



UvA-DARE (Digital Academic Repository)

Clinical, economic and methodological studies in elderly patients with dementia and their informal caregivers

MacNeil Vroomen, J.L.

Publication date

2015

Document Version

Final published version

[Link to publication](#)

Citation for published version (APA):

MacNeil Vroomen, J. L. (2015). *Clinical, economic and methodological studies in elderly patients with dementia and their informal caregivers*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



Comparing Multiple Imputation strategies for zero-inflated cost data: which method works best?

Janet MacNeil Vroomen
Iris Eekhout
Marcel G. Dijkgraaf
Hein van Hout

Sophia E. de Rooij
Martijn W. Heymans
Judith E. Bosmans

Submitted

ABSTRACT

Background

Cost and effect data are prone to missing data because economic evaluations are often “piggy-backed” onto clinical studies where cost data are rarely the primary outcome. The objective of this article was to investigate which multiple imputation strategy is most appropriate to use for missing cost-effectiveness data in a pragmatic randomized controlled trial (RCT).

Methods

Three incomplete datasets were generated from a complete reference dataset with 17%, 35% and 50% missing data in effects and costs. The strategies evaluated included complete case analysis (CCA), multiple imputation with predictive mean matching (MI-PMM), MI-PMM on log-transformed costs (Log MI-PMM), and a two-step MI. Mean cost and effect estimates, standard errors and incremental net benefits were compared with the results of the analyses on the complete reference dataset.

Results

The CCA, MI-PMM, and the two-step MI strategy deviated more from the results for the reference dataset when the amount of missing data increased. In contrast, the estimates of the Log MI-PMM strategy remained stable regardless of the amount of missing data. MI provided better estimates than CCA in all scenarios.

Conclusion

With low amounts of missing data the MI strategies appeared equivalent but with missing data greater than 35%, we recommend using the Log MI-PMM.

1. INTRODUCTION

Researchers should aim for collecting high quality and complete data as missing data may lead to loss of information in epidemiological and clinical research (1). However, missing data are unavoidable when performing trials where data is collected through self-report by the participants. Cost data are prone to missing data because economic evaluations are often “piggy-backed” onto clinical studies where cost data are rarely the primary outcome. Moreover, one missing cost measurement results in a missing total cost estimate, because costs are summed over all measurements.

Complete Case Analysis (CCA) is the default strategy to deal with missing data although it is known for biased estimates, wide standard errors and decreased power. Oostenbrink et al. (2) and Briggs et al. (3) showed that multiple imputation techniques performed better than CCA and simple imputation techniques (conditional mean imputation, single imputation with predictive mean matching (PMM), hot decking and expectation maximization) (2,3)

Recently, multiple imputation has been recommended as the most appropriate way for handling missing data (1,4-7). Multiple imputation can be a powerful tool to estimate missing data (5), but there are some important points to consider when specifying the multiple imputation model. First, the imputation model should include all variables that explain missing values. Second, it should include all variables included in the analysis model and third the imputation model must account for the distribution of the data. This assumption may not be met when imputing cost data in trials because of the distributional issues posed by cost data including constrained positive values, a large amount of zero values, and right-handed tail skew.

Multiple imputation with predictive mean matching (PMM) can be a helpful tool to deal with the skewed distribution of cost data, because PMM preserves the distribution of the data and, therefore, is robust against violations of the normality assumption (5). Another commonly recommended approach to deal with skewed data is to take the log of the skewed variables before imputation and then back transform the variables to their original scale before the target analysis. (4,5,8). Lee and Carlin (8) compared multiple imputation with transformation and PMM to deal with non-normality in continuous variables. They recommended transformation of skewed variables to a symmetric distribution to avoid the introduction of biases of study results. Another alternative is to impute missing data in two separate steps. In the first step, the probability of having costs is imputed which takes care of the zero inflation, and in the second step, an actual cost value is imputed for individuals that are predicted to have costs. In the second step, the skew of the cost data is taken into account by using the PMM algorithm to impute the cost values for the people that are predicted to have costs using only the observed cost data (9).

It is unclear which method to deal with imputation of skewed data is the most appropriate in economic evaluations. Therefore, the objective of this article was to investigate which imputation strategy is most appropriate to impute missing cost and effect data in an

economic evaluation alongside a pragmatic randomized controlled study. The strategies compared include complete case analysis (CCA), MI with predictive mean matching (MI-PMM), MI with predictive mean matching on log-transformed costs (Log-MI-PMM), and two step multiple imputation with predictive mean matching (two-step-MI).

2.2 Missing data

IE generated missing data in the complete dataset using R statistical software (13). Three incomplete datasets were created with 17%, 35% and 50% missing data to investigate the effect of the rate of incomplete data on the performance of the imputation methods. Missing data points were created in the QALY variable and several cost variables. The probability of missing data was related to other variables in the data to satisfy a missing at random (MAR) assumption for the missing data. Center, location, age, administering of a second interview, and abstinence were predictors of missingness in the utility and cost outcome variables for the data set with 17% missing data. In the data set with 35% missing data, predictor variables were treatment group, centre, sex, age, and occurrence of a second interview. In the dataset with 50% missing data, the predictor variables were treatment group centre, age and occurrence of a second interview. Table 2 presents all key cost variables with missing data for the different missing data scenarios.

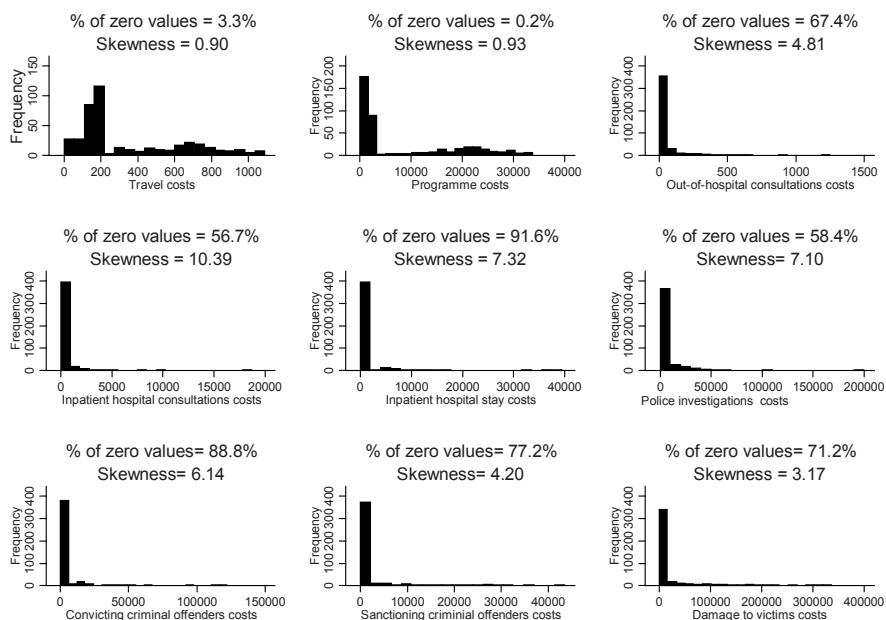


Figure 1 | Frequency distribution of sub costs from reference dataset

Table 1 | Baseline characteristics of reference dataset

	Metadone alone (n=237)	Co-prescribed heroin (n=193)
Explanatory variables:		
% male (n)	55.1 (190)	44.9 (155)
age (sd)	38.9 (5.7)	39.7 (5.8)
% injected (n)	56.3 (98)	43.7 (76)
% completed (n)	60.2 (204)	39.8 (135)
% abstinent (n)	59.3 (80)	40.7 (55)
% second interview performed (n)*	55.7 (59)	44.3 (47)
baseline utility (sd)	0.731 (0.273)	0.739 (0.272)

*Those included early in the trials also completed the questionnaire in the second month.

2.3 Missing data strategies

CCA

In CCA, analysis was restricted to participants with complete cost and effect data. This resulted in smaller sample sizes than in the reference dataset (see Table 2).

Multiple imputation procedure

Multiple imputation was done using fully conditional specification. Fully conditional specification or chained equations is a flexible multivariate model that does not rely on the assumption of multivariate normality (5). Regression models are specified for each variable with missing values, conditional on all of the other variables in the imputation model. Imputations are generated by drawing from iterated conditional models (5).

The imputed values were estimated using the predictive mean matching (PMM) algorithm. PMM is an algorithm that matches the missing value to the observed value with the closest predicted estimate (White et al., 2011). The predicted mean is estimated in a regression equation where a random residual term is added to the estimate in order to account for missing data uncertainty. In PMM, instead of using the predicted estimate, the imputed value is randomly selected from observed values that are closest to the predicted estimate. For example, an older single man misses a measurement for blood pressure and the value for this man is estimated to be 102.34 mmHg by regressing blood pressure on age and sex. Five other older single men have observed blood pressures of 103; 103; 102; 101, and 104 mmHg, respectively. The missing value is then imputed with a random draw from these five blood pressures. PMM has several advantages when imputing cost data. It is more robust against non-normal data as it uses the observed distribution of the data. Furthermore, it imputes only plausible values because it randomly draws from observed values. The process of estimating imputed values is repeated in sequential cycles, each time using the updated data with the imputed estimates from the previous cycle. These cycles are called iterations. One of these iterations (e.g. the 100th) is selected and used as an imputed dataset until 'm' datasets were selected in total. We used 200 imputations to minimize internal variation so that the imputation variation would not affect the performance of each imputation method (1,14-16). We performed MI using the chained command in Stata 12, which uses fully conditional specification to perform the multiple imputations (Statscorp, 2011).

Table 2 | Frequency and description of reference dataset, datasets with 17%, 35% and 50% missingness

description	QALY (99) on a 0 to 1 scale	costs of travel	total program							costs of police investigations of criminal offenders	costs of convicting criminal offenders	costs of sanctions of criminal offenders	costs of damage to victims
			costs including housing, rebuilding and ch and cm	costs of out-of-hospital consults	costs of inpatient hospital consults	costs of inpatient hospital stay	costs of	costs of	costs of				
reference dataset													
mean	0.76	350.04	8,692.99	45.97	316.12	778.96	6,004.28	3,172.95	1,854.53	23,602.36			
standard deviation	0.22	292.79	10,315.04	124.06	1,206.92	3,680.04	14,497.13	12,627.55	5,935.78	59,945.50			
percentage of zeroes	N/A	3.3	0.2	67.4	56.7	91.6	58.4	88.8	77.2	71.2			
skew	-1.28	0.90	0.93	4.81	10.39	7.32	7.10	6.14	4.20	3.17			
18% missing													
mean	0.75	350.04	8,266.09	43.89	323.21	727.68	6,304.52	3,260.05	1,751.09	24,391.48			
standard deviation	0.23	292.79	10,022.88	109.55	1,267.93	3,446.62	15,219.27	12,995.47	5,797.58	61,348.19			
percentage of zeroes	N/A	3.3	0.3	67.2	58.1	91.8	58.9	88.9	77.8	71.0			
skew	-1.30	0.90	0.99	3.95	10.01	7.36	6.81	6.06	4.45	3.14			
35% missing													
mean	0.75	350.04	8,218.24	44.99	336.41	794.57	6,462.09	3,027.35	1,731.97	24,611.35			
standard deviation	0.23	292.79	10,075.05	112.93	1,318.40	3,711.62	15,665.51	12,569.37	5,783.11	62,221.27			
percentage of zeroes	N/A	3.3	0.3	66.0	57.4	91.6	58.8	89.6	78.1	70.7			
skew	-1.23	0.90	1.02	4.00	9.69	7.09	6.75	6.40	4.56	3.17			
50% missing													
mean	0.74	350.04	7,907.28	49.09	338.09	853.85	6,523.91	2,586.35	1,749.48	23,280.58			
standard deviation	0.23	292.79	9,995.16	118.29	1,357.34	3,972.24	15,794.79	11,541.70	5,873.06	60,485.85			
percentage of zeroes	N/A	3.3	0.3	63.7	57.7	91.0	57.7	90.5	78.2	72.5			
skew	-1.20	0.90	1.11	3.82	9.60	7.00	6.99	7.31	4.58	3.27			

We performed the multiple imputations stratified by treatment group to maintain the possible group effect in the data. For all multiple imputation strategies we checked the convergence plots to see if iterations were free from trend and imputations were successful. To solve any occurring convergence problems, we merged highly correlated variables together. For this reason, travel costs were merged together with total programme costs (correlation coefficient > 0.9). In-patient hospital consultations and inpatient length of hospital stay were also highly correlated and were therefore merged together as well.

Three multiple imputation strategies were compared and are described below.

MI-PMM: In the first multiple imputation strategy we performed multiple imputation with predictive mean matching on the raw data.

Log MI-PMM: In the second multiple imputation strategy, we applied the predictive mean matching algorithm to the log transformed cost data. This was done by first adding a constant to the raw cost data in order to circumvent problems when transforming zero values, and next the log was taken. After imputation, the complete data were transformed back to their original scale prior to any analyses being performed.

Two-step MI: The third multiple imputation strategy was a conditional two-step approach. We recoded cost variables to dummy variables where subjects were coded as 1 if they had costs and a 0 for no costs. Missing values were left to be multiply imputed with either a 0 or 1 using a logit function. Next, multiple imputation with the PMM algorithm was performed for missing cases with a value 1 on the dummy variables. Only cases with cost estimates higher than zero were used for this imputation step. For variables that did not have a sufficient amount of zeroes to perform the conditional imputation, we chose to apply only the second step on the raw cost variable.

2.4 Statistical Analysis

We used a generalized linear regression model with a gamma distribution and an identity link to estimate mean differences in total costs. The gamma distribution was chosen to take into account the right skewness of the cost data. The generalized linear model for quality adjusted life years (QALYs) was adjusted for baseline utility estimates. Mean differences and standard errors were pooled using Rubin's rules (16).

We estimated the correlation between the incremental total costs and the incremental QALYs in the reference dataset and the imputed datasets. In the multiple imputation strategies, the covariance between total costs and QALYs was calculated based on the Fisher z transformation and was then pooled using Rubin's rules (5,17).

Incremental Cost-Effectiveness Ratios (ICERs) were calculated using the pooled cost and effect estimates. The ICER is calculated as $\frac{\hat{\Delta}_c}{\hat{\Delta}_e}$ where $\hat{\Delta}_c$ is the difference in total costs between the two intervention groups and $\hat{\Delta}_e$ is the difference in QALYs between the two intervention groups.

Incremental net benefit (INB) estimates were calculated using the following formula: $\hat{b}(\lambda) = \hat{\Delta}_e \lambda - \hat{\Delta}_c$ (18,19) where $\hat{\Delta}_e$ is the difference in QALYs between the two

intervention groups, λ is the willingness to pay, and $\hat{\Delta}_c$ is the difference in costs. The variance of INB was calculated using: $V[\hat{b}(\lambda)] = \hat{V}(\hat{\Delta}_e)\lambda^2 + \hat{V}(\hat{\Delta}_c) - 2\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)\lambda$ where \hat{C} is the covariance between the differences in total costs and QALYs (18,19). We set the willingness-to-pay at EUR 30,000 because this is roughly equivalent to the cut-off value mentioned in the Standard National Institute of Clinical Excellence guidelines (20,000-30,000 pounds per QALY) for economic evaluations (20)

Cost-effectiveness acceptability curves (CEAC) were estimated to quantify the uncertainty due to sampling and measurement errors and because lambda is generally unknown. The CEAC is a plot of the probability that co-prescribed heroin compared to methadone maintenance only is cost-effective (y-axis) as a function of the money society might be willing to pay for one additional QALY (x-axis). The pooled coefficients and variance parameters from the regression models were used for the CEACs.

2.5 Comparison of strategies

The estimates from the reference dataset were considered the “true values” and we compared the estimates from the different multiple imputation strategies with these true values. We evaluated the percentage of bias from the reference analysis (RA) in the different imputation strategies for cost and effect differences, standard error estimates, p-values and t-values. The primary outcomes of interest were the value of INB at a willingness to pay of 30,000 EUR per QALY, the standard error of INB and the probability that co-prescribed heroin compared to methadone maintenance at a willingness to pay of 30,000 EUR per QALY. The strategies that gave the closest estimates to the reference dataset were considered the best.

2.6 Sensitivity analysis

Research suggests that it is better to impute at the item and not the total level (21,22). Therefore, we imputed the total cost variable directly as a sensitivity analysis for all missing data strategies.

3. RESULTS

3.1 Costs

Table 3 presents the cost estimates for the reference case, the CCA, and the different imputation strategies for 17%, 35% and 50% missing data. The difference in costs of -€12,792 in the RA fell within the 95% confidence intervals of all multiple imputation strategies for all rates of missing data. The CCA deviated the most from the RA compared to all other strategies specifically with regard to the cost differences and the associated standard errors in all scenarios. For 17% of missing data, the CCA showed a statistically significant difference in costs just as in the reference analysis. However, for 35% or 50% of missing data the cost difference was no longer statistically significant. The multiple imputations strategies gave similar results to each other in the 17% and 35% missing datasets showing smaller differences in costs and larger standard errors when the amount

of missing data increased compared to the reference analysis. The log transformed-PMM deviated the least from the RA in the 50% missing dataset for the cost difference, standard error and p-values. The two-step MI deviated the most from the RA with regard to cost differences and the standard errors in the dataset with 50% missing data.

3.2 QALYs

Table 4 provides the QALY results for the 17%, 35% and 50% missing data. In the 17% missing dataset, all strategies deviations were roughly the same amount for the difference in QALYs and standard error. All imputation strategies, including the CCA showed a statistically significant difference ($p < 0.001$) in QALYs between the two intervention groups.

In the dataset with 35% missing data, the QALY coefficient in the CCA deviated the least and the most deviation occurred in the Log-MI-PMM, but the reference coefficient was still contained in all confidence intervals. The standard error of the CCA deviated the most from the standard error in the RA while the MI-PMM deviated the least. All strategies still showed that co-prescribed heroin was associated with higher QALY scores compared to methadone maintenance.

In the 50% missing dataset, the QALY coefficient deviated the most in the MI-PMM and the least in the CCA but the regression coefficient from the RA was still within all 95% confidence intervals. The standard error for the CCA deviated the most from the reference analysis, but the deviation in all MI strategies was similar. The CCA was the only strategy where the difference in QALYs was no longer statistically significant.

Table 3 | cost table

17% missing data												
	RA		CCA		MI-PMM		Log-MI-PMM (%)		two-step MI (%)		Bias from RA	M+H
	M	M+H	M	M+H	(% Bias from RA)	M	M+H	Bias from RA	M	M+H		
n	237	193	201	154	237	193	237	193	237	193	237	193
mean	50,560	37,767	53,148	38,933	51,369	37,935	51,966	38,137	52,685	38,482	52,685	38,482
standard error	5,359	3,063	6,056	3,744	5,650	3,268	5,652	3,309	5,642	3,394	5,642	3,394
treatment cost difference	-12,792		-14,215		-13,434		-13,829		-13,203		-13,203	
			(11)		(5)		(8)		(3)		(3)	
standard error	6,086		7,077		6,440		6,459		6,506		6,506	
			(14)		(5)		(5)		(6)		(6)	
z for CCA and t for imputed	-2.100		-2.010		-2.090		-2.140		-2.030		-2.030	
p-value	0.036		0.045		0.037		0.032		0.042		0.042	
95% confidence interval lower limit	-24,720		-28,085		-26,057		-26,489		-25,954		-25,954	
95% confidence interval upper limit	-865		-345		-810		-1,169		-452		-452	
35% missing data												
	RA		CCA		MI-PMM		Log-MI-PMM (%)		two-step MI (%)		Bias from RA	M+H
	M	M+H	M	M+H	(% Bias from RA)	M	M+H	Bias from RA	M	M+H		
n	237	193	163	122	237	193	237	193	237	193	237	193
mean	50,560	37,767	52,255	43,176	50,810	39,851	51,434	40,408	51,195	40,426	51,195	40,426
standard error	5,359	3,063	6,953	4,560	5,989	3,448	5,975	3,601	6,052	3,551	6,052	3,551
treatment cost difference	-12,792		-9,080		-10,959		-11,026		-10,769		-10,769	
			(29)		(14)		(16)		(16)		(16)	

Table 3 | Continued

standard error	6,086	8,463 (39)	6,853 (13)	6,988 (15)	6,954 (14)								
z for CCA and t for imputed	-2.100	-1.070	-1.600	-1.580	-1.550								
p-value	0.036	0.283	0.110	0.115	0.122								
95% confidence interval lower limit	-24,720	-25,667	-24,393	-24,725	-24,400								
95% confidence interval upper limit	-865	7,508	2,475	2,673	2,862								
50% missing data													
	RA		CCA		MI-PMMM		Log-MI-PMIM (%)		Two-step MI (% Bias from RA)				
	M	M+H	M	M+H	M	M+H	M	M+H	M	M	M+H	M	M+H
n	237	193	132	91	237	193	237	193	237	237	193	237	193
mean	50,560	37,767	50,160	42,794	48,711	38,335	49,180	38,527	49,110	49,110	39,454	49,110	39,454
standard error	5,359	3,063	7,447	5,336	5,875	3,513	5,857	3,501	5,913	5,913	3,683	5,913	3,683
treatment cost difference (% Bias from RA)	-12,792		-7,366 (42)		-10,376 (19)		-10,653 (17)		-9,656 (25)				
standard error	6,086		9,496 (56)		6,852 (13)		6,764 (11)		6,954 (14)				
z for CCA and t for imputed	-2.100		-0.780		-1.510		-1.570		-1.390				
p-value	0.036		0.438		0.130		0.115		0.165				
95% confidence interval lower limit	-24,720		-25,978		-23,810		-23,912		-23,289				
95% confidence interval upper limit	-865		11,246		3,058		2,607		3,977				

Table 4 | clinical effects

17% missing data		RA		CCA		MI-PMM		Log-MI-PMM (% Bias from RA)		Two-step MI (% Bias from RA)	
QALY model		M	M+H	M	M+H	M	M+H	M	M+H	M	M+H
n		237	193	201	154	237	193	237	193	237	193
mean		0.730	0.798	0.722	0.798	0.728	0.792	0.727	0.791	0.728	0.792
standard error		0.015	0.016	0.017	0.016	0.016	0.016	0.016	0.016	0.020	0.016
QALY difference		0.054		0.060 (11)		0.061 (12)		0.061 (12)		0.061 (12)	0.061 (12)
standard error		0.018		0.020 (12)		0.020 (10)		0.020 (10)		0.020(11)	0.020(11)
z for CCA and t for imputed		2.970		2.950		3.020		3.020		3.000	3.000
p-value		0.003		0.003		0.003		0.003		0.003	0.003
95% confidence interval lower limit		0.018		0.020		0.021		0.021		0.021	0.021
95% confidence interval lower limit		0.090		0.100		0.100		0.100		0.100	0.100
35% missing data											
QALY model		M	M+H	M	M+H	M	M+H	M	M+H	M	M+H
n		237	193	163	122	237	193	237	193	237	193
mean		0.730	0.790	0.715	0.790	0.718	0.790	0.717	0.791	0.718	0.790
standard error		0.015	0.018	0.020	0.018	0.017	0.016	0.017	0.016	0.017	0.016
QALY difference		0.054		0.068 (24)		0.069 (27)		0.071 (30)		0.069 (27)	0.069 (27)
standard error		0.018		0.023 (27)		0.021 (17)		0.022 (18)		0.022 (20)	0.022 (20)
z for CCA and t for imputed		2.970		2.910		3.230		3.260		3.150	3.150
p-value		0.003		0.004		0.001		0.001		0.002	0.002
95% confidence interval lower limit		0.018		0.022		0.027		0.028		0.026	0.026
95% confidence interval lower limit		0.090		0.113		0.111		0.113		0.112	0.112

Table 4 | Continued

50% missing data

QALY model	RA		CCA		MI-PMM		Log-MI-PMM (% Bias from RA)		Two-step MI (% Bias from RA)	
	M	M+H	M	M+H	M	M+H	M	M+H	M	M+H
n	237	193	132	91	237	193	237	193	237	193
mean	0.730	0.782	0.717	0.782	0.705	0.785	0.708	0.784	0.706	0.784
standard error	0.015	0.021	0.021	0.021	0.018	0.017	0.018	0.018	0.018	0.018
QALY difference		0.054		0.047 (13)		0.077 (41)		0.074 (36)		0.075 (38)
standard error		0.018		0.026 (43)		0.024 (29)		0.024 (30)		0.024 (31)
z for CCA and t for imputed		2.970		1.820		3.260		3.110		3.140
p-value		0.003		0.069		0.001		0.002		0.002
95% confidence interval lower limit		0.018		-0.004		0.031		0.027		0.028
95% confidence interval lower limit		0.090		0.098		0.123		0.120		0.122

3.3 Cost-utility analysis

Figure 2 and Table 5 show the ICERs, INB, its variance, and the probability that co-prescribed heroin compared to methadone maintenance is cost-effective at a threshold value of 30,000 €/QALY for the 17%, 35% and 50% missing datasets. The CCA showed the largest deviation from the RA for the INB and its standard error, and the ICER in the 17% missing data scenario. The INBs in the two-step MI strategy deviated the least from the INB in the reference analysis. The standard error deviated similarly for all imputation strategies. The reference value of INB was contained in the confidence intervals of all imputation strategies. The probability of co-prescribed heroin compared to methadone maintenance being cost-effective was 99 % for a willingness-to-pay threshold value of EUR 30,000 for one-unit gain in QALY score regardless of the imputation strategy.

In the 35% missing data scenario, the CCA deviated the most from the RA for the ICER, INB and its standard error, and the probability that the intervention was cost effective. The MI-PMM deviated least from the RA for the INB standard error compared to the other imputation strategies. The probability of co-prescribed heroin being cost-effective compared with methadone maintenance was 97% for a willingness-to-pay threshold value of EUR 30,000 for one-unit gain in QALY score for all multiple imputation strategies versus 99% for the RA (CCA was 90%).

In the scenario with 50% missings, the INB was no longer statistically significant for the CCA. The Log-MI-PMM showed the least deviation from the RA in the INB coefficient and its standard error, and the probability that the intervention was cost effective. The probability of co-prescribed methadone being cost-effective compared with methadone maintenance at 30,000 €/QALY was 97%.

The reference INB was within the 95% confidence intervals for all imputation strategies (see Figure 2). For the CCA, INB was no longer statistically significant with 35% and 50% of missing data. INB decreased with higher rates of missing data and the uncertainty was larger as evidenced by the larger standard errors and wider confidence intervals in all strategies. The Log-MI-PMM showed the least uncertainty around INB in all missing data scenarios. Figure 3 presents the CEAC curves for the different strategies with 50% of missing data. This figure shows that there are pronounced differences between the strategies in this scenario. It shows that the probability that co-prescribed heroin is cost-effective when the threshold value is zero, is 98% for the reference analysis, 94% for the Log-MI-PMM and MI-PMM, 92% for the two-step-MI and 78% for the CCA. This increases to 99%, 97%, 97%, 96% and 82% for the RA, MI-PMM, Log MI-PMM, two-step MI and CCA, respectively at a threshold value of 30,000 €/QALY.

3.4 Sensitivity analysis

The imputation procedure was applied to the total costs directly for the MI-PMM and the log-MI-PMM. The results showed that the precision decreased resulting in wider standard errors and increased percentage of bias in the cost difference from the reference analysis when applying multiple imputation to the total costs compared with imputation of sub-cost variables (data not shown).

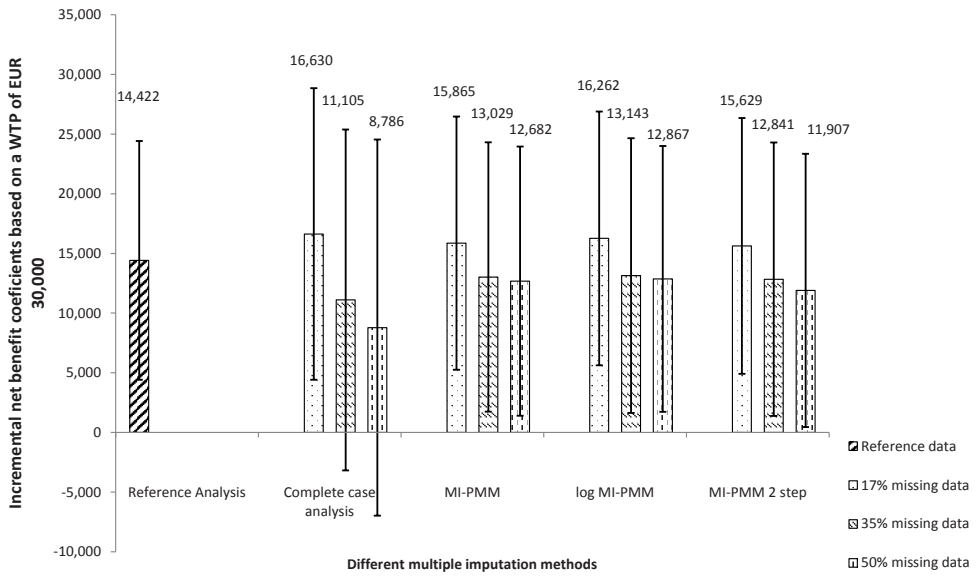


Figure 2 | INB coefficients for a willingness-to-pay of EUR 30,000 based on the amount of missing data and imputation method

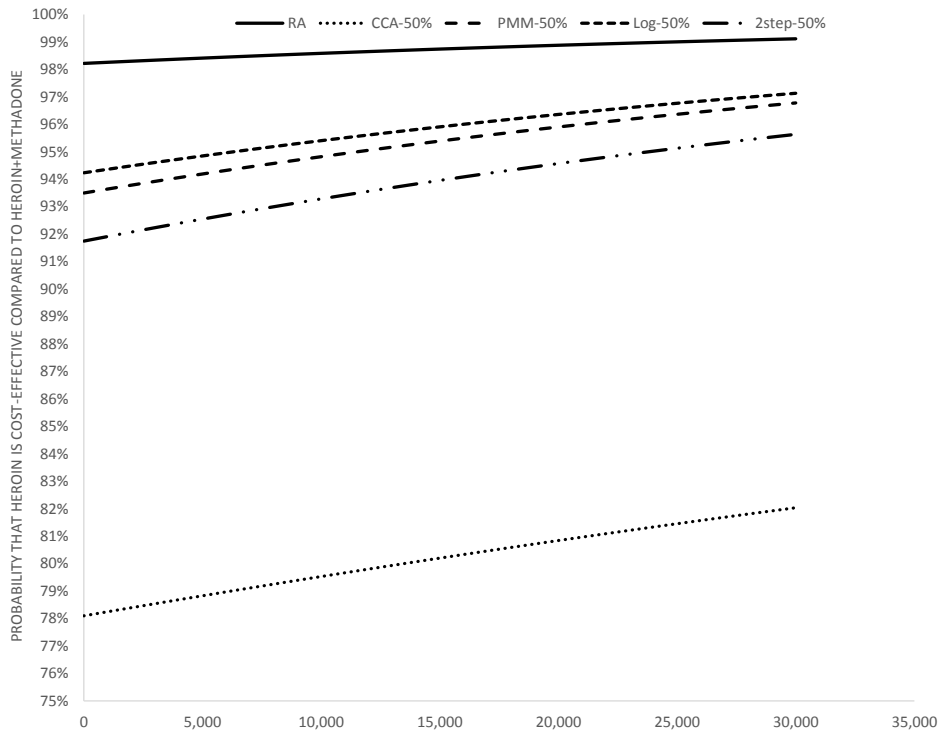


Figure 3 | CEAC curves for a willingness to pay of EUR 30,000 based the 50% missing data scenario

Table 5 | cost-effectiveness output

17% Missing data		CCA	MI-PMM	Log-MI-PMM	Two-step MI
CEA analysis	RA	(% Bias from RA)	(% Bias from RA)	(% Bias from RA)	(% Bias from RA)
Correlation between utility and costs	0.0507	0.0591	0.0517	0.0509	0.0487
Covariance	5.6	8.6	6.7	6.6	6.4
INB	14,422	16,026 (11)	15,257 (6)	15,654 (9)	15,023 (4)
Standard error of INB	6,083	7,270 (20)	6,438 (6)	6,457 (6)	6,504 (7)
CI Lower limit	4,417	4,069	4,669	5,034	4,324
CI Upper limit	24,427	27,983	25,846	26,274	25,721
Prob C-E	0.99	0.99 (0)	0.99 (0)	0.99 (0)	0.99 (0)
ICER	-235,472	-235,448 (0)	-220,988 (6)	-227,410 (3)	-217,656 (8)
35% Missing data		CCA	MI-PMM	Log-MI-PMM	Two-step MI
CEA analysis	RA	(% Bias from RA)	(% Bias from RA)	(% Bias from RA)	(% Bias from RA)
Correlation between utility and costs	0.0507	0.0251	0.0300	0.0292	0.028
Covariance	5.6	4.9	4.4	4.4	4.4
INB	14,422	11,105 (23)	13,029 (10)	13,143 (9)	12,841 (11)
SE of INB	6,083	8,685 (43)	6,864 (13)	7,000 (15)	6,966 (15)
CI Lower limit	4,417	-3,181	1,738	1,629	1,383
CI Upper limit	24,427	25,390	24,319	24,656	24,299
Prob C-E	0.99	0.90 (9)	0.97 (2)	0.97 (2)	0.97 (2)
ICER	-235,472	-134,488 (43)	-158,857 (33)	-156,289 (34)	-155,935 (34)
50% Missing data		CCA	MI-PMM	Log-MI-PMM	Two-step MI
CEA analysis	RA	(% Bias from RA)	(% Bias from RA)	(% Bias from RA)	(% Bias from RA)
Correlation between utility and costs	0.0507	0.0223	0.0433	0.0436	0.0406
Covariance	5.6	5.5	7.0	7.0	6.7
INB	14,422	8,786 (39)	12,682 (12)	12,867 (11)	11,907 (17)
Standard error of INB	6,083	9,584 (58)	6,858 (13)	6,770 (11)	6,962 (14)
CI Lower limit	4,417	-6,978	1,401	1,731	456
CI Upper limit	24,427	24,551	23,962	24,003	23,358
Prob C-E	0.99	0.82 (17)	0.97 (2)	0.97 (2)	0.96 (3)
ICER	-235,472	-155,561 (34)	-134,979 (43)	-144,317 (39)	-128,670 (45)

4. DISCUSSION

4.1 Main findings

In this study, we evaluated the performance of different multiple imputation strategies and CCA for scenarios with varying rates of missing data in costs and effects in a pragmatic economic evaluation. We found that for all rates of missing data, multiple imputation strategies performed better than CCA. The results of the CCA, MI-PMM and the two-step MI were all influenced by the amount of missing data. With a larger amount of missing data, the Log MI-PMM deviated the least from the RA for the cost difference, cost standard error, INB and its standard error, and the probability that the co-prescribed heroin treatment was cost effective in comparison with methadone maintenance at a willingness to pay of 30,000 euros per QALY. Therefore, the Log MI-PMM is considered most appropriate to impute missing cost and effect data. However, when considering QALYs the MI-PMM performed best since it deviated the least from the RA with increasing amounts of missing data. Overall, the Log-MI-PMM was least affected by the amount of missing data.

Our results imply that addressing only the right-skewness of the data by using a log transformation in combination with PMM is enough and that strategies to deal with zero inflation such as our two-step PMM are not needed. Beforehand, we expected that the two-step MI strategy would have performed better because it controls for the large amount of zeroes and the skewness in the data. However, in practice there were no relevant differences with the other multiple imputation strategies and the two-step MI was more difficult to apply than the Log MI-PMM. Not all software packages have incorporated a comprehensive way to apply the two-step MI strategy, whereas the Log MI-PMM is easily applied and available in software packages like SPSS, Stata, SAS and R.

4.2 Comparison with existing literature

Our study adds to the findings from other studies that multiple imputation is better than CCA to deal with missing data in economic evaluations (2,3,23,24). However, in contrast to Briggs et al(3), Oostenbrink et al(2) and Burton et al. (23), we had information on the observed values of the missing data, because we created the missing data ourselves using the MAR assumption. This allowed us to estimate the deviation of the different imputation strategies from the original complete dataset.

Yu et al. (24) showed in a simulation study that predictive mean matching in R and STATA performed reasonably well and maintained the underlying distribution of the resource use data (24). However, they did not evaluate the effect of the different imputation strategies on the cost-effectiveness estimates.

4.3 Strengths and limitations

Our study adds to previous studies by focussing on estimation of both incomplete costs, QALYs and cost-effectiveness and by comparing different MI strategies using the MICE (pmm) in STATA. Additionally we use a correlation after multiple imputation between costs

and utilities using Fisher's Z transformation to calculate the cost-effectiveness (5,17). We used the fully conditional specification with PMM which gave us more flexibility around assumptions of normality (25).

Other strengths of this study were its systematic and applied approach using real data to examine the performance of different multiple imputation strategies in situations with varying amounts of missing data. To our knowledge, this is one of the first studies to compare the two-step MI strategy with other multiple imputation strategies for cost-effectiveness evaluations.

As we used only one dataset we were limited in our evaluation parameters for direct comparisons to the true coefficients instead of averages over simulations. We did perform a small simulation pilot study repeating the imputation procedures to verify the stability of the methods. This was done by repeatedly drawing samples of 100 cases from each of our incomplete datasets and applying our method to these small samples. We simulated 1000 times and used 15 imputations and 20 iterations. For each method and incomplete data condition the average over the 1000 simulations was taken and compared to the complete reference data results. This simulation confirmed the relative differences between the performances of the methods presented in this study. Future research should perform a larger simulation study and to vary the proportion of zeroes in a future study to see how that affects the performance of the missing data methods. It is possible that with a greater amount of zeroes the two part model becomes more beneficial over the other methods. We assumed the same missing mechanism in both treatment arms, and future simulations this probably should be changed for some simulated data.

4.4 Implications for further research

Prospective economic evaluations alongside trials play an important role in providing decision makers cost-effectiveness information to inform reimbursement decisions. It is important that economic evaluations provide robust and unbiased information. The consequences of using different imputation strategies can affect policy decisions. In this study, we considered co-prescribed heroin treatment to be cost-effective in comparison with methadone maintenance in all strategies evaluated although the uncertainty increased. The decision may change depending on the imputation procedure chosen in situations with smaller differences between groups.

In conclusion, we recommend the use of the Log MI-PMM because of its ease to use and its reliable results with increased amounts of missing data. Log MI-PMM also appears to perform well for zero-inflated data providing a constant is used in place of the zero in the data.

REFERENCES

1. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*. 2009;338:b2393.
2. Oostenbrink JB, Al MJ, Rutten-van Molken MP. Methods to analyse cost data of patients who withdraw in a clinical trial setting. *Pharmacoeconomics*. 2003;21(15):1103-1112.
3. Briggs A, Clark T, Wolstenholme J, Clarke P. Missing... presumed at random: cost-analysis of incomplete data. *Health Econ*. May 2003;12(5):377-392.
4. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*. Feb 20 2011;30(4):377-399.
5. van Buuren S. *Flexible Imputation of Missing data*. New York: Chapman & Hall/CRC; 2012.
6. Klebanoff MA, Cole SR. Use of multiple imputation in the epidemiologic literature. *Am J Epidemiol*. Aug 15 2008;168(4):355-357.
7. Nietert PJ, Wahlquist AE, Herbert TL. Characteristics of recent biostatistical methods adopted by researchers publishing in general/internal medicine journals. *Stat Med*. Jan 15 2013;32(1):1-10.
8. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol*. Mar 1 2010;171(5):624-632.
9. Javaras KN, Van Dyk DA. Multiple imputation for incomplete data with semicontinuous variables. *Journal of the American Statistical Association*. Sep 2003;98(463):703-715.
10. Dijkgraaf MG, van der Zanden BP, de Borgie CA, Blanken P, van Ree JM, van den Brink W. Cost utility analysis of co-prescribed heroin compared with methadone maintenance treatment in heroin addicts in two randomised trials. *Bmj*. Jun 4 2005;330(7503):1297.
11. Brooks R. EuroQol: the current state of play. *Health Policy*. Jul 1996;37(1):53-72.
12. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. Nov 1997;35(11):1095-1108.
13. *R: A language and environment for statistical computing*. [computer program]. Vienna, Austria 2008.
14. Horton NJ, Lipsitz SR. Multiple Imputation in Practice: Comparison of Software Packages for Regression Models with Missing Variables. *The American Statistician*. 2001;55(3):244-254.
15. Enders CK. *Applied Missing Data Analysis*. New York Guilford Press; 2010.
16. Rubin DB. Inference and Missing Data. *Biometrika*. 1976;63(3):581-590.
17. Schafer JL. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall; 1997.
18. Nixon RM, Wonderling D, Grieve RD. Non-parametric methods for cost-effectiveness analysis: the central limit theorem and the bootstrap compared. *Health Econ*. Mar 2010;19(3):316-333.

19. Willan AR, Briggs AH. *Statistical Analysis of Cost-effectiveness Data*: John Wiley & Sons Ltd; 2006.
20. NICE. Measuring effectiveness and cost effectiveness: the QALY. 2010; <http://www.nice.org.uk/newsroom/features/measuringeffectivenessandcosteffectivenessstheqaly.jsp>. Accessed 26.09.2013, 2013.
21. Zwaanswijk M, van der Ende J, Verhaak PF, Bensing JM, Verhulst FC. Help-seeking for child psychopathology: pathways to informal and professional services in the Netherlands. *J Am Acad Child Adolesc Psychiatry*. Dec 2005;44(12):1292-1300.
22. Zwaanswijk M, Verhaak PF, van der Ende J, Bensing JM, Verhulst FC. Consultation for and identification of child and adolescent psychological problems in Dutch general practice. *Fam Pract*. Oct 2005;22(5):498-506.
23. Burton A, Billingham LJ, Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. *Clin Trials*. 2007;4(2):154-161.
24. Yu LM, Burton A, Rivero-Arias O. Evaluation of software for multiple imputation of semi-continuous data. *Stat Methods Med Res*. 2007;16(3):243-258.
25. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*. Jun 2007;16(3):219-242.