February 2007

Attrition Bias

Richard B. Miller

Cody S. Hollist
University of Nebraska Lincoln, chollist2@unl.edu

Published in *Encyclopedia of Measurement and Statistics*. Ed. Neil Salkind. 3 vols. Thousand Oaks: Sage Reference, 2007. Vol. 1. Pages 57-60.

ATTRITION BIAS

When data are collected over two or more points in time, it is common for some participants to drop out of the study prematurely. The attrition of the original sample can occur in longitudinal research as well as in experimental designs that include pretest, posttest, and follow-up data collection. In longitudinal research, which often lasts many years, some participants move between data points and cannot be located. Others, especially older persons, may die or become too incapacitated to continue participation in the study. In clinical treatment studies, there may be barriers to continued participation in the treatment program, such as drug relapse or lack of transportation.

Attrition of the original sample represents a potential threat of bias if those who drop out of the study are systematically different from those who remain in the study. The result is that the remaining sample becomes different from the original sample, resulting in what is known as attrition bias. However, if sample attrition over time is not systematic, meaning that there are no unique characteristics among those who drop out, then there is no attrition bias, even though the sample has decreased in size between waves of data collection. It is important, then, for researchers who collect multiple waves of data to check for attrition bias.

Attrition bias is one of the major threats to multiwave studies, and it can bias the sample in two ways. First, attrition bias can affect the external validity of the study. If some groups of people drop out of the study more frequently than others, the subsequent longitudinal sample no longer resembles the original sample in the study. As a result, the remaining sample is not generalizable to the original population that was sampled. For example, a longitudinal sample examining the grieving process of women following the death of a spouse may fail to retain those participants who have become too distraught to fill out the questionnaire. The nonparticipation of this group may bias the findings of the study toward a minimization of depressive symptomatology as a component of the grieving process. In other words, the composition of the sample changes to the point that the results are no longer generalizable to the original population of widows.

Second, systematic, as opposed to random, attrition can negatively affect the internal validity of the study by altering the correlations among the variables in the study. This problem occurs in longitudinal research because the subsamples that are dropping out of the study at a higher rate are underrepresented in the longitudinal sample, which may lead to correlations between variables that are different from the true correlations in the original sample. For example, the underrepresentation of widows with depressive symptomatology in the second or third wave of a study may alter the correlation between insomnia and length of time since the death of the spouse.

Selective attrition affects the internal validity of experimental research when there are differential dropout rates between the treatment and control groups. In a clinical trial of a depression treatment, if the participants in the treatment group drop out at a higher rate than do the participants of the control group, the results of the study will be biased toward showing artificially successful treatment effects, thus compromising the internal validity of the study. However, if the dropout rates are comparable, the threats to internal validity due to attrition are minimal.

Preventing Attrition

Because of the threat of attrition bias to the external and internal validity of studies, it is important to minimize sample attrition when conducting multiwave research. Researchers who have conducted experimental and longitudinal research have made a number of recommendations and suggestions to reduce sample attrition. Mason emphasized the importance of creating a project identity, offering cash and other incentives, developing a strong tracking system to constantly identify the location and status of participants, and keeping follow-up interviews brief. Others recommend collecting detailed contact information about participants to increase the likelihood of locating them for the second and subsequent interviews. Follow-up postcards and telephone reminders also help retain participants in the sample.

Detecting Attrition Bias

Differences in characteristics between those who prematurely drop out a study ("droppers") and those who remain in the sample ("stayers") can be assessed by conducting a logistical regression analysis. Because both groups participated in the first wave of the study, data are available on which to compare the two groups. A dichotomous dependent variable is created with 1 representing the stayers and 0 representing the droppers. Variables from the first wave of data are used as independent variables in the analysis. These variables should include key demographic variables, such as race, income, age, and education, as well as substantive variables that are salient in the study, such as depression, drug abuse, or marital quality. A Statistically significant coefficient for any of the variables means that there is a difference between the stayers and the droppers, indicating attrition bias.

Threats to internal validity due to attrition bias can be tested by comparing the first-wave correlation matrices of the overall sample and the longitudinal sample, which includes only the stayers. This can be done in two ways:

1. Each of the correlation coefficients (for example, the correlation between age and level of depression) is compared using Fisher's *z* Statistical test. A significant *z* score means that the two coefficients are Statistically significantly different, indicating attrition bias.
2. A structural equation modeling program, such as LISREL or AMOS, can be used to test whether the two correlation matrices are invariant, that is, the same. If the test for invariance is nonsignificant, then the two matrices are assumed to be equivalent, with no apparent attrition bias.

Correcting Attrition Bias

Although the strategies used to detect attrition bias are straightforward, there is substantial debate about appropriate strategies to correct attrition bias. Despite the lack of consensus, though, the need for correcting the problem of attrition bias is crucial and continues to motivate Statisticians to pursue solutions.

Correction of nonrandom attrition can be broken into two categories. The first category is correction of data when the mechanism of dropping out is known, or in other words, when the researcher knows which characteristics are related to dropping out of the study. The second category is attrition whose causes the researcher does not know.

Known Cause of Attrition

When the cause of attrition is known, the researcher can take steps to control the data analysis procedure to account for the missing data. A model has been developed that simultaneously calculates the research question and the mechanism for missing data. This model is a sample selection model in which two simultaneous regression models are calculated. The first model is a regression model that addresses the research question, with the hypotheses of the study being examined by the regression of the dependent variable on the key independent variables in the study. The second model includes the variables that are causing attrition, with the dependent variable being a dichotomous variable indicating either continued participation or nonparticipation in the study. The error terms of the substantive dependent variable in the first regression model and the participation dependent variable in the second regression model are correlated. A significant correlation between the two error terms indicates attrition bias. If the correlation is significant, the inclusion of the second model provides corrected regression coefficients for the first, substantive regression model. Thus, the inclusion of the second model that examines attrition bias serves as a correction mechanism for the first, substantive model and enables the calculation of unbiased regression coefficients.

Unknown Cause of Attrition

Heckman proposed a two-step procedure to correct for attrition bias when the cause of the attrition is not readily apparent. He conceptualized the issue of attrition bias as a specification error, in which the variable that accounts for systematic attrition in the study is not included in the regression equation. This specification error results in biased regression coefficients in the analysis. His solution is to first create a proxy of the variable that explains attrition. This is done by conducting a logit regression analysis, similar to the one described in the section on detecting attrition bias. The dependent variable is whether or not each participant participated in the second wave of data collection, and the independent variables are possible variables that may explain or predict dropout. This first step not only tests for attrition bias but also creates an outcome variable, which Heckman calls λ (lambda). Thus, a λ value is computed for all cases in the study, and it represents the proxy variable that explains the causation of attrition in the study.

The second step of Heckman's procedure is to merge the λ value of each participant into the larger data set and then include it in the substantive analysis. In other words, the λ variable is included in the regression equation that is used to test the hypotheses in the study. Including λ in the equation solves the problem of specification error and leads to more accurate regression coefficients.

While Heckman's model has been used by longitudinal researchers for many years, some concerns have arisen regarding its trustworthiness.

Stolzenberg and Relles argue that Heckman's model has been shown to compute inaccurate estimates, and they suggest several cautions when using his model. Nevertheless, Heckman's model offers a possible solution when systematic attrition threatens to bias the results of a study.

—Richard B. Miller and Cody S. Hollist

See also **Longitudinal/Repeated Measures Data**

Further Reading

Boys, A., Marsden, J., Stillwell, G., Hatchings, K., Griffiths, P., & Farrell, M. (2003). Minimizing respondent attrition in longitudinal research: Practical implications from a cohort study of adolescent drinking. *Journal of Adolescence*, *26*, 363–373.

Cuddeback, G., Wilson, E., Orme, J. G., & Combs-Orme, T. (2004). Detecting and Statistically correcting sample selection bias. *Journal of Social Service Research*, *30*, 19–33.

Goodman, J. S., & Blum, T. C. (1996). Assessing the nonrandom sampling effects of subject attrition in longitudinal research. *Journal of Management*, *22*, 627–652.

Graham, J. W., & Donaldson, S. I. (1993). Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow up data. *Journal of Applied Psychology*, *78*, 119–128.

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

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

Jacobson, J. O. (2004). Place and attrition from substance abuse treatment. *Journal of Drug Issues*, *34*, 23–50.

Mason, M. (1999). A review of procedural and Statistical methods for handling attrition and missing data in clinical research. *Measurement and Evaluation in Counseling and Development*, *32*, 111–118.

Miller, R. B., & Wright, D. (1995). Correction for attrition bias in longitudinal analyses. *Journal of Marriage and the Family*, *57*, 921–929.

Stolzenberg, R. M., & Relles, D. A. (1997). Tools for intuition about sample selection bias and its correction. *American Sociological Review*, *73*, 142–167.