Qual Quant (2010) 44:277–287
DOI 10.1007/s11135-008-9196-5

A comparison of multiple imputation with EM algorithm and MCMC method for quality of life missing data

Ting Hsiang Lin

Published online: 11 September 2008

Abstract This study investigated the performance of multiple imputations with Expectation-Maximization (EM) algorithm and Monte Carlo Markov chain (MCMC) method in missing data imputation. We compared the accuracy of imputation based on some real data and set up two extreme scenarios and conducted both empirical and simulation studies to examine the effects of missing data rates and number of items used for imputation. In the empirical study, the scenario represented item of highest missing rate from a domain with fewest items. In the simulation study, we selected a domain with most items and the item imputed has lowest missing rate. In the empirical study, the results showed there was no significant difference between EM algorithm and MCMC method for item imputation, and number of items used for imputation has little impact, either. Compared with the actual observed values, the middle responses of 3 and 4 were over-imputed, and the extreme responses of 1, 2 and 5 were under-represented. The similar patterns occurred for domain imputation, and no significant difference between EM algorithm and MCMC method and number of items used for imputation has little impact. In the simulation study, we chose environmental domain to examine the effect of the following variables: EM algorithm and MCMC method, missing data rates, and number of items used for imputation. Again, there was no significant difference between EM algorithm and MCMC method. The accuracy rates did not significantly reduce with increase in the proportions of missing data. Number of items used for imputation has some contribution to accuracy of imputation, but not as much as expected.

Keywords Missing data · Imputation · EM algorithm · MCMC · Quality of life

T. H. Lin (✉)
Department of Statistics, National Taipei University, 67, Section 3, Ming-Sheng East Road, Taipei 10433, Taiwan, ROC
e-mail: tinghlin@mail.ntpu.edu.tw

1 Introduction

Missing data frequently arise in real world application and they can lead to misleading and potentially dangerous conclusions. For example, missing data in health-related assessment such as quality of life instruments may draw fallacious conclusion in drug development or new intervention in clinical trials (Olschewski et al. 1994; Curran et al. 1998a; Fayers and Machin 2000). Missing data can introduce bias into studies and obscure implication might be imbedded in the missing-ness. Therefore, it is important that appropriate and effective methods available to resolve the problems of missing data. There are two types of missing data are commonly discussed in literatures (Fayers et al. 1998; Curran et al. 1998b). The first type is referred as "unit non-response" when a whole questionnaire is completely left unanswered or unreturned (Curran et al. 1998b); the second type is "item non-response" when one or more items are omitted in a questionnaire (Fayers et al. 1998).

The impact of missing data and the ways to handle incomplete data depend much upon the patterns of incompleteness. A set of definitions for missing data mechanisms has been provided by Little and Rubin (1987), including missing completely at random (MCAR), missing at random (MAR), and non-ignorable missing data (MNAR). MCAR occurs when the missing values on variable Y are independent of all other observed variables and the values of Y itself. In other words, the missing and observed distributions of Y are identical, or it can be expressed as P(Y | y missing) = P(Y | y observed). For example, even if people who refuse to report their education also invariably refuse to provide responses to their income, it's still possible that the data could be missing completely at random. However, MCAR is a very strong assumption and can be impractical for real data (Muthén et al. 1987). The second pattern "missing at random" (MAR) provides a more realistic condition. Under MAR, the probability that an observation is missing on variable Y depends on other observed variables X, but not on the values of Y itself, and the observed values are not necessary a random sample of the hypothetically complete data set. The missing and observed distributions of Y conditional on some other observed variables X are identical, or it can be expressed as P(Y | y missing, y oserved, X) = P(Y | y observed, X). MCAR and MAR are both ignorable when the parameters governing the missing data process are not related to the parameters of interest, and therefore it is not required to model the missing-ness as part of the estimation process. Another missing data pattern is non-ignorable. The missing and observed values of Y are not the same under any conditions, or it can be expressed as P(Y | y missing) $\neq$ P(Y | y observed).The estimates of Y or the relationships between Y and other variables are likely to be biased. More data collection is needed to resolve the problems.

Several methods have been proposed to handle missing data, including complete case method, pair-wise deletion, simple mean imputation, model based imputation (Little and Rubin 1987), and we will address some frequently used methods. The first method is complete case method that ignores and removes all cases with any incomplete data in the analysis. It is also referred as "list-wise deletion" or "case-wise deletion". The method requires MCAR assumption and only under MCAR the parameter estimates are consistent (Arbuckle 1996; Brown 1994; Wothke 2000) and it can yield biased parameter estimates if MCAR conditions does not hold. The problem is that it ignores potentially systematic differences between complete and incomplete cases, and does not produce minimum squared errors for a given sample size. It can also lead to improperly larger variance and higher type II error when testing statistical hypotheses.

The second method is pair-wise deletion and it computes the summary statistics based on all available cases, and the summary statistics can be used to estimate the parameters of interest. The method is applied in many linear models, including regression, factor

analysis, or even in more sophisticated models such as Structural Equation Modeling (SEM). Under MCAR, pair-wise deletion produces consistent and unbiased parameter estimates in large samples. Compared with list-wise deletion, pair-wise deletion yields more efficient estimates and has smaller standard errors in linear regression models when the correlations among the variables are relatively low, while list-wise deletion does better when the correlations are high (Glasser 1964; Haitovsky 1968; Kim and Curry 1977). However, the estimates may be seriously biased if the data are only MAR. In real world situation, the consequence of pairwise deletion in multivariate analysis may lead to loss of sample size and decrease in statistical power.

The third commonly used method is called 'simple mean imputation'. If an item or items of the same domain are missing, each missing value will be replaced with the mean value of other answered items from that domain or related items. This method assumes that at least a fixed number of items on that domain are answered. Mean substitution is a good solution when data is missing at random and normally distributed, and the mean estimate is consistent. However, this method will produce biased estimates of variance and covariance (Haitovsky 1968). When the proportion of missing data increases, it can reduce the variance and leads to a larger R^2 in regression.

Model-based imputation method for missing data is also a popular alternative. Multiple imputations provide an alternative for dealing with data sets with missing values. Rubin's (1987) multiple imputation procedure replaces each missing value with a set of plausible values that represent the uncertainty about the values to be imputed. The multiple imputed data sets are then analyzed by using standard procedures for complete data and combining the results from these analyses. No matter which complete-data analysis is used, the process of combining results from different imputed data sets is essentially the same. This results in valid statistical inferences that properly reflect the uncertainty due to missing values.

Multiple imputation inference involves three distinct phases. First, the missing data are filled in m times to generate m complete data sets. Second, the m complete data sets are analyzed by using standard procedures. Finally, the results from the m complete data sets are combined for the inference. There are several methods that are often used with multiple imputation, including Expectation Maximization (EM-algorithm) (Dempster et al. 1997; Schafer 1997; McLachlan and Krishnan 1997) and Monte Carlo Markov chain (MCMC) (Gilks et al. 1995) method. The Expectation Maximization is a two-step iterative approach to estimate the parameters in the model. It finds maximum likelihood estimates by repeating Expectation (E-step) and Maximization (M-step) steps in parametric models for incomplete data. An E-step finds the distribution for the missing data based on the known values for the observed data and the current estimates of the parameters; and an M-step substitutes the missing data with the expected values. Under SEM, given a set of parameter estimates such as mean vector μ and covariance matrix Σ for a multivariate normal distribution, the E-step calculates the conditional expectation or covariance of the missing data given the observed data and the parameter estimates, in other words, it calculates $E(y_{i\ miss}|y_{i\ obs};\ \widehat{\mu},\ \widehat{\Sigma})$ and Cov $(y_{i\ miss}\ y_{i\ obs};\ \widehat{\mu},\ \widehat{\Sigma})$ for each case i. The values obtained from the E-Step are used to obtain new values of μ and Σ and the M-step finds the parameter estimates to maximize the complete-data log likelihood from the E-step. These two steps are repeated until $(\widehat{\mu}_{K+1},\ \widehat{\Sigma}_{K+1})$ are essentially the same as $(\widehat{\mu}_{K},\ \widehat{\Sigma}_{K})$ or the iteration converges. If there are G distinct missing patterns, the observed-data log likelihood function being maximized can be expressed as:

$$\ln L(\theta|Y_{obs}) = \sum_{g=1}^{G} \ln L_g(\theta|Y_{obs})$$

where $\ln L_g(\theta|Y_{obs})$ is the observed-data log likelihood from the g_{th} pattern, and it can be denoted as:

$$\ln L_g(\theta|Y_{obs}) = -\frac{n_g}{2}\ln|\Sigma_g| - \frac{1}{2}\sum_{i_g}^{n_g}(y_{ig} - u_g)'(y_{ig} - u_g)$$

where n_g is the number of observations in the g_{th} pattern; u_g is the corresponding mean vector, Σ_g is the covariance matrix, and y_{ig} is a vector of observed values corresponding to observed variables for case i.

Monte Carlo Markov chain (MCMC) method is based on pseudo-random draws and allows researchers obtain several imputed data sets. MCMC can be used with both arbitrary and monotone patterns of missing data, and it requires either MAR or MCAR assumption. A Markov Chain is a sequence of random variables in which the distribution of each element depends only on the value of the previous one. In MCMC simulation, one constructs a long Markov Chain to establish a stationary distribution, which is the distribution of interest. By repeatedly simulating steps of the chain, the method draws imputed estimates from the distribution. In general, there are several steps for MCMC method:

1. Starting value. Compute mean and covariance matrix based on all observed data. Use these initial values to estimate the prior distribution.
2. Imputation step. Simulate values for missing data items by randomly selecting a value from the available distribution of values. Repeat the procedure until the difference of the mean vector and covariance matrix between two iterations is smaller than a pre-specified criterion or the distribution is stationary. If the iterations are enough, use imputations from final iteration to form a data set that has no missing values.
3. Posterior step (P-step). If there are not enough iterations, re-compute mean vector and covariance matrix with the imputed estimates from the imputation step, and this is the posterior distribution.

In LISREL, the estimates of μ and Σ obtained from the EM algorithm are used as initial parameters of the distribution used in the first step of the MCMC procedure. In the I-step of LISREL, the missing values given the observed value for each case are simulated from conditional normal distributions with parameters based on μ_k and Σ_k. In the P-step, an estimate μ_k of μ and an estimate Σ_k of Σ are simulated from a multivariate normal and an inverted Wishart distribution. The missing values are then replaced with simulated values, and the next set of μ_k and Σ_k are calculated by the new completed data set. The procedure repeats P-step and I-step n times. In LISREL, missing values are replaced by the average of the simulated values over the n draws (Jöreskog and Sörbom 2004).

In terms of missing data imputation in health-related instruments, the issue of which items should be included for imputation has rarely been addressed. This often happens in most quality of life assessment as well, and it is conventional to use simple mean method with other items of the same domain for imputation. Moreover, other imputation approaches have not been discussed or applied in quality of life assessment yet. This paper focuses on the issues involved in handling forms which contain one or more missing items, and reviews the alternative procedures. Specifically, we will investigate the performance of EM algorithm and MCMC method, and compare the imputation results from domain as well as scale level and examine the impact of missing data rate, and number of items included for imputation.

2 Method

2.1 Data and subjects

The World Health Organization (WHO) originally developed the WHOQOL-100 (World Health Organization Quality of Life Survey) which consists of 100 items used to measure the quality of life. The WHOQOL-100 was later shortened to WHOQOL-BREF (World Health Organization Quality of Life Survey Abbreviated Version) and comprises 26 items plus some national items. The WHOQOL-BREF is one component in 2001 National Health Interview Survey (NHIS) (Lin et al. 2003) in Taiwan which is intended to provide nation-wide estimates for health conditions, health behaviors, and usage of medical resources. The sampling scheme used for NHIS is a multistage stratified systematic sampling design. Face-to-face personal interviews with structured questionnaires were used at the subjects' residences by trained interviewers for data collection. The WHOQOL-BREF Taiwan version contains twenty-six items plus two national items reflecting cultural importance. These two national questions are: "Do you feel respected by other", and "Are you usually able to get the things you like to eat". There are four domains in WHOQOL-BREF: physical health, psychological health, social relationship, and environment. The physical health contains seven items, including facets of pain, medication, energy, mobility, sleep, activities of daily living, and work. The psychological health has six items measuring positive feelings, spirituality, thinking, bodily image, self-esteem, and negative feelings. The social relationship has four items on facets of personal relationship, sexual activity, social support, and one national item on being respected/accepted. Nine items on environment included the facets of physical safety, physical environment, financial resources, information availability, leisure activities, living environment, health and medical care, transport, and one national item on food/eating. The domain scores are scaled in a positive direction (i.e., higher scores indicate higher quality of life). The negative items have been reversely coded so that all items in the scale indicate a positive direction, and 1 represents "not satisfied at all" or strongly disagree with the statement, and 5 represents "extreme satisfied" or strongly agree with the description of that item. There are three negatively worded items: item 3 (pain), item 4 (medication), and item 26 (negative feeling). According to the WHOQOL-BREF user's manual, if there are more than two items are missing from the domains, the domain scores should not be calculated with the exception of social relationship that allows at most one missing item. When more than 20% of data is missing from the instrument, the case should be discarded. When an item is missing, the mean of other items in the domain is substituted.

The eligible respondents are between 20 and 65 years old, and were selected in NHIS and volunteered to participate the WHOQOL-BREF. The sample analyzed in this study must satisfy the scoring criteria as above. A total of 13017 individuals were included in this study. Forty-nine percents of the subjects is male, and the average age is 38.11 years with a standard deviation of 11.32. Near thirty percents of the respondents have received beyond high school education, and the average years of education received is 11.59 year. There are 63% of the subjects are married, 2.95% divorced, 2.12% widowed, and 27.51% are single. The PRELIS 2 computer program (Jöreskog and Sörbom 1996) was used for imputation.

3 Study design

We examined the research questions in two ways. In order to study the effects of the number of items within a domain and missing data rates simultaneously, we chose two extreme

domains: social relationship and environmental domain. The social relationship represents a domain with fewest items among the four, and the item being imputed has highest missing data rate; while the environmental just represents the opposite: a domain with most items and the item to be imputed has lowest missing data rate. In the first empirical analysis, we examined social relationship from both item level and domain level. In the second study, we chose environmental domain and conducted a simulation analysis to examine the effect of the following variables: EM algorithm and MCMC method, missing data rates, and number of items used for imputation.

We first analyzed social relationship domain that has fewest items out of the four domains in WHOQOL-BREF, and we chose item 21 (How satisfied are you with your sex life?) as the item to be imputed since it has the highest missing values rate among the four. We imputed the missing values by including different amount of data: on item levels, we used all other three items of social relationship domain for imputation and also included all other 27 items of the entire WHOQOL-BREF scale to impute the missing values of item 21. On the domain level, we use summation scores of all three other subscales to impute the missing summation scores of the social relationship. The imputation has been performed by both EM algorithm and MCMC method.

In the second analysis, we conducted a Monte Carlo simulation study to assess the performance of EM algorithm and MCMC method. In order to best reserve the properties and nature of quality of life data, we conducted the study based on real data instead of simulated data. We used NHIS set as described above, and we selected one item for imputation. We chose the environmental domain with most items, and the item chosen for imputation is item 14 (To what extent do you have the opportunity for leisure activities?) since it has the lowest missing data rate among all the items in the environmental domain. A set of 2, 5, and 10 percents of data are drawn by simple random sampling method from the NHIS datasets, or the sample sizes equal 260, 650 or 1,300, respectively. Once the cases were drawn, item 14 of these cases are replaced by missing values, i.e., these data are "artificial missing data". Ten datasets are generated for each configuration. We first included the nine items in environmental domain only for imputation, and then we repeated the imputation again by using all other 27 items. The imputations are conducted by both EM algorithm and MCMC method.

4 Results

In the first analysis, we examined social relationship from both item level and domain level and Table 1 displays the imputed value of item 21. The total number of missing observations

Table 1 The observed and imputed values for item 21[a] (in percentages)

Method	No. of items used for imputation	Imputed values					
		1[b]	2	3	4	5	Mean (SD)
EM	4	0.00	0.51	42.23	57.26	0.00	14.103 (2.087)
EM	28	0.00	0.59	41.94	56.38	1.10	14.104 (2.086)
MCMC	4	0.00	1.17	41.72	55.72	1.39	14.104 (2.085)
MCMC	28	0.00	2.13	41.42	54.77	1.69	14.102 (2.083)
Observed	28	1.69	5.85	34.99	53.38	4.09	14.096 (2.087)

[a] Item 21 (How satisfied are you with your sex life?) is contained within the social relationships domain
[b] The number 1 represents "not satisfied at all" or strong disagreement with the item statement and the number 5 represents "extremely satisfied" or strong agreement with the item statement

Table 2 Accuracy rates of EM algorithm by number of items and missing data rates

Observed value	Types of accuracy rates	Missing data rates					
		9 Items			All items		
		2%	5%	10%	2%	5%	10%
1	Exact	6.19	2.27	5.06	8.85	6.17	9.27
	Loose	69.91	64.61	64.08	72.57	70.78	71.67
2	Exact	32.42	35.85	34.47	33.59	35.32	35.28
	Loose	95.29	95.49	95.00	96.21	96.28	95.96
3	Exact	81.06	80.25	80.93	80.75	80.30	80.84
	Loose	100.00	100.00	100.00	99.90	100.00	99.87
4	Exact	26.35	26.24	25.43	28.74	31.08	30.37
	Loose	94.76	96.89	96.47	96.11	97.06	97.14
5	Exact	0.00	0.00	0.00	3.30	3.72	3.73
	Loose	59.34	51.16	51.28	63.74	59.07	62.48

The rows for "loose accuracy rate" represent the percentages when an actual response was imputed exactly as itself or its 'neighboring values' (i.e. the difference between the observed and imputed values was 1 or 0): for example, when the actual response value was 1 and it was imputed as 1 or 2, when the actual response value was 2 and it was imputed as 1, 2 or 3, or when the actual response value was 5 and it was imputed as 4 or 5

being imputed is 1364. The percentages of the actual observed values of 1 through 5 are: 1.69, 5.85, 34.99, 53.38, and 4.09. As it indicates in Table 1, among the imputed values, there is no significant difference between EM algorithm and MCMC method. It has also shown that number of items used for imputation has little impact regardless whether 4 items or 28 items were included for imputation. Compared with the observed values, the middle responses of 3 and 4 are over imputed, and the extreme responses of 1, 2 and 5 are under represented.

The similar patterns occurred for domain imputation. We first calculated the summation scores of all four domains, and imputed social relationship domain score with all three domains for those cases with missing score in social relationship. Then we compared these results with the summation score from Table 1. The imputed value of social relationship domain by EM algorithm has a mean score of 14.102 with a standard deviation 2.032; by MCMC method, we obtained a mean score of 14.101 and standard deviation equals 2.040. The results show the difference between MCMC and EM are quite trivial.

In the second analysis, we conducted a simulation study to assess the performance of EM algorithm and MCMC method. Tables 2 and 3 summarize the accuracy of the EM algorithm and MCMC method by various missing data rates and different number of items. In both tables, the "exact accuracy rates" display the percentages of exact agreement between the imputed values and the actual observed data values, i.e., the percentages of when the actual response was 1, and it was also imputed as 1 (same for responses 2 through 5). The "loose accurate accuracy" represents the percentages when the actual response was imputed as itself or its "neighboring values". For example, when the actual response value is 1, and it was imputed as 1 or 2; when the actual response value is 2, and it was imputed as 1, 2 or 3; when the actual response value is 5, and it was imputed as 4 or 5.

When the original actual response is missing, it is excluded from calculating the accuracy rate. As it indicates in Tables 2 and 3, there are several major findings. First, the results of the EM algorithm and MCMC method are similar, and the accuracy rate for the imputation is about the same. The number of items has some contribution to the accuracy, but not as much as expected. When adding number of items used for imputation from nine to twenty-eight

Table 3 Accuracy rates of MCMC method by number of items and missing data rates

Observed value	Types of accuracy rates	Missing data rates					
		9 Items			All items		
		2%	5%	10%	2%	5%	10%
1	Exact	8.85	3.90	6.58	14.16	10.39	9.44
	Loose	67.26	63.96	64.25	73.45	69.16	70.49
2	Exact	31.76	35.74	34.12	37.25	36.80	35.71
	Loose	92.94	93.57	92.78	94.51	94.64	94.19
3	Exact	74.19	73.55	73.24	74.51	72.95	72.12
	Loose	100.00	99.92	99.85	99.90	99.75	99.74
4	Exact	25.75	28.43	28.82	31.29	32.41	32.71
	Loose	93.86	95.16	94.70	95.06	95.21	95.58
5	Exact	1.10	2.33	3.54	3.30	7.44	7.27
	Loose	52.75	52.56	52.26	62.64	56.74	59.33

The rows for "loose accuracy rate" represent the percentages when an actual response was imputed exactly as itself or its 'neighboring values' (i.e. the difference between the observed and imputed values was 1 or 0): for example, when the actual response value was 1 and it was imputed as 1 or 2, when the actual response value was 2 and it was imputed as 1, 2 or 3, or when the actual response value was 5 and it was imputed as 4 or 5

items, the largest increase of the accuracy rates across all conditions is about 8%. The extreme response values such as 1 and 5 are worst imputed, and it has the lowest accuracy rate. The exact accuracy rates are not quite satisfactory. However, the loose accurate accuracy rates are very high.

5 Discussion

This study investigated the performance of two methods with some missing data conditions in multiple imputations. The effects of EM algorithm and MCMC method, number of items, and missing data rates were examined. The results of the simulation study were somewhat unexpected. The imputing behavior of EM algorithm and MCMC method are quite similar, and there is no significant difference between EM algorithm and MCMC method in terms of accuracy for imputation. One possible reason is that in LISREL, the estimates of μ and Σ obtained from the EM algorithm are used as initial parameters of the distributions in the first step of the MCMC procedure, so there was no substantial difference between these two methods.

Number of items has some contribution to the missing value imputation, but not as mush as expected. With almost three times increase in the number of items, the improvement of accuracy was trivial. One can argue this is due to items from different domains have low correlations with these two items being imputed. However, for item 21, the correlation with all other 27 items range from 0.138 (with item 3) to 0.456 (with item 19); for item 14, the correlations range from 0.104 (item 3) to 0.454 (item 13). Our best explanation for the lack of difference is because the incompleteness of item 14 and 21 has little relationship with items from other domains. Therefore it is not particularly helpful to include other items of different domains into imputation.

The missing data rates have little impact on the accuracy either. Presumably the accuracy would decrease with higher missing data rates, but it was the case here. The results do not support there is a relationship between missing data rates and accuracy of imputation. Our findings are similar to Enders and Bandalos's (2001) simulation. Both studies showed that

multiple imputations are quite robust to the number of missing cases and the estimations were unaffected. It is possible the highest missing data rates from both studies are 25%, which does not have substantial impact on the imputation. Higher missing data rates should be tested in future studies.

One limitation of this study is the findings can only be generalized to item non-response in cross-sectional data. When unit non-response occurs, data imputation become more complicated, especially in a longitudinal study where quality of life is expected to change over time. An approach often used is available case analysis when dealing with longitudinal quality of life data. The method is to include all the QoL information available at that assessment time point when we are comparing two treatments with respect to QoL at specific time points. The disadvantage is that sample size may vary and different sets of patients are included in the analysis at each assessment time point.

Likelihood-based method is another popularly used approach that has been discussed by several authors for missing data problems (Brown 1983; Little and Rubin 1989; Neale et al. 1999). Specifically, full information maximum likelihood (FIML) has also been used as a method for estimating means and covariance matrices from incomplete data in Structural Equation Modeling (Graham et al. 1996; Graham and Hofer 2000; Rovine 1994; Verleye 1996). SEM utilities FIML method in handling missing data and it has been recognized as a theory based approach. The advantage of likelihood-based methods is the results will not be biased even if the data are not missing completely at random. Another advantage of likelihood estimator is its efficiency and easy implementation in statistical software. The approach has been built in several programs such as Mx (Neale et al. 1999), Amos (Arbucle 1995), and LISREL (Jöreskog and Sörbom 2004). FIML assumes that the data has a multivariate normal distribution and it maximizes the log-likelihood of the theoretical model given the observed data. FIML is also robust to data that do not exactly follow multivariate normal distribution. The disadvantage with likelihood based methods is that they require relatively large sample. Hence, it may be a problem to apply FIML when the data sets are small (Boomsma 1983).

Missing items are always likely to occur, and the issues have to be addressed. Unfortunately, most reports in QoL assessments in clinical trials frequently ignore the problems of bias arising from non-random patterns of missing data. It is useful to perform a sensitivity analysis by examining the potential impact of different levels of bias drawn from the observed data. Inferences from the incomplete data are not as convincing as those based on a complete data set regardless how well the analysis was conducted and how scrupulous assumptions were made. The best way to avoid bias and loss of sample size is to avoid missing data to occur. In most quality of life instrument, the proportion of missing items is usually small except some atypical or sensitive questions such as sexuality. Standard procedures should be constructed to ensure the items and forms are complete. It should be emphasized to respondents that they should complete all questions if possible. Sufficient care and attention should be taken at the design stage of a study to ensure an adequate infrastructure, including appropriate personnel and material to carry out the study.

Acknowledgements This study is based (in part) on data from the National Health Interview Survey original database provided by the Bureau of Health Promotion, Department of Health and National Health Research Institutes. The interpretation and conclusions contained herein do not represent those of Bureau of Health Promotion, Department of Health or National Health Research Institutes. The author thanks the Bureau of Health Promotion, Department of Health and National Health Research Institute in Taiwan for providing the data.

References

Arbucle, J.L.: Amos User's Guide. Smallwaters, Chicago (1995)

Arbuckle, J.L.: Full information estimation in the presence of incomplete data. In: Marcoulides, G.A., Schumacker, R.E. (eds.) Advanced Structural Equation Modeling. Lawrence Erlbaum Publishers, , Mahwah (1996)

Boomsma, A.: On the Robustness of LISREL (Maximum Likelihood Estimation) Against Small Sample Size and Non-normality. Sociometric Research Foundation, Amsterdam (1983)

Brown, C.H.: Asymptotic comparison of missing data procedures for estimating factor loadings. Psychometrika **48**, 269–291 (1983)

Brown, R.L.: Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. Struct. Eq. Model. **1**, 287–316 (1994)

Curran, D., Fayers, P.M., Molenberghs, G., Machin, D.: Analysis of incomplete quality-of-life data in clinical trials. In: Staquet, M. (ed.) Quality of Life Assessment in Clinical Trials: Methods and Practice. Oxford University Press, Oxford (1998a)

Curran, D., Molenberghs, G., Fayers, P.M., Machin, D.: Incomplete quality of life data in randomized trials: missing forms. Stat. Med. **17**, 697–709 (1998b)

Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B Methodol. **39**, 1–38 (1997)

Enders, C.K., Bandalos, D.L.: The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Struct. Eq. Model. **8**(3), 430–457 (2001)

Fayers, P., Machin, D.: Quality of Life. Assessment, Analysis and Interpretation. Wiley, Chichester (2000)

Fayers, P.M., Curran, D., Machin, D.: Incomplete quality of life data in randomized trials: Missing items. Stat. Med. **17**, 679–696 (1998)

Gilks, W., Richardson, S., Spiegelhalter, D.: Markov Chain Monte Carlo in Practice. Chapman and Hall (1995)

Glasser, M.: Linear regression analysis with missing observations among the independent variables. J. Am. Stat. Assoc. **59**, 834–844 (1964)

Graham, J.W., Hofer, S.M.: Multiple imputation in multivariate research In: Little, T.D. Schnabel, K.U., Baumert J. (eds.) Modeling Longitudinal and Multilevel Data: Practical Issues Applied Approaches and Specific Examples. Lawrence Erlbaum Associates, Mahwah (2000)

Graham, J.W., Hofer, S.M., MacKinnon, D.P.: Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. Multivar. Behav. Res. **31**, 197–218 (1996)

Haitovsky, Y.: Missing data in regression analysis. J. R. Stat. Soc. B **30**, 67–81 (1968)

Jöreskog, K.G., Sörbom, D.: PRELIS 2: User's Reference Guide. Scientific Software International, Chicago (1996)

Jöreskog, K., Sörbom, D. LISREL 8.7 for Windows. Scientific Software International, Inc., Lincolnwood (2004)

Kim, J.O., Curry, J.: The treatment of missing data in multivariate analysis. Sociol. Methods Anal. **6**, 215–240 (1977)

Lin, T.H., Chang, H.Y., Weng, W.S., Chen, Y.J., Cho, E.Y., Hsiung, C.A., Liu, J.P.: The National Health Interview Survey Information System: an overview. J. Taiwan Pub. Health **22**(6), 431–440 (2003)

Little, R., Rubin, D.: Statistical Analysis with Missing Data. Wiley, New York (1987)

Little, R., Rubin, D.: The analysis of social science data with missing values. Sociol. Methods Res. **18**, 292–326 (1989)

McLachlan, G.J., Krishnan, T.: The EM Algorithm and Extensions. Wiley, New York (1997)

Muthén, B., Kaplan, D., Hollis, M.: On structural equation modeling with data that are not missing completely at random. Psychometrica **52**, 431–462 (1987)

Neale, M.C., Boker, S.M., Xie, G., Maes, H.H.: Mx: Statistical Modeling (5th ed.). Department of Psychiatry, Richmond (1999)

Olschewski, M., Schulgen, G., Schumacher, M., Altman, D.G.: Quality of life assessment in clinical cancer research. Br. J. Cancer **70**, 1–5 (1994)

Rovine, M.J.: Latent variable models and missing data analysis. In: von Eye, A., Clogg, C.C. (eds.) Latent Variable Analysis: Applications for Developmental Research. Sage Publications Thousand Oaks (1994)

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

Verleye, G. (1996). Missing at random data problems in attitude measurement using maximum likelihood structural equation modeling. Unpublished dissertation. Frije Universiteit Brussels, Department of Psychology

World Health Organization. International Classification of Impairments, Disabilities and Handicaps. WHO, Geneva (1980)

Wothke, W. : Longitudinal and multi-group modeling with missing data. In: Little, T.D., Schnabel, K.U., Baumert, J. (eds.) Modeling Longitudinal and Multiple Group Data: Practical Issues, Applied Approaches and Specific Examples. Lawrence Erlbaum Associates, Inc., Mahwah (2000)