COMPARING IMPUTATION OF ENTIRE SUBSCALES VERSUS INDIVIDUAL ITEMS IN A STUDY OF QUALITY OF LIFE FOLLOWING BREAST CANCER

Thomas R. Belin, Moctar Datt, Kate Desmond, Patricia A. Ganz, UCLA
Thomas R. Belin, 51-267 UCLA Center for Health Sciences, Los Angeles, CA, 90095-1772

Key words: Missing data, multivariate model, ridge prior, composite variables

Abstract

In analyses that use variables comprised of several individual items, missing item values force a decision about whether to treat composite variables as missing or to use responses on available component items. In our motivating example, an initial approach was based on a threshold rule: if a critical number of items was observed on the subscale, then the mean of the available items was treated as the observed value for that variable, otherwise the variable was treated as missing. A more sophisticated alternative that exploits information about associations between component items presents technical challenges because of the large number (183) of variables in the study. We compare the results from regression models fitted to (i) a data set where threshold rules were applied and remaining missing items were multiply imputed in a model with 35 composite variables, and (ii) a data set where imputation models were used to accommodate missingness on any of 183 items. Although most inferences from the alternative approaches were similar, each model featured results where statistical significance at conventional levels was affected by the treatment of missing data.

1. Introduction

In analyses that use variables comprised of several individual items, missing item values force a decision about whether to treat composite variables as missing or to use responses on available component items. The present work was motivated by a study of sexuality in breast cancer survivors, with particular interest focusing on sexual functioning and its relationship to a number of candidate predictor variables (Ganz, Desmond, Belin, Meyerowitz, and Rowland 1999). Data were obtained on 472 women who were at least one year beyond their initial cancer diagnosis and who agreed to complete a lengthy survey. Although most women answered every question, and although a number of items were answered by all respondents, there was a small percentage of missing data on a number of items.

An initial approach to handling missing responses in this study was based on a threshold rule: if a critical number of items was observed on the subscale, then the mean of the available items was treated as the observed value for that variable, otherwise the variable was treated as missing. A more sophisticated alternative that exploits information about associations between component items presents technical challenges because of the large number (164) of variables in the study. We compare the results from regression models fitted to (i) a data set where threshold rules were applied and remaining missing items were multiply imputed in a model with 35 composite variables, and (ii) a data set where imputation models were used to accommodate missingness on any of 164 items. We were interested in whether the alternative approaches would produce similar or different results. Although we do not favor interpreting research results based on a fixed significance level for all questions, we focused on whether conclusions about the statistical significance of various predictors was affected by the treatment of missing data using a conventional significance level (0.05) as a point of reference.

2. Methods

2.1 Available variables

Both the individual-item analysis and the entire-subscale analysis used a core set of 17 items capturing demographic and clinical information on patients. These variables were: age; indicators for African American and other non-Caucasian ethnicity; years since breast cancer diagnosis; indicators for mastectomy with breast reconstruction and mastectomy without reconstruction (with lumpectomy being the reference level); present height, present weight, and weight at time of diagnosis; indicators for any chemotherapy, any hot flashes, any current tamoxifen use, and any serious health condition other than breast cancer; current satisfaction with sexual relationship and self-report of overall impact of breast cancer on sexual functioning; and indicators for whether subject has new sexual partner and whether partner has any of a set of sexual problems. The remaining variables in these analyses were obtained from widely used quality-of-life research instruments. Some of the questions from established scales were omitted to avoid redundancy and to reduce respondent burden.

Several sets of questions were each used as single scales and were not subdivided into subscales. These included the 20-item CES-D depression scale (Radloff 1977), 27 items of the Mental Health Index from the

Medical Outcomes Study (MOS; Stewart, Ware, Sherbourne, and Wells 1992), the 14-item Revised Dyadic Adjustment Scale (Busby, Christensen, Crane, and Larson 1995) measuring marital adjustment, the 17-item Watts Sexual Functioning Scale (Watts 1982), and 12 items from the MOS Social Support scale (Sherbourne and Stewart 1991). The Cancer Rehabilitation Evaluation System, or CARES (Ganz, Schag, Lee, and Sim 1992) contributed subscales on body image (3 items), marital affection (4 items), marital communication (6 items), sexual dysfunction (4 items), sexual interest (4 items), and an overall sexual summary score. The RAND SF-36 health index (Hays, Sherbourne, and Mazel 1993; Ware and Sherbourne 1992) contributed subscales on general health, emotional well-being, physical functioning, social functioning, energy, pain, emotional role limitations, and physical role limitations. Overall, these scales consist of 147 items. When combined with the 17 demographic and clinical characteristics, there were 164 variables included.

2.2 Ad hoc rules for characterizing scales as observed or missing

A number of published reports on well-established quality-of-life measures outline rules for characterizing the outcome as observed or missing based on whether a critical threshold is met regarding the number of component items of the subscale that are observed. For example, for the CES-D scale, which is calculated as a sum of 20 items ranging in score from 0 to 3 with higher scores indicating worse depression, published guidelines (ref) suggest treating the entire scale as missing if more than 4 items are missing. If 1 to 4 items are missing, the guidelines recommend treating missing values as zero scores. In contrast, for the subscales of the RAND-36 quality of life measure (general health, physical functioning, emotional well-being, pain, energy, emotional role limitations, physical role limitations), published references suggest calculating the mean score if any item on the subscale is observed (ref). The discrepancies between these approaches illustrate the ad hoc nature of these rules for handling incomplete item data.

2.3 Multivariate normal imputation

To address the incomplete data problem, we adopted a strategy of producing multiple imputations for missing values and obtaining inferences using the framework of Rubin (1987). We used software for multiple imputation under a multivariate normal model developed by J.L. Schafer, which is available at http://www.stat.psu.edu/~jls. Extensive discussion of this approach is provided in Schafer (1997). Below we describe two different approaches to handling incomplete data, one approach based on modeling subscale scores and the other based on modeling individual items.

In general, suppose the imputation model includes p variables. If, for individual i we label these variables $X_{i,1}, X_{i,2}, \ldots, X_{i,p}$, then the complete data model is:

$$X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,p}) \sim iid \quad N_p(\mu, \Sigma)$$

for $i = 1, 2, \ldots, n$, where n is the total number of subjects. The primary rationale for including all p variables in the imputation model, including both outcomes and predictors, is that the goal of the imputation modeling is to make use of associations present in the observed portion of the data to produce plausible imputed values (Rubin 1996). In particular, omitting outcome variables from the imputation model corresponds to an implicit assumption of zero association between predictors and outcomes when imputing for missing values on predictors, which can lead to attenuation of effects when analyses are done on completed data sets. A collateral benefit of including all of the variables in the imputation model is that the imputation modeling only needs to be done once rather than having to implement separate imputation models for separate analyses.

An attractive feature of a multivariate normal imputation model is that all of the conditional distributions of one variable given the others are linear regression relationships. That is, suppose that we partition X_i into $X_i^{(1)}$ of length p_1 and $X_i^{(2)}$ of length p_2, where $p_1 + p_2 = p$, the dimension of X_i. It follows that

$$\begin{pmatrix} X_i^{(1)} \\ X_i^{(2)} \end{pmatrix} \sim iid \quad N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

where μ_1 and μ_2 are subvectors of μ of length p_1 and p_2, respectively, and Σ_{11}, Σ_{12}, Σ_{21}, and Σ_{22} are the corresponding submatrices of Σ. Furthermore,

$$X_i^{(2)} | X_i^{(1)} \sim iid \quad N_{p_2}(\alpha_{2\cdot1} + \beta_{2\cdot1} X_i^{(1)}, \Sigma_{22\cdot1})$$

where

$$\alpha_{2\cdot1} = \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1,$$

$$\beta_{2\cdot1} = \Sigma_{21}\Sigma_{11}^{-1},$$

$$\Sigma_{22\cdot1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12},$$

which are well-known results from regression analysis. This makes it easy to produce imputed values for continuously scaled variables.

For the data set with 35 variables, there are 595 correlation parameters in Σ, and with 164 variables, there are 13,366 correlation parameters in Σ. To stabilize estimation, we utilized the ridge prior described by Schafer (1997, pp. 155-157). The idea is

that when the sample covariance matrix is singular or nearly singular, estimates will be more stable and computational difficulties can be overcome by smoothing the estimated covariance matrix toward a diagonal matrix, which corresponds to an assumption of a priori independence. Here, we smoothed the estimated covariance matrix toward the diagonals of the sample covariance matrix using a weight corresponding to the information content of three observations (as compared to the 472 observations in the overall data set).

A number of the measures in the present study were categorical. The multivariate normal model is not tailored to this situation, but we employed a strategy to improve the basis for using a multivariate normal model as an approximation to the actual joint distribution. All of the categorical variables here were expressed as binary variables coded 0 or 1. In this context, the multivariate normal model implies that the probability of a 1 is a linear function of the other predictors. In producing multiple imputations in the present setting, we draw binary outcomes with the corresponding predicted probability, with an ad hoc approach of always predicting a 1 if the predicted value under the linear model is greater than 1 and always predicting a 0 if the predicted value under the linear model is less than 0. We consider the impact of this modeling choice in the Discussion section.

3. Results

Here, we contrast results from two regression analyses. The outcome variable in both analyses is a transformed version of the CARES summary score for sexual functioning, specifically log(1 + CARES_SX). The first analysis includes main effect predictors only, and the second analysis added two interaction effects between the CARES body image scale and type of surgery (mastectomy with breast reconstruction or mastectomy without reconstruction, with lumpectomy as the reference level). For each of these models, we present the results from "available-case analysis" (Little and Rubin 1987), where all items must either be observed or singly imputed according to ad hoc criteria for a case to be used in the estimation of model parameters. In these examples, 460 cases were used to fit the regression models in the available-case analyses rather than the entire set of 472 cases. Alongside the available-case analysis, we present the results from the scale-based imputation and the item-based imputation.

The main effects included in both models are age; indicators for African American (Black) or other non-Caucasian ethnic origin (Other); years since cancer diagnosis (YrsOut); a summary of social support (SocSpt); an indicator for any partner problem (APP); indicators for mastectomy with reconstruction (MastRec) and mastectomy without reconstruction (MastNot); indicators for tamoxifen use (Tamox), any chemotherapy (HadChemo), and any history of hot flashes (HotFlash); an indicator for any other serious health condition (AnyCond); a summary of general health status (GnrlHlth); an indicator for having a new sexual partner (NewPart); and summaries from the revised Dyadic Adjustment Scale (RDAS), Mental Health Index (MHI), and CARES Body Image (BodyImg) scales.

The variables Age, Black, Other, YrsOut, APP, MastRec, MastNot, Tamox, HotFlash, and AnyCond were fully observed. The variable HadChemo was missing for one subject (0.21%), and NewPart was missing for two subjects (0.42%). Among the scale variables when deterministic rules were applied, GnrlHlth, MHI, and BodyImg were treated as fully observed; the outcome CARES_SX was missing for 9 individuals; the RDAS was missing for two subjects (0.42%); and SocSpt was missing for one subject (0.21%). However, there were scattered missing values on subscale items, particularly on questions with sexual content. Still, the maximum rate of missingness on any item in this analysis was 3.6% for a few items included in the CARES_SX summary scale. (Some of the items on the Watts scale had rates of missingness between 4% and 5%, but they were used here only in the imputation model and not in the regression models for evaluation purposes.)

Table 1 displays the results of the regression of log(1+CARES_SX) on the main effects of the variables listed above using available-case analysis, scale-based imputation, and item-based imputation. Most of the predictors that appear significant in the available-case analysis (APP, HadChemo, HotFlash, NewPart, RDAS, MHI, BodyImg) remain significant in the scale-based and item-based imputation models. However, there are some differences across the various approaches. Specifically, the predictor variable Age is significant only in the scale-based imputation approach, while the variable YrsOut is significant in the available-case analysis as well as in the item-based imputation approach, but not in the scale-based imputation approach.

Table 2 displays the results of the regression of log(1+CARES_SX) on the main effects of the variables listed above along with interactions between the BodyImg indicator and the indicators of type of surgery (MastRec and MastNot). The results for the main effects are similar to those in Table 1. There is an additional discrepancy regarding the significance of the interaction between BodyImg and MastRec, with this predictor being significant at the 0.05 level in the available-case and item-based imputation analyses but

only borderline significant (p = 0.0750) in the scale-based imputation analysis.

Figures 1(a) and 1(b) plot the estimated percentages of missing information about the regression coefficients for the item-based imputation analysis versus the scale-based imputation analysis from the respective models. Although the percentages of missing information are not uniformly higher in the item-based analysis, as expected the percentages of missing information are generally higher in the item-based analysis.

4. Discussion

Despite the small percentage of missing items in this example, the findings in the regression analyses we performed were somewhat sensitive to the choice of imputation strategy. Rubin (1996) favors reflecting as accurately as possible the full complexity of the data at hand while noting that it can be technically demanding to do so. Here, we used available technology for fitting high-dimensional multivariate normal models, adapting the results to the present setting through the use of techniques for converting continuously-scaled imputations into binary values where appropriate. The approximations used in that step certainly could influence the results as well; the fact that the same approach was used for item-based and scale-based imputation suggests, however, that there are still is the potential for diverging results depending on the choice of imputation model. The item-based imputation approach is attractive both because of the way it incorporates observed data on items from subscales and because it is apt to generate larger variances and hence to be more conservative. Development of flexible multivariate approaches for handling the high-dimensional data seen in behavioral and quality-of-life research continues to be of considerable interest.

References

Busby, D.M., Christensen, C., Crane, D.R., Larson, J.H. (1995), "A Revision of the Dyadic Adjustment Scale for Use with Distressed and Nondistressed Couples: Construct Hierarchy and Multidimensional Scales," *Journal of Marital and Family Therapy*, 21:289-308.

Ganz, P.A., Desmond K.A., Belin, T.R., Meyerowitz, B.E., Rowland, J.H. (1999), "Predictors of Sexual Health in Women After a Breast Cancer Diagnosis," *Journal of Clinical Oncology*, to appear.

Ganz, P.A., Schag, C.A.C., Lee, J.J., Sim, M.S. (1992), "The CARES: A Generic Measure of Health-Related Quality of Life for Cancer Patients," *Quality of Life Research*, 1:19-29.

Hays, R.D., Sherbourne, C.D., Mazel, R.M. (1993), "The RAND 36-item Health Survey 1.0," *Health Economics*, 2:217-227.

Little, R.J.A., Rubin, D.B. (1987), *Statistical Analysis with Missing Data*, New York: John Wiley and Sons.

Radloff, L.S. (1977), "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population," *Applied Psychological Measurement*, 1:385-401.

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

Rubin, D.B. (1996), "Multiple Imputation after 18+ Years," *Journal of the American Statistical Association*, 91:473-489.

Schafer, J.L. (1997), *Analysis of Incomplete Multivariate Data*, New York: Chapman and Hall.

Sherbourne, C.D., Stewart, A.L., (1991), "The MOS Social Support Survey," *Social Science and Medicine*, 32:705-714.

Stewart, A.L., Ware, J.E. Jr., Sherbourne, C.D., Wells, K.B. (1992), "Psychological Distress/Well-Being and Cognitive Functioning Measures," in *Measuring Functioning and Well-Being: The Medical Outcomes Study Approach*, A.L. Stewart and J.E. Ware Jr., eds., Durham, NC: Duke University Press, 102-142.

Ware, J.E. Jr., Sherbourne, C.D. (1992), "A 36-item Short Form Health Survey (SF-36): I. Conceptual Framework and Item Selection," *Medical Care*, 30:473-483.

Watts, R.J. (1982), "Sexual Functioning, Health Beliefs, and Compliance with High Blood Pressure Medications," *Nursing Research*, 31:278-283.

Table 1. Comparison of methods for main effects model, outcome = log (1 + CARES_SX)									
	Available-case analysis			Scale-based imputation			Item-based imputation		
Predictors	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
Intercept	.9839	.1620	**.0001**	1.0139	.1569	**.0001**	1.0100	0.160	**.0001**
Age	.0029	.0020	.1486	.0041	.0019	**.0317**	.0030	.0019	.1185
Black	.0156	.0483	.7474	.0141	.0465	.7614	.0109	.0476	.8187
OtherEth	-.0500	.0570	.3809	-.0311	.0555	.5751	-.0447	.0568	.4310
YrsOut	.0286	.0128	**.0262**	.0211	.0125	.0909	.0262	.0126	**.0380**
SocSpt	-.0013	.0012	.2674	-.0019	.0011	.0994	-.0015	.0012	.1901
APP	.1575	.0363	**.0001**	.1441	.0352	**.0001**	.1487	.0358	**.0001**
MastRec	-.0003	.0433	.9941	.0170	.0421	.6870	.0028	.0428	.9475
MastNot	-.0307	.0443	.4880	-.0152	.0419	.7157	-.0301	.0434	.4880
Tamox	-.0002	.0352	.9946	.0006	.0341	.9871	.0103	.0349	.7680
HadChemo	.0998	.0359	**.0057**	.1023	.0346	**.0032**	.1048	.0353	**.0030**
HotFlash	.0777	.0359	**.0310**	.0783	.0350	**.0251**	.0771	.0356	**.0301**
AnyCond	.0555	.0350	.1140	.0464	.0341	.1731	.0525	.0346	.1293
GnrlHlth	-.0016	.0010	.1112	-.0013	.0010	.1782	-.0013	.0010	.1700
NewPart	-.2648	.0788	**.0008**	-.2300	.0754	**.0023**	-.2559	.0773	**.0009**
RDAS	-.0053	.0025	**.0376**	-.0066	.0025	**.0073**	-.0059	.0025	**.0201**
MHI	-.0039	.0016	**.0145**	-.0036	.0016	**.0197**	-.0039	.0016	**.0138**
BodyImg	.0891	.0172	**.0001**	.0855	.0166	**.0001**	.0874	.0170	**.0001**

Boldface type indicates result significant at $p < .05$.

Table 2. Comparison of methods for main effects model, outcome = log (1 + CARES_SX)									
	Available-case analysis			Scale-based imputation			Item-based imputation		
Predictors	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
Intercept	.9835	.1635	**.0001**	1.0180	.1589	**.0001**	1.0081	.1619	**.0001**
Age	.0028	.0020	.1525	.0041	.0019	**.0315**	.0030	.0019	.1170
Black	.0129	.0483	.7896	.0119	.0466	.7986	.0079	.0475	.8679
OtherEth	-.0657	.0573	.2522	-.0403	.0559	.4704	-.0593	.0570	.2980
YrsOut	.0289	.0128	**.0245**	.0208	.0125	.0953	.0261	.0126	**.0378**
SocSpt	-.0012	.0012	.2938	-.0018	.0011	.1095	-.0014	.0012	.2216
APP	.1629	.0362	**.0001**	.1476	.0352	**.0001**	.1539	.0357	**.0001**
MastRec	.0751	.0560	.1808	.0778	.0546	.1543	.0796	.0554	.1504
MastNot	.0046	.0570	.9355	-.0004	.0547	.9935	-.0011	.0556	.9839
Tamox	-.0009	.0351	.9806	-.0013	.0341	.9706	.0095	.0348	.7848
HadChemo	.0911	.0360	**.0117**	.0982	.0347	**.0046**	.0972	.0353	**.0059**
HotFlash	.0746	.0358	**.0379**	.0750	.0350	**.0320**	.0733	.0355	**.0388**
AnyCond	.0485	.0351	.1673	.0412	.0342	.2286	.0455	.0346	.1886
GnrlHlth	-.0017	.0010	.0846	-.0014	.0010	.1472	-.0015	.0010	.1307
NewPart	-.2746	.0786	**.0005**	-.2363	.0754	**.0017**	-.2649	.0771	**.0006**
RDAS	-.0057	.0025	**.0263**	-.0068	.0025	**.0055**	-.0062	.0025	**.0134**
MHI	-.0038	.0016	**.0181**	-.0036	.0016	**.0215**	-.0038	.0016	**.0170**
BodyImg	.1354	.0283	**.0001**	.1151	.0268	**.0001**	.1323	.0281	**.0001**
BodyImg × MastRec	-.0929	.0409	**.0235**	-.0700	.0393	.0750	-.0929	.0403	**.0211**
BodyImg × MastNot	-.0536	.0365	.1423	-.0288	.0353	.4143	-.0487	.0362	.1785

Boldface type indicates result significant at $p < .05$.