**Missing data in longitudinal studies: Comparison of multiple imputation methods in a real clinical setting**

Journal of
Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

# Missing data in longitudinal studies: comparison of multiple imputation models in a real clinical setting

| | |
|---|---|
| Journal: | *Journal of Evaluation in Clinical Practice* |
| Manuscript ID | Draft |
| Wiley - Manuscript type: | Original Papers |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | Rosato, Rosalba; University of Turin<br>Pagano, Eva; "Città della Salute e della Scienza Hospital", Unit of Cancer Epidemiology and CPO Piemonte<br>Testa, Silvia; University of Turin, Department of Psychology<br>Zola, Paolo; University of Turin<br>Di Cuonzo, Daniela; University of Turin |
| Keywords: | medical research |
| Abstract: | Rationale, aims and objectives. Missing data represent a challenge in longitudinal studies. The aim of the study is to compare the performance of the multivariate normal imputation and the fully conditional specification methods, using real dataset with missing data partially completed two years later.<br>Method. The data used came from an ongoing randomized controlled trial with five-year follow-up. At a certain time, we observed a number of patients with missing data and a number of patients whose data were unobserved because they were not yet eligible for a given follow-up. Both unobserved and missing data were imputed. The imputed unobserved data were compared with the corresponding real information obtained two years later.<br>Results. Both imputation methods showed similar performance on the accuracy measures and produced minimally biased estimates.<br>Conclusion. Despite the large number of repeated measures with intermittent missing data and the non-normal multivariate distribution of data, both methods performed well and was not possible to determine which was better. |
| | |

# Missing data in longitudinal studies: comparison of multiple imputation models in a real clinical setting

Rosalba Rosato[1,2*], Eva Pagano[2], Silvia Testa[3], Paolo Zola[4], Daniela Di Cuonzo[1,2]

**Running title: Missing imputation in longitudinal studies**

[1] Department of Psychology, University of Turin, Italy

[2] Unit of Cancer Epidemiology, "Città della Salute e della Scienza" Hospital-University of Turin and CPO Piemonte, Turin, Italy.

[3]

[4] Department Surgical Sciences, University of Turin, Italy

***Corresponding author:**

Rosalba Rosato

Department of Psychology

University of Turin,

Via Verdi, 10 10126 Turin

Tel. +39 011670.2923

Fax: +39 011670.2795

Email: rosalba.rosato@unito.it

**Abstract**

**Rationale, aims and objectives**. Missing data represent a challenge in longitudinal studies. The aim of the study is to compare the performance of the multivariate normal imputation and the fully conditional specification methods, using real dataset with missing data partially completed two years later.

**Method.** The data used came from an ongoing randomized controlled trial with five-year follow-up. At a certain time, we observed a number of patients with missing data and a number of patients whose data were unobserved because they were not yet eligible for a given follow-up. Both unobserved and missing data were imputed. The imputed unobserved data were compared with the corresponding real information obtained two years later.

**Results.** Both imputation methods showed similar performance on the accuracy measures and produced minimally biased estimates.

**Conclusion.** Despite the large number of repeated measures with intermittent missing data and the non-normal multivariate distribution of data, both methods performed well and was not possible to determine which was better.

**Keywords:** Missing data, fully conditional specification, multivariate normal imputation, quality of life.

## Introduction

Missing data is a common problem in longitudinal studies, in which it is highly probable that some respondents' information will be missed at some pre-specified times. It has long been recognized that missing data may affect the validity of results and the power of studies. Moreover, in presence of a high percentage of missing values the statistical power of a study can be reduced and may cause selection biases if observations with missing data are excluded from the analysis [1, 2]. The risk of biased results is related to the missing mechanism, i.e. the reason why the data are missing. The missing mechanisms are commonly classified as missing completely at random (MCAR) if the probability that the variable is missing is unrelated to itself or other observed variables; missing at random (MAR) if other variables (but not the variable itself) can be used to predict missingness on a given variable; or missing not at random (MNAR) if it is neither MCAR nor MAR. Multiple imputation (MI) is a statistical method widely used for handling missing data in which the observed data distribution is used to generate a set of plausible values for the missing data [3].

Although MI can be implemented with data not missing at random (MNAR) [2], standard implementations assume that data were missing at random (MAR). MI will give an efficient inference if data are MAR and the imputation model includes auxiliary variables that both predict the incomplete variable and predict whether the incomplete variable is missing [4-6].

In longitudinal studies, if we assume that respondents missing data are related to their own previous values, all available measurements can be used to impute missing values [7]. Most MI implementation methods are designed without taking into account the temporal order in which the data were collected, but ignoring longitudinal nature of the observed data could produce biased parameter estimates [8].

In the presence of repeated measurement data, the multivariate normal imputation (MVNI) method has been proposed [3, 9]. MVNI requires the specification of a comprehensive model for the joint distribution of all the variables and assumes the multivariate normality; moreover, imputed values are not usually used as predictors to impute other variables [9]. As observed by Demirtas "researchers rarely witness that the MVN assumption holds with the real data" [10]. Never-the-less, in presence

of normality departure, both in terms of asymmetry or kurtosis, a simulation study has shown that, MVNI yields unbiased estimates for the mean value, whereas the procedure seems suffering in terms of variance estimates in presence of lower sample size and high proportion of missing values [10].

More recently the fully conditional specification (FCS) method (also named sequential generalized regression, multiple imputation by chained equations) has been proposed [7, 8, 11]. The FCS method is based on a set of imputation models, one for each variable with missing values; in the presence of repeated measures, values previously imputed can be used to impute the next variable values [12]. The FCS approach is very flexible in creating multivariate models without assuming multivariate normality of variables. The FCS approach allows to take advantage of the temporal ordering of the repeated measurements and to avoid overfitting and collinearity problems [7].

Among the studies that have investigated MI in longitudinal data, few compared the results of multiple imputation using the MVNI and FCS approaches [13-16]. The accuracy of imputation methods has mainly been evaluated using simulated data or real data with simulated missing patterns. In these approaches the true value of the missing datum is known, but the main issue is that missing pattern are simulated. In the present paper we apply the MVNI and FCS methods using real data, a real missingness pattern and a known true value. Researchers rarely witness that the MVN assumption holds with real data. Because it is impossible to formulate analytical tools and computational routines for every situation, practitioners generally rely on the MVN assumption when imputing continuous data. Hence, an assessment of MVNI and FCS MI methods performance is warranted via real data.

The main aim is to compare the performance of the MVNI and the FCS multiple imputation methods on self-rated Quality of Life (QoL) and Psychological Well Being (PWB) measures, in longitudinal real data, with different degrees of violation of the multivariate normality assumption. Using an ongoing cohort of endometrial cancer patients, missing data and those not yet observed due to short-term follow up are imputed with both models. Subsequently they are compared with the real collected data once completely available.

# Material and Methods

### Setting and sample

The real data used in this paper were collected in a randomized controlled trial on women with a surgically treated endometrial cancer, the "TOTEM Study". The trial compares two follow up regimens with different tests intensity: an intensive follow-up versus a minimal one. The trial was retrospectively registered at www.clinicaltrials.gov (June 8, 2009), identifier: NCT00916708.

Recruitment started in 2008 and is ongoing. Participants are followed for a period of five years from the end of primary treatment (surgery or adjuvant therapy). Clinical and QoL evaluations are planned at baseline, 6, 12, 24, 36, 48, and 60 months. Data about follow-up regimens are not available in this study. Age and severity of endometrial carcinoma (grading) were used in the imputation process as auxiliary variables, as they may be correlated to QoL scores [17].

All procedures were in agreement with the principles of the Helsinki Declaration and its later amendments. The study protocol was approved by the local ethics committee and all participants have provided their written informed consent to participate in the study.

### Study design

All participants recruited between September 2008 (first patient enrolled) and May 2014 were included in the present study (N=812). We have access to two versions of the data, one that includes data through May 2014, and one that includes data through May 2016.

Patients who dropped out were imputed as missing data. Five subjects died during the follow-up and were excluded from the analyses as they were assumed to be in different health conditions from those who dropped out for other reasons and, from this point of view, the patients' death during follow-up is not a missing data problem.

Two kinds of missingness were considered: missing data from respondents who failed to completely fill in the questionnaires and unobserved data from patients not yet eligible for a given follow-up

time. We are interested in the observations for which data are unobserved as of 2014, and observed as of 2016. The imputed values of the unobserved data were compared to the corresponding true values actually observed two years later. For example, consider a participant who enters the study in May 2012. In 2014, we have observed the baseline data and the 6, 12, and 24 month follow-up, but we have not yet observed the 36, 48, and 60 month follow-up. In 2016, we have now observed the 36 and 48 month follow-up, but not yet the 60 month follow-up. This design imputes all missing values in the 2014 data, then compares the imputed values for the 36 and 48 month follow-up to the observed values in the 2016 data.

**Measures**

QoL by way of the Short Form 12-item survey (SF12) [18, 19] and psychological wellbeing by way of a short version of the Psychological General Well-Being (PGWBS) [20] were evaluated at each time point.

The 12 items of SF12 are weighted and summed to provide a Physical and a Mental Health Composite Score (PCS and MCS) which are rescaled in the range 0-100, with higher scores indicating a better health condition. Respondents who did not completed anyone of the 12 items were classified as missing in PCS and MCS [19]. We did not impute single missing items. The PGWBS provides a general evaluation of self-perceived psychological health and well-being covering six domains: anxiety, depressed mood, positive well-being, self-control, general health and vitality. The PGWBS used in the "TOTEM Study" consists of the six items [20] plus another item for each domain. Following Grossi et al. (2006), the composite score was computed if at least one item was filled and single missing items were not imputed.

Age and endometrial cancer grading FIGO (International Federation of Gynecology and Obstetrics) at baseline were also used as auxiliary variables [5] in the imputation analyses. FIGO grading was preferred to the cancer TNM (tumor, nodes, metastasis) stage because it was more associated with the QoL measurements in our data. FIGO grading ranging from 1 (confined to the organ of origin) through 4 (distant metastasis). As a rule, the lower the number, the less the cancer has spread.

In this paper, we focus on the first five measurements (from baseline to 36 months) because of the low number of patients for whom a complete set of follow-up data was available at the time of the analyses.

**Statistical methods - imputation methods**

Multivariate normal distribution for each composite score was tested using Mardia's skewness and kurtosis tests [21] and the Henze-Zirkler test [22].

The MVNI and FCS imputation methods were used. The MVNI uses a multivariate normal regression to impute missing value: $y_i = \Theta_i' z_i + \varepsilon_i$ where $y_1, y_2, \ldots, y_N$ are a random sample from a p-variate normal distribution recording values of p imputation variables (PCS, MCS and PWG in our data); $\Theta$ is a matrix of regression coefficients, $z_i$ is a vector of independent variables from observation *i* (in our data: age, grading, PCS, MCS and PGWBS) and $\varepsilon_i$ is a vector of random error from a p-variate normal distribution.

Consider the partition $y_i = (y_{i(m)}, y_{i(o)})$ corresponding respectively to missing and observed values of imputation variables in observation i for i = 1, . . . , N. The process consists of two steps, an imputation step (I) and a posterior step (P), performed in several iterations. At each iteration of the I step, missing values in $y_i$ are replaced with draws from the conditional posterior distribution of $y_{i(m)}$ given observed data and current values of model parameters independently for each i = 1, . . . , N. We used a Bayesian approach with a Markov Chain Monte Carlo algorithm to obtain imputed values from a multivariate normal distribution. From the posterior predictive distribution of the missing data, the MNVI, during the P step, provides a new random draw of model parameters, given the observed data and the data imputed in the previous I step. The posterior distribution was assumed to be normal. Based on the studies by Schafer and Yucel, inference by MI is quite robust to departures from model assumptions used to derive the posterior predictive distribution of missing data [9, 23].

The FCS approach does not require the multivariate normality assumption of data. A separate regression model for each variable with missing data is specified. Imputation arising by estimating

each conditional distribution in turn uses both observed cases and the values imputed in the previous turns. For each time point, a linear model was set for the single QoL composite scores (i.e. PCS, MCS and PGWBS). The predictive mean matching (PMM) algorithm was used, as it is indicated in case of non-normally distributed continuous variables. The PMM imputes missing value of y with the property that values are sampled only from the observed values of y. More specifically, the following steps are involved: 1) a linear regression of $y_{i(o)}$ on $z_{i(o)}$ is estimated, producing a set of coefficients θ; 2) a new set of coefficients θ* are obtained, making a random draw from the posterior predictive distribution of θ; 3) using θ* and $z_i$, predicted values ($\hat{y}$) are generated for all cases; 4) for each case with missing y, a set of cases with observed y – whose $\hat{y}$ are close to the $\hat{y}$ for the case with missing data – are identified and from among these nearest cases, one is randomly chosen and its observed y value is assigned as the imputed value. This process is repeated for all variables with missing values in turn and, in order to stabilize the results, the procedure is repeated for several cycles. We included age and grading as auxiliary variables in both imputation methods in order to have comparable models with respect to predictors. In addition to auxiliary variables, in FCS imputation, QoL scores at time *t* were assumed to depend on the corresponding QoL composite scores (i.e. PCS, MCS and PGWBS) at the previous (subscribe *1,2, ... t-1*) and next times (subscribe *t+1, t+2… T, whit T as last time observation*) [24, 25]. In the imputation methods we included the previous and next times Qol scores assuming the existence of an autoregressive correlation structure among times.

The model for the PGWBS included the PCS and the MCS, whereas the model for the PCS included the PGWBS but not the MCS because it comes from the same questionnaire (the SF12) and the missing data were the same. For the same reason, the model for the MCS did not include the PCS. In other words, for example, we assumed that the PCS score depends on the previous and following PCS scores, the previous and present PGWBS scores, the baseline characteristics age and grading.

The equations used for the imputation were:

$$PCS_{it} = \sum_{j=1}^{t-1} PCS_{ij} + \sum_{j=t+1}^{T} PCS_{ij} + \sum_{j=1}^{T} PGWBS_{ij} + age_i + grading_i + random\ variation_i$$

$$MCS_{it} = \sum_{j=1}^{t-1} MCS_{ij} + \sum_{j=t+1}^{T} MCS_{ij} + \sum_{j=1}^{T} PGWBS_{ij} + age_i + grading_i + random\ variation_i$$

$$PGWBS_{it} = \sum_{j=1}^{t-1} PGWBS_{ij} + \sum_{j=t+1}^{T} PGWBS_{ij} + \sum_{j=1}^{T} PCS_{ij} + \sum_{j=1}^{T} MCS_{ij} + age_i + grading_i + random\ variation_i$$

It should be noted that *t-1* and *t+1* do not exist, respectively, for the first and the last compilation of data collection. The random variation is assumed to have a normal distribution.

Imputed values were used to predict the missing data of other variables. Considering the proportion of missing cases in the TOTEM study (0.59), in a conservative approach 75 imputation were performed, according to the recommendation of creating more datasets than the percentage of missing data, to have more stable parameter estimates, and better standard error estimates [4].

**Comparison of methods**

In order to assess the quality of the process of imputation we used the questionnaires completed by the patients themselves over the following two years. The true values used in the accuracy evaluation were the updated QoL scores.

In the imputation phase both real missing and unobserved data were treated assuming that there were no significant differences between the two types of missing. In order to test this assumption, the t-test and chi-square were used to compare age, and grading distribution before the imputation.

The performance of the two methods of imputation was evaluated by using the root mean square deviation (RMSD), the mean absolute deviation (MAD), and the BIAS, calculated on the differences between the imputed data and the real data available in the updated database after 24 months. The indexes were calculated as follows [26]:

1) $RMSD = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \hat{x}_i)^2}{m}}$

2) $MAD = \frac{\sum_{i=1}^{m} | x_i - \hat{x}_i |}{m}$

3) $BIAS = \frac{\sum_{i=1}^{m}(x_i - \hat{x}_i)}{m}$

where $\hat{x}_i$ (i=1,…, m) is the mean of the imputed values, $x_i$ is the true value updated 24 months later; m is the number of missing values for which the real value was available 2 years later.

An additional comparison between the two techniques was made comparing, at each time, the mean QoL scores available at the date of 31 May 2014 with the mean of the imputed datasets obtained with the two models.

Imputations were performed using the MICE algorithm in Stata 13 [27], all other computations and analyses were performed in SAS 9.4.

## Results

The cohort includes 812 cases, five patients died during the first 36 months of follow-up. Their average age is 62.94 years (SD=10.41). Almost 72% of them are in cancer TNM stage 1 (i.e. localized cancer that has not spread to nearby lymph nodes or to distant sites). FIGO grading is equal to 1 (confined to the organ of origin) for 41.4% of patients, 2 (invasion of surrounding organs or tissue) for 41% and 3 (spread to distant nodes or tissue) for the remaining ones.

Table 1 shows the description of data collection according to the first five waves. The number of respondents, missing data, number of subjects who have dropped out or died are presented. The patients' status of non-respondent (missing) or not yet available in the current round (unobserved data) could change over time; for example one could have been classified as non-respondent at the beginning of this study (May 31, 2014) and then could have become respondent two years later (in

the update of May 31, 2016). For each QoL measure the increase in the number of completed questionnaires after 24 months was reported.

Eighty-nine different patterns of missingness, i.e. combinations of observed, missing and not yet available data during time of measurement were observed in the data. A summary of ten of the most representative response patterns for both QoL indexes is provided in Appendix 1. The first line identifies subjects who had completed all the questionnaires (4.26% for the SF12 and 4.99% for the PSWBS), the last two lines are examples of non-monotone missing data patterns while in the rest of the table monotone missing patterns are presented. Overall, the presence of non-monotone missing observations is 35.09% for FS-12 and 34.84% for PGWBS.

Before imputation, the multi-normality tests revealed a skewness and non-normality of the distribution for the SF12 physical and mental component scores and for the PGWBS summary score ($p<0.05$ for Mardia's skewness and kurtosis test and the Henze-Zirkler test) (Table 2). In Appendix 2 the univariate distribution of the raw scores was plot by time points for each QoL measures.

There is not a statistical significant difference between cases with unobserved data and missing data, in fact age and grading distribution of the real missing data were similar to those of the unobserved data at all times of measurement in both the SF12 and PSWBS scores (Appendix 3).

When we compare the imputed values of the unobserved data to the corresponding true values actually observed two years later, according to the accuracy measures (i.e. RMSD, MAD, BIAS), for each QoL score and time of measurement, the MNVI and FCS techniques show similar performance (Table 3). Results at 6-months should be interpreted carefully due to the small sample size.

For each index and QoL score, we calculate the mean value and 95% confidence interval from the four times of measurement. The results showed the presence of a relatively high variability among time for the statistics used to evaluate the performance of the imputation method, despite this there were no significant differences between the two imputation methods (Appendix 4).

The last comparison between MVNI and FCS imputation methods refers to the mean QoL scores of the observed values and the mean of the imputed data at each time of measurement. The differences

between the two models were negligible, with the main deviations being observed in the later follow-up period, although these were less than one (Table 4).

## Discussion

MI is a powerful and feasible approach for handling missing data in longitudinal studies.

However, choosing an imputation algorithm that is appropriate for the particular data structure at hand is not straightforward. The aim of our study was to compare the performance of the two techniques on real data, which do not conform to multivariate normality and show varying distributional characteristics such as various degrees of skewness. Although our study presented a large number of repeated measures with intermittent missing data and heterogeneous patterns of missingness, the two methods appeared to produce minimally biased estimates: the difference between the scores calculated using imputed values and those calculated on real values never exceeded 1 percent.

Most previous studies, using multiple imputation for missing data, described the results of simulation studies to investigate the performance of different imputation approaches; the conclusions do not agree on the best method [8, 10, 13-16, 28].

In a previous comparison between the two techniques, Van Buuren [16] showed that a conditional approach was more reliable when no realistic joint distribution of observed values can be specified. Moreover, it was suggested to use the MVNI to impute longitudinal data with an unstructured correlation structure, while FCS is more appropriated with an autoregressive correlation structure [14].

In the literature, several paper investigated how departure from normality assumption affects the MI estimates. Demirtas and colleagues have found that biases and coverage rates (i.e. the percentage of replications in which the confidence interval covers the true parameter value) can be conditioned by lower sample size and high proportion of missing values [10]. In accordance with Lee and Carlin

(2010) and De Silva and colleagues (2017), we found no evidence that the MNVI method performed less well, despite the unrealistic multivariate normal distribution assumption.

To the best of our knowledge, this study is the first to use "real data" to validate and compare two techniques of imputation on longitudinal records with a large number of waves and a high frequency of missing data.

In an ongoing study with several time point measurements, not all patients have yet completed the follow-up study and, consequently, missing data due to this reason are added to those caused by non-participation. The assessment of which method "performed well" is hinged on the "true" values of missing data available in the update of May 2016 for those patients who had not yet been eligible for a given follow-up evaluation two years earlier.

A potential limitation of our study is related to nature of data, in fact we used an ongoing trial that prevent us to use the treatment arm information, so we included in the imputation model a limited number of patient covariates. Another potential limitation lies in the fact that we imputed real missing values together with unobserved data completed by the patients themselves over the following twenty-four months assuming that missingness was at random. As is well known, persons with missing data are usually different from those with known data. Assuming that data are missing at random implies that outcomes for patients who completed the questionnaires are expected to be similar to outcomes for non-respondents with similar baseline characteristics. For this purpose, we included age, grading, the other QoL scores and the previous values on the same information available in the imputation process. In addition, persons with unobserved data may be different from those with missing data, but we treated these populations in the same way after testing that there were no significant differences among them. Therefore, the results obtained on unobserved data could be extended to real missing data and allow considering the two imputation techniques as almost equivalent.

The findings from this study based on real data with several time point measurements and heterogeneous patterns of missingness, showed that MVNI and FCS methods produce similar results

and minimally biased estimates on QoL measures. Despite the limited amount of available data with an update in May 2016, particularly in the first twelve months, we believe that this research, comparing imputed values with their corresponding real information obtained two years later, provide a useful comparison between the MVNI and FCS approaches.

## References


1. Fairclough, D.L., H.F. Peterson, and V. Chang, *Why are missing quality of life data a problem in clinical trials of cancer therapy?* Stat Med, 1998. **17**(5-7): p. 667-77.
2. Curran, D., et al., *Incomplete quality of life data in randomized trials: missing forms.* Stat Med, 1998. **17**(5-7): p. 697-709.
3. Rubin, D.B., *Multiple imputation for nonresponse in surveys.* 1987.
4. White, I.R., P. Royston, and A.M. Wood, *Multiple imputation using chained equations: Issues and guidance for practice.* Stat Med, 2011. **30**(4): p. 377-99.
5. Collins, L.M., J.L. Schafer, and C.M. Kam, *A comparison of inclusive and restrictive strategies in modern missing data procedures.* Psychol Methods, 2001. **6**(4): p. 330-51.
6. Pedersen, A.B., et al., *Missing data and multiple imputation in clinical epidemiological research.* Clin Epidemiol, 2017. **9**: p. 157-166.
7. Nevalainen, J., M.G. Kenward, and S.M. Virtanen, *Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification.* Stat Med, 2009. **28**(29): p. 3657-69.
8. Welch, C.A., et al., *Evaluation of two-fold fully conditional specification multiple imputation for longitudinal electronic health record data.* Stat Med, 2014. **33**(21): p. 3725-37.
9. Schafer, J., *Analysis of Incomplete Multivariate Data.* 1997.
10. Demirtas, H., S. Freels, and R. Yucel, *Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: a simulation assessment.* J Stat Comput Simul, 2008. **78**: p. 69-84.
11. Saraceno, L., et al., *The relationship between childhood depressive symptoms and problem alcohol use in early adolescence: findings from a large longitudinal population-based study.* Addiction, 2012. **107**(3): p. 567-77.
12. Van Buuren, S.B.J., *Fully conditional specification in multivariate imputation.* J Stat Comput Simul, 2006. **76**: p. 1049-64.
13. Lee, K.J. and J.B. Carlin, *Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation.* Am J Epidemiol, 2010. **171**(5): p. 624-32.
14. Kalaycioglu, O.C., A King, M Omar, RZ., *A comparison of multiple imputation methods for handling missing data in repeated measurements observational studies.* R Stat Soc Ser. A-Stat. Soc., 2016. **179**: p. 683-706.
15. De Silva, A.P., et al., *A comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time: a simulation study.* BMC Med Res Methodol, 2017. **17**(1): p. 114.
16. Van Buuren, S., *Multiple imputation of discrete and continuous data by fully conditional specification.* Stat Methods Med Res, 2007. **16**(3): p. 219-42.
17. Jordhoy, M.S., et al., *Quality of life in advanced cancer patients: the impact of sociodemographic and medical characteristics.* Br J Cancer, 2001. **85**(10): p. 1478-85.
18. Ware, J., Jr., M. Kosinski, and S.D. Keller, *A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity.* Med Care, 1996. **34**(3): p. 220-33.
19. Kodraliu, G., et al., *Subjective health status assessment: evaluation of the Italian version of the SF-12 Health Survey. Results from the MiOS Project.* J Epidemiol Biostat, 2001. **6**(3): p. 305-16.
20. Grossi, E., et al., *Development and validation of the short version of the Psychological General Well-Being Index (PGWB-S).* Health Qual Life Outcomes, 2006. **4**: p. 88.
21. Mardia, K., *Measures of multivariate skewness and kurtosis with applications.* Biometrika, 1970. **57**: p. 519-30.
22. Henze, N. and B. Zirkler, *A class of invariant consistent tests for multivariate normality.* Commun Stat-Theor M, 1990. **19**: p. 3595-617.
23. Yucel, R.D., H., *Impact of non-normal random effects on inference by multiple imputation: A simulation assessment.* Comput Stat Data Anal, 2010. **54**: p. 790-801.

24. Biering, K., M. Frydenberg, and N.H. Hjollund, *Self-reported health following percutaneous coronary intervention: results from a cohort followed for 3 years with multiple measurements.* Clin Epidemiol, 2014. **6**: p. 441-9.
25. Biering, K., N.H. Hjollund, and M. Frydenberg, *Using multiple imputation to deal with missing data and attrition in longitudinal studies with repeated measures of patient-reported outcomes.* Clin Epidemiol, 2015. **7**: p. 91-106.
26. Engels, J.M. and P. Diehr, *Imputation of missing longitudinal data: a comparison of methods.* J Clin Epidemiol, 2003. **56**(10): p. 968-76.
27. Royston, P. and I. White, *Multiple Imputation by Chained Equations (MICE): Implementation in Stata.* J Stat Softw, 2011. **45**(4): p. 20.
28. Kropko, J., et al., *Multiple imputation for continuous and categorical data: Comparing joint multivariate normal and conditional approaches.* Polit Anal 2014. **22**: p. 497-519.



**Conflict of interests**

All the co-authors declare no financial competing interests, nor any no relevant conflicts of interests associated with this work

**Funding**

The TOTEM study was funded by the Piedmont Regional Cancer Network.

**Authors' contributions**

- The study is based on an idea developed by RR who participated in the statistical analyses (execution and organization, review and critique) and wrote the paper.
- DC shared the project idea, performed the statistical analyses (execution, organization and discussion), participated in writing the paper and created the infographics.
- EP shared the project idea and participated in writing the paper.
- ST shared the project idea and participated in writing the paper.
- The TOTEM trial was ideated by PZ who shared the project idea and participated in writing the paper.

**Table 1 An overview of response patterns (May 2014) and the increase in complete data after 24 months (May 2016).**

| Description | Baseline | 6 months | 12 months | 24 months | 36 months |
|---|---|---|---|---|---|
| Available in current round (n) | 812 | 764 | 656 | 446 | 266 |
| Overall mortality (n) | NA | 1 | 3 | 4 | 5 |
| Overall dropped out for other reasons (n) | NA | 5 | 42 | 83 | 105 |
| Not yet available in current round (n)[a] | NA | 42 | 111 | 279 | 436 |
| **Questionnaire specific response** | | | | | |
| SF12 PCS/MCS | | | | | |
| Complete (n) | 559 | 529 | 406 | 192 | 111 |
| Real Missing (n) | 253 | 235 | 250 | 254 | 155 |
| | | | | | |
| Increase in complete data after 24 months (n) | NA | 37 | 91 | 165 | 171 |
| PGWBS | | | | | |
| Complete (n) | 617 | 575 | 451 | 221 | 110 |
| Real Missing (n) | 195 | 189 | 205 | 225 | 156 |
| | | | | | |
| Increase in complete data after 24 months (n) | NA | 39 | 93 | 184 | 176 |

**Notes**:[a] Patients who have not yet completed the follow-up study (not yet reached the current round of data collection).
**Abbreviations:** NA, not applicable; SF12, Short Form 12-item survey; PCS, physical component summary; MCS, mental component summary; PGWBS, short version of the Psychological General Well-Being Index.

**Table 2 Results of Mardia's skewness test and kurtosis test and of the Henze-Zirkler test on the SF12 summary scores and the PGWBS summary score.**

| | | Value | Probability |
|---|---|---|---|
| **PCS** | Skewness Mardia | 73.23 | 0.0002 |
| | Kurtosis Mardia | 1.31 | 0.1915 |
| | Henze-Zirkler T | 1.02 | 0.0070 |
| **MCS** | Skewness Mardia | 62.56 | 0.0028 |
| | Kurtosis Mardia | 1.00 | 0.3162 |
| | Henze-Zirkler T | 1.27 | <.0001 |
| **PGWBS** | Skewness Mardia | 98.34 | <.0001 |
| | Kurtosis Mardia | 4.21 | <.0001 |
| | Henze-Zirkler T | 1.25 | <.0001 |

**Abbreviations:** PCS, physical component summary; MCS, mental component summary; PGWBS, short version of the Psychological General Well-Being Index.

**Table 3 Comparison between the Quality of Life scores imputed in the 2014 dataset (not-yet-available data only) and the true values available in the 2016 dataset.**

Panel A) SF12 Physical component summary (PCS)

| | *Measurement points* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | ***6 months*** | | ***12 months*** | | ***24 months*** | | ***36 months*** | |
| ***N*** | 37 | | 91 | | 165 | | 171 | |
| ***PCS mean of observed scores (at May 2016)*** | 45.21 | | 46.78 | | 46.89 | | 48.41 | |
| | ***Imputation method*** | | | | | | | |
| | ***MVNI*** | ***FCS*** | ***MVNI*** | ***FCS*** | ***MVNI*** | ***FCS*** | ***MVNI*** | ***FCS*** |
| ***PCS mean of imputed data*** | 47,99 | 47,90 | 47,03 | 47,18 | 47,98 | 47,94 | 45,86 | 46,13 |
| ***RMSD*** | 7,65 | 7,68 | 7,49 | 7,58 | 8,15 | 8,18 | 7,17 | 7,26 |
| ***MAD*** | 5,88 | 5,99 | 5,31 | 5,48 | 6,17 | 6,22 | 6,09 | 6,19 |
| ***BIAS*** | -2,78 | -2,69 | -0,25 | -0,40 | -1,09 | -1,05 | 2,55 | 2,28 |

Panel B) SF12 Mental component summary (MCS)

| | *Measurement points* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | ***6 months*** | | ***12 months*** | | ***24 months*** | | ***36 months*** | |
| ***N*** | 37 | | 91 | | 165 | | 171 | |
| ***MCS mean of observed scores(at May 2016)*** | 50.54 | | 49.80 | | 49.23 | | 48.93 | |
| | ***Imputation method*** | | | | | | | |
| | ***MVNI*** | ***FCS*** | ***MVNI*** | ***FCS*** | ***MVNI*** | ***FCS*** | ***MVNI*** | ***FCS*** |
| ***MCS mean of imputed data*** | 49,88 | 49,59 | 49,26 | 49,00 | 48,31 | 48,02 | 48,77 | 48,82 |
| ***RMSD*** | 8,21 | 8,16 | 8,04 | 8,26 | 7,86 | 7,79 | 7,46 | 7,59 |
| ***MAD*** | 7,02 | 6,75 | 6,20 | 6,52 | 6,29 | 6,32 | 5,93 | 6,06 |
| ***BIAS*** | 0,66 | 0,95 | 0,54 | 0,80 | 0,92 | 1,21 | 0,16 | 0,11 |

Panel C) Short version of the Psychological General Well-Being Index (PGWBS)

| | *Measurement points* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | ***6 months*** | | ***12 months*** | | ***24 months*** | | ***36 months*** | |
| ***N*** | 39 | | 93 | | 184 | | 176 | |
| ***PGWBS mean of observed scores (at May 2016)*** | 45.92 | | 45.01 | | 44.47 | | 45.67 | |
| | ***Imputation method*** | | | | | | | |
| | ***MVNI*** | ***FCS*** | ***MVNI*** | ***FCS*** | ***MVNI*** | ***FCS*** | ***MVNI*** | ***FCS*** |
| ***PGWBS mean of imputed data*** | 46,06 | 45,67 | 45,11 | 45,09 | 44,69 | 44,50 | 44,46 | 44,62 |
| ***RMSD*** | 6,50 | 6,52 | 6,97 | 7,07 | 7,05 | 7,01 | 6,95 | 6,94 |
| ***MAD*** | 5,24 | 5,24 | 5,02 | 5,03 | 5,15 | 5,08 | 5,36 | 5,39 |
| ***BIAS*** | -0,14 | 0,25 | -0,10 | -0,08 | -0,22 | -0,03 | 1,21 | 1,05 |

**Abbreviations**: SF12, Short Form 12-item survey; MVNI, multivariate normal imputation; FCS, fully conditional specification; RMSD, root mean square deviation; MAD, mean absolute deviation; N, number of patients with unobserved data in 2014.

**Table 4 Comparison between Quality of Life mean scores for imputed (missing and not-yet-available data) and observed data at May 2014.**

Panel A) SF12 Physical Component Summary (PCS)

| *Measurement points* | *N* | *PCS mean score on observed data* | *Mean difference between observed and imputed data* | |
|---|---|---|---|---|
| | | | *MVNI* | *FCS* |
| ***Baseline*** | 559 | 41.09 | 0.07 | 0.02 |
| ***6 months*** | 529 | 45.89 | 0.08 | 0.19 |
| ***12 months*** | 406 | 46.75 | 0.50 | 0.56 |
| ***24 months*** | 192 | 48.14 | 0.64 | 0.53 |
| ***36 months*** | 111 | 46.39 | 0.91 | 0.58 |

Panel B) SF12 Mental Component Summary (MCS)

| *Measurement points* | *N* | *Observed mean SF12 MCS score* | *Mean difference between observed and imputed data* | |
|---|---|---|---|---|
| | | | *MVNI* | *FCS* |
| ***Baseline*** | 559 | 43.79 | 0.06 | 0.01 |
| ***6 months*** | 529 | 47.33 | 0.23 | 0.4 |
| ***12 months*** | 406 | 47.94 | 0.02 | 0.05 |
| ***24 months*** | 192 | 47.36 | 0.02 | -0.16 |
| ***36 months*** | 111 | 47.81 | -0.49 | -0.64 |

Panel C) Short version of the Psychological General Well-Being Index (PGWBS).

| *Measurement points* | *N* | *Observed mean PGWBI score* | *Mean difference between observed and imputed data* | |
|---|---|---|---|---|
| | | | *MVNI* | *FCS* |
| ***Baseline*** | 617 | 39.53 | 0.19 | 0.09 |
| ***6 months*** | 575 | 43.14 | -0.03 | 0.03 |
| ***12 months*** | 451 | 43.61 | -0.01 | -0.05 |
| ***24 months*** | 221 | 43.36 | -0.3 | -0.63 |
| ***36 months*** | 110 | 43.27 | -0.53 | -0.78 |

**Abbreviations**: SF12, Short Form 12-item survey; MVNI, multivariate normal imputation; FCS, fully conditional specification; N, number of patients with complete data in 2014.

**Appendix 1** Most representative missing patterns (September 2008-May 2014).

| Pattern of missingness | Baseline | 6 months | 12 months | 24 months | 36 months | Percentage |
|---|---|---|---|---|---|---|
| **SF-12** | | | | | | |
| **1** | ● | ● | ● | ● | ● | 4.26 |
| **2** | ● | ● | ● | × | × | 7.06 |
| **3** | ● | ● | ● | ○ | × | 4.74 |
| **4** | ● | ● | ● | ● | ○ | 4.26 |
| **5** | ● | ● | ○ | × | × | 3.53 |
| **6** | ● | ● | ● | ○ | ○ | 3.16 |
| **7** | ● | ○ | × | × | × | 3.16 |
| **8** | ● | ● | ○ | ○ | ○ | 2.68 |
| **9** | ○ | ● | ● | ○ | ○ | 2.55 |
| **10** | ○ | ● | ○ | ○ | ○ | 2.55 |
| **PGWBSh** | | | | | | |
| **1** | ● | ● | ● | ● | ● | 4.99 |
| **2** | ● | ● | ● | × | × | 8.03 |
| **3** | ● | ● | ● | ● | ○ | 4.62 |
| **4** | ● | ● | ● | ○ | × | 4.14 |
| **5** | ● | ● | ○ | × | × | 4.14 |
| **6** | ● | ● | ● | ○ | ○ | 3.89 |
| **7** | ● | ● | ○ | ○ | ○ | 3.28 |
| **8** | ● | ○ | × | × | × | 3.28 |
| **9** | ○ | ● | ● | ○ | ○ | 2.43 |
| **10** | ○ | ● | ○ | ○ | ○ | 2.07 |

● observed ○ missing × not yet available in current round

**Abbreviations:**; SF-12, Short Form 12-item survey; PGWBSh, short version of the Psychological General Well-Being Index.

Appendix 2 Univariate distribution of QoL raw scores by time points

**Figure A: Histograms of SF12-Mental Composite row scores by time points**

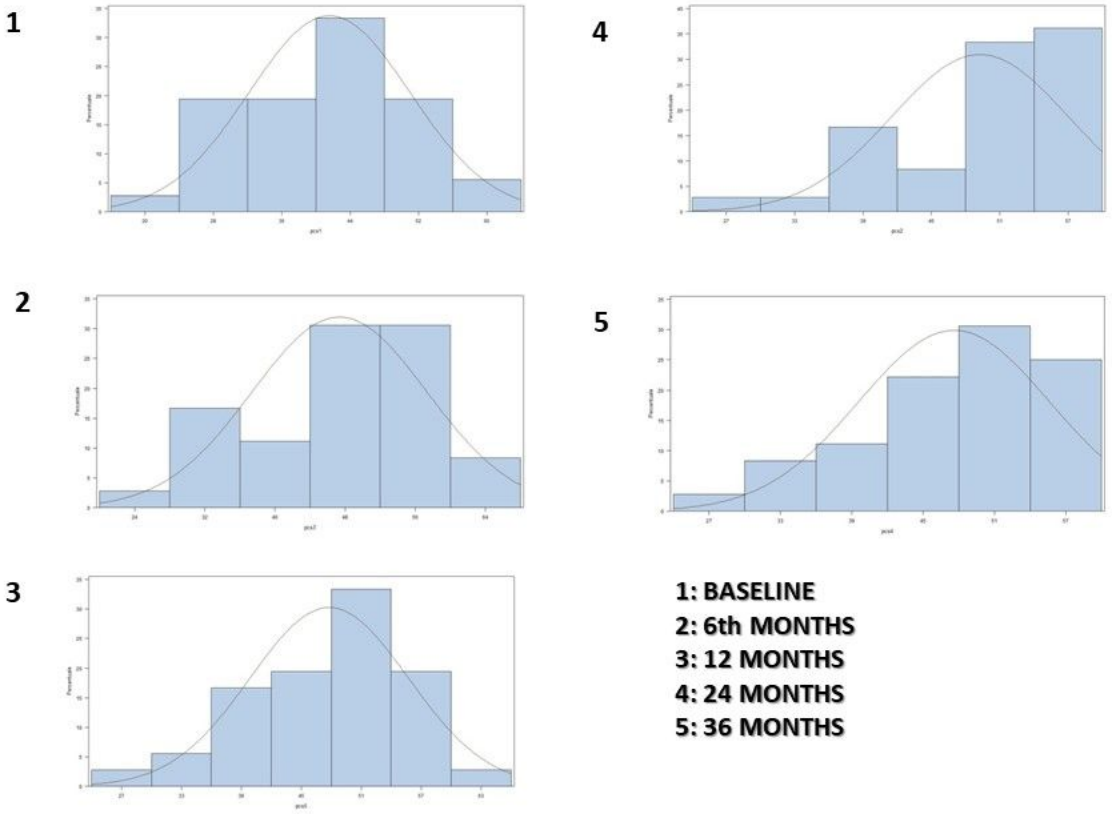


**Figure B: Histograms of SF12-Physical Composite row scores by time points**

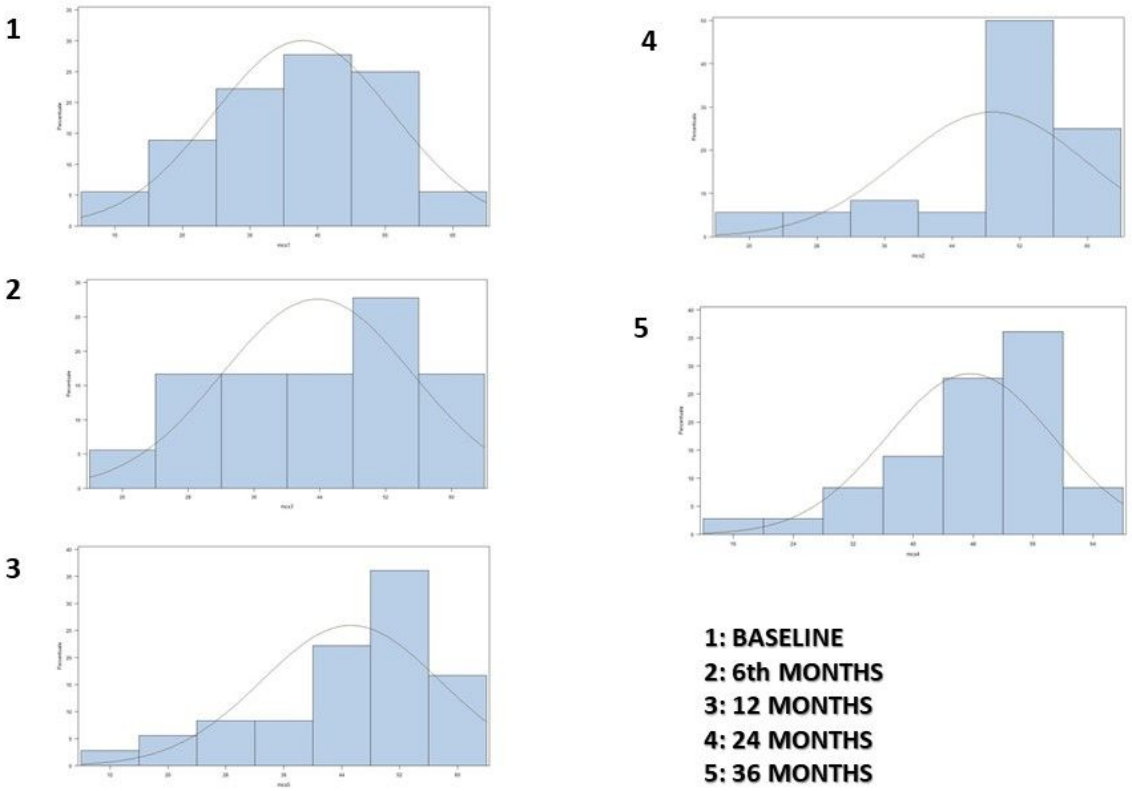
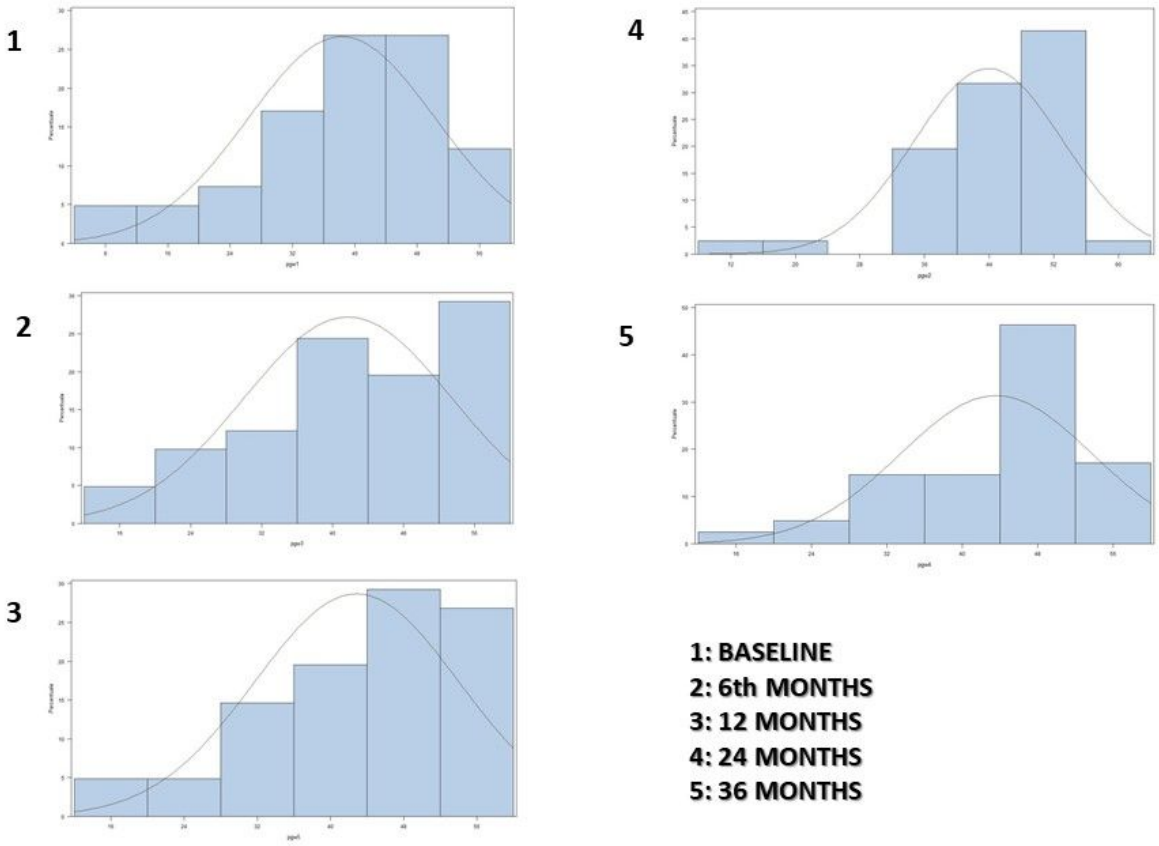


**Figure C: Histograms of Psychological Well-being Composite row scores by time points**



**Appendix 3 A comparison of age and grading between real missing and unobserved data (May 2014).**

| *QoL scores:* | | *6 months* | | | *12 months* | | | *24 months* | | | *36 months* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *REAL MISSING* | *UNOBSERVED DATA NYA* | *P-value* | *REAL MISSING* | *UNOBSERVED DATA NYA* | *P-value* | *REAL MISSING* | *UNOBSERVED DATA NYA* | *P-value* | *REAL MISSING* | *UNOBSERVED DATA NYA* | *P-value* |
| *SF12* | *Age (mean)* | 63.54 | 62.69 | 0.63 [a] | 63.82 | 63.97 | 0.90 [a] | 63.04 | 62.77 | 0.77 [a] | 64.26 | 62.5 | 0.07 [a] |
| | *Grading 1(%)* | 37.87 | 50 | | 39.6 | 49.55 | | 39.37 | 45.88 | | 43.23 | 44.95 | |
| | *Grading 2(%)* | 41.28 | 40.48 | 0.16 [b] | 43.2 | 36.94 | 0.21 [b] | 47.64 | 37.63 | 0.06 [b] | 45.81 | 38.99 | 0.18 [b] |
| | *Grading 3(%)* | 20.85 | 9.52 | | 17.2 | 13.51 | | 12.99 | 16.49 | | 10.96 | 16.06 | |
| *PGWBS* | *Age* | 63.4 | 62.7 | 0.70 [a] | 63.74 | 63.97 | 0.85 [a] | 63 | 62.67 | 0.81 [a] | 63.5 | 62.5 | 0.30 [a] |
| | *Grading 1(%)* | 37.04 | 50 | | 42.93 | 49.55 | | 39.56 | 45.88 | | 43.59 | 44.95 | |
| | *Grading 2(%)* | 42.33 | 40.48 | 0.15 [b] | 43.9 | 36.94 | 0.46 [b] | 47.56 | 37.63 | 0.08 [b] | 44.23 | 38.99 | 0.37 [b] |
| | *Grading 3(%)* | 20.63 | 9.52 | | 13.17 | 13.51 | | 12.88 | 16.49 | | 12.18 | 16.06 | |

**Notes**:[a] t-test; [b] chi-square test.
**Abbreviations:** NYA, not yet available; SF12, Short Form 12-item survey; PGWBS, short version of the Psychological General Well-Being Index.

**Appendix 4 95% confidence intervals for the mean of the Root Mean Square Deviation (RMSD), Mean Absolute Deviation (MAD) and BIAS by imputation methods and QoL scores.**