Imputing Attendance Data in a Longitudinal Multilevel Panel Data Set

April 2015

SHORT REPORT

Baby FACES 2009

This page is left blank for double-sided printing.

Imputing Attendance Data in a Longitudinal Multilevel Panel Data Set

OPRE Report 2015-19

April 2015

Submitted to:
Amy Madigan, *Project Officer*
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Submitted by:
Jaime Thomas
Pia Caronongan
Bethany Simard
Cheri A. Vogel
Kimberly Boller

Mathematica Policy Research

Project Director:
Cheri A. Vogel Mathematica
Policy Research P.O. Box 2393
Princeton, NJ 08543-2393

Contract Number: HHSP23320072914YC
Mathematica Reference Number: 06432.140

Suggested citation:
Thomas, J., P. Caronongan, B. Simard; C.A. Vogel, and K. Boller (2015). Imputing Attendance Data in a Longitudinal Multilevel Panel Data Set. OPRE Report 2015-19. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

This page is left blank for double-sided printing.

ABSTRACT

Given the intensive demands that the collection of attendance data places on program staff, it can often be challenging to collect and may result in a fair amount of missing data, which can compromise the reliability and validity of attendance estimates. Little is known about which methods for handing missing data generate the most accurate estimates of attendance. In order to address this issue, we simulate data on children's weekly child care center attendance over the course of a year and compare different methods of estimating attendance. The results indicate that when data are missing on one variable and at one level only, complete case analysis produces accurate estimates of average weekly attendance, regardless of the amount or type of missingness. When estimating total yearly attendance, complete case analysis is inaccurate, but both mean replacement and multiple imputation produce reasonable estimates. A lesson learned from this exercise is that when the desired estimates are simple univariate descriptive statistics, single imputation techniques such as mean replacement can perform as well as more complicated techniques such as multiple imputation.

IMPUTING ATTENDANCE DATA IN A LONGITUDINAL, MULTILEVEL PANEL DATA SET

I. Introduction

Research has shown that the duration and intensity of participation in early care and education is related to cognitive and socio-behavioral outcomes (see for example, Loeb et al. 2008 and Caughy et al. 1994). Proper linking of child care dosage to developmental outcomes requires accurate data on attendance. However, given the intensive demands that the collection of attendance data places on program staff, it can often be challenging to collect and may result in a fair amount of missing data.

Missing data make it difficult to study attendance as well as other aspects of early childhood programs because "missingness" can lead to biased, inefficient, and unreliable estimates of parameters of interest (Schafer and Graham 2002). A lack of consensus on how best to deal with missing data generates inconsistent approaches that can influence the validity of findings and whether a particular study can replicate relations demonstrated in prior research. Methods for addressing missingness range from the simple and computationally straightforward, such as listwise deletion or complete case analysis, in which an entire record is excluded from analysis if any single value is missing (Allison 2001), to computationally intensive, such as full information maximum likelihood (FIML) and multiple imputation (MI).

This paper focuses on missingness in a multilevel panel data set that includes longitudinal data. We simulate data on children's child care center attendance over the course of a year, in which weekly attendance observations are nested within children who are in turn nested within child care programs, which creates data sets with different patterns and degrees of missingness. To determine which method of handling missing data generates the most accurate estimates of attendance, we compare estimates from the different methods and provide guidance on how to implement them.

This study provides meaningful methodological contributions with a wide range of applications for settings such as child care and early education in which frequently collected attendance data are important. Our goal is to contribute to the development of best practices in handling missing data to facilitate the replication of research findings.

Section II of this report provides background on missingness in longitudinal and multilevel settings, and discusses how missingness can affect estimates of attendance in child care and educational settings. Section III describes the longitudinal, multilevel panel data set on which we base our simulated data sets. Section IV discusses our methods for creating simulated data sets with different missingness mechanisms—that is, the sources, or types, of missingness—and amounts of missingness. This section also discusses the different methods we used to deal with missing data. Section V describes the attendance estimates resulting from each imputation method under different amounts of missingness and missingness mechanisms. Section VI contains our conclusions.

II. Background

Both the mechanism leading to missingness and the amount of missingness in a dataset can affect parameter estimates. In general, as the percentage of missing data increases, the choice of approach to deal with missing data in order to minimize bias becomes more important (Newman 2003). The likelihood of bias is lowest when data are missing completely at random (MCAR), in other words, when the probability of missingness is the same for all units (Gelman and Hill 2007). MCAR data would result if a survey respondent decided to answer a question or not based on a coin flip.[1] When data are missing at random (MAR), the probability that a variable is missing depends only on available (that is, measured or observed) information. In this case, unbiased estimation is possible as long as all the variables that affect the probability of missingness are included in the analysis (Gelman and Hill 2007). MAR data would result if, for example, women were less likely to answer a question than men, and researchers included gender in the analysis (assuming that gender was the only predictor of response). Finally, when data are missing not at random (MNAR), the probability that a variable is missing depends on unobserved predictors or on the variable itself. MNAR data would result if less sociable people were less likely to answer a question, but sociability was not measured by the researchers. Another example would be if people with higher earnings were less likely to reveal their earnings. To minimize bias in parameter estimates when using MNAR data, the missing data mechanisms can be modeled, or additional variables that help to predict missingness can be included in analyses, but the potential for bias remains (Gelman and Hill 2007).

Longitudinal data are important in education and early childhood settings because they allow researchers to examine changes over time in developmental outcomes and other measures. Factors such as burden on respondents, however, can increase the potential for missingness as the number of waves of data collection increase. Researchers have discussed advantages and disadvantages of different methods of addressing missingness in longitudinal data sets. Newman (2003) conducted a simulation exercise, creating data sets with different levels of missingness (25 to 75 percent) and different missingness patterns (MCAR, MAR, and MNAR). He compared several techniques of dealing with missing data: listwise and pairwise deletion, (single) regression imputation, the expectation-maximization (EM) algorithm, FIML, and MI, and found that FIML and MI techniques outperform other methods. Kristman et al. (2005) also considered longitudinal data with different amounts and patterns of missingness, comparing listwise deletion, regression imputation, weighting, and MI. They found that, with MCAR or MAR data at all levels of missingness, listwise deletion performed as well as the other methods, and that no method performed well on MNAR data with moderate or high levels of missingness.

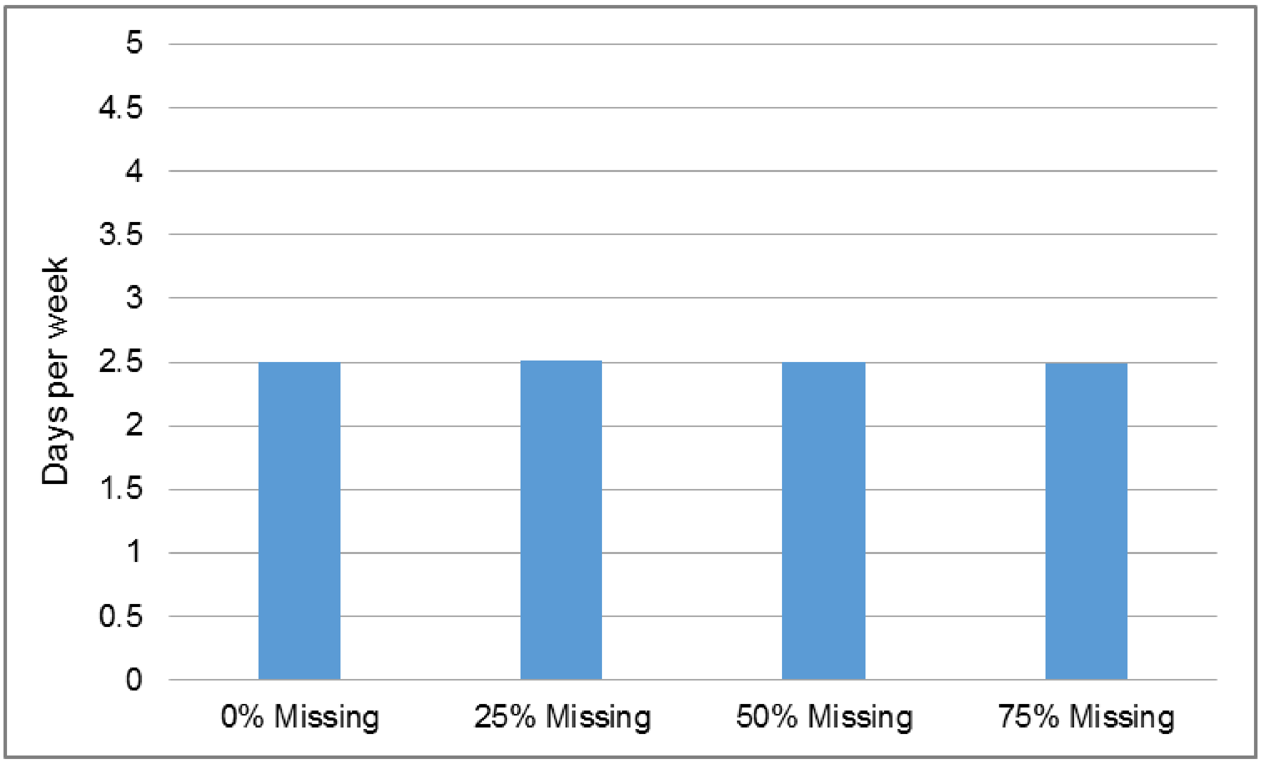
The importance of multilevel modeling to reflect the nesting of children within educational settings has long been understood (see, for example, Raudenbush and Bryk 2002). He et al. (2010) implicitly use multilevel modeling techniques in that their data consist of repeated measures on the same set of subjects. They developed an MI approach that incorporates covariate information as well as the temporal patterns of the time series. Instead of imposing parametric assumptions about the time series (for example, that the observations over time follow a linear pattern), the authors estimate missing values using less restrictive nonparametric methods. Other researchers focus explicitly on missing data techniques for multilevel data. Maas and Snijders (2003) discussed the advantage that multilevel modeling can be easily used to analyze repeated measures if the data are incomplete. Yucel (2008) tailored widely used multilevel models to multiply impute missing values in variables observed at any level of hierarchy.

Dosage in general and child care center attendance in particular are important factors in evaluations of early childhood interventions (see, for example, Caughy et al. 1994). Missing data can

[1] This is an unrealistic example because data are rarely MCAR. Even a seemingly random occurrence like nonresponse due to a postal questionnaire being lost in the mail could be nonrandom since the loss could be related to the area in which the post office is located, for example (example adapted from www.missingdata.org.uk).

pose a greater problem for estimates of total attendance than for average weekly attendance. Figures 1 and 2 are a simple depiction of how the amount of missingness could affect attendance estimates. To create these figures, we simulated 52 weeks of child care attendance data under four different scenarios. Each week's attendance was a random draw from the integers 0 through 5. The simulation had 100 replications. In the first scenario, there were no missing data. In the second, 25 percent of the data were missing—that is, we had attendance data for 39 of 52 weeks. In the third and fourth scenarios, 50 and 75 percent of the data were missing, respectively. In all cases with missing data, data are MCAR. In this simple example, we consider measures of attendance with no missing data to be the true values. Any deviations from these values reflect bias. Figure 1 shows that average weekly attendance is quite similar regardless of the amount of missingness. Figure 2 shows that total yearly attendance is, as expected, highly dependent on the amount of missing data: estimates of total yearly attendance decrease as the percentage of missing data increases. These figures show one way in which irregular reporting of attendance data can lead to bias in estimates of total attendance and thus make child care dosage difficult to measure. In this paper, we discuss how best to "fill in" missing weeks to obtain accurate estimates of total attendance.

Figure 1. Example estimates of average weekly attendance with different amounts of missing data



Source: Simulated data.

Figure 2. Example estimates of total yearly attendance with different amounts of missing data

Source: Simulated data.

III. Data source

We simulate data based on child care center attendance information collected as part of the Early Head Start Family and Child Experiences Survey (Baby FACES). Early Head Start provides comprehensive child development and family support services to low-income pregnant women, infants, and toddlers and their families, offering center-based, home-based, and other program options. Baby FACES is a longitudinal descriptive study of Early Head Start that captures family- and child-level information in addition to program-level characteristics. From a nationally representative sample of 89 programs, 976 parents of children who were in two age cohorts were enrolled in spring 2009: 194 newborns, including pregnant women and children up to 8 weeks old, and 782 one-year-olds, including children aged 10 to 15 months. The study collected data yearly until children left the program early or transitioned out of the program at the age of 3. The Baby FACES study team gathered detailed information from program directors, targeted data on participant families from parent interviews, reports on study children by their teachers or home visitors, weekly data on center attendance and home visit receipt, and direct assessments of social-emotional and cognitive development and parent-child interactions.

We use Monte Carlo simulations to create data sets for this paper, basing the simulations on data from children in the one-year-old cohort who were in center-based care exclusively for one year (from spring 2009 to spring 2010) and who did not leave their Early Head Start programs over this period. We consider only a few of the many variables collected as part of Baby FACES. Weekly attendance observations over the course of a year make up level 1. These level 1 observations are nested within children (level 2). We include at level 2 basic demographic indicators (gender, race/ethnicity, and family receipt of public assistance). Children, in turn, are nested within programs (level 3). At level 3, we include basic program characteristics (total enrollment and urban locale), available for all programs in the sample. Table 1 contains descriptive statistics of these variables.

Table 1. Descriptive statistics of Baby FACES variables

Variable	Number of non-missing values	Number of missing values	Mean	SD	Min	Max
Level 1: Center attendance						
Number of days per week	7619	4445	3.90	1.39	0	7
Level 2: Child characteristics						
Male	232	0	0.50	0.50	0	1
White race/ethnicity	224	8	0.28	0.45	0	1
Family receives public assistance	195	37	0.66	0.47	0	1
Total yearly attendance	230	2	129.24	63.76	0	251
Number of nonmissing weekly attendance observations	232	0	32.84	15.24	0	52
Level 3: Early Head Start program characteristics						
Total enrollment	64	0	179.66	87.09	45	370
Urban locale	64	0	0.69	0.47	0	1

Source: Baby FACES.

Note: Data are from children in the one-year-old cohort who were in center-based care exclusively for one year (from spring 2009 to spring 2010) and who did not leave their Early Head Start programs over this period. The sample consists of 232 children over 52 weeks in a balanced panel data set with 12,064 total observations.

A. Descriptive statistics and missingness

The sample consists of 232 children over 52 weeks in a balanced panel data set with 12,064 total observations at level 1. Of these, 4,445 attendance observations are missing, for a missingness rate of 37 percent. On average, children attended nearly four center days per week. At the program level, there were no missing data. Of the 64 programs in the sample, 69 percent were located in an urban area. Average total enrollment per program was 180.

At the child level, missingness varied by child characteristic. Child gender had no missing observations, and 50 percent of children were male. Child race/ethnicity (white or nonwhite) had eight missing observations at the child level, for a missingness rate of 3 percent. Twenty-eight percent of the children in the sample were white. For public assistance receipt, a proxy for family socioeconomic status (SES), 37 values were missing, for a missingness rate of 16 percent. Sixty-six percent of families in the sample received public assistance.

For each child, we also summed across weeks with nonmissing attendance data to measure total yearly attendance. On average, children had 33 weeks, or 63 percent, of nonmissing attendance data. The number of missing weeks ranged from zero to 52. Two children in the sample had no weekly attendance data. Of the 230 children with at least one nonmissing week of attendance data, the average total yearly attendance was 129 days.

To create simulated data sets with different amounts and sources of missingness but similar characteristics, we obtain—from this highly restricted sample and variable set—means, variances, and covariances of the variables we consider. Table 2 presents the pairwise correlations between each pair of variables. Correlations were generally low. Considering correlations above 0.1,[2] weekly attendance was positively correlated with total yearly attendance, and being white was negatively correlated with receiving public assistance and attending a program in an urban locale. Being on public assistance was positively correlated with attending a program in an urban locale. Total program enrollment was also positively correlated with urban locale. An exception to the low correlations was the relationship between total yearly attendance and number of nonmissing weeks of attendance data: the correlation was greater than 0.9, which indicates a large potential for bias in estimates of total yearly attendance if we do not take missing weeks into account.

IV. Methods for creating simulated data sets with different levels, mechanisms, and methods of handling missingness

Based on the Baby FACES data set described above, we used Monte Carlo simulation to create 1,000 three-level data sets with no missing observations. Each data set had 52 weekly attendance observations nested within five children per program nested within 50 programs, for a total of 13,000 observations. We manipulated these to create data sets with different levels of missingness arising from different types of missingness. We considered three levels of missingness: 25, 50, and 75 percent (in addition to zero percent missing data), and three sources of missingness: MCAR, MAR, and MNAR.

Table 2. Pairwise correlations between Baby FACES variables

	Days per week	Total yearly attendance	Non-missing weeks	Male	White	Public assistance	Total enrollment	Urban locale
Days per week	1.00							
Total yearly attendance	0.27	1.00						
Non-missing weeks	0.01[a]	0.92	1.00					
Male	0.09	0.04	-0.02	1.00				
White	0.06	0.21	0.19	0.04	1.00			
Public assistance	0.06	0.01[a]	-0.02	0.04	-0.14	1.00		
Total enrollment	0.01[a]	-0.06	-0.07	0.02	-0.08	0.09	1.00	
Urban locale	0.05	-0.13	-0.18	0.07	-0.17	0.20	0.13	1.00

Source: Baby FACES data.

Note: Data are from children in the one-year-old cohort who were in center-based care exclusively for one year (from spring 2009 to spring 2010) and who did not leave their Early Head Start programs over this period. The sample consists of 232 children over 52 weeks in a balanced panel data set with 12,064 total observations. All pairwise correlations are significantly different from zero, $p < 0.05$, except where noted.

[2] Because the sample size is so large, though the magnitudes are small, all but the correlation between center attendance and total enrollment are significantly different from zero, $p < 0.05$.

[a] Not significantly different from zero.

For a given level and type of missingness, we used three different methods of handling missing data. The simplest is complete case analysis, in which we considered only those children with nonmissing attendance data for all 52 weeks within the year. Next, we considered two different methods of imputing missing weekly attendance data. In the first, we used information solely from level 1 by filling in missing observations using the mean of nonmissing attendance observations (that is, mean replacement or mean imputation). In the second, we used MI to fill in missing attendance observations at level 1 using information from levels 2 and 3. We had a total of 28 simulated data sets: three levels of missingness * three sources of missingness * three methods of handling missing data = 27 data sets + one data set with no missing data. Within each simulated data set, we computed average weekly attendance and total yearly attendance for each child. In Section IV.A, we discuss how we created data sets with MCAR, MAR, and MNAR data. In Section IV.B, we discuss how we addressed missing data using complete case analysis, mean replacement, and MI.

A. Missingness mechanisms

MCAR. We randomly deleted 25, 50, and 75 percent of weekly attendance observations. To do this, we generated a random order for the observations using a random uniform distribution over (0, 1) and sorted the observations from lowest to highest. For each level of missingness, we set the variable to missing for the last *x* percent of observations, where *x* corresponds to 25, 50, or 75 percent of the variable's observations.

MAR. We determined which values of weekly center attendance to delete by modeling missingness using variables in the data set. We used the following linear relationship to model the missing probability:

(1) Probability of missingness = 0.28
+ 0.02*Male
− 0.1*White
− 0.02*Public assistance
+ 0.0002*Program enrollment
+ 0.1*Urban locale
+ Error[3]

We obtained the coefficients from a regression of an indicator that weekly center attendance was missing on a constant and the variables in equation (1) using actual Baby FACES data. Using the logistic function, we converted this value into a probability and sorted the observations from lowest to highest probability of missingness. For each level of missingness, we set the variable to missing for the last *x* percent of observations, where *x* corresponds to 25, 50, or 75 percent of the variable's observations.

MNAR. To create MNAR data, we determined which values of weekly center attendance to delete by modeling missingness as a function of weekly center attendance itself. We modeled the probability of missingness as negatively related to weekly attendance—that is, lower values of weekly attendance were more likely to be missing. As mentioned previously, we sorted the observations

[3] The error term follows a standard normal distribution.

from lowest to highest probability of missingness. For each level of missingness, we set the variable to missing for the last *x* percent of observations, where *x* corresponds to 25, 50, or 75 percent of the variable's observations.

B. Methods of handling missing data

Complete case analysis. The simplest way to handle missing data is to drop observations that are missing values of any variables in the analysis. The main advantage of complete case analysis is its simplicity. Beyond the obvious disadvantage of decreased sample sizes, another disadvantage is that, because missingness can be related to observed and unobserved characteristics, dropping observations that are missing values on any variable can bias estimates of interest. For example, if children with lower attendance were more likely to be missing attendance records, using complete case analysis would bias estimates of average weekly and total yearly attendance. In particular, the estimates would be biased upwards, because observations of children with low attendance would be dropped.

In this analysis, data were missing at level 1 only (weekly attendance observations). Variables at the child level, such as race/ethnicity or receipt of government assistance, had no missing data; nor did variables at the program level, such as total program enrollment. To implement complete case analysis, we dropped any observations that were missing weekly attendance data. The amount of missing data ranged from 0 to 75 percent, as described previously.

Mean replacement. Mean replacement (1) allows the researcher to retain observations that would have been dropped in a complete case analysis, (2) is easy to implement, and (3) can accurately predict missing values (Schafer and Graham 2002). Its main disadvantage (as with all single imputation methods) is that it understates the uncertainty researchers have about the missing values—by choosing one value to replace the missing value, researchers act as if they know the true value with certainty. In particular, mean replacement distorts relationships between variables by understating their correlations (Gelman and Hill 2007).

To implement mean replacement as a method of imputing missing weekly attendance observations, we computed the average of nonmissing weekly attendance observations for each child and replaced any missing observations with that average.

Multiple imputation. Like mean replacement, multiple imputation (1) allows the researcher to retain observations that would have been dropped in a complete case analysis, and (2) can accurately predict missing values. Unlike mean replacement, multiple imputation takes into account uncertainty about the missing values by replacing a missing value with not one, but several imputed values (Gelman and Hill 2007).

In this analysis, we filled in the missing weekly attendance observations using linear regression. The regression model predicted weekly attendance as a function of the child-level variables gender, race/ethnicity, and receipt of government assistance, and the program-level variables program enrollment and urban locale. We created five imputed data sets in each iteration of the Monte Carlo simulation.[4] Within each imputed data set, we estimated average weekly and total yearly attendance at the child level, then combined these estimates across the five multiply imputed data sets.

V. Results

Our findings consist of estimates of average weekly and total yearly attendance, using three methods to handle missing data: complete case analysis, mean replacement, and MI. We applied these methods to simulated data with varying degrees of missingness and with different missingness mechanisms (MCAR, MAR, and MNAR). Table 3 presents the results for estimates of average weekly attendance. The top row of the table gives the average weekly attendance estimate in the benchmark case with 0 percent missing data. The benchmark level of average weekly attendance at the child level is just over 4 days per week. For each combination of missingness mechanism and level of missingness, we reported the mean difference from the benchmark case and the *p*-value of this difference. Across all levels of missingness and missingness mechanisms, all three methods of handling missing data produce accurate estimates of average weekly attendance.

The situation is quite different when we consider estimates of total yearly attendance (Table 4). At the child level, the benchmark level of total yearly attendance is an average of 209.5 days per year. Complete case analysis grossly understates total yearly attendance, ranging from estimates that are about 52 days lower than the benchmark when 25 percent of attendance observations were missing, to understating attendance by about 157 days when 75 percent of the observations were missing. As mentioned previously, this is a mechanical relationship—a sum of weekly attendance observations, each of which can range from 0 to 5, that contains fewer observations will yield a lower total than a sum that contains more observations. Mean replacement and multiple imputation both result in reasonably accurate estimates of total yearly attendance, but mean replacement performs better than multiple imputation in our simulations—regardless of type and level of missingness, the mean replacement estimate was not significantly different from the benchmark case. This is not surprising as mean replacement does not change the sample mean of a variable, whereas multiple imputation can result in a different sample mean.

VI. Conclusions

Our findings provide guidance on which method to use to handle missing data in the context of accurately estimating average and cumulative measures of attendance within longitudinal, multilevel panel data. When data are missing on one variable and at one level only, complete case analysis produced accurate estimates of average weekly attendance, regardless of the amount or type of missingness. When estimating total yearly attendance, complete case analysis was inaccurate, but both mean replacement and multiple imputation produced reasonable estimates. A lesson learned from this exercise is that, when the desired estimates are simple, univariate descriptive statistics, single imputation techniques such as mean replacement can perform as well as more complicated techniques such as multiple imputation. However, when accurate estimates of the relationships between variables are desired, methods such as multiple imputation that take into account uncertainty about the imputed values are preferred. Directions for future research include exploring

[4] To more accurately represent uncertainty about missing values, researchers can create more missing data sets. We chose a relatively low number to reduce processing time and because our estimates of interest are means of single variables and not regression coefficients. A sensitivity check using 50 imputed datasets yielded similar results.

how best to impute missing values in a multilevel data set with missingness at more than one level, and which imputation method to use when desired estimates involve multivariate analysis.

Table 3. Average weekly attendance estimates using different methods of handling missing data

	Benchmark (0% missing) = 4.03 days per week		
	Mean difference from benchmark *p*-value of difference		
Type of missingness	25% missing	50% missing	75% missing
Complete case analysis			
MCAR	0.00 0.959	0.00 0.305	0.00 0.667
MAR	0.00 0.082	0.00 0.112	0.00 0.211
MNAR	0.00 0.576	0.00 0.240	0.00 0.639
Mean replacement			
MCAR	0.00 0.805	0.00 0.537	0.00 0.398
MAR	0.00 0.307	0.00 0.222	0.00 0.333
MNAR	0.00 0.862	0.00 0.086	0.00 0.150
Multiple imputation			
MCAR	0.01 0.000	0.04 0.000	0.00 0.277
MAR	0.00 0.908	0.04 0.000	-0.02 0.000
MNAR	0.01 0.000	0.04 0.000	0.00 0.299

Source: Simulated data.

Note: For each level of missingness, type of missingness, and method of handling missingness, we simulated 1000 3-level data sets (52 weekly attendance observations nested within 5 children nested within 50 programs) based on actual relationships among variables observed in Baby FACES Family Service Tracking data. Data were missing at level 1 only (that is, weekly attendance observations). The benchmark scenario, reported in the top row of the table, has zero missing data. For each level of missingness, type of missingness, and method of handling missingness, we report the result of a two-sample t-test comparing mean attendance across 1,000 simulated data sets from the benchmark scenario (0 percent missing) to mean attendance across 1,000 simulated data sets with the specified level, type, and method of handling missingness. Underneath each mean difference, we report the associated *p*-value.

MAR = missing at random; MCAR = missing completely at random; MNAR = missing not at random.

Table 4. Average total yearly attendance estimates using different methods of handling missing data

	Benchmark (0% missing) = 209.45 days per year		
	Mean difference from benchmark *p*-value of difference		
Type of missingness	25% missing	50% missing	75% missing
	Complete case analysis		
MCAR	-52.37 0.000	-104.77 0.000	-157.09 0.000
MAR	-52.48 0.000	-104.80 0.000	-157.12 0.000
MNAR	-52.40 0.000	-104.78 0.000	-157.10 0.000
	Mean replacement		
MCAR	-0.02 0.805	-0.05 0.537	-0.08 0.398
MAR	-0.09 0.307	-0.11 0.222	-0.09 0.333
MNAR	-0.02 0.862	-0.15 0.086	-0.13 0.150
	Multiple imputation		
MCAR	0.70 0.000	2.25 0.000	0.07 0.277
MAR	-0.01 0.908	1.93 0.000	-0.79 0.000
MNAR	0.70 0.000	2.25 0.000	0.06 0.299

Source: Simulated data.

Note: For each level of missingness, type of missingness, and method of handling missingness, we simulated 1000 3-level data sets (52 weekly attendance observations nested within 5 children nested within 50 programs) based on actual relationships among variables observed in Baby FACES Family Service Tracking data. Data were missing at level 1 only (that is, weekly attendance observations). The benchmark scenario, reported in the top row of the table, has zero missing data. For each level of missingness, type of missingness, and method of handling missingness, we report the result of a two-sample t-test comparing mean attendance across 1,000 simulated data sets from the benchmark scenario (0 percent missing) to mean attendance across 1,000 simulated data sets with the specified level, type, and method of handling missingness. Underneath each mean difference, we report the associated *p*-value.

MAR = missing at random; MCAR = missing completely at random; MNAR = missing not at random.

REFERENCES

Allison, P.D. (Ed.). *Missing Data* (no. 136). Thousand Oaks, CA: Sage, 2001.

Caughy, M.O.B., J.A. DiPietro, and D.M. Strobino. "Day-care Participation as a Protective Factor in the Cognitive Development of Low-Income Children. *Child Development,* vol. 65, no. 2, 1994, pp. 457–471.

Gelman, A., and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2007.

He, Y., R. Yucel, and T.E. Raghunathan. "A Functional Multiple Imputation Approach to Incomplete Longitudinal Data." *Statistics in Medicine,* vol. 30, no. 10, 2011, pp. 1137–1156.

Kristman, V.L., M. Manno, and P. Côté. "Methods to Account for Attrition in Longitudinal Data: Do They Work? A Simulation Study." *European Journal of Epidemiology,* vol. 20, no. 8, 2005, pp. 657–662.

Loeb, S., M. Bridges, D. Bassok, B. Fuller, and R. Rumberger. "How Much Is Too Much? The Influence of Preschool Centers on Children's Social and Cognitive Development. *Economics of Education Review,* vol. 26, no. 1, 2007, pp. 52–66.

Maas, C.J., and T.A. Snijders. "The Multilevel Approach to Repeated Measures for Complete and Incomplete Data." *Quality and Quantity,* vol. 37, no. 1, 2003, pp. 71–89.

Newman, D.A. "Longitudinal Modeling with Randomly and Systematically Missing Data: A Simulation of Ad Hoc, Maximum Likelihood, and Multiple Imputation Techniques." *Organizational Research Methods,* vol. 6, no. 3, 2003, pp. 328–362.

Raudenbush, S.W., and A.S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods* (vol. 1). Thousand Oaks, CA: Sage, 2002.

Schafer, J.L., and J.W. Graham. "Missing Data: Our View of the State of the Art." *Psychological Methods,* vol. 7, no. 2, 2002, p. 147.

This page is left blank for double-sided printing.

This page is left blank for double-sided printing.