



Mixed model variance adjustments when missing values are multiply imputed

Hrishikesh Chakraborty

Correspondence: rishic@mailbox.sc.edu

Department of Epidemiology and Biostatistics, Arnold School of Public Health, The University of South Carolina, Columbia, USA.

Abstract

Studies frequently have missing values for both the dependent and independent variables. Multiple imputation (MI) techniques are approaches to replacing missing values so that complete-data analysis methods may be employed. Often, mixed models are used to analyze data with missing independent variable values, but in mixed models, having a large number of independent variables with missing values will exclude a significant number of data vectors from the analysis, creating a major problem when attempting to draw meaningful conclusions from an analysis. Our goal was to use multiple imputation techniques to impute missing values for independent variables in a longitudinal study setting, and then to use the complete data matrix to fit mixed models independently for each of the imputed datasets. In this paper, we present a method to combine multiple estimates and inferential statistics generated from multiply imputed datasets using the same mixed model, which has not previously been done. Additionally, we are incorporated two sets of variance-covariance matrices for each imputed set and also adjusted degrees of freedom. In the example we compared estimates using complete and multiply imputed data using mixed models. We conclude that in some situations it is desirable to use multiple imputation techniques and mixed models together to draw conclusions.

Keywords: Multiple imputations, mixed models, missing values, vectors

Introduction

Multiple imputation (MI) techniques are commonly accepted methods for replacing missing values with imputed values based on the underlying model. Data is imputed prior to analysis. Statistical methods are then used to analyze the imputed datasets. Over the years there has been a remarkable improvement in statistical methods that address the different types of missing values. Additionally, statistical software is readily available for data analysis with missing observations in many modeling contexts. Some previous studies conducted research on missing data and mixed models, but none of them expanded it to missing data multiple imputation our study. One study showed that in studies with a high percentage of missing values [1], the mixed model approach without any ad hoc imputation is more powerful, and another simulation study [2] examined the MI of missing repeated outcome measurements did not increase precision in the estimated rate of change in the end point to linear mixed-effects models. One recent study [3] presented a comparison of Mixed-Effect Model Repeated model versus last observation carried forward using 25 NDA data sets. However, the application of inferential statistics within mixed models using imputed data is not fully developed, nor is software readily available to support such an approach. We became interested in using the multiple imputation method to impute missing values of independent variables to make use of maximum available information for mixed model analysis.

In mixed models, independent variables are treated as known

values, but in practice we encounter missing values for both dependent and independent variables. When conducting data analysis using mixed models, an entire row of the data matrix is excluded if there is only one missing value for independent variables. This results in a fitted model based on fewer subjects or time points, or a reduced data matrix, therefore excluding a lot of information from the model. Since so much information can be excluded from the model, it can be difficult to draw meaningful conclusions.

In this paper, we present a way to use the entire dataset by using appropriate partial multiple imputations (imputing what are scientifically meaningful values for the independent variables) with mixed models to fit a model. First, we imputed the missing independent variable values five times using an appropriate multiple imputation technique, thereby creating five imputed datasets. We then used a mixed model to analyze the imputed datasets. We combined the estimates (one set of estimates from each imputed dataset) for unique interpretation by accounting for within-and between-imputation variability in a mixed model setting, while incorporating several variance-covariance components and adjusting the degrees of freedom for the significant tests.

We use two types of data to illustrate the technique: (1) cross-sectional data with clustering and (2) longitudinal repeated-measures data. We imputed independent variables only if it made sense scientifically and there was enough information to impute the missing values. We presented the model estimates

from different multiply imputed datasets and also compared the original estimates (using datasets with the missing values) with combined estimates (from five models based on the five, different partially imputed datasets). We also calculated the variance inflation factor for each variable.

Methods

Inference from complete data is typically based on a point estimate for a parameter θ , a variance-covariance matrix U , and a normal reference distribution. In the presence of missing values, multiple imputation is a state-of-the-art method to impute the missing values [4]. Different imputation methods are available to impute the missing values. Employing an appropriate imputation method will yield m (> 1) sets of imputed data.

To get m sets of estimates and variance-covariance matrix estimates, we used appropriate mixed models (based on the research question) on each of the imputed datasets. The mixed model equation for the j^{th} subject can be written as $\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{d}_j + \mathbf{V}_j$ with the assumptions: $\mathbf{d}_j \sim \text{NID}(\mathbf{0}, \boldsymbol{\Delta})$, $\mathbf{V}_j \sim \text{NID}(\mathbf{0}, \sigma^2 \mathbf{I})$, and the covariance matrix for the model equation is $V(\mathbf{Y}_j) = \boldsymbol{\Sigma}_j = \mathbf{Z}_j \boldsymbol{\Delta} \mathbf{Z}_j' + \mathbf{V}_j$. This model can be constructed using the "usual" linear model methods $E(\mathbf{Y}_j) = \mathbf{X}_j \boldsymbol{\beta}$, where \mathbf{Y}_j represents the vector of measurements from the j -th subject through all periods, \mathbf{X}_j is the fixed-effects design matrix for the j^{th} subject, $\boldsymbol{\beta}$ is the fixed-effect parameter for all subjects, \mathbf{Z}_j is the random effects design matrix for the j^{th} subject, \mathbf{d}_j is the random effects coefficient for j^{th} subject (\mathbf{d}_j contains increments to population intercepts and slopes), and \mathbf{V}_j represents the vector of random "measurement errors" for j^{th} subject.

The design matrix, \mathbf{X}_j and fixed-effect parameter effects, elements of $\boldsymbol{\beta}$, are similar to the design matrix, \mathbf{X} , and the regression parameters in a typical multiple regression, ANOVA, or ANCOVA model, in that $E(\mathbf{Y}_j) = \mathbf{X}_j \boldsymbol{\beta}$. Thus, an element of $\boldsymbol{\beta}$ may represent the "slope" of a regression surface with respect to a covariate, or a treatment effect, or similar quantity. The random effects design matrix, \mathbf{Z}_j , and the subject-specific random effects coefficient, \mathbf{d}_j , represent random deviations about $E(\mathbf{Y}_j) = \mathbf{X}_j \boldsymbol{\beta}$ that are associated with data from the j^{th} subject. The vector \mathbf{V}_j is the vector of random deviations, or "measurement errors," about the expected value of data from the j -th subject. It is very similar to the vector of random deviations, usually denoted ϵ_j or \mathbf{e}_j in a multiple regression, ANOVA, or ANCOVA model.

The combined estimate of θ from m mixed models with imputed datasets is $\hat{\theta} = m^{-1} \sum_{k=1}^m \hat{\theta}_k$ (where $k = 1, 2, 3, \dots, m$),

the average of the m estimates and the variance-covariance matrix estimates associated with θ has within- (\hat{U}) and between- (\hat{B}) imputation components, and the total variance-covariance matrix estimate is

$$\hat{T} = \hat{U} + (1 + m^{-1})\hat{B} \text{ where, } \hat{U} = m^{-1} \sum_{k=1}^m \hat{U}_k \text{ and}$$

$$\hat{B} = (m - 1)^{-1} \sum_{k=1}^m (\hat{\theta}_k - \hat{\theta})(\hat{\theta}_k - \hat{\theta})'$$

[15]. From each mixed model run we estimated the variance-covariance matrix of

$$\hat{U}_k = \begin{bmatrix} X\hat{R}^{-1}X & X\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + G^{-1} \end{bmatrix}^{-1}$$

where Z is the known design matrix, G is the random effects covariance matrix, and R is the error variance.

To test linear hypotheses ($H_0 : L\theta = c$) [where $L\theta$ is a vector of coefficients for the linear hypotheses, and c is a vector of constants] about the parameters, we used two different test statistics depending on whether univariate or multivariate inference was required. For Univariate inference, we used a t-test $t_{v_i} = \frac{L\theta}{\sqrt{L\hat{T}L'}}$ for univariate inference (when

the rank of L is equal to 1) where, $v_i = (m-1) \left[1 + \frac{\hat{U}}{(1+m^{-1})\hat{B}} \right]^2$. The quantity $r = \frac{(1+m^{-1})\hat{B}}{\hat{U}}$ is the relative increase in variance due

to missing values, and the quantity $\hat{\lambda} = \frac{\eta+2/(v+3)}{\eta+1}$ is called

the fraction of missing information about θ [15]. For multivariate inferences (when the rank of L is greater than 1), we use an F-test

$$F_{p,v} = \frac{\theta'L'(L'\hat{T}L)^{-1}L\theta}{\text{rank}(L)} \text{ where rank}(L)=p, \text{ and } v = \left(\frac{m-1}{p} \right) \left(1 + \frac{1}{r} \right)$$

$$\text{where } r = \frac{1}{p} \left(1 + \frac{1}{m} \right) \text{trace}(\hat{B}\hat{U}^{-1})$$

is an average relative variance increase due to nonresponse [4,5]. In some situations, specifically, for small m , between-imputation covariance matrix B is unstable and does not have full rank [5]. The suggestion for handling an unstable between-imputation covariance matrix is to assume that the population between and within imputation covariance matrices are proportional to one another. Then a more stable estimate of total variance can be calculated, which leads to a different test statistic and a change in the degrees of freedom. One suggestion is to change the degrees of freedom v as [6].

$$v = \frac{1}{2}(p+1)(m-1) \left(1 + \frac{1}{r} \right)^2 \text{ for } p(m-1) \leq 4, \text{ and}$$

$$v = 4 + [p(m-1) - 4] \left[1 + \frac{1}{r} \left(1 - \frac{2}{p(m-1)} \right) \right]^2 \text{ for } p(m-1) > 4$$

A number of statistical software packages are able to impute missing values. In this paper we used SAS [7] Proc MI to

generate the multiply imputed datasets, which uses the EM algorithm to do the imputation, and assumes a multivariate normal distribution and missingness at random (MAR). Next we used Proc MIXED to generate estimates and variance covariance matrices where we used the random intercept and slope. There is no SAS procedure available to accommodate two different covariance matrices (G and R) in the analysis and to make inferences. SAS Proc MIANALYZE can be used to generate combined inferences in simple situations. We used Proc IML to program the procedure to obtain the average estimates, the within-between combined standard errors, the approximate degrees of freedom, and to generate inferential results.

Examples

Cross sectional data with clustering

For this analysis, we merged databases from the National Center for Education Statistics [8] and the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (SECCYD). The SECCYD was conducted by the NICHD Early Child Care Research Network supported by NICHD through a cooperative agreement that calls for scientific collaboration between the grantees and the NICHD staff. The NCES database contains descriptive information on all U.S. public schools and their students, teachers, curricula, and factors related to the children's classroom environment [9]. Details of the original sampling procedure and characteristics of the SECCYD database have been described in several publications [10-14]. The SECCYD database contains longitudinal information collected from an initial sample of 1,364 children born in 10 study sites [6]. Data from the 54-months and first grade time points were selected from the SECCYD database. Data corresponding to the children's year in first grade were selected from the NCES database. These variables were used to model the children's experience in first grade, based on economic resources at the family level and the school level [13-17].

The NCES database contained information only on public schools, but the SECCYD database included children in both public and private schools. Out of 1,364 children from the SECCYD database, 709 were successfully mapped to their respective public schools, after excluding 57 children from the analysis because only one child was selected from each classroom. Therefore, each child in the combined dataset represents a unique classroom, resulting in clusters of unique classrooms in each school district.

The dependent variable used in this example reflects classroom instructional quality in first grade. Independent variables used in these models are related to the children's socio-demographic, academic, classroom, and school-related attributes at 54 months and in first grade. We did not impute all the missing values for all independent variables. Instead we selectively imputed the missing values based on their scientific validity. For example, we chose not to impute

independent variables related to instructional spending or percentage of nonwhite students for each school in a school district because there was no way to scientifically estimate such values from the NCES database. Additionally, categorical variables were not imputed. Given the structure of these data, clustering was taken into account at the school-district level in the mixed model analysis.

The dependent variable for the model, with only one missing value, was classroom instructional quality at first grade. Independent variables were spending for instruction per child, percentage of nonwhite students, total number of children in class at first grade, teacher's experience, teacher's education, income-to-needs ratio at first grade, mother's education, class persistence of students at first grade, teacher's perception of barriers at first grade class, WJ-R mean score of cognitive at 54 months, WJ-R mean score of achievement at 54 months, and percentage in daycare from 6 to 54 months.

Longitudinal data

Data from the SECCYD study were used again in the second example. The data were structured longitudinally and were collected at the 24-, 36-, and 54-months and during the first grade. A total of 1364 children were included in the analysis, each potentially having four cognitive and language data points. Of the 5456 observations in the analysis dataset, 998 were excluded due to missing dependent variable values from 659 children, and 6 additional observations were deleted due to variable values being nonimputable, as in the previous example.

Independent variables at the four time points included household income, race, age, maternal education, maternal depression, maternal sensitivity, presence of husband/partner in the household and number of hours in childcare. Household income was based on a calculation of median income-to-needs ratio at two time periods up to first grade. Income-to-needs was dichotomized, after imputations, as low and high at the two age periods, resulting in four categories of income: chronically poor, early poor, late poor, and never poor. Mother's education was measured in years. Maternal depression was measured with the Center for Epidemiological Studies Depression scale (CES-D). Maternal sensitivity represents constructs of measurements related to sensitivity to nondistress, positive regard, and intrusiveness at 24 months, and supportive presence, respect for autonomy, and hostility at 36 and 54 months and first grade time points. The dependent variable was based on the children's cognitive and language development at 24, 36, and 54 months and first grade [18]. A mixed model with individual random intercept and slope (unstructured covariance matrix for the intercept and slope) was used to model the child's cognitive outcome. (Table 3) contains the percentages of missing values for the analysis dataset, (Table 4) contains the parameter estimates and significances for the mixed models with imputed values, and (Table 5) compares the combined estimates based on the

Table 1. Estimates from mixed models Using different imputed datasets.

Variable Name	Percent Missing	Model Estimates from Imputed Datasets				
		1	2	3	4	5
Spending for instruction per child	19.5	0.00	0.00	0.00	0.00	0.00
Percentage of nonwhite students	19.4	-0.03*	-0.03*	-0.03*	-0.03*	-0.03*
Total number of children in class at Grade 1	2.2	0.05	0.06	0.06	0.05	0.05
Teacher's experience at Grade 1	1.4	-0.03	-0.03	-0.02	-0.02	-0.02
Teacher's education at Grade 1	1.5	0.17	0.14	0.14	0.16	0.17
Income to needs ratio at Grade 1	7.6	-1.09*	-0.67	-1.10*	-0.97*	-1.18*
Mother's education	0.0	0.07	0.11	0.08	0.12	0.09
Class persistence of students at Grade 1	2.8	4.20*	4.05	3.96	3.72	3.78
Teacher's perception of barriers at Grade 1	0.0	0.02	-0.09	-0.05	-0.08	-0.01
WJ-R mean score of cognitive at 54 months	3.4	-0.02	-0.02	-0.02	-0.03	-0.02
WJ-R mean score of achievement at 54 months	3.8	0.02	0.03	0.03	0.03	0.02
Percentage in daycare from 6 to 54 months	1.1	-0.00	-0.00	-0.00	-0.00	-0.00

* = p-value < .05. Response variable is classroom instructional quality at first grade.

Table 2. Comparison of Combined Estimates to Original Model.

Variable Name	Combined Estimates (n=699)	Estimates from Original Model (n=585)	Relative Increase in Variance
Spending for instruction per child	0.00	0.00	0.00
Percentage of nonwhite students	- 0.03	-0.03*	0.01
Total number of children in class at Grade 1	0.05	0.05	0.02
Teacher's experience at Grade 1	-0.02	-0.01	0.01
Teacher's education at Grade 1	0.14	0.03	0.01
Income to needs ratio at Grade 1	-1.03	-1.12*	0.22
Maternal education	0.10	0.07	0.06
Class persistence of students at Grade 1	4.21*	3.12	0.01
Teacher's perception of barriers at Grade 1	0.03	0.12	0.02
WJ-R mean score of cognitive at 54 months	-0.02	-0.01	0.06
WJ-R mean score of achievement at 54 months	0.02	0.03	0.09
Percentage in daycare from 6 to 54 months	0.00	0.00	0.01

* = p-value < .05. Response variable is classroom instructional quality at first grade.

imputed datasets to the original model.

Results and summary

(Table 1) displays the percentage of missing values for independent variables in the database, and the model estimates for the five imputed datasets of the cross-sectional model. The NCES variables have a maximum of nearly 20% missing

values, whereas the SECCYD data have a maximum of nearly 8% missing. We assumed that the missing values are missing at random. The five imputed datasets correspond to the five sets of model estimates showing the between imputation variability among the parameter estimates. The estimates demonstrate slightly different values and significance levels. For example, the estimate of the *income-to-needs ratio* for Model 2 is lower than the other models and possesses a *p*-value greater than 0.05.

(Table 2) compares the differences in parameter estimates and significance levels of the original dataset and the combined results of the five imputed datasets. "Original model" implies that the same mixed models were used to generate the estimates by using the dataset before imputation. The model based on the original dataset uses 585 data vectors, but the combined model based on the imputed datasets uses 699, an increase of 114 data vectors with the MI technique. Noteworthy is the lack of significance of *percentage of nonwhite students* and *income-to-needs ratio* and the significance of *class persistence of students* for the combined estimates. The variable *income-to-needs ratio* has the greatest relative increase in variance, 0.22.

(Table 3) shows the percentage of missing values for four categories of longitudinal measures: maternal sensitivity, maternal depression, partner/husband in the household, and hours per week in child care. Missing values abound for these variables, ranging from 9 to 26% at each measurement point. Because missing values may not coincide on the same data vector, a high potential that an appreciable percentage of observations will be excluded from the model exists.

(Table 4) contains the estimates from the five imputed datasets. As in the cross-sectional example, differences exist in parameter estimates among the five models. There do not seem to be differences in significance, at the *p* < 0.05 level, among the models. However, differences in estimates and significance are evident when the original model is compared

Table 3. Percentage of Missing Values at Four Time Points.

Variable Name	Percent Missing
Maternal education	0.1
Maternal sensitivity at 24 months	14.1
Maternal sensitivity at 36 months	14.9
Maternal sensitivity at 54 months	23.8
Maternal sensitivity at Grade 1	26.4
Maternal depression at 24 months	18.0
Maternal depression at 36 months	11.9
Maternal depression at 54 months	21.0
Maternal depression at Grade 1	26.0
Husband or partner lives with mother at 24 months	11.5
Husband or partner lives with mother at 36 months	10.9
Husband or partner lives with mother at 54 months	20.5
Husband or partner lives with mother at K1	21.8
Hours/week in care at 23 months	9.2
Hours/week in care at 34 months	9.9
Hours/week in care at 53 months	16.7
Income group	23.2

Note: Gender, race, and age were not imputed, but were included in the model. Response variable is classroom instructional quality at first grade.

Table 4. Estimates from Mixed Models using Different Imputed Datasets.

Variable Name	Estimates from Imputed Datasets				
	1	2	3	4	5
Maternal education	1.21*	1.21*	1.21*	1.22*	1.20*
Maternal sensitivity	0.92*	0.96*	0.98*	0.68*	0.85*
Maternal depression	-0.06*	-0.06*	-0.05*	-0.05*	-0.05*
Hours/week in care	0.00	0.00	0.00	0.00	0.00
Income group: Always poor	92.51*	92.45*	92.46*	92.30*	92.32*
Income group: Late poor	95.23*	95.23*	95.25*	95.17*	95.19*
Income group: Early poor	94.78*	94.84*	94.81*	94.85*	94.76*
Income group: Never poor	97.71*	97.76*	97.71*	97.74*	97.75*
Husband or partner lives with mother	-0.84	-1.07	-0.92	-0.86	-1.22

* = p-value < .05. Response variable is classroom instructional quality at first grade.

to the combined imputed models (Table 5). Using the multiple imputation technique, the effective number of data vectors used in the model increased from 2833 (original model) to 4452. The estimate of maternal sensitivity decreased from 1.33 in the original model to 0.88 in the combined model. The estimate of maternal depression changed from -0.03 in the original model to -0.05 in the combined model and became significant, at the $p < 0.05$ level, in the combined model. Maternal sensitivity, which had 14% to 26% missing data, experienced the greatest relative increase in variance at 0.34.

Discussion and conclusions

The use of multiple imputations in combination with mixed models is ideal when a significant number of data vectors are not included in the final model. It is very important to look

Table 5. Comparison of Combined Estimates to Original Model.

Variable Name	Combined Estimates (n=4452)	Estimates from Original Model (n=2833)	Relative Increase in Variance
Maternal education	1.21*	1.14*	0.00
Maternal sensitivity	0.88*	1.33*	0.34
Maternal depression	-0.05*	-0.03	0.01
Husband or partner lives with mother	-0.98	-0.85	0.08
Hours/week in care	0.00	0.01	0.00
Income group: Always poor	92.41*	93.1*	0.01
Income group: Late poor	95.22*	95.5*	0.00
Income group: Early poor	94.81*	94.9*	0.00
Income group: Never poor	97.73*	97.9*	0.00

* = p-value < .05. Response variable is classroom instructional quality at first grade.

at the nature of the missing values and type of variable and judge the appropriateness of the imputation before imputing any missing values. Only when it is scientifically logical should missing dependent variable values be filled in by appropriate imputation methods. Using appropriate analysis methods, correctly specifying the variance-covariance components, and properly adjusting degrees of freedom in a mixed model scenario from a partially or completely multiply imputed datasets improves the acceptability of the results to the scientific community. We also found that the estimates and inference value changes in the combined model compared with the original model are not imputation dependent.

Whether fitting a cross-sectional or longitudinal model, a multiple imputation technique can impact model estimates and their corresponding significance levels. The examples demonstrated that even though there is a relative increase in variance associated with a combined estimate, the inclusion of more data vectors can offset this increase and achieve significance where there was none in the original model. The original model with missing values may not be representative of the underlying population model because it did not include all available information. The MI technique is equally difficult to interpret because imputed values are based on an assumed distribution and model. For these reasons it is desirable to compare both the model estimates from the original and multiple imputed datasets on a scientific basis to come up with expected desirable estimates and draw inferences.

Competing interests

The author declares that he has no competing interests.

Acknowledgements

The authors would like to thank the National Institute

of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (SECCYD) for permission to use the data.

Publication history

Editor: Guy Nathaniel Brock, University of Louisville, USA.

EIC: Max K. Bulsara, University of Notre Dame, Australia.

Received: 01-Jul-2013 Revised: 25-Sep-2013

Accepted: 01-Oct-2013 Published: 31-Oct-2013

Citation:

Chakraborty H. **Mixed model variance adjustments when missing values are multiply imputed.** *J Med Stat Inform.* 2013; 1:2. <http://dx.doi.org/10.7243/2053-7662-1-2>

References

1. Chakraborty H. and Gu H. **A Mixed Model Approach for Intent-to-Treat Analysis in Longitudinal Clinical Trials with Missing Values.** RTI Press publication No. MR-0009-0903. Research Triangle Park, NC: RTI International 2009. | [Book](#)
2. Peters SA, Bots ML, den Ruijter HM, Palmer MK, Grobbee DE, Crouse JR, 3rd, O'Leary DH, Evans GW, Raichlen JS, Moons KG and Koffijberg H. **Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models.** *J Clin Epidemiol.* 2012; **65**:686-95. | [Article](#) | [PubMed](#)
3. Siddiqui O, Hung HM and O'Neill R. **MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets.** *J Biopharm Stat.* 2009; **19**:227-46. | [Article](#) | [PubMed](#)
4. Rubin DB. **Multiple Imputation for Nonresponse in Surveys.** New York: John Wiley & Sons, Inc. 1987.
5. Schafer JL. **Analysis of Incomplete Multivariate Data.** New York: Chapman and Hall. 1997. | [Book](#)
6. Li KH, Raghunathan TE, and Rubin DB. **Large-Sample Significance Levels from Multiply Imputed Data Using Moment-Based Statistics and F Reference Distribution.** *Journal of the American Statistical Association.* 1991; **86**:1065-1073. | [Article](#)
7. SAS Institute, Inc., *SAS release 9.2*; Cary, North Carolina 2005.
8. **National Center for Education Statistics (NCES).**
9. Glimpse W.G. **Decennial Census School District Project Planning Report.** Working Paper No. 98-07, Washington, D.C.: U.S. Department of Education, National Center for Education Statistics 1998.
10. National Institute of Child Health and Human Development (NICHD) and Research Triangle Institute (RTI). **The NICHD Study of Early Child Care and Youth Development.** November 9, 2004. | [Website](#)
11. NICHD Early Child Care Research Network. **Characteristics of infant child care: Factors contributing to positive caregiving.** *Early Childhood Research Quarterly.* 1996; **11**: 269-306. | [Article](#)
12. NICHD Early Child Care Research Network. **Child care in the first year of life.** *Merrill-Palmer Quarterly.* 1997; **43**:340-360.
13. NICHD Early Child Care Research Network. **The NICHD Study of Early Child Care.** *Psychiatric Times.* 1998; **15**:71-72.
14. **Relations between family predictors and child outcomes: are they weaker for children in child care?** NICHD Early Child Care Research Network. *Dev Psychol.* 1998; **34**:1119-28. | [Article](#) | [PubMed](#)
15. **Child-care and family predictors of preschool attachment and stability from infancy.** *Dev Psychol.* 2001; **37**:847-62. | [Article](#) | [PubMed](#)
16. NICHD Early Child Care Research Network. **Before Head Start: Income and ethnicity, family characteristics, child care experiences and child development.** *Early Education and Development.* 2001; **12**:545-576. | [Article](#)
17. **Do children's attention processes mediate the link between family predictors and school readiness?** *Dev Psychol.* 2003; **39**:581-93. | [Article](#) | [PubMed](#)
18. NICHD Early Child Care Research Network. **Child care and child development: The NICHD Study of Early Child Care.** In S. L. Friedman & H. C. Haywood (Eds.), *Developmental follow-up: Concepts, domains, and methods* (pp. 377-396). New York: Academic Press 1994.