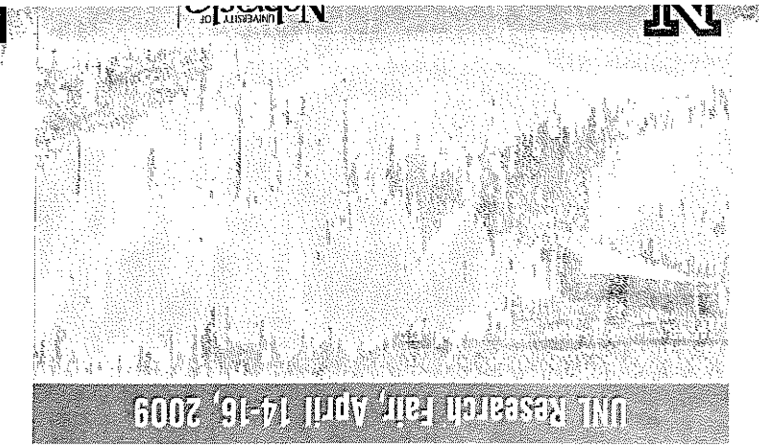

Lawrence Erlbaum Associates, Inc., Publishers
10 Industrial Avenue
Mahwah, NJ 07430
www.erlbaum.com

Cover design by Kathryn Houghtaling

Cover graphics by Elizabeth K. McConnell and Todd D. Little

Library of Congress Cataloging-in-Publication Data

Modeling Contextual Effects in Longitudinal studies

ISBN 978-0-8058-5019-2–ISBN 0-8058-5019-8 (cloth)
ISBN 978-0-8058-6207-2–ISBN 0-8058-6207-2 (pbk)
ISBN 978-1-4106-1587-9–ISBN 1-4106-1587-1 (e book)

Copyright information for this volume can be obtained by contacting the Library of Congress.

Books published by Lawrence Erlbaum Associates are printed on acid-free paper, and their bindings are chosen for strength and durability.

The final camera copy for this book was prepared by the author and therefore the publisher takes no responsibility for consistency or correctness of typographical style.

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

CHAPTER TWO

Statistical Analysis With Incomplete Data: A Developmental Perspective

Scott M. Hofer
Oregon State University

Lesa Hoffman
University of Nebraska

One of the thorniest problems that developmental researchers must face is that of *missing* or *incomplete data*. Incomplete data can take many forms, such as item or scale nonresponse, participant attrition (i.e., study drop-out), and mortality within the population of interest (i.e., lack of initial inclusion or incomplete follow-up due to death). Statistical analysis in general is aimed at providing inferences regarding level or state, subgroup differences, variability, and construct relations within a population, and incomplete data complicate this process. In order to make appropriate population inferences about development and change, it is important not only to consider thoroughly the processes leading to incomplete data, but also to obtain measurements of these selection and attrition processes to the greatest extent possible. A developmental, ecological perspective is useful in this regard by providing a framework in which to consider the impact of many static and dynamic individual and contextual factors on selection and attrition processes in addition to the impact these factors have on developmental processes of interest.

In this chapter, we outline some of the general issues for making statistical inferences to populations when data are incomplete. We begin by briefly reviewing statistical theory as to the different types of incomplete data, as well as their likely sources within a developmental context. An overview of common approaches for analysis with incomplete data is then provided, including two-stage multiple imputation and single-stage likelihood-based estimation procedures. Current approaches for the analysis with incomplete data rely on the statistical assumption that the data are at least *Missing at Random* (where the probability of missing information is related to covariates and previously measured outcomes), and in the latter parts of this chapter we emphasize what this means in terms of measurements and modeling approaches for developmental studies. We describe several statistical methods through which the impact of contextual variables on selection and attrition processes can be incorporated properly. In the final section, we discuss complications that can result from heterogeneity in the timing and sources of incomplete data when the goal is to make inferences about developmental processes.

Methods for addressing incomplete data and their statistical complexities remain a very active area of research. Thus, we aim to provide a conceptual understanding of these methods, directing readers to current resources for analysis with incomplete data, rather than to provide a complete tutorial. Excellent overviews of current methods for addressing incomplete data can be found, among others, in Allison (2002), Diggle, Liang, and Zeger (1994), Graham, Cumsille, and Elek-Fisk (2003), and Schafer and Graham (2002), as well as the larger works of Little and Rubin (1987, 2002) and Schafer (1997). Our emphasis in this chapter is on the importance of considering multiple types of processes that may lead to incomplete data in order to achieve appropriate population inferences about developmental processes.

STATISTICAL THEORY FOR THE ANALYSIS OF INCOMPLETE DATA

We have seen large strides in recent decades in the development of methods with which to address incomplete data in statistical analyses. However, while these methods may offer a ray of hope to the analyst faced with copious amounts of missing data, in exchange one must be willing to make certain assumptions regarding the missing data mechanisms, or reasons why certain values are missing. If these assumptions are not plausible, one could end up making inferences that are at best underpowered, and at worst, wrong. In this section we provide a brief overview of the types of mechanisms that can lead to incomplete data, as was first explicated by Rubin (1976) (see also Allison, 2002; Little & Rubin, 1987, 2002). Despite the somewhat nonintuitive nature of these terms, they are standard among methodologists and users of missing data treatments, and as such will be used here as well.

Missing Completely at Random (MCAR)

The most restrictive assumption that can be made regarding the nature of the missingness in one's data is known as *Missing Completely at Random* (MCAR), in which the probability of data being observed does not depend on the value of the missing information or that of any other variables in the dataset. Despite the use of the term random, systematic patterns of missing data can still qualify as MCAR if the mechanism generating the missing data is unrelated to the outcomes of interest. For example, items skipped accidentally on a questionnaire or data lost to equipment, computer, or experimenter error could likely be considered MCAR. Similarly, missingness could also be considered MCAR if it results from participants being absent from data collection for reasons unrelated to the variables of interest.

In contrast to unintentional missingness, incomplete data may be intentionally introduced by design in order to minimize response burden or fatigue effects yet maximize statistical power and breadth of measurement. In this approach, often referred to as *planned missingness* (e.g., Graham, Hofer, & MacKinnon, 1996; Graham, Hofer, & Piccinin, 1994; Graham, Taylor, & Cumsille, 2001; McArdle, 1994), different items or variables are collected purposively across separate subsets of the same sample, and missing data methods are used to analyze data across the subsets as if it were complete in the full sample. Planned missingness is a superior alternative to unintentional missingness, in that the causes of missingness are known and unrelated to participant characteristics in the former case, but may be due to unmeasured characteristics of the participant in the latter. Any data that are missing due to planned non-administration (e.g., random assignment of different forms) would be considered Missing Completely at Random, provided that data were indeed collected from everyone as intended.

Missing at Random (MAR)

A less-restrictive assumption is that of *Missing at Random* (MAR), also known as *Ignorable, Accessible,* or *Non-Informative Missingness.* Simply put, in this scenario the probability of having missing information on a given variable is unrelated to the missing values themselves *after* controlling for other variables that are related to the missingness on that variable. The assumption of MAR is often realistic provided that individual or contextual covariates are collected that predict the probability of missingness. For example, consider nonresponse to a survey question about annual income. It is possible that the question was skipped accidentally, but it is also possible that persons with lower or higher incomes would be less likely to answer questions regarding their annual incomes.

If the researchers have collected other variables from the participants that are likely to be related to annual income, such as educational attainment or type of employment, they could make use of those "proxy" variables in order to continue to make appropriate inferences regarding the distribution of the annual income variable or the relationship of annual income with other variables in the sample. Within a longitudinal context, missingness at a given occasion of measurement could be considered MAR if the values of the missing variable could be predicted from those obtained on previous occasions.

Not Missing at Random (NMAR)

Finally, the least restrictive assumption regarding possible mechanisms of missingness is that of *Not Missing at Random* (NMAR; also referred to as *Missing Not at Random*), also known as *Non-Ignorable*, *Non-Accessible*, or *Informative Missingness*. NMAR represents the worst-case scenario for a data analyst, in that the probability of missingness on a given variable is related to the missing values themselves after controlling for other relevant variables. Let us reconsider the earlier examples, If the probability of skipping a given item on a questionnaire was related to the response of that item (e.g., persons with lower incomes were less likely to report their incomes) and no other information was available that was related to this nonresponse, the missingness would be considered NMAR. Similarly, in computer-administered timed tasks, if computer error was more likely on trials with longer response times, and no other data was available that was related to response time for that task, the missingness would be considered NMAR. Finally, if the probability of missing a measurement occasion is related to the values that would have been obtained at that occasion or in the future (i.e., unobserved variables), but was not predictable from previous observations, the missingness would again be considered NMAR. When incomplete data arise from NMAR processes, the estimation of parameters will be biased in unknown ways when analyzed using methods that assume the data are at least MAR.

SOURCES OF INCOMPLETE DATA: TYPES OF ATTRITION PROCESSES

In evaluating which mechanism or mechanisms might be responsible for incomplete data within a given study, one must consider the substantive processes that could be involved. An important first step is to distinguish developmental processes of substantive interest (i.e., as measured in study outcomes) from other developmental processes that may be related to item nonresponse (e.g., fatigue, lack of motivation, embarrassment) and unit nonresponse (e.g., study attrition, mortality). We next consider the case of unit nonresponse or study attrition from a developmental orientation and the extent to which such incomplete data can be considered as mere nuisance, as a natural process, or as a problem of threshold or censoring.

Attrition as a Nuisance

In many cases, the processes leading to nonresponse or attrition may not be of direct interest. For example, the moving away of participants from a longitudinal data collection project is an unfortunate event, but may not be substantively noteworthy, assuming the move was due to reasons unrelated to the areas under investigation. Similarly, computer or equipment malfunction is a process one typically wishes to minimize, not explore further. Incomplete data is assumed to arise from nuisance processes such as these in many substantive analyses. In these cases, the emphasis is on the best use of all available data in order to achieve appropriate population inferences in the presence of missing data. Covariates for missingness (i.e., variables related to the probability of missingness) may not be available or relevant, and as a result may not be incorporated into the method of addressing the missingness. The statistical analysis procedures reviewed in the next section are well-suited for this scenario, and permit estimation of unbiased and efficient population parameters when data are Missing Completely at Random (i.e., when the probability of missingness is unrelated to the process under study) or Missing at Random (i.e., unrelated after accounting for covariates related that can predict the probability of missingness).

Attrition as a Natural Process

In contrast to the previous examples in which missing data are largely seen as a hurdle to overcome, incomplete data can also be the natural result of developmental and population aging processes. For example, attrition in studies of aging is often nonrandom, or *selective*, in that it is likely to result from mortality or declining physical and mental functioning of the participants over the period of observation (Cooney, Schaie, & Willis, 1988; Rabbitt, Watson, Donlan, Bent, & McInnes, 1994; Riegel, Riegel, & Meyer, 1967; Siegler & Botwinick, 1979; Streib, 1966). In diary studies of daily experiences, participants may vary in compliance as a function of individual characteristics as well as the types of experiences they are recording (Bolger, Davis, & Rafaeli, 2003). In studies of adolescent substance abuse, high levels of alcohol or drug use at the previous time point may be related to attrition at the next time point (Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997). In long-term intervention studies, participants who do not feel they are benefiting from the treatment (i.e., those in the placebo group) as well as those who feel they have already improved 'enough' may opt to discontinue participation (Hedeker & Gibbons, 1997). In these examples the fact that the values are missing is likely to be informative of the value that would have been obtained (Diggle & Kenward, 1994).

These scenarios can present an important inferential problem for the analysis of longitudinal studies, in that at each new wave of assessment the sample becomes less and less representative of the population from which it originated. As such, generalizations from the sample of continuing participants to the initial population may become difficult to justify (e.g., Nesselroade, 1988; Vaupel & Yashin, 1985). It is important to note, however, that the problem of nonrandom attrition is not unique to longitudinal studies. In cross-sectional studies, nonrandom population attrition (i.e., mortality; morbidity) manifests itself as nonrandom initial sample selection. Particularly pertinent to research on aging, cross-sectional samples of individuals of different ages are necessarily comprised of individuals who are the surviving members of the population, making inferences to a single population of "aging" individuals difficult to justify. Such differential population selection (mortality, morbidity) is often related to the processes of interest and cannot be evaluated in cross-sectional studies, given that data are typically not collected on nonparticipants. In longitudinal studies, however, information about the nonreturning participants prior to their departure may be available and the relationship with the probability of missingness can be carefully considered (e.g., Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997). Inferences to population subgroups or parameters conditional on both age and survival are possible in such cases.

In scenarios of selective nonresponse or attrition such as these, it is critically important to consider possible sources of missing data during the early design stages of a study in order to obtain measurements on individual and contextual covariates that are likely to be related to the probability of missingness. It is only through the appropriate inclusion of such covariates in the subsequent models that the assumption on which most statistical methods of addressing missing data are based, that of Missing at Random (MAR), can possibly be satisfied.

Problems of Threshold or Censoring

In addition to missing data arising from nuisance or substantive processes, incomplete data can also be the result of problems of range restriction within a measurement instrument. Sometimes measurement instruments may not be appropriate for all individuals within a defined population, or the phenomenon of interest may not apply to all individuals. For example, the Mini-Mental Status Exam (Folstein, Folstein, & McHugh, 1975) is often given as a measure of general cognitive status in older adults, but most healthy adults will score at ceiling (i.e., will be right-censored), with deficits observed only for those with severely declining abilities, such as in advancing dementia. Thus, the measure may only be useful within certain subsets of the sample. Similarly, in studies of substance abuse, adolescents who do not smoke or drink will necessarily score at the floor of a measure of amount or intensity of use (i.e., will be left-censored), resulting in a zero-inflated distribution of substance use, given that only those who engage in the activity can logically vary in amount or intensity (Brown, Catalano, Fleming, Haggerty, & Abbot, 2005).

In these instances, although the censored data are not truly missing, they are also often not informative about individual characteristics for portions of the sample or for certain time points (i.e., for which measurements away from ceiling or floor could not be obtained). Under such conditions, analyses may be performed for the subsample of individuals with noncensored responses, an approach that is less than optimal. An alternative approach has been recently developed by Olsen and Schafer (2001); (see also Brown, Catalano, Fleming, Haggerty, & Abbot, 2005), that of two-part latent growth curve models. In this approach, responses on a single censored outcome are modeled as two distinct variables: A dichotomous indicator for whether or not each case is censored (e.g., whether or not the participant smokes), and a continuous indicator of the value if not censored (e.g., number of cigarettes smoked per day). Predictors of each outcome can then be evaluated simultaneously. Although use of the two-part latent growth curve models thus far has largely been limited to substance use research, they are likely to have many other applications as well.

STATISTICAL APPROACHES FOR ANALYSIS WITH INCOMPLETE DATA

The options for addressing incomplete data within a statistical analysis depend largely on the hypothetical reasons for missingness. If the source of the missingness for a given variable can be considered Missing Completely at Random (MCAR), then many options exist for addressing this missingness. These include *listwise deletion* (i.e., complete cases analysis), the old standby and default in many statistical analysis packages, as well as newer methods based on maximizing likelihoods or multiple imputation, as presented in the next section. In the case of MCAR, regardless of which method is used, model parameters (e.g., means, variances, correlations, or regression weights) from analyses including the variable with missingness are likely to be *unbiased*, which means that the obtained parameter estimate will be a close match to the value that would have been obtained had the data been complete. Parameter estimates may vary across missing data methods in their *efficiency*, the extent to which the standard error around the parameter estimate is as small as it would have been if the data were complete (Graham, Hofer, & MacKinnon, 1996; Schafer & Graham, 2002). This lack of efficiency translates directly into a loss of statistical power and a greater likelihood of a Type II error (i.e., failing to reject the null hypothesis when it should be rejected). Thus, in the case of MCAR, maximum likelihood and multiple imputation methods can offer considerable improvements over listwise deletion.

If the missingness on a given variable is not Missing Completely at Random (MCAR), then the use of listwise deletion will likely lead to biased as well as inefficient estimates, and is generally not recommended. It is important to note that single imputation methods, such as mean-based or regression-based imputation, are *not* recommended. These procedures are known to result in biased estimates if the data are not MCAR although with very low proportions of missing values the bias may be negligible (but the magnitude of bias remains unknown unless more appropriate procedures are used for comparison). A major problem with these approaches is that there is no appropriate statistical basis with which to obtain standard errors of the parameter estimates (Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997; Graham, 2003).

When it appears that either Missing at Random (MAR) or Not Missing at Random (NMAR) is applicable, then the analyst must carefully consider the various mechanisms behind the probability of missingness and incorporate all individual or contextual covariates that could be related to the probability of missingness for each variable within the missingness model (as discussed in the next section). Maximum likelihood and multiple imputation methods for treating missing data carry with them the assumption that the missingness is at least Missing at Random (MAR). Unfortunately, unlike other statistical assumptions such as multivariate normality, one cannot empirically evaluate the extent to which the missingness in the to-be-analyzed data can be considered ignorable (MAR) or nonignorable (NMAR). Thus, the appropriateness of missing data treatments with regard to the inferences that can be made to the target population depends largely on principled argument and the availability of measured covariates at the individual and contextual levels that can presumably capture the missingness processes. We return to this point later in the chapter.

In the next section, we describe briefly the available statistical approaches for the analysis of incomplete data, assuming data are Missing at Random. These include two-stage approaches, the *expectation-maximization (EM) algorithm* and *multiple imputation*, in which the missing data model is generated separately from the substantive model, such that a complete analysis requires both steps. In contrast, in single-stage models that make use of *full information maximum likelihood* (FIML) procedures, the missing data model and the substantive model can be estimated simultaneously (i.e., in one step). We also briefly describe alternative approaches that do not assume Missing at Random.

Expectation-Maximization (EM) Algorithm

Dempster, Laird, and Rubin (1977)(see also Little & Rubin, 1987, 2002) described the utility of the *expectation-maximization* (EM) algorithm for analysis with incomplete data. A general form of the EM algorithm is used to obtain sufficient statistics, such as covariances and means, on which other forms of statistical analysis can be performed. In a typical EM algorithm, one begins with user-specified starting values for the variances, covariances, and means of the variables in the dataset or rely on listwise deletion to provide starting values. In the *expectation* step, the "best guess" is filled in for any missing value based on regression equations in which each variable with missing values serves as an outcome and all other variables serve as predictors. In the *maximization* step, means for the newly completed data are calculated in the typical manner, but the variances and covariances are calculated with additional components of variance added to them in order to correct for underestimation. The variances, covariances, and means are then compared to those given as starting values. The new estimates of these parameters are then used to update the regression equations for use in a second expectation step, followed by a second maximization step, and the updated estimates are again compared to those from the previous run. The EM algorithm continues to repeat until the estimates change a negligible amount between iterations.

The most typical implementation of the EM algorithm is the generation of maximum likelihood estimates of variances, covariances, and means for continuous variables that may then be analyzed with other statistical analysis programs (e.g., general linear models, structural equation models). The estimates produced using the EM algorithm are unbiased and efficient under the assumption of MAR (Graham et al., 1994; Graham et al., 1997). However, standard errors for the estimates must be generated separately, such as with a bootstrap procedure (Efron & Tibshirani, 1998), given the different sample sizes on which each of the parameters (e.g., covariances) are based. Nevertheless, this approach is useful for providing maximum likelihood estimates of variances, covariances, and means for reporting summary descriptive statistics even when single-stage methods of model estimation are used (as described shortly). Several software programs are available that use this method, including EMCOV (Graham & Hofer, 1993), NORM (Schafer, 1997), S-PLUS, SAS Proc MI, and SPSS Missing Value Analysis.

Multiple Imputation (MI)

Multiple imputation (MI; Rubin, 1987) permits the analysis of "complete" data sets within standard statistical models (e.g., general linear models), with the additional strength that standard errors for model parameters can be obtained that properly account for both between-imputation and within-model variability. The MI procedure accounts for missing data in an initial step, referred to as the *missing data model*. The missing data model need not be the same as the substantive model of interest, and should contain all covariates believed to relate to the probability of missingness across variables. Missing values are "filled in" or *imputed* based on regression-predicted values (in which all other variables in the missing data model serve as predictors) along with a random error term. A series of data imputations is performed, in which multiple "complete" data sets are generated from the missing data model. Usually between 5 and 10 imputations are sufficient, but up to 20 imputations may be necessary when large proportions of data are missing (see Schafer, 1997). These MI models can be performed on mixtures of continuous and categorical outcomes and covariates.

The next step is to estimate the statistical model of interest within each imputed or "complete" data set. Because the issue of informativeness (i.e., nonrandom missingness) has already been addressed by the covariates within the missing data model, only covariates of substantive interest need to be included in the substantive model. Finally, the parameter estimates and their standard errors from each of the substantive models need to be combined appropriately. Although the parameter estimates can simply be averaged across models, the aggregation of their standard errors requires the application of Rubin's (1987) rules, which include both between-imputation and within-model variability in arriving at the final standard errors. Several software programs make data imputation and the combining of results with Rubin's rules quite easy, including NORM (Schafer, 1997), the SAS programs of Proc MI and Proc MI Analyze, SPLUS, and LISREL. Detailed examples for multiple imputation with NORM are provided by Graham and Hofer (2000) and Graham, Cumsille, and Elek-Fisk (2003). Allison (2002) provides detailed examples using the SAS MI procedures.

Full Information Maximum Likelihood (FIML)

In contrast to the two-stage approaches in which missing data and substantive statistical models are considered in separate stages, in *full information maximum likelihood* (FIML; also known as *Direct ML*) methods, substantive model parameters can be estimated from incomplete data in a single step without any additional iterations or calculations (Little & Rubin, 1987, 2002). Maximum likelihood estimates of the variances, covariances, and means can be generated for reporting purposes as well. Although much more convenient than multiple imputation, the use of FIML carries with it the assumption of Missing at Random (MAR), or that the probability of missingness is unrelated to what the missing values would have been. Thus, all covariates related to the probability of missingness (e.g., previous observations, individual or contextual characteristics) need to be included in the substantive model, which can be difficult to accomplish in practice. In particular, cases with any missing covariates will not be included in the substantive model within certain types of multilevel modeling programs (e.g., HLM, SAS Proc Mixed, MLwin), although this requirement can be relaxed to include cases with partially observed covariates in other general programs for multilevel and/or structural modeling that also use FIML (e.g., Mplus, Mx, AMOS, LISREL, EQS).

Within the context of general structural equation modeling, Graham (2003) discussed the issue of how to properly include covariates for missingness that are not of substantive interest. He presented two FIML-based structural models, the extra dependent variable model and the saturated correlates model, through which one can include covariates for missingness in such a way so as not to distort the substantive model. In the *extra dependent variable model*, covariates for missingness are specified as dependent variables (i.e., as endogenous variables predicted by the exogenous variables), and their residual variances are correlated with those of other endogenous variables. In the *saturated correlates model*, covariates for missingness are specified as independent variables that are correlated directly with other exogenous variables, and correlated with the residual variances of endogenous variables. Covariates for missingness are allowed to correlate with each other in both models. Simulation results revealed that both models performed as well as two-stage approaches (e.g., multiple imputation, EM algorithm) in terms of recovering parameter estimates and standard errors. With regard to assessing model fit, however, the saturated correlates model resulted in model fit statistics equivalent to the substantive model (i.e., without covariates for missingness), and was to be preferred over the extra dependent variable model, for which discrepancies were found.

An older, related method is that of multiple group structural equation modeling (see Allison, 1987; Graham, Hofer, & Piccinin, 1994; Muthèn, Kaplan, & Hollis, 1987). Essentially, each missing data pattern represents a different group in the model, and equality constraints on the model parameters are placed across groups for variables that are present. The main limitation of the multiple group approach is that each pattern of missingness must be defined as a separate group, which quickly becomes unwieldy in complex models and can result in a loss of information because sample sizes for particular patterns may be insufficient.

Extensions of the multiple group structural equation modeling approach for incomplete data have been used for cohort sequential analyses (e.g., Duncan, Duncan, & Hops, 1996, McArdle & Hamagami, 1992, Miyazaki & Raudenbush, 2000), in which longitudinal trajectories of participants who began the study at different ages are pieced together to form a single aggregate trajectory under the assumption of Missing at Random (i.e., that observations for the ages before and after the time of study are simply Missing at Random). This approach, however, requires the assumption of *age convergence*, or that estimates of between-person differences will converge onto estimates of within-person changes. In other words, age convergence models assume that the only characteristics that separate individuals are chronological age, with no additional processes operating that create differences across birth cohorts that would lead to lack of age convergence, such as nonrandom selection, attrition, or mortality. As such, age convergence models may not be tenable in many longitudinal applications, particular with samples from the latter parts of the lifespan.

Alternatives to Missing at Random

Despite our best efforts to predict probability of missingness with observed covariates, the pattern of missingness may still be informative about the outcomes of interest, a scenario that falls within the category of nonignorable missingness (i.e., Not Missing at Random or NMAR; Little, 1995). Two approaches may be used within this scenario: Pattern-mixture models and selection models. In *pattern-mixture models* (Hedeker & Gibbons, 1997; Little, 1993, 1995), subgroups are identified based on patterns of missing data, and the analysis includes indicators of the subgroup membership. Thus, results are conditional on missing data patterns, although the mechanisms thought to be responsible for the different patterns are not considered explicitly. *Selection models* (Diggle & Kenward, 1994; Verbeke & Molenberghs, 2000) require one to first specify a distribution for the outcomes (e.g., multivariate normal), and then to specify the manner in which the probability of missingness is related to the outcomes. Details about these statistically complex models go beyond the scope of this chapter, but they are an important area of continuing development.

INFERENTIAL ISSUES FOR THE ANALYSIS OF INCOMPLETE DATA

Single Versus Conditional Populations

As reviewed in the last section, great strides have been made in the options for addressing incomplete data in terms of the quality of the parameters that can be obtained from statistical models. Yet the resulting population inferences from those parameters remain problematic conceptually within developmental studies. Most notably, these methods and their corresponding assumptions are based on the notion of a single, accessible population of individuals. That is, they assume that model parameters from the aggregate sample can be used to infer about an "average" individual across time. Some forms of nonparticipation do logically permit inference to a single population, such as with substance abuse in adolescence (i.e., in most cases, drop-outs are still members of the general population of adolescents even if they are not in school or available for measurement).

Yet in other forms of nonparticipation, such as that of mortality in aged populations, inference to individuals within a single, stationary population over time is logically impossible because deceased individuals have left the population of interest, and thus the population is continually being redefined. In cross-sectional studies, sample-level means across age are comprised of distinct groups of individuals (i.e., those who were available to be measured when the study was conducted out of everyone who was originally could have lived to that age), and as such, initial sample selection is already confounded with population mortality. Thus, aggregate-level model parameters cannot be used to make inferences to individual-level change processes in cross-sectional studies. In longitudinal studies, however, individual-level change processes can be evaluated directly. As discussed in the following sections, however, aggregate-level model parameters in longitudinal studies with selective morality must be defined as conditional on the probability of surviving or remaining in the study at a given time point, and not in reference to an immortal population (e.g., "all older adults") that has no real-world counterpart (DuFouil, Brayne, & Clayton, 2004; Harel, 2003; Kurland & Heagerty, 2004, 2005; Ribaudo, Thompson, & Allen-Mersh, 2000).

Temporal Spacing of Observations in Measuring Sources of Missingness

As discussed previously, multiple imputation (MI) and full information maximum likelihood (FIML) approaches require the assumption of Missing at Random (MAR). The extent to which MAR is satisfied depends largely on the availability of covariates that adequately capture the missingness process. It is important to note that results obtained from using either MI or FIML will be asymptotically equivalent provided that the missing data model is the same in each, or if the same covariates of missingness used in the imputation model are also included in the analysis model estimated with FIML (i.e., as saturated correlates in a structural model or as predictors in a multilevel model). Thus, the choice of one method over another may be based on practical considerations. It is generally easier to include covariates for missingness within an imputation model than as extra variables in a substantive model, but this advantage may be offset by the additional effort needed to conduct substantive analyses on each imputed dataset and then combine the results appropriately within MI, a process unnecessary within FIML.

Simulation research has repeatedly shown that the inclusion of covariates related to missingness within MI and FIML will result in parameter estimates and standard errors comparable to what would have been observed with complete data (e.g., Graham, 2003; Graham, Hofer, & Piccinin, 1994; Graham, Hofer, & MacKinnon, 1996; Graham & Schafer, 1999). Further, it appears there is no downside to also including covariates not related to missingness in the hope of satisfying MAR (Collins, Schafer, & Kam, 2001). Although the assumption of MAR is not testable, many scholars have expressed optimism about the utility of MAR-based methods in real-world data, noting that, "...In many psychological research settings the departures from MAR are probably not serious" (Schafer & Graham, 2002, p. 154), and that, "...With MAR missingness, although there is bias when the causes of missingness are not included in the model, the bias is much less of a problem than previously thought (Graham, Cumsille, & Elek-Fisk, 2003), even preferring MAR-based methods to alternatives for non-ignorable missingness: "...The MAR assumption has been found to yield more accurate predictions of the missing values than methods based on the more natural NMAR mechanism" (Little & Rubin, 2002, p. 19).

Within longitudinal research, however, one must consider not only which events or processes related may be related to nonresponse but more specifically, the timing and spacing of the measurements of those processes. Although it is often stated that the inclusion of covariates in the missingness model can render the assumption of MAR tenable, the measurement time frame of dynamic covariates is not often explicitly considered. Probabilities, general patterns, and sources of missingness may change over time as a result of other changes occurring during the observation period. At issue, then, is how best to obtain measures of the dynamic and heterogeneous nature of these missing data processes.

Cohen (1991) and Gollob and Reichardt (1987, 1991) consider the issue of temporal spacing within the context of the measurement of time-varying covariates. Essentially, causal mechanisms need time for their influences to be exerted, and the size of the effect will vary with the time interval between the causal influence and the outcome. Thus, if one statistically controls for a covariate measured at a time before it exerts it causal influence, resultant model parameters may still be biased by the covariate. Time-varying covariates must be measured within the time frame in which they are exerting their influence in order to provide adequate representations of the causal, time-dependent processes that result in participant nonresponse in a viable missing data model. However, deciding on what an appropriate time frame might be is not an easy task, and may not be informed by previous longitudinal studies, given that the data collection intervals from many studies are determined by logistical and financial factors, rather than theoretical expectations about the timing of developmental processes (Cohen, 1991).

The model for missing data must be at least as rich as the substantive model of interest. For group comparison and interaction, effects can be maintained in the imputation model by imputing within group or modeling the group by outcome interaction effect within the imputation model. When researchers are interested in nonlinear influences of context and the moderated effects at different levels of contextual variables, such interaction and higher order effects must be included in the imputation model as well.

Heterogeneity in the Timing and Causes of Nonresponse

Given the importance of covariates that relate to the probability of missingness in satisfying the assumption of Missing at Random (MAR) that is implicit in most of the current approaches to analysis with incomplete data, how might covariates be included that capture changing probabilities of missingness over time as a function of different mechanisms? One way in which attrition can be considered as a dynamic process is in the context of an important event that may lead to or increase the likelihood of nonresponse over time. When time to or from a significant event is an important predictor of an attrition process, a covariate of time-to-event may also be included within the statistical model. For example, in studies in which differential mortality is an issue, a covariate of time-to-death could be included, such that population inference would then be conditional on not only chronological age but also on remaining age, permitting further examination of the observed selection process over time (e.g., Johansson et al., 2004). Similarly, when selective attrition is thought to be related to a disease process, a covariate of time-since-diagnosis could be included (e.g., Sliwinski, Hofer, Hall, Buschke, & Lipton, 2003). In separating aging-related changes from other mechanisms of change (e.g., mortality, disease-related change), inferences can then be defined as conditional on the probability of surviving or remaining in the study.

Although time-to-event covariates can easily be included when complete data for the event are available (e.g., age at death or diagnosis), other methods may be required when time to event information is not known for all participants. One such method was used by Harel and colleagues (Harel, 2003; Harel, Hofer, & Schafer, 2003, March; see also Rubin, 2003), that of *two-stage multiple imputation*, in which imputation for missing age at death values was performed first, followed by imputation for the other missing values using the just-imputed age at death values. Additional alternatives have been presented by Guo and Carlin (2004), who estimated joint survival and growth curve models using Bayesian methods, and by DuFouil, Brayne, and Clayton (2004), who proposed a conditional model that incorporates both study attrition and death. Although evaluation of these approaches is still ongoing, the results thus far have been encouraging with regard to the treatment of distinct processes leading to non-response.

CONCLUSIONS

The last decade has seen considerable theoretical and computational advances in conducting statistical analyses in the presence of incomplete data. Methods such as multiple imputation and full information maximum likelihood are widely available in custom and commercial software, and have shown great promise in their ability to provide model parameters that are unbiased and efficient (or at the very least, less biased and less inefficient compared with older approaches such as listwise deletion and mean-based or regression-based imputation). However, the utility of these newer methods is contingent upon the extent to which individual and contextual covariates and scores obtained at the current and previous occasions can predict the probability of missingness, an untestable assumption known as *Missing at Random*.

Within developmental studies, study attrition may not simply be a nuisance that complicates examination of developmental processes, but rather may be an indicator of or a natural result of a developmental process in and of itself. One must consider that the forces leading to incomplete data are likely to differ in their origin and timing across persons, and include temporally-relevant covariates for those processes within the analysis framework to the greatest extent possible. With regard to the inferences that can be made from analyses of incomplete data, one must also consider the extent to which the impact of incomplete data on the results can safely be "ignored" if appropriate analyses are performed, or whether the results must instead be considered conditional (i.e., nonignorable) on the processes leading to incomplete data. This is primarily a problem of whether one can make inferences to a single, accessible population of developing individuals, or whether the population itself is undergoing dynamic changes (i.e., mortality), such that individuals comprising the population are different across age strata or different developmental periods. In the latter case, population inference may instead need to refer to multiple or conditional populations (e.g., of individuals who remain in the population at a given developmental period) rather than to a single, stationary, and nonexistent population.

REFERENCES

Allison, P. D. (1987). Estimation of linear models with incomplete data. In C. Clogg (Ed.), *Sociological methodology 1987* (pp. 71-103). San Francisco: Jossey Bass.

Allison, P. D. (2002). *Missing data (Sage university papers series on quantitative applications in the social sciences, series no. 07-136).* Thousand Oaks, CA: Sage.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579-616.

Brown, E. C., Catalano, C. B., Fleming, C. B., Haggerty, K. P., & Abbot, R. D. (2005). Adolescent substance use outcomes in the raising healthy children project: A two-part latent growth curve analysis. *Journal of Consulting and Clinical Psychology.*

Cohen, P. (1991). A source of bias in longitudinal investigations of change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 18-25). Washington, DC: American Psychological Association.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclu-sive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6 (4)*, 330-351.

Cooney, T. M., Schaie, K. W., & Willis, S. L. (1988). The relationship between prior functioning of cognitive and personality dimensions and subject attrition in longitudinal research. *Journal of Gerontology: Psychological Sciences, 43*, 12-17.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, B39*, 1-38.

Diggle, P. J., & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics, 43*, 49-93.

Diggle, P. J., Liang, K., & Zeger, S. L. (1994). *Analysis of longitudinal data.* Oxford, England: Clarendon Press.

DuFouil, C., Brayne, C., & Clayton, D. (2004). Analysis of longitudinal studies with death and dropout: A case study. *Statistics in Medicine, 23*, 2215-2226.

Duncan, S. C., Duncan, T. E., & Hops, H. (1996). Analysis of longitudinal data within accelerated longitudinal designs. *Psychological Methods, 1*, 236-248.

Efron, B., & Tibshirani, R. J. (1998). *An introduction to the bootstrap.* New York, NY: Chapman and Hall.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189-198.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58*, 80-92.

Gollob, H. F., & Reichardt, C. S. (1991). Interpreting and estimating in-direct eects assuming time lags really matter. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions.* (pp. 243-259). Washington, DC: American Psychological Association.

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling, 10*, 80-100.

Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Eds.), *Research methods in psychology* (pp. 87-114). New York: Wiley.

Graham, J. W., & Hofer, S. M. (1993). EMCOV.EXE user's guide. textupUnpublished manuscript.

Graham, J. W., & Hofer, S. M. (2000). Multiple imputation in multivariate research. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 201-218). Mahwah, NJ: Lawrence Erlbaum Associates.

Graham, J. W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P., & Schafer, J. L. (1997). Analysis with missing data in prevention research. In K. Bryant, M. Windle, & S. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 325-366). Washington, DC: American Psychological Association.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research, 31*, 197-218.

Graham, J. W., Hofer, S. M., & Piccinin, A. M. (1994). Analysis with missing data in drug prevention research. In L. M. Collins & L. Seitz (Eds.), *Advances in data analysis for prevention intervention research.* National Institute on Drug Abuse Research Monograph Series (#142). Washington, D.C.: National Institute on Drug Abuse.

Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 1-29). Thousand Oaks, CA.: Sage.

Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing-data designs in analysis of change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 335-353). Washington, DC: American Psychological Association.

Guo, X., & Carlin, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician, 58*, 16-24.

Harel, O. (2003). *An introduction to the bootstrap.*Unpublished doctoral dissertation. Unpublished doctoral dissertation, Department of Statistics, University Park, PA.

Harel, O., Hofer, S. M., & Schafer, J. L. (2003, March). *Analysis of longitudinal data with missing values of two qualitatively different types.* Paper presented at the annual meeting of the International Biometric Society (Eastern North American Region). Tampa, Florida.

Hedeker, D., & Gibbons, R. D. (1997). Application of the random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2(1)*, 64-78.

Johansson, B., Hofer, S. M., Allaire, J. C., Piccinin, A. M., Berg, S., & Pedersen, N. L. (2004). Change in cognitive capabilities in the oldest old: The effects of proximity to death in genetically related individuals over a 6-year period. *Psychology and Aging, 19*, 145-156.

Kurland, B. F., & Heagerty, P. J. (2004). Marginalized transition models for longitudinal binary data with ignorable and non-ignorable drop-out. *Statistics in Medicine, 23*, 2673-2695.

Kurland, B. F., & Heagerty, P. J. (2005). Directly parameterized regression conditioning on being alive: Analysis of longitudinal data truncated by deaths. *Biostatistics, 6*, 241-258.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association, 88*, 125-134.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association, 90*, 1112-1121.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data* (1st ed.). New York: Wiley.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research, 29*, 409-454.

McArdle, J. J., & Hamagami, F. (1992). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental Aging Research, 18*, 145-166.

Miyazaki, Y., & Raudenbush, S. W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological Methods, 5*, 44-63.

Muthèn, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52*, 431-462.

Nesselroade, J. R. (1988). Sampling and generalizability: Adult development and aging research issues examined within the general methodological framework of selection. In K. Schaie & R. Campbell (Eds.), *Methodological issues in aging research* (pp. 13-42). New York, NY: Springer.

Olsen, M. K., & Schafer, J. L. (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association, 96*, 730-745.

Rabbitt, P., Watson, P., Donlan, C., Bent, N., & McInnes, L. (1994). Subject attrition in a longitudinal study of cognitive performance in community-based elderly people. *Facts and Research in Gerontology*, 29-34.

Ribaudo, H. J., Thompson, S. G., & Allen-Mersh, T. G. (2000). A joint analysis of quality of life and survival using a random effects selection model. *Statistics in Medicine, 19*, 3237-3250.

Riegel, K. F., Riegel, R. M., & Meyer, G. (1967). A study of the drop-out rates in longitudinal research on aging and the prediction of death. *Journal of Personality and Social Psychology, 4*, 342-348.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581-592.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Rubin, D. B. (2003). Nested multiple imputation of NMES via partially incompatible MCMC. *Statistica Neerlandica, 57*, 3-18.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* New York: Chapman and Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177.

Siegler, I. C., & Botwinick, J. (1979). A long-term longitudinal study of intellectual ability of older adults: The matter of selective subject attrition. *Journal of Gerontology, 34*, 242-245.

Sliwinski, M. J., Hofer, S. M., Hall, C. B., Buschke, H., & Lipton, R. B. (2003). Modeling memory decline in older adults: The importance of preclinical dementia, attrition, and chronological age. *Psychology and Aging, 18*, 658-671.

Streib, G. (1966). Participants and drop-outs in a longitudinal study. *Journal of Gerontology, 21*, 200-209.

Vaupel, J., & Yashin, A. (1985). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *American Statistician, 39*, 176-185.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer-Verlag.