Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2006 : Methodological Issues in Measuring Population Health

2006

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2006
Methodological Issues in Measuring Population Health

The Effect of Model Specification on Multiply Imputed Data: Lessons Learned from Project DC-HOPE

Marie G. Gantz and M. Nabil El-Khorazaty for the NIH-DC Initiative[1]

Abstract

The District of Columbia Healthy Outcomes of Pregnancy Education (DC-HOPE) project is a randomized trial funded by the National Institute of Child Health and Human Development to test the effectiveness of an integrated education and counseling intervention (INT) versus usual care (UC) to reduce four risk behaviors among pregnant women. Participants were interviewed at baseline and three additional time points. Multiple imputation (MI) was used to estimate data for missing interviews. MI was done twice: once with all data imputed simultaneously, and once with data for women in the INT and UC groups imputed separately. Analyses of both imputed data sets and the pre-imputation data are compared.

KEY WORDS: Multiple Imputation, Intention to Treat, Randomized Trial, Repeated Measures.

1. Introduction

The District of Columbia Healthy Outcomes of Pregnancy Education (DC-HOPE) project is part of the National Institutes of Health (NIH)-DC Initiative, the purpose of which is to reduce infant mortality among minority populations in Washington, DC. Project DC-HOPE, funded by the National Institute of Child Health and Human Development (NICHD), is a randomized trial to evaluate the effectiveness of an integrated education and counseling intervention to reduce four risk behaviors among pregnant women. Women were eligible to participate if they self-identified themselves as African American (AA) or Latina and were at least 18 years old, English-speaking, and less than 29 weeks estimated gestational age (EGA) at the time of enrollment. Participants were recruited at six DC clinics. Women seeking prenatal care at the clinics were asked to complete a brief computerized screening interview (Audio-Computer Assisted Self Interview: A-CASI) (see Thornberry et al., 2003). Eligible women who screened positive for the risk of smoking, environmental tobacco smoke exposure (ETSE), depression, or intimate partner violence victimization (IPV) were invited to participate in the study. Participants were randomly assigned to receive an education and counseling integrated intervention which would coincide with their prenatal care visits (INT) or usual care (UC).

Between July 2001 and October 2003, 2,913 women were screened, and 1,070 enrolled in DC-HOPE. Because the vast majority of women enrolled identified themselves as AA, the 22 who did not were excluded from the final analysis. In addition, four women were deemed ineligible because they failed to complete the requisite BL interview prior to the termination of pregnancy. Of the 1,044 women included in the final analysis, 521 were randomized to the INT group and 523 to the UC group.

Baseline (BL) telephone interviews were conducted an average of 3-4 weeks after recruitment. There were no significant differences in the baseline characteristics of women in the two care groups (see Table 1). Participants were also interviewed at up to three additional time points. Follow-up 1 (F1) took place at 22-26 weeks EGA, Follow-up 2 (F2) occurred 34-38 weeks EGA, and the Postpartum (PP) interview was administered 6-10 weeks after

[1] Marie G. Gantz, Ph.D., RTI International, 3040 Cornwallis Road, Research Triangle Park, North Carolina, USA, 27712; M. Nabil El-Khorazaty, Ph.D., RTI International, 6110 Executive Boulevard, Suite 902, Rockville, Maryland, USA, 20852; The NIH-DC Initiative is a collaboration between Children's National Medical Center, George Washington University SPHHS, Georgetown University Medical Center, Howard University, NICHD and RTI International.

delivery. Unit nonresponse ranged from 43% for F1 to 20% for PP (see Table 2). Missingness was partially due to ineligibility for a given interview resulting from late enrollment, premature delivery, etc. (see Figure 1). The rate of item-level missing data for each interview was less than two percent.

Characteristic	Intervention (N=521)	Usual Care (N=523)	P-value (INT vs. UC)
Age (years)	24.4 ± 5.5	24.8 ± 5.3	0.25
EGA at baseline (weeks)	19.3 ± 7.0	18.6 ± 6.8	0.11
Household size	3.9 ± 1.9	4.0 ± 1.9	0.24
Married/Living with partner	125 (24.0%)	122 (23.3%)	0.80
High school graduate	362 (69.5%)	366 (70.0%)	0.86
Employed	185 (35.5%)	196 (37.6%)	0.49
Medicaid recipient	409 (79.1%)	401 (76.8%)	0.37
Alcohol use during pregnancy (self report)	111 (21.4%)	112 (21.4%)	0.98
Illicit drug use during pregnancy (self report)	67 (12.9%)	56 (10.7%)	0.28
Smoking	106 (20.4%)	92 (17.6%)	0.26
ETSE	365 (71.4%)	377 (73.4%)	0.49
Depression	229 (44.0%)	234 (44.7%)	0.80
IPV	169 (32.4%)	167 (31.9%)	0.86

Table 1. Baseline characteristics of participants in the Intervention and Usual Care groups

Interview	Timing	% of Participants Missing Interview
Baseline (BL)	3-4 weeks after recruitment	0
First follow-up (F1)	22-26 weeks EGA and a minimum of 4 weeks after BL	43
Second follow-up (F2)	34-38 weeks EGA	31
Postpartum (PP)	6-10 weeks after delivery	20

Table 2. Percent of Project DC-HOPE participants missing each interview

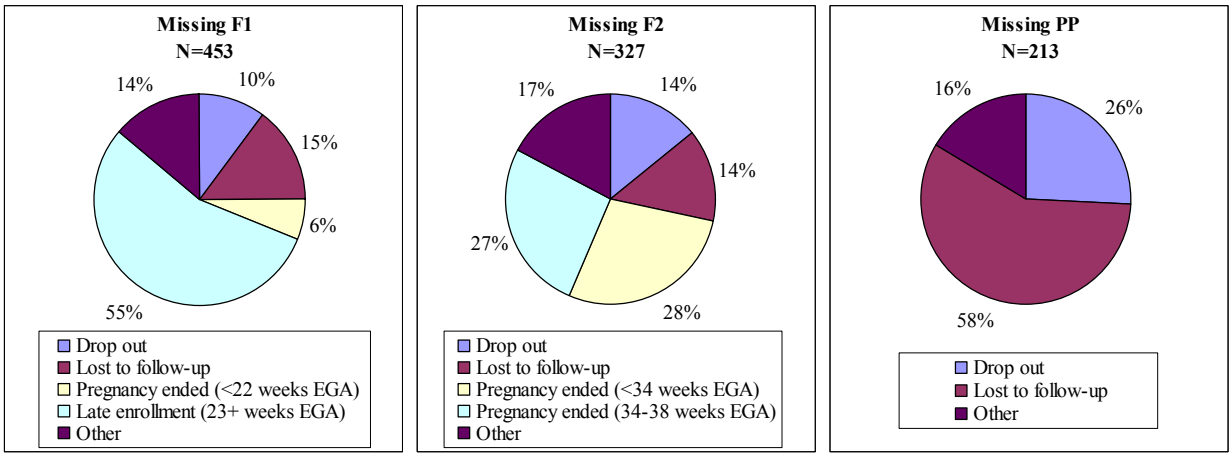

Figure 1. Reasons for missing interviews

Women who were missing one of the follow-up interviews during pregnancy were more likely to be older (OR=1.04 for +1 year, P=0.01), to have smoked more than 100 cigarettes in their lifetimes (OR=1.53, P<0.01), and to be more depressed at BL (OR=1.31 for +1 unit in mean Hopkins Symptom Checklist (HSCL) Score [Derogatis et al., 1973; Derogatis et al., 1974], P=0.01). Participants missing the postpartum interview were more likely to be single (OR=1.49, P=0.04), to receive Temporary Assistance for Needy Families (TANF) (OR=1.50, P<0.01), and to have smoked more than 100 cigarettes in their lifetimes (OR=1.41, P=0.03). In addition, women in the INT group who did not complete the follow-up interviews during pregnancy attended an average of 3.5 fewer intervention sessions than those who did have the interviews (P<0.01). Women in the INT group who did not participate in the PP interview had an average of 3.1 fewer intervention sessions compared to those with the interview (P<0.01).

2. Methods

Our goal for data analysis was to use an Intention-to-Treat (ITT) approach in which participants are analyzed according to their randomized care group (CG) assignment, regardless of actual receipt of intervention or withdrawal from the study. The ITT approach avoids biases that may be introduced when non-compliant individuals are excluded from analysis, and it requires that outcomes be non-missing for all subjects. In order to make an ITT analysis possible, multiple imputation (MI) was used to fill in the missing interview data. MI was accomplished using IVEware imputation and variance estimation software developed by the Survey Methodology Program at the University of Michigan's Survey Research Center, Institute for Social Research. The software and documentation are available for download at http://www.isr.umich.edu/src/smp/ive/. IVEware performs MI using the sequential regression imputation method described by Raghunathan, et al. (2001). Linear, logistic, Poisson, or polytomous regression models are used to impute continuous, binary, count, or categorical missing values. Each time MI was performed, five complete data sets were produced. Variability among the imputed data sets reflects the uncertainty inherent in predicting unknown values.

The entire MI process was repeated twice, using two different approaches. First, MI was conducted on the data set as a whole (total imputation). By default, the regression model IVEware uses to impute missing values for a given variable includes all other variables in the data set as predictors. Due to multicollinearity in our data set, however, we limited predictors to those that increased the R-square of the regression model by 0.135. In the second MI approach, subjects in the INT group were separated from those in the UC group, and MI was conducted for each CG individually (group imputation), using the same set of predictors. This method accounted implicitly for the effect of CG and interactions between CG and the other predictors, thus allowing for more complex imputation models. In the group imputation approach, the minimum R-square criterion for predictive variables was 0.155.

3. Results

For group imputation (in which the data for the two care groups were imputed separately), the imputed values tended to be comparable to the non-missing data for the CG. In contrast, for total imputation (when data for both care groups were imputed together), the imputed values for the INT group tended to be more similar to those of the UC group (see Figure 2). For both imputation approaches, smoking rates were higher than they were in the original data. This is consistent with the fact that women who were missing the F1, F2 and PP interviews were more likely to have smoked more than 100 cigarettes in their lifetimes. Similarly, increased rates of depression in the imputed data sets reflected the fact that women who were more depressed at BL were less likely to complete the follow-up interviews.

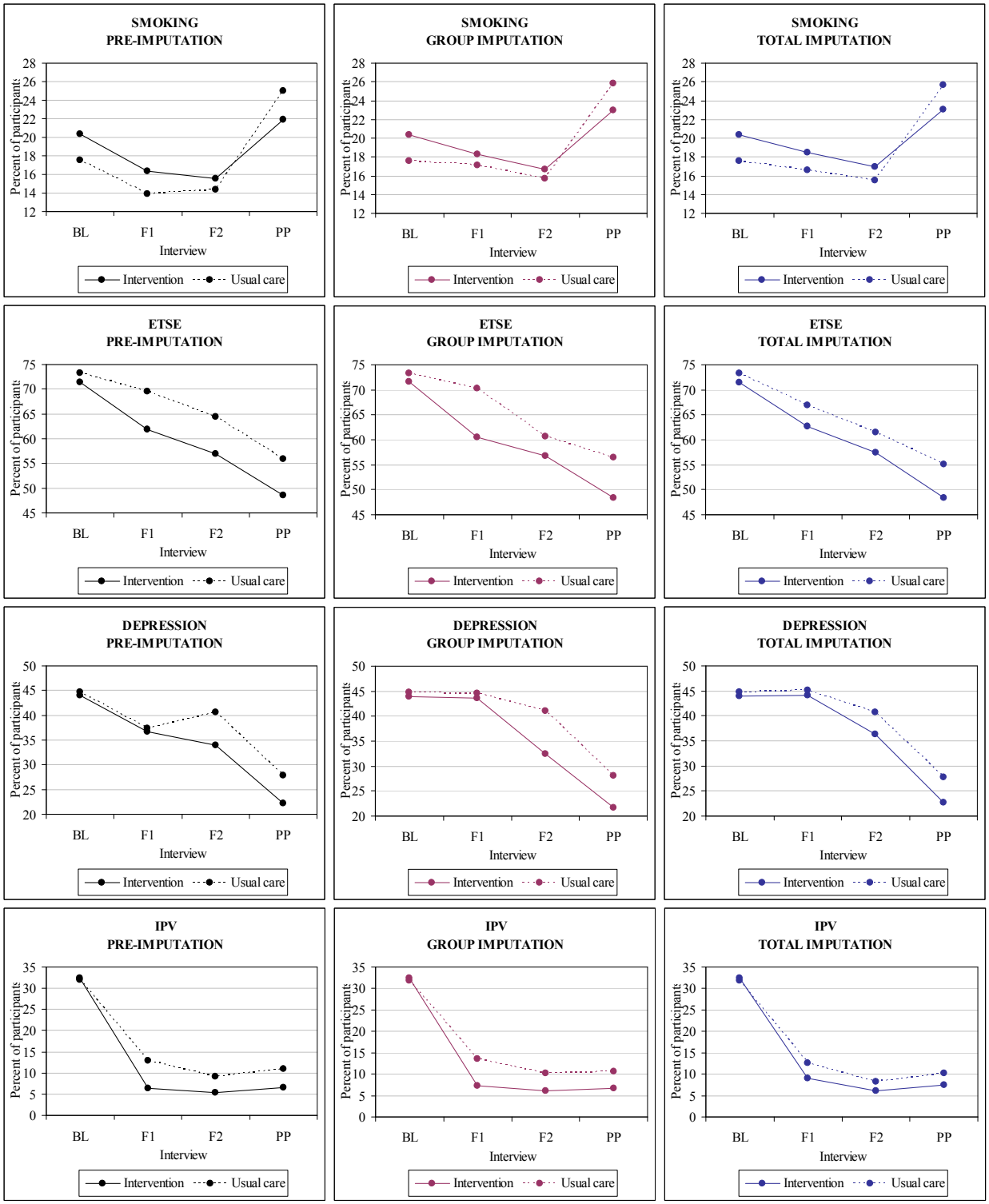

Figure 2. Percent of women reporting risks, in pre-imputation and imputed data sets

Ordinal regression models were created to predict the number of risks participants reported at F2 and at PP. The outcome consisted of four categories: 0 risks, 1 risk, 2 risks and 3-4 risks. Baseline characteristics listed in Table 1 were included as covariates if the effects were significant in the regression model. Analysis of the multiply imputed data involved running the model for each of the five imputations separately and then combining the results into one overall inference that accounted for both between- and within-imputation variance (Rubin, 1987). SAS version 9.1

(SAS Institute, Inc., Cary, NC) software was used for all statistical analysis. The MIANALYZE procedure was used to obtain the multiple imputation inference.

Whether total or group imputation was done, imputation resulted in reduced intervention effects compared to the pre-imputation data (see Table 3). The difference between care groups was smallest for total imputation.

Data Set	Effect of CG (UC vs. INT)			
	Number of Risks at F2		Number of Risks at PP	
	OR	P	OR	P
Pre-imputation	1.65	<.01	1.90	<.01
Group imputation	1.55	<.01	1.74	<.01
Total imputation	1.36	0.04	1.60	<.01

Table 3. Results of ordinal regression model predicting being in a higher risk category (0,1,2,3-4 risks)

Longitudinal methods were used to test for the effect of the intervention over time (F1-PP). Generalized linear mixed models were created for each of the four risk factors, incorporating significant BL covariates. For all four of the risk factors, the model for the data set with the INT and UC groups imputed together (total imputation) had the smallest odds ratio (OR) for the effect of intervention. In the models to predict smoking, the OR for the effect of the intervention on smoking was greatest in the pre-imputation data, but the effect was only significant in the group imputed data (see Table 4). Even then, the significance was not strong (P=0.04). The effect of the intervention on ETSE was significant in all three data sets, though it was least significant when the care groups were imputed together. For depression, only the group imputed data showed a significant intervention effect. The effect of the intervention on IPV was significant in all three data sets.

Data Set	Effect of CG on Risks Over Time (UC vs. INT)							
	Smoking		ETSE		Depression		IPV	
	OR	P	OR	P	OR	P	OR	P
Pre-imputation	1.56	0.75	1.44	<.01	1.37	0.03	1.73	<.01
Group imputation	1.41	0.04	1.35	<.01	1.35	<.01	1.82	0.02
Total imputation	1.24	0.18	1.22	0.04	1.22	0.07	1.48	<.01

Table 4. Results of generalized linear mixed models predicting risks over time (F1-PP)

4. Conclusions

The two approaches to data imputation yielded somewhat different results, yet there were also similarities. Both total and group imputation schemes recognized the fact that smokers and more depressed participants were underrepresented in the follow-up interviews. This fact is encouraging, as it indicates that covariates associated with the missing outcomes were successfully identified by the imputation software and incorporated into the predictive models used to estimate the missing values.

As noted previously, the group imputation approach is akin to explicitly accounting for the effect of CG and all possible interactions between CG and other covariates. The disadvantage of this approach is it assumes that women in the INT group who did not respond to the follow-up interviews were likely to behave in a similar fashion to those who did complete the interviews. This assumption may very well be faulty. While it is possible to adjust the imputation models for known differences between respondents and nonrespondents, such as differing likelihoods of smoking or being depressed, there may be other, unmeasured differences as well. Creating separate imputation models for women in each CG might have resulted in an overestimation of the intervention effect. This bias was potentially mitigated by the presence of a covariate for the number of intervention sessions attended, however, this variable was only included in the regression models if it increased the R-square by 0.155. As a result, the group imputation approach represents a “best case scenario” for the intervention effect that could be detected if there were no missing data.

In contrast, the total imputation approach might represent an overly conservative approach. Recall that the regression models used to estimate missing values for each variable included only those predictors that increased the model R-square by 0.135. Thus, although CG and key interactions involving CG were potential predictors, if those terms did not meet the minimum R-square criterion, they were not included in the regression models. As a result, it is possible that the intervention effect was underestimated in the total imputation data. On the other hand, the similarity in imputed values for the INT and UC groups could be an indication that nonrespondents in the INT group were less likely to reap the benefit of the intervention. This finding would be consistent with the fact that women in the INT group who attended fewer intervention sessions were less likely to complete the follow-up interviews.

In summary, the total imputation approach probably provides a more conservative, yet more accurate, view of the intervention effect. Because women who did not complete the follow-up interviews were less likely to have attended intervention sessions, both the group-imputed and pre-imputation data likely overestimate the effect of being assigned to the INT group. Thus, total imputation is more appropriate for an ITT analysis which evaluates the effect of being randomized to care groups without regard to compliance or drop-out.

References

Derogatis, L. R., Lipman, R. S. and L. Coci (1973), "An outpatient psychiatric rating scale-preliminary report," *Psychopharmacology Bulletin*, pp. 13-28.

Derogatis, L., Rickels, K., Uhlenhuth, E. H. and L. Covi (1974), "The Hopkins symptom checklist: a measure of primary symptom dimensions," in P. Pichot (ed.) *Psychological Measurements in Psychopharmacology: Problems in Psychopharmacology*, Basel: Kargerman, pp. 79-110.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., and P. Solenberger (2001), "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models," *Survey Methodology*, 27, pp. 85-95.

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, Inc.

Thornberry, J. S., El-Khorazaty, M. N., El-Mohandes, A. E., Johnson, A. A., Joseph, J. G., Sivasubramanian, K. N., Kiely, M., and the NIH-DC Initiative to Reduce Infant Mortality in Minority Populations (2003), "Acceptability of Audio Computer-Assisted Self-Interview (A-CASI) as a screening tool for risk factors in women receiving prenatal care," paper presented at the Pediatric Academic Societies' Annual Meeting, Seattle, Washington.