Advances in Research on Participant Attrition in Prevention Intervention Studies

John W. Graham, Lauren E. Connell, and Michael L. Hecht
Penn State University

This paper describes a promising approach to dealing with participant attrition, a methodological challenge common in prevention intervention research. Attrition undermines the internal validity of studies evaluating the efficacy/effectiveness of preventive interventions. Reducing the impact of attrition is key to these essential evaluations.

Graham (2012) suggested several strategies for studying attrition, and outlined a taxonomy of eight cases of attrition that consider all possible combinations of the treatment (vs. control: T), the dependent variable (Y), and the interaction between them (TY) as causes of missingness. Attrition is not a problem in two cases (case 1 = not T, not Y, not TY as causes; case 2 = T only as cause), because missingness is either MCAR or MAR. Previous research (Collins, Schafer, & Kam, 2001; Graham et al., 2008; Graham, 2012) has shown minimal attrition bias in a third case for the regression coefficient for T predicting Y (βYT) (Case 3: Y only as cause). The remaining five cases have not been studied.

This paper describes results of a Monte Carlo simulation that extends the work of Collins, Schafer, & Kam (2001) to cover the five cases of the taxonomy that have not been studied previously in depth. Our simulation focuses on case 8, in which T, Y, and the TY interaction are all responsible, to varying degrees, for the observed missingness on Y. Our work focused on what Collins et al. (2001) referred to as MNAR-Linear missingness, and we created missingness using the same strategy they used. We studied several key factors relating to missingness: overall % missing; the T effect (difference in % missing between T and C groups); the overall Y effect (using a quantity we call "range": the difference in the probability of Y missing between the 4th and 1st quartiles of Z); and the TY effect (difference in "range" between T and C groups). We also examined two values for the quantity ρZY, the correlation between the cause of missingness, Z, and the main dependent variable, Y. Our work makes use the definition suggested by Collins, Schafer, & Kam (2001) for missing data bias that is not significant in a practical sense; we judged bias not to be of practical significance as long as the absolute value of the standardized bias was less than 40. Results of our simulation show clear patterns of demarcation between combinations of factors that do and do not show bias that is significant in a practical sense. My presentation will present these results.

Estimating the various quantities (% missing; T effect; Y effect, TY effect; ρZY) in empirical data is key to estimating the impact of bias due to attrition. The % missing, T effect, and ρZY quantities are all easily estimated in empirical studies with longitudinal follow-up data. The Y and TY effects are more difficult to estimate, but with longitudinal follow-up data, plausible estimates of these quantities are possible. Making judicious use of the estimates of all these quantities in empirical data, in conjunction with careful sensitivity analyses, one can make judgements in a particular empirical study about the likelihood that attrition has caused bias that is significant in a practical sense.

Graham (2012) also suggested data collection methods for reducing the impact of attrition. One strategy involves asking attrition relevant questions (e.g., whether the participant will be available for the next wave of measurement). The benefits of this strategy have not been studied previously in empirical research. Using our simulation results, and our strategies for estimating the various quantities in empirical data, we demonstrate the benefits of using the attrition-relevant measures to reduce attrition bias.