Performance of Standard Imputation Methods for Missing Quality of Life Data as Covariate in Survival Analysis Based on Simulations from the International Breast Cancer Study Group Trials VI and VII

Marion Procter[1] and Chris Robertson[2]

[1]Frontier Science Scotland, Kincraig, Kingussie, UK

[2]Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK

Corresponding author: Marion Procter, Frontier Science (Scotland) Ltd, Grampian View, Kincraig, Kingussie, PH21 1NA, UK

e-mail: marion.procter@frontier-science.co.uk

Running head: Performance of Imputation Methods for QoL Scores

Keywords: quality of life; simulation study; imputation methods; missing data mechanism; time-dependent Cox model

Presented in part at ISCB 30, Prague, Czech Republic, August 2009

Abstract

Imputation methods for missing data on a time-dependent variable within time-dependent Cox models are investigated in a simulation study. Quality of life (QoL) assessments were removed from the complete simulated datasets, which have a positive relationship between QoL and disease-free survival (DFS) and delayed chemotherapy and DFS, by missing at random (MAR) and missing not at random (MNAR) mechanisms. Standard imputation methods were applied before analysis. Method performance was influenced by missing data mechanism, with one exception for simple imputation. The greatest bias occurred under MNAR and large effect sizes. It is important to carefully investigate the missing data mechanism.

1. Introduction

The question of whether good quality of life is associated with good prognosis is of clinical interest in breast cancer clinical trials (e.g. Epplein et al. 2011; Kenne Sarenmalm et al. 2009; Coates et al. 2000). However, traditional endpoints such as DFS and overall survival do not reflect the patient's sense of well-being. Thus, it is becoming increasingly common for quality of life to be assessed throughout the study (Fairclough 2010, p.1).

In practice, quality of life assessments often contain missing observations. The potential problems associated with missing observations, such as missing quality of life assessments, include bias of the parameter estimates and loss of power to detect clinically important differences among treatment groups over time (Fairclough 2010, chapter 6; Little and Rubin 2002, chapter 1 and 3). Methods for dealing with analysis of data with missing observations, such as imputation-based procedures, where the missing values are filled-in and the

completed data are analyzed by standard methods, have been proposed in the statistical literature (e.g. Rubin 1987; Little and Rubin 2002; Molenberghs and Kenwood 2007).

Standard imputation methods were applied to missing quality of life assessments in the IBCSG dataset before analysis in a time-dependent Cox model. There was no evidence of a statistically significant or clinically important relationship between quality of life, as measured by coping score, and DFS (Procter 2016; Procter and Robertson 2017). The parameter estimates for the square root of the coping score (S_Pacis) and delayed chemotherapy following the standard imputation methods were similar to those from the all available analysis. The small increase in the standard error of the parameter estimates compared to the all available analysis was similar following each of the standard multiple imputation methods (Procter and Robertson 2017). It is possible that the performance of the standard imputation methods in the setting of the IBCSG dataset is influenced by the fact that there was no evidence of a relationship between quality of life and DFS. Here, we investigate the performance of the standard imputation methods where such a relationship exists.

Complete simulated datasets were generated with a positive relationship between quality of life and DFS and a positive relationship between delayed chemotherapy and DFS, using the algorithm described by MacKenzie and Abrahamowicz (2002). Here, a high quality of life was associated with improved DFS and delayed chemotherapy was associated with improved DFS. Simulated datasets with missing data were generated by artificially removing coping scores from the complete simulated datasets. There were 5 different methods of artificially removing coping scores considered, representing 4 different scenarios for the missing data mechanism. Standard imputation methods were then applied to missing quality of life assessments in the simulated datasets before analysis in a time-dependent Cox model. An

overview is provided in Figure 1. From the time-dependent Cox model analysis, the performance of the standard imputation methods was compared given different combinations of positive relationship between quality of life and DFS and positive relationship between delayed chemotherapy and DFS. The focus was on any bias in the parameter estimates and the probability of finding the relationship between delayed chemotherapy and DFS.

2. Patients and Methods

2.1. IBCSG Trial VI and VII and Time-Dependent Cox Model Analysis

IBCSG Trial VI was designed to examine different durations and timing of adjuvant chemotherapy in premenopausal and perimenopausal patients. In postmenopausal patients, tamoxifen alone or together with different durations and timing of chemotherapy was compared in IBCSG Trial VII. Between July 1986 and April 1993, 1554 premenopausal and perimenopausal patients were randomized to Trial VI and during the same time period 1266 postmenopausal patients were randomized to Trial VII. The patient's self-assessed quality of life were prospectively collected throughout the study. Baseline quality of life was assessed on, or as close as possible to, the first day of adjuvant therapy. Quality of life was recorded approximately 3 months after randomization, then every 3 months until 24 months, and again at 1 and 6 months after recurrence. Coping/perceived adjustment ("coping score") was measured on a linear analogue scale, ranging from 0 ('no effort at all' to cope with illness) to 100 ('a great deal of effort' to cope with illness).

The IBCSG dataset was the initial reference for creating the simulated datasets, which have a positive relationship between quality of life and DFS and a positive relationship between delayed chemotherapy and DFS. As with the IBCSG dataset (Procter and Robertson 2017), time-dependent Cox model analysis of coping scores in the simulated datasets took place after imputation of missing coping scores. The parameters in the time-dependent Cox model

analysis were again the square root of the coping score (S_Pacis) together with an indicator for delayed chemotherapy.

2.2. Standard Imputation Methods

The standard imputation methods applied to IBCSG dataset were also applied to the simulated datasets. Technical details of all methods are described in Procter and Robertson (2017). The standard simple imputation methods applied were:

i) last observation carried forward (LOCF)

ii) median imputation by patient

iii) linear regression with previous coping score(s)

The standard multiple imputation methods applied were:

i) bootstrapping: subgroups defined by baseline coping score and subgroups defined by previous coping score

ii) nearest neighbor imputation

iii) predictive mean matching

iv) pattern mixture models – Curran's analytical technique (Curran 2000)

2.3. Simulated Datasets

Method for Simulating Data

Complete simulated datasets were generated with a positive relationship between quality of life and DFS and a positive relationship between delayed chemotherapy and DFS. Here, a high quality of life was associated with improved DFS and delayed chemotherapy was associated with improved DFS. These associations were from the parameters β_{sp} and β_{del} of the time-dependent Cox model respectively. Four combinations of β_{sp} and β_{del} were considered. Low coping scores correspond to high quality of life. The complete simulated

datasets considered the 2231 patients from the IBCSG dataset with an observed baseline coping score (approximately randomization). The combinations considered a value for β_{sp} of 0.1 or 0.4 and a value for β_{del} of -0.165 or -0.195. The values for β_{sp} of 0.1 and 0.4 lead to a relationship between quality of life and DFS being found in all the complete simulated datasets (Table S1; online appendix only). The values for β_{del} of -0.165 and -0.195 approximately correspond to conventional values for the power of the hypothesis tests set in clinical trials (0.8 and 0.9) (Table S1). For each of the 4 combinations, 150 simulated datasets were generated. The time-dependent Cox model analysis of the complete simulated datasets is shown in Table S1.

In the IBCSG dataset, the median follow-up time is 12.3 years. The DFS survival times (time to DFS event and follow-up time for patients with no DFS event) approximately followed Weibull distributions. The algorithm described by MacKenzie and Abrahamowicz (2002) was used to simulate a positive relationship between quality of life and DFS and a positive relationship between delayed chemotherapy and DFS (see Table S2, online appendix only).

Hazard Function from the Weibull Distribution and Creating a Time-Dependent Process

The MacKenzie and Abrahamowicz algorithm for randomly generating time-to-event data arises from an interpretation of the expression for the partial likelihood. The method for simulating a positive relationship between quality of life and DFS and a positive relationship between delayed chemotherapy and DFS according to this algorithm is summarised as follows. A matrix of coping scores based on the patients' coping scores in the IBCSG dataset is considered. Simulated DFS times (event or censored) are simulated from Weibull distributions. The simulated DFS times are considered in ascending order and matched to a

patient. The risk set of patients who have yet to be matched to a DFS time is identified. The probability of selection is calculated for each patient in the risk set of patients, and the patient to match the DFS time to is selected. For times to DFS event, this selection probability is based on the covariates i) the centred S_Pacis and ii) indicator for delayed chemotherapy. To create a time- dependent process, the centred S_Pacis at the appropriate time period is used when calculating the selection probability, as shown in in Table S3 (online appendix only). For censored DFS times, the selection probability is equal for all patients in the risk set. The patient matched is removed from the risk set and the steps repeated until all patients have been matched to a DFS time.

Artificially Removing Data

There are three major categories of missing data: i) missing completely at random (MCAR), ii) missing at random (MAR) and iii) informative missing data (Rubin 1976; Little and Rubin 1987, Chapter 1; Little and Rubin 2002, Chapter 1). For each of the complete simulated datasets 5 different methods of artificially removing coping scores were considered. These 5 methods represent 4 different scenarios and each of the 3 different categories for the missing data mechanism in the IBCSG dataset. Details are shown in Table S4 (online appendix only):

i) Higher coping scores (lower quality of life) have a higher probability of being missing: **informative missing data** (Method 1 and 5 Table S4)

ii) Lower coping scores (higher quality of life) have a higher probability of being missing: **informative missing data** (Method 2 Table S4)

iii) Later time periods have a higher probability of being missing: **MAR** (Method 3 Table S4)

iv) Coping scores missing at random (approximately 30%): **MCAR** (Method 4 Table S4)

The most likely scenario in the IBCSG dataset was that higher coping scores (poorer quality of life) have a higher probability of being missing. Two methods of artificially removing data under this scenario were considered. Each of the methods were derived in order that in the simulated datasets approximately 30% of the expected coping score were missing, similar to the IBCSG dataset.

2.4. Technical Details of Patients Considered in Time-Dependent Cox Model Analysis of Simulated Datasets

The status of coping scores for time-dependent Cox model analysis from the 600 simulated datasets, 150 simulated datasets in each of the 4 combinations of β_{sp} and β_{del}, with coping scores artificially removed according to a particular method is described in Table 1. As with the IBCSG dataset (Procter and Robertson 2017), the assumption of proportional hazards did not raise concerns about the time-dependent Cox model (partly shown in Figure S1, online appendix only).

When performing the imputation by LOCF and by linear regression using previous coping scores, the patients with the baseline coping score artificially removed could not be considered in the time-dependent Cox model analysis. In the case of Method 5, for example, these imputation methods were performed on 1498 patients rather than 2231.Patients with no observed coping scores could not be considered in the time-dependent Cox model analysis when performing median imputation by patient.

3. Results

In comparing the performance of the standard imputation methods, the focus was on comparisons of the bias in the parameter estimate of the time-dependent variable and differences in the probability of finding the relationship between delayed chemotherapy and DFS associated with the missing value mechanism and type of imputation method.

Simple Imputation Methods

The summary of findings from simple imputation is shown in Figure 2, with further details for LOCF in Table S5 (online appendix only).The performance of the standard simple imputation methods was better when considering the combination of weak positive relationship between quality of life and DFS and weak positive relationship between delayed chemotherapy and DFS than in the other combinations of β_{sp} and β_{del} (Figure 2; bias in parameter estimate for delayed chemotherapy partly shown in Table S5.2). The parameter estimate for S_Pacis was robust when considering the weak relationship between quality of life and DFS. It was also robust following linear regression using previous coping scores when considering the strong relationship between quality of life and DFS (Figure 2). The trend was for the parameter estimate to be biased towards 0 following LOCF or median imputation by patient when considering the strong relationship between quality of life and DFS. The bias was most extreme when later time periods were associated with a higher probability of being missing (Figure 2).The time-dependent Cox model analysis of the complete simulated datasets indicates a lack of precision in the estimates of β_{del} and led to a wide range of parameter estimates from each of the completed simulated datasets. The imprecision is reflected in the fact the mean standard error of the parameter estimate was at least 0.055 (partly shown in Table S5.2). As with the parameter estimate for S_Pacis, the

parameter estimate for delayed chemotherapy was robust when considering the combination of weak positive relationship between quality of life and DFS and weak positive relationship between delayed chemotherapy and DFS. For the remaining combinations of β_{sp} and β_{del}, the trend was for the parameter estimate for delayed chemotherapy to be biased towards 0 (partly shown in Table S5.2).

A relationship between quality of life and DFS was found in all the completed simulated datasets (partly shown in Table S5.1, column "Number of 95% CIs for hazard ratio containing 1"). The probability of finding significance of association between delayed chemotherapy and DFS following simple imputation is shown in Figure 3, with further details for LOCF in Table S5. Patients who had a missing baseline coping score could not be considered in the time dependent Cox model analysis following LOCF and linear regression using previous coping scores. This lead to i) a larger mean standard error in the parameter estimates [details not shown] and ii) lower probability of finding the probability of the relationship between delayed chemotherapy and DFS compared to median imputation by patient (Figure 3). Here, the probability of finding the relationship between delayed chemotherapy and DFS was lowest when higher coping scores were associated with a higher probability of being missing. This probability was highest when later time periods were associated with a higher probability of being missing when considering the strong relationship between delayed chemotherapy and DFS (Figure 3).

The performance of the standard simple imputation methods was influenced by the missing data mechanism except when the weak positive relationship between quality of life and DFS and weak positive relationship between delayed chemotherapy and DFS was considered. However, there was no suggestion that the performance of the standard simple imputation

methods was noticeably better when coping scores were missing at random compared to other missing data mechanisms (Figure 2).

Multiple Imputation Methods

The summary of findings from multiple imputation is shown in Figure 4 and the probability of finding significance of association between delayed chemotherapy and DFS is shown in Figure 5. Further details for bootstrapping, subgroups defined by baseline coping score are shown in Table S6 (online appendix only). A relationship between quality of life and DFS was found in all the complete simulated datasets. However, the fact that the standard multiple imputation methods led to i) a biased parameter estimate of S_Pacis closer to 0 than the theoretical parameter value (Figure 4), especially for a strong relationship between coping score and disease free survival, and ii) a reduction in the probability of finding the relationship between delayed chemotherapy and DFS (Figure 5) in almost all cases indicates that they did not perform well. The bias of the parameter estimate of S_Pacis was most extreme when later time periods were associated with a higher probability of being missing (Figure 4). This most extreme bias was higher following bootstrapping than following other multiple imputation methods, around 30% to 40% compared to around 17% to 27%. Similarly to standard simple imputation methods, the time-dependent Cox model analysis of the completed simulated datasets indicates a lack of precision in the estimates of β_{del} and led to a wide range of mean parameter estimates based on repetitions of multiple imputation for each of the simulated datasets in each scenario. The imprecision is reflected in the fact that the standard error of the parameter estimate was around 0.065 and around 0.068 when considering the weak positive relationship and the strong positive relationship between quality of life and DFS respectively (partly shown in Table S6.2).

The largest influence of the performance of the standard multiple imputation methods was the combination of the positive relationship between quality of life and DFS and positive relationship between delayed chemotherapy and DFS. The relative bias in the parameter estimates of S_Pacis was lower when there was a weak relationship between quality of life and DFS than when there was a strong relationship between quality of life and DFS (Figure 4). The probability of finding a relationship between delayed chemotherapy and DFS was lowest when considering the combination of strong relationship between quality of life and DFS and weak relationship between delayed chemotherapy and DFS (Figure 5).

Further, the bias in the parameter estimate of S_Pacis was largest when later time periods were associated with a higher probability of being missing (Figure 4). The parameter estimate of delayed chemotherapy could only be considered robust, in one combination, when higher coping scores (lower quality of life) had a higher probability of being missing according to method 1. The applicable combination was the combination of the weak relationship between DFS and weak relationship between delayed chemotherapy and DFS (partly shown in Table S6.2). In the remaining combinations, the bias in parameter estimate of delayed chemotherapy was lowest when coping scores were missing at random (partly shown in Table S6.2).

The standard errors of the parameter estimates of S_Pacis and delayed chemotherapy following standard multiple imputation methods (partly shown in Table S5.1 and Table S5.2 respectively; online appendix only) were larger than from the complete simulated datasets (Table S1), reflecting the uncertainty in the imputed values. However, bias in the parameter estimate for S_Pacis was more apparent than the standard simple imputation methods than the standard simple imputation methods. Considering the bias in the parameter estimates,

there was no indication that the standard multiple imputation methods were more useful than the standard simple imputation methods in the context of the simulation study.

4. Summary

There was a suggestion that the performance of the standard simple imputation methods was influenced by the missing data mechanism except when the combination of weak positive relationship between quality of life and DFS and weak positive relationship between delayed chemotherapy and DFS was considered. The performance of the standard simple imputation methods was poorest when higher coping scores (lower quality of life) were associated with a higher probability of being missing.

The performance of the standard multiple imputation methods was influenced by the missing data mechanism. The standard multiple imputation methods did not perform well in the context of the simulation study. While none of the standard imputation methods could be recommended for the scenarios considered in the simulation study, there were differences in the bias seen in the parameter estimate of S_Pacis. The bias in the parameter estimate for S_Pacis was largest when following bootstrapping, subgroups defined by baseline coping score. The probability of finding a relationship between delayed chemotherapy and DFS was general lowest following bootstrapping, subgroups defined by baseline coping score. The fact that the standard multiple imputation methods did not perform well and the suggestion that the performance of the standard imputation methods was influenced by the missing data mechanism illustrates the importance of carefully investigating the missing data mechanism when performing imputation techniques.

In the simulations approximately 30% of the data on the time-dependent covariate were assumed to be missing. This is on the large side and is representative of the underlying IBCSG trial. In other situations we might expect less missing data and if the percentage missing is as low as 5%-10% then we would expect that all methods would be more applicable and the bias would be less. In some scenarios, 10%-20% of the missing data will have little or no effect on the results of the study. The reason for the missing data needs to be considered as well as the amount of missing data (Fairclough 2010, p.126-127; Little and Rubin 2002, p.41-42). Special care should be taken if imputation is being applied when more than 30-50% of the data are missing (White, Royston, and Wood 2011).

The work in this manuscript focused on standard imputation methods. It could be extended by applying further multiple imputation methods. In particular, chained equations, implemented in the statistical software MICE, is becoming more common. Of note, the statistical literature on correctly specifying the imputation model continues to be developed (e.g. White, Royston, and Wood 2011). Though chained equations were not considered, they are related to multiple imputation methods that were applied: neighbour imputation and predictive mean matching.

Implications of Findings from Applying Simple Imputation Methods

- Simple imputation methods have limitations; the main limitation is the underestimation of the variance of the parameter estimate
- There are only limited circumstances when it is appropriate to draw inferences from the parameter estimate resulting from simple imputation; if the parameter estimates are considered, then justification should be provided
- The simple imputation methods may provide information as part of a sensitivity analysis into the sensitivity of results to the assumptions about the missing data

- The influence of the missing data mechanism on the performance the standard simple imputation methods in the simulation study illustrates the importance of carefully investigating the missing data mechanism

Implications of Findings from Applying Multiple Imputation Methods

- Multiple imputation methods generally assume the data are MAR; pattern mixture models were developed to analyze informative missing data
- A monotone missing data pattern was required to implement pattern mixture models – Curran's analytical technique and the remaining standard multiple imputation techniques apart from bootstrapping
- In the context of the simulation study, the standard multiple imputation methods did not perform well; this is influenced by the fact
  i) for two of the scenarios the simulated datasets have informative missing data
  ii) all the simulated datasets have a general missing data pattern
- The bias of the parameter estimate for S_Pacis was most extreme when later time periods were associated with a higher probability of missingness
- The influence of the missing data mechanism on the performance of the standard multiple imputation methods in the simulation study again illustrates the importance of carefully investigating the missing data mechanism

When appropriately performed, imputation allows valid inferences from standard procedures. Development of imputation-based procedures continues, while recognising that imputation-based procedures are not always the best approach to analyzing missing data. When appropriately performed, imputation allows valid inferences from standard procedures. However, it is important to investigate why observations are missing and to give careful consideration to the final choice of imputation method used as imputation methods involve untestable assumptions. While statistical methods for dealing with missing data exist, it is always preferable to have the actual data and it is important to minimise the amount of missing data in a clinical trial.

Acknowledgements

The authors thank Jürg Bernhard, Karen Price, Richard Gelber, Meredith Regan and the IBCSG for providing us with the breast cancer data.

Appendices containing supplementary summary tables

Full details of the results of the simulation study can be found in Appendix E and Appendix F of Procter M.J. (2016). *Influence of missing explanatory variables and longitudinal assessments in breast cancer clinical trials*. University of Strathclyde. Dept. of Mathematics and Statistics PhD thesis and available from the University of Strathclyde library website https://www.strath.ac.uk/library

References

Coates, A.S., Hürny, C., Peterson, H.F., Bernhard, J., Castiglione-Gertsch, M., Gelber, R.D., Goldhirsch, A. for the International Breast Cancer Study Group (2000). Quality-of-life scores predict outcome in metastatic but not early breast cancer. *Journal of Clinical Oncology* 18: 3768-3774. doi: 10.1200/JCO.2000.18.22.3768.

Curran, D (2000). Analysis of Incomplete Longitudinal Data. PhD diss., Linsburg Universitair Centrum, as cited in Fairclough, D.L. (2010). *Design and Analysis of Quality of Life Studies in Clinical Trials 2nd edition*. Chapman and Hall: Boca Raton.

Epplein, M., Zheng, Y., Zheng, W., Chen, Z., Gu, K., Penson, D., Lu, W. and Shu, X.-O. (2011). Quality of life after breast cancer diagnosis and survival. *Journal of Clinical Oncology* 29: 406-412. doi: 10.1200/JCO.2010.30.6951.

Fairclough, D.L. (2010). *Design and Analysis of Quality of Life Studies in Clinical Trials, 2nd edition.* Boca Raton: Chapman and Hall.

Kenne Sarenmalm, E., Odén, B., Öhlén, J., Gaston-Johansson F. and Holmberg S.B. (2009). Changes in health-related quality of life may predict recurrent breast cancer. *European Journal of Oncology Nursing* 13: 323-329. doi: 10.1016/j.ejon.2009.05.002.

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

Little, R.J.A. and Rubin, D.B. (2002). Statistical Analysis With Missing Data, 2nd edition. Hoboken: John Wiley & Sons Inc.

MacKenzie, T. and Abrahamowicz, M. (2002). Marginal and hazard ratio specific random data generation: Applications to semi-parametric bootstrapping. *Statistics and Computing* 12: 245-252. doi: 10.1023/A:1020750810409.

Molenberghs, G. and Kenward, M.G. (2007). *Missing Data in Clinical Studies*. Chichester: John Wiley & Sons.

Procter, M.J. 2016. Influence of missing explanatory variables and longitudinal assessments in breast cancer clinical trials. PhD diss., University of Strathclyde.

Procter, M. and Robertson, C. (2017). Imputing Missing Quality of Life Data as Covariate in Survival Analysis of the International Breast Cancer Study Group Trials VI and VII. Communication of Statistics - Simulation and Computation. Advance online publication. doi: 10.1080/03610918.2017.1390123.

Rubin, D.B. (1976). Inference and missing data, *Biometrika*, 63, 581-592.

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

White, I. R., Royston, P., and Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine*, 30, 377-399. doi:10.1002/sim.4067