jensen (1985). For examples of the application of selection bias correction methods to models with censored and/or truncated continuous dependent variables, see Treno, Gruenewald, and Johnson (1998) and Vartanian (1999). Lastly, there is an extensive literature specific to the issue of attrition bias (e.g., for some examples see Foster & Bickman, 1996; Hensher, 1987; McGuigan, Ellickson, Hays, & Bell, 1997; Miller & Wright, 1995; Polak, 1999).

DISCUSSION AND CONCLUSION

We have presented an easy-to-use method for detecting and statistically correcting selection bias when the outcome of interest is a dichotomous variable, and we have briefly noted similar procedures for modeling a wide range of other types of dependent variables. Thus, we offer one solution to addressing the ubiquitous problem of less than 100% participation encountered by social work researchers.

In this paper we have focused on the use of sample selection methods to detect and statistically correct for selection bias resulting from non-participation. However, these statistical methods can be used more generally to detect and statistically correct for selection bias. For example, they can be used to estimate treatment effects in the presence of non-random assignment to groups (Winship & Mare, 1992; Shadish, Cook & Campbell, 2002). Or, more generally, they can be used to model systematic preexisting between-group differences, and then in turn to statistically correct for such differences in estimating between-group differences in outcomes (e.g., Grogan-Kaylor, 2001; Hughes, 1997; Littell, 1997; Reynolds & Temple, 1995).

Finally, given that much if not most of the research in social work is subject to selection or attrition bias, social work researchers should expand their knowledge and implementation of the available methods for detecting and correcting selection bias. By neglecting these methods, social work researchers run the risk of obtaining results of unknown generalizability and misestimating the effects of independent variables on substantive variables of interest.

REFERENCES

Ards, S., Chung, C., & Myers, S. L., Jr. (1998). The effects of sample selection bias on racial differences in child abuse reporting. *Child Abuse & Neglect, 22* (2), 103-115.

Berk, R. A. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review, 48* (3), 386-398.

Brooks, D., & Barth, R. P. (1999). Adult transracial and inracial adoptees: Effects of race, gender, adoptive family structure, and placement history on adjustment outcomes. *American Journal of Orthopsychiatry, 69* (1), 87-99.

Cohen, E. H., Mowbray, C. T., Bybee, D., Yeich, S., Risbisl, K., & Freddolino, P. P. (1993).Tracking and follow-up methods for research on homelessness. *Evaluation Review, 17* (3), 331-352.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design and analysis issues for field settings.* Chicago, IL: Rand McNally.

Courtney, M. E., Piliavin, I., & Wright, B. R. E. (1997). Transitions from and returns to out-of-home care. *Social Service Review, 71* (4), 652-667.

Devine, J. A., Brody, C. J., & Wright, J. D. (1997). Evaluating an alcohol and drug treatment program for the homeless: An econometric approach. *Evaluation and Program Planning, 20* (2), 205-215.

Foster, E. M., & Bickman, L. (1996). An evaluator's guide to detecting attrition problems. *Evaluation Review, 20* (6), 695-723.

Goodfellow, M., Kiernan, N., Ahern, F., & Smyer, M. A. (1988). Response bias using two-stage data collection: A study of elderly participants in a program. *Evaluation Review, 12* (6), 638-654.

Greene, W. H. (1981). Sample selection bias as a specification error: Comment. *Econometrica, 49* (3), 795-798.

Greene, W. H. (1995). *LIMDEP user's manual*. Plainview, NY: Econometric Software.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Greenwell, L., & Bengston, V. L. (1997). Geographic distance and contact between middle-aged children and their parents: The effects of social class over 20 years. *Journal of Gerontology, 52B* (1), 513-526.

Grogan-Kaylor, A. (2001). The effect of initial placement into kinship foster care on reunification from foster care: A bivariate probit analysis. *Journal of Social Service Research, 27* (4), 1-31.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimation for such models. *Annals of Economic and Social Measurement, 5* (4), 475-492.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica, 46* (6), 931-959.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47* (1), 153-161.

Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives, 9* (2), 85-110.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies, 64*, 605-654.

Hensher, D. A. (1987). Issues in the pre-analysis of data. *Transportation Research-A, 21A* (4/5), 265-285.

Jaeger, D. A. (1993). Sample selection-corrected estimation ("Heckit") [Computer program macro]. SAS Institute [On-line]. Available: http://ftp.sas.com/techsup/download/stat/heckman.html

Jensen, E. (1985). Desired fertility, the "Up to God" response, and sample selection bias. *Demography, 22* (3), 445-454.

Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons, Inc.

Lauby, J., Kotranski, L., & Halbert, J. (1996). Effects of intervention attrition and research attrition on the evaluation of an HIV prevention program. *Journal of Drug Issues, 26* (3), 663-667.

Lee, B., & Marsh, L. C. (2000). Sample selection bias correction for missing response observations. *Oxford Bulletin of Economics and Statistics, 62* (2), 305-323.

Leonard, D. K., & Jiang, J. (1999). Gender bias and the college predictions of the SATs: A cry of despair. *Research in Higher Education, 40* (4), 375-407.

Littell, J. H. (1997). Effects of duration, intensity, and breadth of family preservation services: A new analysis of data from the Illinois Family First experiment. *Children and Youth Services Review, 19* (1/2), 17-39.

McDonald, T. P., Moran, J. R., & Garfinkel, I. (1990). Absent father's ability to pay more child support. *Journal of Social Science Research, 13* (4), 1-18.

McGuigan, K. A., Ellickson, P. L., Hays, R. D., & Bell, R. M. (1997). Adjusting for attrition in school-based samples: Bias, precision, and cost trade-offs of three methods. *Evaluation Review, 21* (5), 554-567.

Miller, R. B., & Wright, D. W. (1995). Detecting and correcting attrition bias in longitudinal family research. *Journal of Marriage and the Family, 57* (4), 921-929.

Moffitt, R. (1991). Program evaluation with non-experimental data. *Evaluation Review, 15* (3), 291-314.

Orme, J. G., & Buehler, C. (2001). Introduction to multiple regression for categorical and limited dependent variables. *Social Work Research, 25*(1), 49-61.

Orme, J. G., Buehler, C., McSurdy, M., Rhodes, K. W., Cox, M. E., & Patterson, D. A. (in press). Parental and familial characteristics of family foster care applicants. *Children and Youth Services Review.*

Orme, J. G., Buehler, C., McSurdy, M., Rhodes, K. W., Cox, M. E., & Cuddeback, G. (in press). Parental and familial characteristics used in the selection of family foster care applicants. *Children and Youth Services Review.*

Polak, J. (1999). Empirical analysis of attrition and underreporting in mailback and personal interview panel surveys. *Transportation Research Record, 1660*, 164-171.

Reynolds, A. J., & Temple, J. A. (1995). Quasi-experimental estimates of the effects of a preschool intervention. *Evaluation Review, 19*(4), 347-373.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi- experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

StataCorp. (2001). *Stata Statistical Software: Release 7.0*. College Station, TX: Stata Corporation.

Stolzenberg, R. M., & Relles, D. A. (1997). Tools for intuition about sample selection bias and its correction. *American Sociological Review, 62* (3), 494-507.

Treno, A. J., Gruenewald, P. J., & Johnson, F. W. (1998). Sample selection bias in the emergency room: An examination of the role of alcohol. *Addiction, 93* (1), 113-129.

Vartanian, T. P. (1999). Adolescent neighborhood effects on labor market and economic outcomes. *Social Service Review, 73* (2), 142-167.

Vella, F., & Verbeek, M. (1999). Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics, 90* (2), 239-263.

Winship, C., & Mare, R. D. (1992). Models for sample selection bias. *Annual Review of Sociology, 18*, 327-350.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology, 25*, 659-706.

York, R. O. (1998). *Conducting social work research.* Boston: Allyn & Bacon.

RECEIVED: 05/02
REVISED: 01/03
ACCEPTED: 01/03