Comparing Methods for Handling Missing Cost and Outcome Data in Clinical Trial-based Cost-effectiveness Analysis

Modou Diop (diop@go.ugr.es)  
University of Granada

David Epstein  
University of Granada

Research Article

Keywords: Longitudinal missing outcome, repeated measure, mixed model, multiple imputation, complete-case-analysis, Bayesian parametric approach, cost-effectiveness analysis

DOI: https://doi.org/10.21203/rs.3.rs-479996/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Comparing methods for handling missing cost and outcome data in clinical trial-based cost-effectiveness analysis

Authors

Modou Diop (MD)
Research Fellow
Department of Applied Economics
University of Granada
Campus de Cartuja, 18071
GRANADA
Spain
Telephone: +34 958 242826
Email diop@go.ugr.es, Corresponding Author
Orcid: 0000-0003-0934-4005

David Epstein (PhD)
Department of Applied Economics
University of Granada
Campus de Cartuja, 18071
GRANADA
Spain
Telephone: +34 958 242826
Email davidepstein@go.ugr.es

ACKNOWLEDGEMENTS: We are grateful to the EVRA trial members for the availability of the data: Manjit S. Gohel, MD, Cambridge University Hospitals NHS Foundation Trust, United Kingdom, Jocelyn Mora, MSc, Francine Heatley, BSc, Alun H. Davies, DSc, Department of Surgery and Cancer, Imperial College London, United Kingdom; Matyas Szigeti, MSc, Jane Warwick, PhD, Imperial Clinical Trials Unit, School of Public Health, Imperial College London, United Kingdom; Andrew Bradbury, MD, University of Birmingham, United Kingdom; Richard Bulbulia, MD, Keith R. Poskitt, MD, Gloucestershire Hospitals National Health Service Foundation Trust, United Kingdom; Nicky Cullum, PhD, University of Manchester & Manchester University National Health Service Foundation Trust, United Kingdom; Isaac Nyamekye, MD, Worcestershire Acute Hospitals National Health Service Trust, United Kingdom; Sophie Renton, MS, North West London Hospitals National Health Service Trust, United Kingdom.

ADDITIONAL CONTRIBUTION: We thank the patient focus group who helped ascertain the importance of the EVRA research question, identify the most important outcome measures, and confirm the acceptability of trial interventions and follow-up protocols. As the Early Venous Reflux Ablation Trial Group, we thank the National Health Service centers and participating principal investigators and their colleagues for recruiting and monitoring trial participants: Addenbrooke’s Hospital: Manjit Gohel, MD, D. Read, P. Hayes, S. Hargreaves, K. Dhillon, M. Anwar, A. Liddle, and H. Brown; Bradford Royal Infirmary, Bradford: K. Mercer, F. Gill, A. Liu, W. Jeppson, W. Cowen, W. Rafferty, and K. Storton. Charing Cross & St Mary’s Hospitals, London: A.H. Davies, K. Dhillon, R. Kaur, E. Solomon, K. Sritharan, R. Velineni, C. S. Lim, A. Busuttil, R. Booton, C. Bicknell, M. Jenkins, T. Lane, and E. Serjeant. Cheltenham General Hospital: K. Poskitt, R. Bulbulia, J. Walordn, G. Wolfrey, F. Slim, C. Davies, L. Emerson, M. Grasty, M. Wymans, C. Wakeley, A. Cooper, J. Clapp, N. Hogg, J. Howard, J. Dyer, S. Lyes, D. Teemul, K. Harvey, M. Pride, A. Kindon, H. Price, L. Flemming, G. Birch, H. Holmes, and J. Weston. Cumberland Infirmary: T. Joseph, R. Eiffel, T. Ojimba, T. Wilson, A. Hodgson, L. Robinson, J. Todhunter, D. Heagerty, A. Mckeane, and R. McCarthy. Derriford Hospital, Plymouth: J. Barwell, C. Northcott, A. Elstone, and C. West. Frimley Park Hospital: P. Chong, D. Gerrard, A. Croucher, S. Levy, C. Martin, and T. Craig. Hull Royal Infirmary: D. Carradice, A. Firth, E. Clarke, A. Oswald, J. Sinclair, I. Chetter, J. El-Sheikha, S. Sandhra, C. Leung, and J. Hatfield. Leeds General Infirmary: J. Scott, N. Dewhirst, J. Woods, D. Russell, R. Darwood, M. Troxler, J. Thackeray, D. Bell, D.
Watson, L. Williamson, and M. Todd. Musgrove Park Hospital, Taunton: J. Coulston, P. Eyers, K. Darvall, I. Hunter, A. Stewart, A. Moss, J. Rewbury, C. Adams, L. Vickery, L. Foote, H. Durman, F. Venn, P. Hill, K. James, F. Luxton, D. Greenwell, K. Roberts, S. Mitchell, M. Tate, and H. Mills. New Cross Hospital, Wolverhampton: A. Garnham, D. McIntosh, M. Green, K. Collins, J. Rankin, P. Poulton, V. Isgar, and S. Hobbs. Northwick Park Hospital, Harrow: S. Renton, K. Dhillon, M. Trivedi, M. Kafeza, S. Parsapour, H. Moore, M. Najem, S. Connarty, H. Albon, C. Lloyd, J. Trant, and S. Chhabra. Queen Elizabeth Hospital, Birmingham: R. Vohra, J. McCormack, J. Marshall, V. Hardy, R. Rogoveanu, W. Goff, and D. Gardiner. Russells Hall Hospital, Dudley: A. Garnham, R. Gidda, S. Merotra, S. Shiralkar, A. Jayatunga, R. Pathak, A. Rehman, K. Randhawa, J. Lewis, S. Fullwood, S. Jennings, S. Cole, and M. Wall. Salisbury District Hospital: C. Ranaboldo, S. Hulin, C. Clarke, R. Fennelly, R. Cooper, R. Boyes, C. Draper, L. Harris, and D. Mead. Solihull Hospital (part of the University Hospitals Birmingham National Health System Foundation Trust): A. Bradbury, L. Kelly, G. Bate, H. Davies, M. Popplewell, M. Claridge, M. Gannon, H. Khaira, M. Scriven, T. Wilmink, D. Adam, and H. Nasr. Northern General Hospital, Sheffield: D. Dodd, S. Nawaz, J. Humphreys, M. Barnes, J. Sorrell, D. Swift, P. Phillips, H. Trender, N. Fenwick, H. Newell, and C. Mason. Royal Bournemouth General Hospital: D. Rittoo, S. Baker, R. Mitchell, S. Andrews, S. Williams, J. Stephenson, and L. Vamplew. Worcester Royal Hospital: I. Nyamekye, S. Holloway, W. Hayes, J. Day, C. Clayton, and D. Harding. York Hospital: A. Thompson, A. Gibson, Z. Murphy, T. Smith, and J. Whitwell. We thank members of our 2 oversight committees, the Trial Steering Committee (Professor Julie Brittenden [Chair]; Miss Rebecca Jane Winterborn [Consultant Vascular Surgeon]; Professor Andrea Nelson [Head of School and Professor of Wound Healing]; Dr Richard Haynes [Research Fellow and Honorary Consultant Nephrologist] and Mr Bruce Ley-Greaves [lay member] who provided invaluable input and advice as the independent lay member over the course of the study) and the Data Monitoring Committee (Professor Gerard Stansby [Chair, Professor of Vascular Surgery]; Professor Frank Smith [Professor of Vascular Surgery & Surgical Education]; Professor Marcus Flather [Professor of Medicine - Clinical Trials]; Dr Ian Nunney [Medical Statistician]) for their support and guidance. There was no financial compensation for these contributions.
Comparing methods for handling missing cost and outcome data in clinical trial-based cost-effectiveness analysis

ABSTRACT

OBJECTIVES: This study compares methods for handling missing data to conduct cost-effectiveness analysis in the context of a clinical study.

METHODS: A long-term clinical trial with staggered recruitment was used as a case study (EVRA, NIHR-HTA project 11/129/197). Patients had between 1 year and 5.5 years (median 3 years) of follow-up under “early” or “deferred” treatment. The methods compared were Complete-Case-Analysis (CCA), multiple imputation using linear regression (MILR) and using predictive mean matching (MIPMM), Bayesian parametric approach using the R package missingHE (BPA) and repeated measures mixed model (RMM). The outcomes were total mean costs and total mean quality-adjusted life years (QALYs) at different time horizons (1 year, 3 years and 5 years).

RESULTS: All methods found no statistically significant difference in cost at the 5% level in all years, and all methods found statistically significantly greater mean QALY at year 1. At year 3, BPA showed a statistically significant difference while other method did not at later time horizons. Standard errors differed substantially between the methods employed.

CONCLUSION: MIPMM seemed to perform better than MILM, confirming findings from simulation studies. RMM and BPA might be feasible options, though they did not perform better than MIPMM in this dataset. Further simulation studies and applications should continue to compare these methods.

KEYWORDS

Longitudinal missing outcome, repeated measure, mixed model, multiple imputation, complete-case-analysis, Bayesian parametric approach, cost-effectiveness analysis
Comparing methods for handling missing cost and outcome data in clinical trial-based cost-effectiveness analysis

1. Introduction.

Missing data occurs when one or all variables are missing for a given subject. This often occurs in longitudinal studies and can particularly be a problem in within-study cost-effectiveness analysis (CEA) because accurate estimates of total mean cost and quality-adjusted life years require full data to be collected on each subject at each follow-up time point (1–3).

Handling missing resource use and outcomes requires assumptions about the missing data mechanism, specifically, the factors that might have influenced data being missing and whether individuals with missing data are similar to, or in some way different from, those with fully observed data. An inappropriate method can bias the results, make inefficient use of the data available or reduce the overall statistical power of the study (4–6). Cost-effectiveness analysis adds extra dimensions that need to be considered. First, because total costs and total quality-adjusted life years (QALYs) are a weighted sum of multiple item measurements at each follow-up, estimation of these variables over the total time horizon for any subject requires complete data over the full follow-up, so methods which discard subjects with any missing data or do not make use of all the observations are highly inefficient. Second, any analytic method has to take account that there are two dependent variables (costs and QALYs) that will have a correlated, bivariate distribution.

The purpose of this study is to compare methods for handling missing data in cost-effectiveness analysis, using as a case-study a 5-year clinical trial: The Early Endovenous Ablation in Venous Ulceration (EVRA) (7). With a given dataset, there are several ways in which the chosen missing data approach might influence the results: different subjects used in the analysis, different number of observations used per subject, different statistical models of the missing data mechanism and the latent correlation between observed and missing observations, or different estimation model to estimate total mean costs and QALYs and the correlation between them. The aim of this study is to compare results using different
methods for handling missing data. The dataset provides an interesting case study because despite very complete overall follow-up of patients, there was considerable item missingness arising from the way the trial was designed. In each case we use a common estimation model for costs and QALYs, and standard off-the-shelf software. Previous tutorial papers (1–3) compared case-analysis (CCA) (8–10), multiple Imputation (MI) (8,10–12) and repeated measure mixed model (RMM) (13). This study extends this previous methodological work by including a Bayesian parametric approach using the R package missingHE (BPA) (14) and examining the results over different time horizons (and hence with different quantities of missing data) of 1, 3 and 5 years. In each case we calculate the mean incremental total cost and QALY, standard errors, the incremental cost-effectiveness ratio (ICER) and the cost-effectiveness acceptability curve (CEAC).

The key assumption for CCA to be unbiased is that individuals with complete data are representative of those with missing data, conditional on the variables included in the analysis model. This assumption is known as “missing completely at random” (MCAR). Furthermore, CCA is inefficient in studies with more than one follow-up because all the information from individuals with at least one assessment missing is discarded (list-wise deletion). Also, CCA is not ‘intention-to-treat’ because some randomised patients with follow-up data are excluded (1,15). MI operates within the more realistic “missing at random” (MAR) framework and is usually fairly straightforward to implement using standard software. Furthermore, it allows considerable flexibility about how the posterior distributions of missing data are specified, including semi-parametric (posterior mean matching, PMM) approaches, and separates the imputation of missing data from the estimation of the analysis model, which helps practitioners understand the intuition behind the approach. However, some have criticised the rather “ad-hoc” approach of MI and argue that this method does not account for the complexities of the data or the time dependence between the responses and does not correspond to a valid probabilistic model (16). BPA using the package missingHE attempt to provide a more explicit mathematical framework for making inferences from incomplete data, i.e., it presumes that assumption about missing data as well as the uncertainty it carries out can be made explicit through prior distributions(17). The package missingHE also provides models to explore missing not at random (MNAR) situations. Nevertheless, the use of missingHE is still not frequent in CEA due to the unfamiliarity of software among practitioners and the reliance on a limited range of parametric models. MI replaces each missing observation with a set of plausible imputed (predicted) values, drawn from a posterior predictive distribution of the missing data given the observed
data. BPA has some similarities with MI, but does not explicitly create multiple datasets, but instead treats missing items as if they were parameters and estimates posterior distributions directly for each of them.

RMM, also known as a variance-components models, random effects model, hierarchical linear model, or multi-level model, relies on maximum likelihood to estimate the covariance within subject, and if specified correctly are valid for data that are MAR. In the mixed model, information from observed data is used to implicitly impute the missing data (18). In this paper we use the parameters estimated from this model to calculate total incremental mean costs and total mean QALY for each treatment group. This is achieved by dividing the follow-up into a series of discrete time periods, estimating the mean incremental cost per person (treatment-by-time interaction) for each time period using the mixed model, and then estimating total mean incremental cost as the sum of these coefficients over the whole-time horizon. The process is repeated in the same way to estimate total mean incremental QALY. The correlation between total mean incremental cost and total mean incremental QALY is preserved by repeatedly running the routines using resampling with replacement (bootstrap). In this paper, we assume the simplest form of RMM, with linear functional form and unstructured covariance. More complex error structures and functional forms are possible, (19–25) but are not explored here. RMM and related approaches such as Generalized Estimating Equations are widely used for unbalanced longitudinal data but, with few exceptions (13), have not been widely used for cost-effectiveness analysis.

The methods differ in whether the data are input to the model in the ‘wide’ or ‘long’ formats, and hence the number of outcome observations employed. CCA uses data in the wide format (one cost outcome variable per subject and one QALY outcome variable), and employs N – M observations, where N is the number of randomised subjects and M is the number with any missing data. BPA based on missingHE also uses the wide format and employs N cost outcome variables and N QALY outcomes. Missing outcome variables are treated as parameters to estimate. Period costs and EQ-5D values at other time points are included as covariates in the BPA. MI can employ either the wide or long format. Crucially, outcome variables for all time points for all subjects are used in the missing data model. Hence the number of outcome observations for costs and QALYs is \( \sum_{i=1}^{N} T_i \) where \( T_i \) is the number of follow-up time points for subject i. The mean cost and QALY for each subject i over the required time horizon is then imputed passively in each imputed dataset. RMM is a hierarchical combined statistical and
estimation model which also uses $\sum_{i=1}^{N} T_i$ observations. As RMM only consists of an analysis model, with no separate missing data model, total mean costs and QALYs are not estimated for each individual. Instead, the incremental mean costs and QALYs (conditioned on the baseline covariates in the model) are predicted aggregately. See Supplementary Material for further explanation. Table 1 summarises the approaches.

Table 1 (summary of the approaches) about here

2. Methods

2.1 Data

The Early Endovenous Ablation in Venous Ulceration (EVRA) randomised clinical trial evaluated the cost-effectiveness of “early” versus “deferred” endovenous ablation to treat venous leg ulcers. The trial methods and patients are described elsewhere (7). Briefly, resource use items in hospital, primary and community care and medications related to the treatment of venous ulceration, adverse events or complications were collected by case note review and questionnaires completed at baseline and monthly thereafter up to one year, plus one further telephone follow up between October 2018 and March 2019.

The baseline covariates included in all the estimation models were: TREAT is treatment randomised (“early” coded as 1 or “delayed” coded as zero). The variable WEEK_t is the time variable (coded as a set of categorical (dummy or factor) variables) representing the week after randomisation at which data are observed, from t=0 (baseline) to t=16 (week 260). SIZE, AGE and DURATION are the ulcer size (cm²), subject’s age (years) and length of time with ulcer (years), respectively, measured at baseline and centred at the means. SITE was coded as a factor variable.

Each item of resource use was multiplied by UK unit costs obtained from published literature, NHS reference costs, and manufacturers’ list prices to calculate overall costs within each of these categories for each patient (7). The costs for each individual over their follow-up (from randomization to date of censoring for that individual) were assigned or apportioned into discrete time periods, that corresponded to 12 monthly periods during the first year (as follow-ups were monthly) and then yearly periods
thereafter. This allowed discounting to be applied (3.5% per year), and facilitated analysis using the MI and mixed model in long format (see below).

EQ-5D-5L was collected at baseline, 6 weeks, 6 months, 12 months, plus one further telephone follow up between October 2018 and March 2019, and a utility index was calculated at each time point using a published tariff (26). Due to rigorous follow-up procedures, there were very few withdrawals or dropouts from the study. Nevertheless, due to the study design, data were not observed for every subject for every period (unbalanced panel) for two reasons. First, patients were followed over different time horizons (from 3 years to 5 years) due to staggered recruitment into the study. Second, to keep the cost of the research study low, quality of life was only collected at one time point after the first year.

Patients who died during the study were assigned zero costs and HRQOL thereafter. Code and example data are available in the Supplementary data, [http://dx.doi.org/10.17632/j8fmdwd4jp.5](http://dx.doi.org/10.17632/j8fmdwd4jp.5).

### 2.2 Examination of pattern of missing data and hypothesis about missing data mechanism

Data are incomplete in this study for three reasons. First, recruitment of the 450 patients into the clinical study across the 20 vascular centres took place between October 2013 and September 2016. The study finalised on March 2019. This “staggered” recruitment into the trial meant that patients had a minimum of 1 years of follow-up and a maximum of 5.5 years (median 3 years). Incomplete data arising from “administrative censoring” is commonly considered independent of other variables (MCAR) (27).

Figure 1 about here (censoring pattern)

Second, all patients had regular and periodically scheduled follow-up during the first year after recruitment, but only one follow-up after one year. Figure 1 shows how this study design influences the missing data pattern. A patient recruited in 2014 will have complete follow-up during the first year, missing data at years 2, 3 and 4, and one follow-up at 5 years (patient A). A patient recruited in 2015 will have complete follow-up during the first year, missing data at years 2 and 3, one follow-up at year 4, and missing data for year 5. A patient recruited in 2016 (patient C) will have complete follow-up during the first year, missing data at year 2, one follow-up at year 3, and missing data for years 4 and 5. As before, it would be reasonable to consider this pattern of missing data to be MCAR, as the choice of date of final
follow-up was independent of other variables. This mainly affected collection of EQ-5D, because in the absence of questionnaire data, most types of resource use could be obtained from case-notes.

Third, a relatively small amount of data were incomplete due to early dropout and item missingness (such as failure to complete an interim questionnaire). These data are usually considered MAR.

The pattern of missingness was examined using descriptive statistics and via the linear logistic model (Equation 1). This can distinguish MAR from MCAR, although cannot rule out the possibility that data are missing not at random (MNAR).

\[
\logit(\pi_{it}) = \gamma_1 TRT_i + \gamma_2 DURATION_i + \gamma_3 AGE_i + \gamma_4 SIZE_i + \gamma_5 Site_i + \gamma_6 WEEK_i \tag{1}
\]

where \( \pi \) denotes the probability that an observation is missing in individual \( i \) at time \( t \).

2.3 Repeated measure mixed model framework

We considered a RMM where the outcome could vary over subject, centre, and time point, after controlling for covariates. A two-level variance component model was specified as Equation 2 (long format data).

\[
Y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 DURATION_i + \beta_3 AGE_i + \beta_4 SIZE_i + \beta_5 Site_i + \beta_6 WEEK_i + \delta TREAT_i \ast WEEK_i + \zeta_i + \epsilon_{it} \tag{2}
\]

\( Y_{it} \) is the outcome variable (one model for costs during each period \( t \) and another for EQ-5D tariff at the end of each period \( t \) for each subject \( i \) at time point \( t \). Hence for the model where the dependent variable is cost, \( Y_{i0} \) is set to be zero for all subjects, \( Y_{i2} \) is the cost for patient \( i \) during the first 4 weeks \( Y_{i1} \) is the cost between the 4th and the 8th week, and so on up to \( Y_{i12} \) (week 52). After that, the periods are set to be yearly, so that \( Y_{i13} \) is the cost between week 52 and week 104 (year 2), and so on up to \( Y_{i16} \) (year 5 or week 260). \( \zeta_i \) is the random deviation of subject \( i \)'s mean costs or EQ-5D tariff from the overall mean \( \beta_0 \) and \( \epsilon_{it} \), often called within-subject residual across time, is the random deviation of \( Y_{it} \) from subject \( i \)'s mean costs or EQ-5D tariff (19,28).

The model estimates of the \( \delta \) are a (vector of) coefficients for the interactions between treatment assignment and period number and hence represents the mean incremental cost of early treatment (versus...
delayed) during period t (in the cost model) or the mean incremental EQ-5D tariff at follow-up time point t (in the EQ-5D model). These analyses were implemented using the mixed command in STATA 15. To estimate total mean incremental cost per patient over a desired time horizon (e.g., 3 years), the relevant period coefficients are simply added up (lincom). Thus, for example, where the dependent variable is cost accrued during the preceding period, and $\hat{\delta}_1$ is the time*treatment interaction coefficient at 4 weeks (~month 1), $\hat{\delta}_2$ at 8 weeks (~month 2), and $\hat{\delta}_3$ at 13 weeks (month 3), then the difference in total mean incremental cost over the first 3 months is $\hat{\delta}_1 + \hat{\delta}_2 + \hat{\delta}_3$. To estimate mean total incremental QALY over a given time horizon, the “area under the curve” applying the trapezium rule is calculated. Hence, using the coefficients from the EQ-5D model over the first 3 months (where $\hat{\beta}_1$ is the difference in EQ-5D at baseline, $\hat{\delta}_1$ at 4 weeks and $\hat{\delta}_2$ at 3 months), the estimated mean total incremental QALY over the first 3 months would be $0.5 \times ((\hat{\beta}_1 + \hat{\delta}_1) \times \frac{4}{52} + (\hat{\delta}_1 + \hat{\delta}_2) \times \frac{9}{52})$. Uncertainty was estimated by bootstrapping incremental mean costs and QALYs (29) and shown by the CEAC.

2.4 Multiple imputation model

MI consists of 3 steps:

1. M imputations (completed datasets) are generated under some chosen imputation model replacing missing values with “plausible” substitutes, based on distribution of the observed data. This process is known as proper imputation because it incorporates all sources of variability and uncertainty in the imputed values, including prediction errors of the individual values and errors of estimation in the fitted coefficients of the imputation model.

2. Completed-data analysis (estimation) step. The desired analysis is performed separately and identically on each imputation $m = 1, \ldots, M$. The analysis task can be done with usual complete-data method, giving us $m$ estimates of the original quantity of interest, $Q_1, \ldots, Q_m$, with their variance–covariance matrices. Results from the $m$ analyses differ because the missing values have been replaced by different imputations.

3. Pooling step. The results obtained from M completed-data analyses are combined into a single multiple-imputation result using Rubin’s rules (30). The pooled estimated mean of a given coefficient is $Q^* = \frac{1}{M} \Sigma_{j=1}^M Q_j$. The combined variance–covariance matrix incorporates both
within-imputation variability (uncertainty about the results from one imputed data set) and between-imputation variability (reflecting the uncertainty due to the missing information).

The same long-format data were used in RMM and MI. Analyses were implemented using the mi suite of commands in STATA 15. Step 1 was performed by multivariate imputation by chained equation (MICE) (also known as fully conditional specification (FCS) (31) or sequential regression multivariate imputation(32)) which is a practical approach to generating imputations based on a set of inter-linked imputation models. The process begins by choosing the first variable to impute, say costs in the first period ($Y_1$). Values for all other variables (both EQ5D at each follow up and period costs) to be imputed are then filled in using a simple rule, such as simple random sampling with replacement from the observed values. Then, $Y_1$ is regressed on all other variables and baseline covariates, and then missing values for $Y_1$ are replaced by simulated draws from the corresponding posterior predictive distribution of $Y_1$. Then, the process is repeated for the next variable (e.g., $Y_2$), which is regressed on all other variables and using the newly imputed values in $Y_1$. Again, missing values in $Y_2$ are replaced by draws from the posterior predictive distribution of $Y_2$. The process is repeated for all other variables with missing values in turn: this is called a cycle. In order to stabilize the results, the procedure was repeated for 20 cycles to produce a single imputed data set, and the whole procedure is repeated M times to give M imputed data sets (33–36). To avoid bias in the analysis model, all covariates that are in the analysis model were included in the imputation model (37).

A second method for MI, predictive mean matching (PMM) was also used. PMM combines the standard linear regression and the closest-neighbour imputation approaches, using the normal linear regression to obtain linear prediction. It then uses the linear predictions as a distance measure to form the set of the nearest neighbours (suitable “donor”) consisting of the complete value. Finally, it randomly draws an imputed value from this set. By drawing from the observed data, PMM preserves the distribution of the observed values in the missing part of the data which makes it more robust than the fully parametric linear approach (38). Possible donors were set with 10 closest neighbours as suggested in Morris et al., (39). M=40 imputations were performed (40). Statistical efficiency was assessed by FMI/M, where FMI is the fraction of missing information. Values of FMI range between 0 and 1. A large FMI (close to 1) indicates high variability between imputed data sets; that is, the observed data in the imputation model do not provide much information about the missing values. Monte Carlo Errors (MCE) were calculated to
indicate the stability of the model. FMI and MCE reflect the variability of MI results across repeated uses of the same imputation procedure and are useful for determining an adequate number of imputations to obtain stable MI results (12).

For each of the m complete datasets, total cost and total QALY over 1 year, 3 years and 5 years for each subject were imputed passively using the same formulas given in the section for RMM. The difference between RMM and MI being that in the RMM approach, estimates of total mean cost and QALY for the group as a whole are made by linear combination (\texttt{lincom}) of the coefficients, while MI imputes a total cost and QALY for each subject, and then proceeds to estimate mean incremental cost and QALYs for the group as a whole using bivariate normal regression (\texttt{sureg} in STATA 15). Coefficients from this regression were then combined across the multiple imputed datasets using Rubin's rules (\texttt{mi estimate}) [34]. The bootstrap was not used with MI as bootstrap combined with multiple imputation can be very complex (41). The CEAC was calculated parametrically from the coefficients and covariance matrix of the bivariate normal regression.

2.5 Complete case analysis.

This method assumes covariate dependent MCAR (missing completely at random) (1). Total cost and total QALY were calculated for each individual \( i \) over the relevant time horizon \( T \) (1, 3 or 5 years) (Equation 3). Any subject with a missing period cost or EQ-5D prior to the relevant time horizon was dropped (as total cost and total QALY for individual \( i \) at time \( T \) cannot be calculated if any prior values up to \( T \) are missing). A bivariate normal regression was performed at each time horizon for total costs and total QALY (equation 3) with data in the wide format. The CEAC was calculated using the bootstrap (parametric estimates were also tried and made no noticeable difference to the results so are not reported).

\[
Y_i = \beta_0 + \beta_1eq5d0_i + \beta_2TREAT_i + \beta_3DURATION_i + \beta_4AGE_i + \beta_5SIZE_i + \beta_6SITE_i + \epsilon_i \quad (3)
\]

2.6 Bayesian Parametric Approach (BPA)

In principle BPA allows to fit the model of interest while simultaneously handling all the different issues related to the missing data as well as to correctly propagate and quantify uncertainty. Each unobserved quantity in the model is handled as if it were a parameter (17,42–44). Costs and QALYS were estimated through a joint distribution based on Markov Chain Monte Carlo (MCMC) using the R function
**selection**, within the missingHE package (17). BPA requires the specification of four models: the first two are the estimation models for the total QALY and total cost variables (\(Y\)) and the last two are the auxiliary models which are fitted (similarly to equation 1) to estimate the probability \(Y\) is missing using logistic regressions. The four models include covariates of treat, duration, size, age, site, and the two last model also include the length of follow-up, as the probability of missingness increases with time since baseline.

The missingness mechanism is assumed MAR and the distribution of costs and QALY is assumed normal. Non-informative priors were used for the precision of the dependent variables, which were varied from 0.001 to 0.01 in sensitivity analysis. Incremental mean costs and QALY were computed from the first two models and the CEAC was calculated parametrically.

### 2.7. Ethics and Consent

The trial was approved by the South West–Central Bristol Research Ethics Committee, and trial oversight was provided by an independent trial steering committee and an independent data and safety monitoring committee. Written informed consent was obtained from all participants. Details of the trial design and implementation are provided in the published protocol of the EVRA study (45). The study was conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly, Helsinki 1964 and later revisions.

### 3. Results

#### 3.1 Pattern of missingness

No baseline data were missing. 74% of subjects had complete data (costs and EQ-5D) at 1 year, 10% at year 3 and 25% at year 5 (Table 2). This pattern arises from the staggered recruitment and because the final questionnaire was administered at a fixed calendar point irrespective of when the subject was recruited.

[Table 2 about here (missing data pattern)]

The logistic model showed the probability that a value is missing in costs and EQ-5D are related to the length of follow-up (week), age at baseline and site (p < 0.0001), see supplementary table S1 and table S2. From this, we assume MAR. However, it cannot be ruled out that data might be MNAR.
Only subjects with complete data were used in CCA: year 1, n=338; year 3, n=44 and year 5, n=147. The BP included all the 450 subjects in the analysis, one observation (row) per subject. The data for RMM and MI included all the 6,861 available observations structured as an unbalanced panel, so that for any subject $i$ at time $t$ with missing dependent variable $Y_t$ (period cost or EQ-5D), data from earlier or later follow-up times informed an estimate of that value, either indirectly through the variance-covariance matrix (RMM) or directly through proper imputation (MI). The data for BPA were in the wide format ($N=450$) and included the period costs and EQ-5D at each follow-up time point as covariates in the analysis models.

3.2 Cost effectiveness analysis

Table 3 shows a summary of the results of the cost-effectiveness-analysis with the five different approaches at each time point. All methods agreed that there was no statistically significant difference in cost at the 5% level at any time horizon. Early intervention was associated with statistically significantly greater mean QALY among all methods at year 1. BPA showed a statistically significant difference at year 3, while other methods showed trend towards greater QALY for the intervention, but this did not reach statistical significance.

At 3 years early intervention dominated according to both RMM and BPA methods. The ICER according to CCA was £14/QALY; MI computed £489/QALY using PMM and £521/QALY using linear regression. All methods suggested that early intervention is cost-effective at a threshold of £20,000 per QALY at 1-, 3- and 5-year time horizons. At a threshold of £20,000/QALY, the estimated probability that the intervention was cost-effective was 93% using RMM and 53% using CCA, see Figure 2.

When we compare the two methods for multiple imputation, MIPMM show a loss of efficiency of 0.03% in costs using M=40 and 0.8% in QALY while MILR shows 0.20% and 1.3% for costs and QALY, respectively. MCE were less than 10% of the standard errors (SE) in both methods, indicating reasonable stability of the models. Imputations with MIPMM seemed to correspond more closely than MILR to the distribution of observed data (Supplementary Figure 1).
RMM showed greatest standard errors, SE (482) at year 1 for incremental mean costs than other methods (Figure 3a). CCA showed the greatest SE at year 3 and BPA at year 5, 769 and 807, respectively. MIPMM showed the lowest SE at all time horizons. Regarding QALY at year 1, CCA and BPA showed greater SE than other methods (Figure 3b). BPA presented the highest SE at year 3 and 5. Other methods showed similar SE for incremental mean QALY at years 1, 3 and 5.

[Figure 2 (ceac) and Figure 3 (SE) about here]

4. Discussion.

Staggered recruitment is a common problem in clinical trial and causes patients to have different lengths of follow-up, meaning that the dataset is an unbalanced panel with a considerable amount of missing data. Any chosen statistical model should aim to use data efficiently, be robust to potential sources of bias, and provide credible estimates of uncertainty.

This paper compared five methods for handling missing data empirically, some in common use and others less so, using a real data set with several follow-up points over a long time period. Since we do not know the true values of the missing data, this approach does not allow us to generalize about which method is “correct”. Nevertheless, based on the results in this paper, we can make some general comments about which methods might be more appropriate to deal with challenges similar to the ones presenting in this case study. We have attempted to use a similar estimation model in each case, so that differences arise mainly from the number of subjects and observations per subject that comprise the data, and the assumed latent correlation between observed and missing data.

CCA is the simplest method to implement. However, it is the least efficient as any subjects with any incomplete observations are discarded. This is seen in the large standard errors, particularly for mean costs. Furthermore, CCA relies on implausible assumptions. BPA offers a principled framework for handling missing data, and would be unbiased under the assumption of MAR. However, the implementation we have used here (the selection function in R) includes all the subjects but does not include the cost and QALY data at each time point for each subject as dependent variables. Hence selection is less efficient than MI or RMM, and this is carried forward into the high degree of uncertainty represented by the relatively large standard errors around mean costs and QALY, and the low CEAC probability compared with other methods.
MI and RMM employ all the available cost and EQ-5D observations in all the subjects, and so are efficient methods. This is important when there is substantial item missingness, as we have in this dataset. Both are straightforward to implement, though perhaps rather slow to compute. RMM would not be a suitable option if there were considerable missing baseline covariates (selection and CCA share this limitation). MIPMM seemed to perform somewhat better than MILR in terms of validation of imputed values against observed values in our dataset and estimated slightly lower standard errors, and greater efficiency measured by FMI/M. Some caution is needed when using MIPMM if there are few donors in the vicinity of an incomplete case, leading to a risk of bias (39). Also, if a donor is selected for many individuals or repeatedly used by the same individual across imputations this will lead to inefficiency, underestimating the between-imputation variance. RMM allows the analyst to compute bootstrap simulations of the analysis model, whereas MI (at least in the implementation we employ here) relies on parametric assumptions. Some analysts may favour semi-parametric methods such as bootstrap when data are not normally distributed. MI is limited in the estimation commands that are available in standard software, because of the need to combine using Rubin’s rules (mi estimate in STATA). For example, fitting random effects for the sites in the estimation model might be desirable but would be difficult to implement. Standard errors for RMM appeared considerably greater than for MIPMM, which leads us to conclude that MIPMM offers greatest precision in this dataset at least.

5. Conclusion

This study has empirically compared methods for handling missing data in cost-effectiveness analysis. Despite rigorous trial procedures for following up subjects by telephone and in person there was still considerable missing data arising from staggered recruitment and item missingness. The longitudinal dataset employed in the case study is likely to be fairly typical of the type of data encountered in other well-managed clinical trials. CCA and BPA did not use all the relevant outcome variables for each subject and hence can be considered relatively inefficient. MIPMM seemed to perform better than MILM, confirming findings from simulation studies(46). RMM seems to be a feasible option and has some points in its favour at least in theory, though it did not perform better than MIPMM in this dataset. Further simulation studies and applications should be undertaken to compare these methods.

LIST OF ABBREVIATIONS
RMM. repeated measure mixed model; CCA. complete-case-analysis; MIPMM. Multiple Imputation using predictive mean matching; MILR. Multiple Imputation using linear regression; BPA. Bayesian parametric approach; QALY. quality-adjusted life years; ICER. incremental cost ratio; CEAC. cost-effectiveness acceptability curve; SE. standard error; MAR. missing at random; MCAR. missing completely at random

DECLARATION

ETHICS: The trial was approved by the South West–Central Bristol Research Ethics Committee, and trial oversight was provided by an independent trial steering committee and an independent data and safety monitoring committee. The study was conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly, Helsinki 1964 and later revisions.

CONSENT: Written informed consent was obtained from all participants

AVAILABILITY OF DATA AND MATERIAL: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. Also, codes of STATA and R used in this study are available in Mendeley Data, V1, http://dx.doi.org/10.17632/j8fmdwd4jp.5.

CONFLICTS OF INTEREST/COMPETING INTERESTS: The authors declare they have no conflicts of interest.

FUNDING: This study was funded by the National Institute for Health Research (NIHR HTA) Programme (EVRA, project number 11/129/197).

AUTHORS CONTRIBUTION: All authors contributed to the study conception design and analysis, and read and approved the final manuscript.

6. References

1. Faria R, Gomes M, Epstein D, White IR. A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials. Pharmacoeconomics. 2014;32(12):1157–70.

2. Myers WR. Handling Missing Data in Clinical Trials: An Overview. Therapeutic Innovation & Regulatory Science. 2000;34(2):525–33.
3. Fitzmaurice G, Laird N, Ware J. Applied longitudinal Analysis (2nd Edition). Wiley. 2011;http://biosun1.harvard.edu/~fitzmaur/ala/.

4. Mavridis D, White IR, Higgins JPT, Cipriani A, Salanti G. Allowing for uncertainty due to missing continuous outcome data in pairwise and network meta-analysis. Statistics in Medicine. 2014;34(5):721–41.

5. Nakai M, Ke W. Review of the methods for handling missing data in longitudinal data analysis. International Journal of Mathematical Analysis. 2011;5(1–4):1–13.

6. Ibrahim JG, Chu H, Chen MH. Missing data in clinical studies: Issues and methods. Journal of Clinical Oncology. 2012;30(26):3297–303.

7. Gohel MS, Mora, MSc J, Szigeti M, Epstein DM, Heatley F, Bradbury A, et al. Long-term Clinical and Cost-effectiveness of Early Endovenous Ablation in Venous Ulceration. JAMA Surgery [Internet]. 2020 Dec 1;155(12):1113. Available from: https://doi.org/10.1001/jamasurg.2020.3845

8. Leurent B, Gomes M, Carpenter JR. Missing data in trial-based cost-effectiveness analysis: An incomplete journey. Health Economics (United Kingdom). 2018;27(6):1024–40.

9. Carroll OU, Morris TP, Keogh RH. How are missing data in covariates handled in observational time-to-event studies in oncology? A systematic review. BMC Medical Research Methodology. 2020;20(1):1–15.

10. Manca A, Palmer S. Handling missing data in patient-level cost-effectiveness analysis alongside randomised clinical trials. Applied Health Economics and Health Policy. 2005;4(2):65–75.

11. Hayati Rezvan P, Lee KJ, Simpson JA. The rise of multiple imputation: A review of the reporting and implementation of the method in medical research Data collection, quality, and reporting. BMC Medical Research Methodology. 2015;15(1):1–14.

12. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine. 2011;30(4):377–99.

13. Cox E, Saramago P, Kelly J, Porta N, Hall E, Tan WS, et al. Effects of Bladder Cancer on UK Healthcare Costs and Patient Health-Related Quality of Life: Evidence From the BOXIT Trial. Clinical Genitourinary Cancer [Internet]. 2020;18(4):e418–42. Available from: https://doi.org/10.1016/j.clgc.2019.12.004

14. Ma Z, Chen G. Bayesian methods for dealing with missing data problems. Journal of the Korean Statistical Society. 2018;47(3):297–313.

15. Patrick R, R. WI. Journal of Statistical Software Multiple Imputation by Chained Equations (MICE): Implementation in Stata. 2011;45(4). Available from: http://www.jstatsoft.org/

16. Gabrio A, Daniels MJ, Baio G. A Bayesian parametric approach to handle missing longitudinal outcome data in trial-based health economic evaluations. Journal of the Royal Statistical Society Series A: Statistics in Society. 2020;183(2):607–29.
17. Daniels MJ, Hogan JW. Missing data in longitudinal studies: strategies for Bayesian modeling and sensitivity analysis. Hogan JW, editor. Boca Raton: Chapman & Hall/CRC; 2008. (Monographs on statistics and applied probability; 109).
18. Bell ML, Fairclough DL. Statistical Methods in Medical Research. 2014;
19. Rabe-Hesketh S. Multilevel and longitudinal modeling using Stata. 2nd ed. Skrondal Anders, editor. College Station, Tex: Stata Press; 2008.
20. Diggle P, Kenward MG. Informative Drop-Out in Longitudinal Data Analysis. Applied Statistics. 1994;43(1):49.
21. Doornkamp JC. Linked references are available on JSTOR for this article: Modeling the Drop-Out Mechanism in Repeated-Measures Studies. The Geographical Journal. 1993;159(2):196–202.
22. Pigott TD. A review of methods for missing data. International Journal of Phytoremediation. 2001;21(1):353–83.
23. Verbeke G, Fieuws S, Molenberghs G, Davidian M. The analysis of multivariate longitudinal data: A review. Statistical Methods in Medical Research [Internet]. 2014 Feb 20;23(1):42–59. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2688697/pdf/nihms-110443.pdf
24. Daniels MJ, Zhao YD. Modelling the random effects covariance matrix in longitudinal data. Statistics in Medicine. 2003;22(10):1631–47.
25. Diggle PJ, editor. Analysis of longitudinal data. 2nd ed. ; Oxford: Oxford University Press; 2004. (Oxford statistical science series; 25).
26. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health [Internet]. 2012;15(5):708–15. Available from: http://dx.doi.org/10.1016/j.jval.2012.02.008
27. Deresa NW, van Keilegom I. A multivariate normal regression model for survival data subject to different types of dependent censoring. Computational Statistics and Data Analysis [Internet]. 2020;144:106879. Available from: https://doi.org/10.1016/j.csda.2019.106879
28. Monsalves MJ, Bangdiwala AS, Thabane A, Bangdiwala SI. LEVEL (Logical Explanations & Visualizations of Estimates in Linear mixed models): Recommendations for reporting multilevel data and analyses. BMC Medical Research Methodology. 2020;20(1):1–9.
29. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. Health Economics. 1997;6(4):327–40.
30. Campion WM, Rubin DB. Multiple Imputation for Nonresponse in Surveys. Journal of Marketing Research. 1989;26(4):485.
31. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in Medicine. 1999;18(6):681–94.
32. Raghunathan T, Lepkowski J, Hoewyk J, Solenberger P. A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. Survey Methodology. 2000 Nov 30;27.

33. Royston P. Multiple imputation of missing values: Update of ice. Stata Journal. 2005;5(4):527–36.

34. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Statistical Methods in Research. 2007;16(3):219–42.

35. Bartlett JW, Seaman SR, White IR, Carpenter JR. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. Statistical Methods in Medical Research. 2015;24(4):462–87.

36. Bartlett JW, Morris TP. Multiple imputation of covariates by substantive-model compatible fully conditional specification. Stata Journal. 2015;15(2):437–56.

37. Schafer JL. Analysis of incomplete multivariate data. Boca Raton [etc]: Chapman and Hall; 2000.

38. Little RJA. Statistical analysis with missing data. Third edit. Rubin DB, editor. Hoboken, NJ: Wiley; 2020. (Wiley series in probability and statistics).

39. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. BMC Medical Research Methodology. 2014;14(1).

40. Bodner TE. What improves with increased missing data imputations? Structural Equation Modeling. 2008;15(4):651–75.

41. Brand J, van Buuren S, le Cessie S, van den Hout W. Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data. Statistics in Medicine. 2019;38(2):210–20.

42. Baio G, Dawid AP. Probabilistic sensitivity analysis in health economics. Statistical Methods in Medical Research. 2015;24(6):615–34.

43. Baio G. Bayesian methods in health economics. Bayesian Methods in Health Economics. 2012. 1–223.

44. Gabrio A, Mason AJ, Baio G. A full Bayesian model to handle structural ones and missingness in economic evaluations from individual-level data. Statistics in Medicine [Internet]. 2019 Apr 15;38(8):1399–420. Available from: http://doi.wiley.com/10.1002/sim.8045

45. Gohel MS, Heatley F, Liu X, Bradbury A, Bulbulia R, Cullum N, et al. A Randomized Trial of Early Endovenous Ablation in Venous Ulceration. New England Journal of Medicine [Internet]. 2018 May 31;378(22):2105–14. Available from: http://www.nejm.org/doi/10.1056/NEJMoa1801214

46. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: A simulation study. BMC Medical Research Methodology. 2010;10.
Figure 1

Schematic relation between recruitment date and missing data pattern for 3 hypothetical patients
Figure 2

Cost-effectiveness acceptability curves at 3 years Note: RMM. Repeated measure mixed model; MIPMM multiple imputation using predictive men matching; MILR. multiple imputation using linear regression; CCA. complete-case-analysis; BPA. Bayesian parametric approach.
Figure 3

Standard errors of a) incremental mean costs b) incremental mean QALY. Note: RMM. repeated measure mixed model; CCA. complete-case-analysis; MIPMM. Multiple Imputation using predictive mean matching; MILR. Multiple Imputation using linear regression; BPA. Bayesian parametric approach.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS1Missingnessmechanismincosts.rtf
- SupplementaryTableS2MissingnessmechanisminEQ5D.docx
- SupplementaryTableS3Treatmentcostsoveralltimeperiods.rtf
- SupplementaryTableS4QALYoveralltimeperiods.docx
- Supplementaryfigure1.Distributionofimputationat3year.docx
- Supplementarymaterial1.docx
• Table1.summaryofmodels.docx
• Table2.Missingdatapattern.docx
• Table3.Resultsofthedifferentmodels.docx