



Universiteit
Leiden
The Netherlands

Patient-centered statistics in clinical prediction, trials and meta-analysis

Thomassen, D.

Citation

Thomassen, D. (2025, June 19). *Patient-centered statistics in clinical prediction, trials and meta-analysis*. Retrieved from <https://hdl.handle.net/1887/4250170>

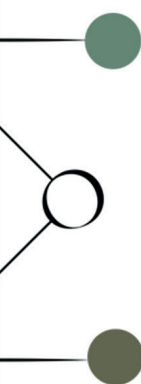
Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4250170>

Note: To cite this publication please use the final published version (if applicable).





Handling missing values in patient-reported outcome data in the presence of intercurrent events

Doranne Thomassen
Satrajit Roychoudhury
Cecilie Delphin Amdal
Dries Reynders
Jamme Z. Musoro
Willi Sauerbrei
Els Goetghebeur
Saskia le Cessie
On behalf of SISAQOL-IMI Work Package 3

*BMC Medical Research Methodology. 2025;
accepted for publication.*



ABSTRACT

Introduction As patient-reported outcomes (PROs) are increasingly used in the evaluation of medical treatments, it is important that PROs are carefully analyzed and interpreted. This may be challenging due to substantial missing values. The missingness in PROs is often closely related to patients' disease status. In that case, using observed information about intercurrent events (ICEs) such as disease progression and death will improve the handling of missing PRO data. Therefore, the aim of this study was to develop imputation models for repeated PRO measurements that leverage information about ICEs.

Methods We assumed a setting in which missing PRO measurements are missing at random given observed measurements, as well as the occurrence and timing of ICEs, and potentially other (baseline or time-varying) covariates. We then showed how these missingness assumptions can be translated into concrete imputation models that also account for a longitudinal data structure. The resulting models were applied to impute anonymized PRO data from a single-arm clinical trial in patients with advanced lung cancer.

Results In our trial example, accounting for death and other ICEs in the imputation of missing data led to lower estimated mean health-related quality of life (while alive) compared to an available case analysis and a naive linear mixed model imputation.

Conclusion Information about the timing and occurrence of ICEs contribute to a more plausible handling of missing PRO data. To account for ICE information when handling missing PROs, the missing data model should be separated from the analysis model.

KEYWORDS

Patient-reported outcomes, missing data, multiple imputation, estimand, intercurrent event, repeated measurements

1 | INTRODUCTION

Patient-reported outcomes (PROs) are increasingly used to provide evidence of efficacy and tolerability of treatments in clinical trials, including oncology¹⁻⁴. As with any other trial outcome, it is important that PROs are carefully analyzed and interpreted, in line with a clear research aim. However, structured approaches to analyze PRO data were found to be lacking in many oncology studies⁵⁻⁸. Previous studies have shown how a PRO-related research question may be translated into a well-defined estimand using the recent framework proposed by ICH E9 (R1)⁹, and which corresponding statistical methods may be used for estimation^{10,11}. However, little attention has been paid to the alignment of the missing PRO data strategy with the PRO estimand in a study¹²⁻¹⁶.

Defining the estimand in a study requires specifying the treatment, the population, the variable of interest, the population-level summary, and a strategy for handling intercurrent events (ICEs)^{9,11}. ICEs refer to events that occur in a trial after the start of treatment and affect the interpretation or existence of measurements of interest. Examples of ICEs are treatment discontinuation and death. Strategies for handling ICEs mentioned in the estimand framework include a while-the-ICE-has-not-occurred strategy, a treatment policy strategy, a hypothetical strategy, a composite variable strategy, and a principal stratum strategy^{9,11}. The estimand framework distinguishes a strategy for handling missing data from the chosen strategy to address intercurrent events⁹. The chosen ICE strategy partly determines which data are required to estimate the estimand and which are considered missing if not collected.

Methods to address missing data should respect the chosen ICE strategy for a coherent analysis and interpretation of results. PROs are often measured at several time points, with the aim of estimating or comparing the mean (change in) PRO value at (a) time point(s) of interest. The traditional way of handling missing longitudinal data is to use (generalized) linear mixed models or mixed models for repeated measures (MMRM) in the analysis^{17,18}. However, such analyses may not align with the chosen ICE strategy. For example, after a patient has died, their PRO data do not exist and are not generally considered missing⁹. Analyses using mixed models assume that missing data are missing at random (MAR) conditional on the model variables^{19,20}. Implicitly, at any time point, the PRO values of those who have died will be considered MAR and the model will handle these 'missing' PRO values based on the observed PRO values in those who are still alive at that time. Therefore, estimated means directly derived from a mixed model correspond to a hypothetical strategy to handle death, where interest is in the PRO in a hypothetical world where patients do not die during the study. This strategy is often applied implicitly but may not always align with the research aim¹¹.

Additionally, there is a relation between ICEs and the occurrence of missing PRO data. ICEs may cause PRO data to be unavailable: when a patient experiences disease progression, they may be too ill to fill out a questionnaire. On the other hand, the occurrence and timing of ICEs may be correlated with the value of missing PRO measurements. It has been shown that, on average, the health-related quality of life (HRQoL) of patients with cancer deteriorates after disease progression^{21,22} and in the weeks leading up to death²³⁻²⁸. As a result, observed information about ICEs may be utilized for a better handling of missing PRO data.

In this SISAQOL-IMI²⁹ study, we therefore aimed to evaluate imputation methods for missing PRO data that

1. take into account information about the occurrence and timing of intercurrent events;

2. are suitable for repeated measurements data;
3. can be aligned with an estimand of choice; and
4. do not impute PROs after death.

We illustrated our methods using anonymized data from a single-arm clinical trial in patients with advanced lung cancer.

2 | METHODS TO HANDLE MISSING PRO DATA

2.1 | Study setting

Suppose that a trial is analyzed where a PRO was scored on a numerical scale. This PRO was scheduled to be measured at baseline and at one or more predefined time points after baseline. We denote with Y_{ij} the PRO reported by patient $i \in \{1, \dots, n\}$ at timepoint $j \in \{1, \dots, m\}$ where $j = 1$ at baseline. Besides the PROs Y_{ij} , baseline variables X_i were recorded for each patient i , and possibly time-varying variables Z_{ij} . Additionally, it was recorded when patients passed away during the study, when they were lost to follow-up and when they experienced non-terminal ICEs such as discontinuation of protocol treatment. We assume that interest is in estimating the mean (or mean change in) PRO at one or several time points after baseline (e.g., for nonrandomized studies), or in differences between the means (or mean changes) in different treatment groups (e.g., for randomized studies).

2.2 | Estimand and data required for estimation

As mentioned previously, defining an estimand requires specification of one or more strategies to address each ICE, respectively. The selected ICE strategies partly determine the data needed to estimate the estimand of interest and the data that are considered missing if not observed.

When addressing death, a terminal ICE, a while alive strategy may be deemed most appropriate, especially when the study has a descriptive aim^{11,30}, as PRO values after death do not exist. An example of a while alive estimand is the mean PRO at a time point of interest in those who are still alive at that time point, which should be combined with an estimate of the probability to survive until that timepoint. For most other ICE strategies, PRO values after death are not relevant either. In a composite variable strategy, a composite endpoint is defined that includes both the PRO and death (e.g., EQ-5D). A principal stratum strategy addresses death by estimating a contrast in PROs between treatment arms in the (counterfactual) subset of patients who would not have died in any of the treatment arms in a study. A treatment policy strategy cannot be applied to terminal ICEs.

The estimation of a hypothetical estimand for death considers PRO values in a hypothetical world where patients do not die during the study. By definition, these values are not available after a patient has died in the observed world. But even before death, it is difficult to determine which PRO values in the observed world would be the same in the hypothetical world where death does not occur. For example, in the weeks leading up to death due to cancer, HRQoL tends to deteriorate²⁴. In a hypothetical world without death, such a deterioration may not take place. Therefore, observed PRO values in the weeks before a patient's death may not be suitable for the estimation of a hypothetical estimand for death. Furthermore, there may be settings in which a hypothetical world without death but with disease progression is difficult to imagine.

For non-terminal ICEs, a treatment policy strategy (reflecting the intention-to-treat principle) requires PRO data after the ICE has occurred and this will be considered missing

if not collected. A while-the-ICE-has-not-occurred estimand (e.g., while on treatment) and a composite estimand are estimated based on PRO data before the ICE (and information about the ICE itself, for the composite outcome). The data required for estimation of a principal stratum estimand depend on the corresponding subgroup of interest.

Some non-terminal ICEs are defined by a conscious action or intervention, for instance, treatment switching or the addition of rescue medication. For such intervention-ICEs, a hypothetical world in which the intervention was not applied might be better imaginable than for ICEs that are not interventions. Methods to estimate hypothetical estimands for intervention-ICEs have been described in the literature recently, with links to imputation methods and causal inference techniques^{12,31}. Overall, it is important to clearly define the hypothetical world targeted by a hypothetical estimand and to assess whether observed data are sufficiently in line with this hypothetical world for estimation. Any extrapolation from the observed world to an unobserved, hypothetical world requires untestable assumptions about the similarity of the two worlds.

2.3 | Notation to describe the occurrence and timing of ICEs

To include information about the occurrence and timing of ICEs in methods to address missing PRO data, this information must first be encoded. Let the variable $D_i \in \{0,1\}$ indicate whether patient i died during the study, where s_i denotes their corresponding survival time ($D_i = 1$) or the time at which they were censored for survival ($D_i = 0$). Further, we suppose that K different types of non-terminal ICE could be observed (for example, treatment discontinuation due to toxicity, disease progression), each event denoted by E_k , $k \in \{1, \dots, K\}$. Analogous to the variables describing survival, the binary variable E_{ik} , $k \in \{0,1\}$ denotes whether E_k occurred for patient i during the study. If E_k was indeed observed for patient i before the end of their follow-up (i.e., when $E_{ik} = 1$), T_{ik} denotes the time at which ICE E_k was observed for patient i . If patient i did not experience E_k during their follow-up (i.e., when $E_{ik} = 0$), then T_{ik} denotes the time at which follow-up for ICE E_k ended for patient i . For example, for disease progression, T_{ik} refers to time of the last assessment of the disease status. In other words, T_{ik} is an event time if E_k is observed, and a censoring time if E_k is not observed.

Based on this information, we define the following time-varying variables for each patient:

- s_{ij} denotes the difference in time units between the survival time s_i and the current timepoint j . As there are no observations after death or censoring for overall survival, s_{ij} is nonnegative.
- Analogously, t_{ijk} denotes the current ‘time-distance to non-terminal ICE E_k for patient i ’: the difference in time units between T_{ik} and the current timepoint j . The difference t_{ijk} is positive when timepoint j is before T_{ik} , $t_{ijk} = 0$ at T_{ik} , and t_{ijk} is negative after T_{ik} .
- E_{ijk} indicates whether ICE E_k has yet been observed at timepoint j . For the terminal ICE death, such an indicator is redundant as there are no observations after death.

As a result, each patient is viewed on several timescales at once: the follow-up time that starts at baseline, as well as a timescale relative to each ICE E_k and death, respectively (Figure S1, Appendix).

2.4 | Missing data assumptions

Often, not all PRO scores are recorded at each of the required time points, leading to missing PRO data. We assume that when a patient did not report their outcome at one timepoint, they may still have reported their outcome at future timepoints (unless, of course, death had occurred).

Information on the occurrence and timing of ICEs may make a MAR assumption conditional on observed information more plausible. Therefore, missing PRO measurements

at timepoint j are assumed to be MAR conditional on the timepoint j , the PRO Y_{ij^*} for other relevant timepoints $j^* \neq j$, the occurrence and timing of ICEs (E_{ik} , E_{ijk} , t_{ijk} , D_i , s_{ij}), treatment (if applicable), any relevant baseline covariates X_i and possibly time-varying covariates Z_{ij} . As ICEs may occur after the missed measurement, this means we assume that future events provide information about what the missed measurement may have been. After death, PRO measurements are not considered missing as they are nonexistent. If an ICE has not occurred during the study, the time of its occurrence may be right-censored at the end of the study. We assume noninformative censoring for simplicity in our examples but methods to account for informative censoring (such as reweighting³²) can be combined with the missing PRO data strategies outlined in this paper.

2.5 | Imputation of missing PRO data

It is important that the chosen method to address missing PRO data accounts for the longitudinal structure of the PRO data, and that it is informed by the occurrence and timing of ICEs, which may be observed after the missing PRO was planned to be measured. For this reason, likelihood-based methods where a mixed model is used for the implicit imputation and analysis of PROs at the same time do not suffice. While the information on (future) ICEs is informative for the missing values, conditioning on future values in an analysis model is inappropriate as this will not target the estimand of interest and yield uninterpretable estimates of the treatment effect. Therefore, as a way to take ICE information into account, missing PRO values should be handled using a missing data model that is separate from the substantive analysis model.

To deal with missing PRO measurements in accordance with our MAR assumptions, we distinguish two types of methods: reweighting methods by the inverse probability of missingness (IPW) and multiple imputation (MI) methods for the missing outcomes¹⁹. Kurland and Heagerty have shown that IPW methods can be combined with generalized estimating equations (GEE) to estimate and perform inference on marginal means of the (change in) PRO over time under a MAR assumption conditional on the time until death³⁰. To obtain such marginal means only in those still alive, an independence working correlation should be specified for the GEE³⁰. Recently, Haug et al. have also proposed IPW weights for missing HRQoL data that account for time of death and combined these weights with a machine learning model to estimate average HRQoL over time³³. These IPW techniques can be extended to include all variables in our MAR assumption described above in the estimation of the IPW weights^{32,34,35}.

For MI, it has been recommended that imputation models are compatible with the intended analysis model, that is, that they include at least all variables from the analysis model as well as their relation presumed in the analysis model (e.g., interactions)³⁶. As the PROs in our case are numerical and longitudinal, we assume that the analysis model will be some mixed model or GEE (depending on the estimand) where the outcome is modelled as a (flexible) function of the timepoint j , treatment, the outcome at other time points, and potentially covariates. The imputation model would then include these variables, as well as the ICE-related variables in our MAR assumption.

Where possible, we recommend modelling the time-from-ICE variables t_{ijk} and s_{ij} flexibly while imputing missing PRO values, for example with splines, to allow for nonlinear relationships with the PRO (e.g., a drop in HRQoL in the final weeks before death). Another choice to make is whether to let the value of t_{ijk} decrease indefinitely after T_{ik} or to keep it constant from some time point, for example, when there are almost no patients with data after ICE E_k . Interactions are preferably included between the time variables for the ICE (t_{ijk} , s_{ij}) and their corresponding event/censoring indicators (E_{ik} , D_i), to model the relationship between the PRO and an event time differently from the relationship between the PRO and a censoring time. If the relation between the PRO and t_{ijk} is expected to be different before and after ICE E_k occurs, interactions between t_{ijk} and the time-varying event indicator E_{ijk} can also

be included. Of course, including splines and interactions in an imputation model requires a sufficiently large dataset and may not always be feasible.

Besides information about ICEs, PROs observed in the same participant at other time points are also likely informative for missing PRO measurements³⁷. We distinguish two approaches to include this information in imputation, following Huque et al³⁸. The first option is a cross-sectional approach where, for each missing PRO measurement, PRO measurements at other relevant time points are included explicitly as variables in the imputation model (longitudinal data in wide format). Alternatively, a multilevel approach may be used, where all PRO measurements are treated as one outcome variable with a within- and between-subject correlation (multilevel data structure) and possibly a relationship with time (longitudinal data in long format). Huque et al. showed in a simulation-based comparison that cross-sectional approaches using standard software implementations often performed well³⁸. However, the cross-sectional approach requires the variables to be imputed if missing to be the same for every participant. When the number of time points at which PROs are to be imputed differs between participants (for example, to avoid imputation after death), multilevel approaches may be better suited.

For the multilevel MI approach, software implementations in R³⁹ exist to model the joint distribution of incomplete variables by iterating through univariate models for each incomplete variable conditional on the other variables (fully conditional specification, ‘MI-FCS’)⁴⁰. Within FCS, we noted two types of approaches in the literature for multilevel data. In the first approach, mixed models (i.e., models with random effects) are used as the univariate conditional models (e.g., ‘2l.norm’, in the mice package⁴¹). The second approach includes using flexible generalized additive models as the univariate conditional models combined with a clustered bootstrap instead of a simple bootstrap to approximate the distribution to impute from (‘aregImpute’ function in the Hmisc package⁴²). While marginal means from a (linear) mixed model imply a hypothetical estimand for death, mixed models can be used to impute *individual* missing PRO values in those patients who are still alive. It has been shown that mixed models have a subject-specific interpretation in the population currently alive⁴³.

3 | ILLUSTRATION IN A SINGLE-ARM LUNG CANCER TRIAL

The implementation of several of the methods mentioned was illustrated and compared in real clinical trial data.

3.1 | Description of the trial

We illustrated our methods in PRO data from a single-arm, multicenter phase 2 trial⁴⁴ evaluating a new anticancer drug in patients with late-stage lung cancer. PRO data were available for 834 patients. Patient-reported HRQoL data, measured with the EORTC QLQ-C30 global QoL scale^{45,46}, were included as a secondary endpoint, with scale scores ranging from 0 (representing the worst) to 100 (best possible score). Henceforth, we will refer to this specific PRO as ‘QoL’.

While on trial medication, participants were asked to report on QoL on the first day of protocol treatment and every three weeks while on protocol treatment. We therefore refer to the timing of QoL measurements with their tri-weekly cycle number and baseline is defined here as the first day of cycle 1. After cycle 10 (30 weeks), the study protocol allowed for completion of the questionnaire on the first day of alternate cycles (i.e., every six weeks).

Three types of ICEs were recorded in the data: progression of disease (PD, E_1), treatment discontinuation (TD, E_2), and death. Although disease progression was often followed by the discontinuation of treatment, the physician and patient could decide together when to stop the treatment. After PD, 345 patients continued the trial medication for at least one month and 241 continued for 3 months or more. Upon discontinuation of trial medication, patients were asked to complete one final questionnaire, after which QoL data collection was stopped, while follow-up for overall survival continued. The sponsor anonymized the measurements

before they were shared with us. Figure 4.1 summarizes the availability of QoL data over time in the trial data. A more elaborate description of the anonymized data can be found in a previous case study¹¹.

3.2 | Estimand

In this single-arm trial setting, our analysis aim was to describe HRQoL over time in the study population. Our endpoint of interest was the absolute numerical value of the global QoL score as reported by each patient, at each cycle. The mean global QoL score at cycles 1 through 40 was chosen as the population-level summary. As the trial participants were patients with cancer in an advanced disease stage, many deaths were observed during the study.

The research objective was to estimate the mean global QoL score from the start of study medication up to treatment cycle 40 or death, in the reality where PD and TD occur as they did in the study. Because HRQoL after death is not meaningful, and a hypothetical situation without deaths was far removed from reality in this case, we opted for a while alive strategy to address death¹¹. For cycles 1-40, the mean QoL was estimated at each cycle, within only those patients who were still alive at that cycle, accompanied by an estimate of the survival probability up to that cycle. For the other intercurrent events (PD and TD), a treatment policy strategy was applied. Estimation of this estimand required QoL measurements from each participant for cycle $j = 1, \dots, 40$, or until their death if death occurred earlier, irrespective of PD or TD. Where such measurements were not available, they were considered missing data.

3.3 | Missing data imputation

3.3.1 | Variables and coding

We stored the QoL data Y_{ij} in long format where each study participant had several rows of data, one row for each predefined QoL measurement cycle j . When a patient died or was censored for overall survival, the patient had no rows in the data after this event. As survival times s_i differed between individuals, the number of rows in the data varied between patients.

Each data row contained a patient identifier i , the number of the measurement cycle j , and the corresponding QoL measurement Y_{ij} (possibly missing). The row also contained time-

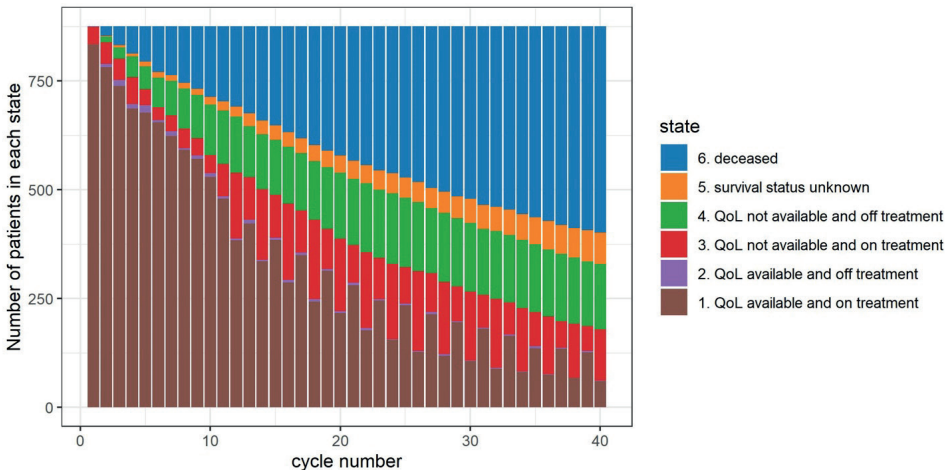


Figure 4.1 Availability of global QoL scores over time, in relation to treatment and survival status. Note that by design of the study, almost no global QoL scores were reported after discontinuation of protocol treatment. As a result, there are no, to very few patients in state 2 at each cycle.

constant variables E_{ik} , T_{ik} , D_i , and S_i ; as well as time-varying variables E_{ijk} , t_{ijk} , s_{ij} . As there were almost no QoL data after TD, t_{ij2} was kept constant at 0 after T_{i2} , assuming the same time-pattern at and at any time after the event. When a measurement was taken after the event, this was encoded by a time-varying indicator for TD. Analogously, for PD t_{ij1} was kept constant from 3 cycles after T_{i1} (at -3). The baseline covariates age, sex and ECOG status were also recorded.

All time variables were modelled flexibly with splines to allow for possible nonlinear time-dependencies of QoL. We preferred restricted cubic splines (rcs) but opted for B-splines in some situations where the near-collinearity of rcs basis functions inhibited the convergence of the imputation model⁴⁷. Discontinuation of treatment was observed in all study participants at some point, as participants only had access to the treatment within the study. As death and PD were not observed in all study participants, we included interactions between the time-to-ICE and the event indicator, so that a different relation was assumed between the PRO and the ICE itself versus the PRO and the censoring of the ICE.

3.3.2 | Imputation models

For comparison, we applied several MI-FCS methods to our data. As there was only one variable with missing values in the data (QoL), only one conditional model needed to be specified (for QoL) and the process of iterating through several conditional models in each imputation was unnecessary. The number of imputations should be large enough to sufficiently approximate the predictive distribution of the missing values given the observed data (i.e., with sufficiently small Monte Carlo errors). While other suggestions have been made⁴⁸ a rule of thumb is that a number of imputations equal to the percentage of missing values in the data will suffice^{49,50}. In our case, 40.5% of QoL measurements were missing. Therefore, the number of imputations was set to 40 in our analyses. The MI-FCS models implemented are described in detail below (see also the R code provided in Online Resource 2).

First, we implemented multiple imputation with chained equations⁴¹ (MICE), for multilevel data using two different approaches. For our first MICE model, (model 1), we specified a linear multilevel model for QoL with heterogeneous within-patient variances and random effects on all predictors ('2l.norm' method in the mice package⁴¹). The model can be expressed in pseudo-code as: $Y_{ij} \sim (1 + \text{bspline}(j) + \text{bspline}(t_{ij1}): E_{i1} + E_{i1} + \text{bspline}(t_{ij2}) + \text{spline}(s_{ij}): D_i + D_i + E_{ij1} + E_{ij2} + \text{age} + \text{sex} + \text{ECOG} | i)$.

Our second MICE model (model 2) was a predictive mean matching algorithm based on a linear mixed model for QoL with a random patient intercept (the '2l.pmm' method in the miceadds package⁵¹). In pseudo-code, the model is defined as: $Y_{ij} \sim (1 | i) + \text{bspline}(j) + \text{bspline}(t_{ij1}): E_{i1} + E_{i1} + \text{bspline}(t_{ij2}) + \text{bspline}(s_{ij}): D_i + D_i + E_{ij1} + E_{ij1}: E_{i1} + E_{ij2} + \text{age} + \text{sex} + \text{ECOG}$.

Both MICE approaches required the use of B-splines as the rcs basis vectors created issues with model convergence due to near-collinearity. The splines for the time-till-ICE variables all had four knots located at 1, 4, 9 and 20 cycles before the ICE occurred. As a default, the mice package requires the conditional imputation models to be specified through a predictor matrix that indicates which other variables in the data to use for imputation of each incomplete variable. To include interactions, splines or other nonlinear terms in the imputation, the user must add variables to the data that encode these terms. For splines, this entails manually computing spline bases and adding them as variable columns to the data. The 'formulas' argument in the 'mice' function may offer an easier way to specify nonlinear terms in an imputation model using standard R formula syntax. However, it is currently unclear how to define a cluster variable for multilevel imputation models in this way, and therefore we relied on model specification through the predictor matrix.

Second, we fitted a flexible linear mixed model to the data with a random patient-intercept using the lme4 package⁵² (model 3). Again, the knots for the time-till-ICE splines were located at 1, 4, 9 and 20 cycles before occurrence of the ICE. We subsequently used

the ‘predictInterval’ function from the package merTools⁵³ to draw imputations from (an approximation of) the conditional predictive distribution of the QoL variable. The model structure was as follows: $Y_{ij} \sim (1 | i) + \text{rcs}(j) + \text{rcs}(t_{ij1}) + \text{rcs}(t_{ij2}) + \text{rcs}(s_{ij}) * D_i + E_{ij1} + E_{ij1} * E_{i1} + E_{ij2} + \text{age} + \text{sex} + \text{ECOG}$. The interaction between time-to-PD (t_{ij1}) and PD (E_{i1}) was dropped because the model was over-parametrized. The advantages of this method compared to MICE were that we could specify the imputation model as a formula instead of a predictor matrix, and that we were able to incorporate restricted cubic splines. An obvious disadvantage is that this method is only applicable when QoL is the only incomplete variable.

Third, we used the ‘aregImpute’ function in the Hmisc package⁴² to implement generalized additive regression models for imputation and clustered bootstrap to account for the multi-level structure of the data in the process of drawing imputations. Two different outcome models were specified: (a) a conditional normal model for QoL (model 4); and (b) conditional predictive mean matching for QoL (model 5). The model structure for both can be translated to pseudo-code as follows: $Y_{ij} \sim \text{rcs}(j) + \text{rcs}(t_{ij1}) * E_{i1} + \text{rcs}(t_{ij2}) + \text{rcs}(s_{ij}) * D_i + E_{ij1} + E_{ij1} * E_{i1} + E_{ij2} + \text{age} + \text{sex} + \text{ECOG}$.

All splines were restricted cubic splines with four knots at automatically selected location.

All corresponding R code is available in Online Resource 2, which includes an example of how we coded the variables in our dataset.

3.3.3 | Imputation results

For each of the methods, the QoL distribution in the imputed datasets was shifted somewhat to the left compared to the available data (Figure S2, Appendix), suggesting that imputed QoL values were on average lower than the available values. A possible reason for this is that our imputation models detected a decrease in QoL in the weeks before death and around PD in the observed data, while patients were less likely to fill out questionnaires during these weeks, which was then accounted for in the imputation. For the mice 2l.pmm method, the distributions of QoL data in the respective imputed datasets differed more between imputations than for the other methods.

3.4 | Statistical analysis

After the imputation of missing data, the mean QoL while alive at each cycle was estimated using GEE with the cycle number as a categorical variable and an independence correlation structure^{11,30}. In the available data (model 0), this led to the same point estimates as simply calculating the sample mean QoL within the group of patients still alive at each respective cycle. We assumed non-informative censoring for overall survival (and other measurements) in our analyses, as most censoring was administrative at study end.

To compare our imputation methods to a reweighting approach (IPW-GEE, model 6), we also estimated the mean QoL while alive using the same GEE as above but on the incomplete data and weighted by the inverse probability of a QoL measurement being available. The weights were estimated using a flexible logistic regression model containing the same independent variables as the imputation models, and were subsequently stabilized and truncated.

Finally, to assess the result of not including information about ICEs in the imputation, we fitted the same GEE (without weights) to a dataset with QoL imputed based on a linear mixed model with only a random patient-intercept and forward-time (cycle number) as a spline, and the baseline covariates age, sex and ECOG status (i.e., a naive linear mixed model where ICE information is not included, model 7).

3.5 | Analysis results

The imputation models 1-5 described in Section 3.2 resulted in very similar estimated mean QoL while alive over time (Figure 4.2). Only the two-level predictive mean matching model

(model 2) showed slightly higher estimates than the others. The results from the IPW-GEE (model 6) closely follow the imputation results in the first 20 cycles but become unstable at later cycles when some patient types become rare and therefore some patients had much larger weights than others. Even in early cycles, there were some patients with larger weights than others (e.g., those who died early or experienced early PD), which led to mean QoL estimates that were slightly different from the imputation methods' results. Importantly, all imputation methods that did account for ICEs (models 1-5) led to lower estimated mean QoL compared to the available data (model 0) and to the imputation without ICE information (model 7). The mean QoL resulting from models 1-5 was similar to the mean QoL in the available data at early cycles but the difference increased to about 5 points in mean QoL at later cycles when there were more missing values and more ICEs had occurred.

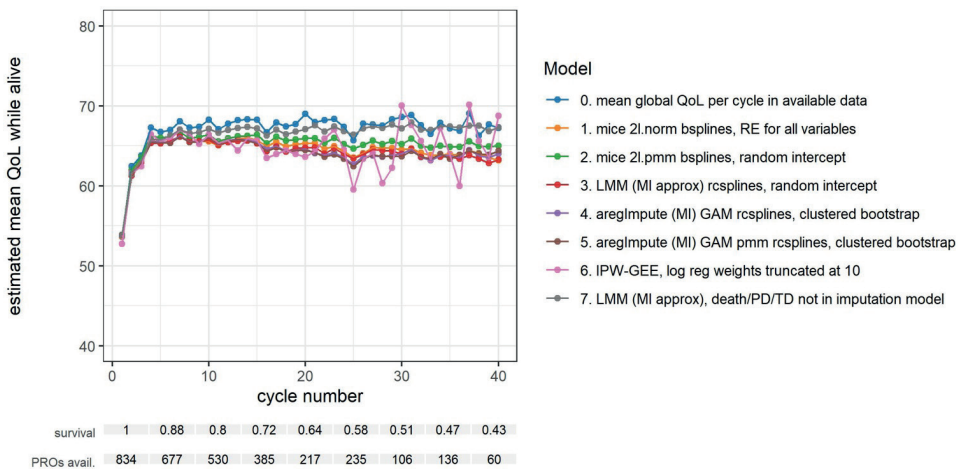


Figure 4.2 Estimated mean global QoL while alive in the available data (blue) and when using various methods to address missing data (other colors). Survival probabilities were estimated using the Kaplan-Meier method. *RE: random effects, LMM: linear mixed model, MI: multiple imputation, GAM: generalized additive model, rcsplines: restricted cubic splines, pmm: predictive mean matching, IPW-GEE: inverse probability weighted generalized estimating equations, log reg: logistic regression, PD: progression of disease, TD: treatment discontinuation.

Regarding statistical efficiency, the standard errors resulting from most imputation methods were similar, providing no clear reason to prefer one over the others (Figure 4.3). The two-level predictive mean matching model (model 2) did result in wider confidence intervals (CIs) than other methods. The IPW-GEE method (model 6) provided CIs similar to the imputation methods in the earlier cycles, but showed large uncertainty at later cycles, in line with the instability in the corresponding point estimates.

4 | DISCUSSION

In this study, we have shown how imputation models for longitudinal PROs can be implemented that incorporate information about ICEs and that do not impute PROs after death. In our example, accounting for death and other ICEs in the imputation of missing HRQoL data led to lower estimated mean HRQoL over time compared to an available case analysis and a naive linear mixed model analysis. The information about the ICEs may (partly) capture that patients in a worse state are less likely to fill out questionnaires while they are more likely to have lower HRQoL values. In general, incorporating information about the timing and occurrence of ICEs when handling missing PRO data in a study is important as it makes a MAR assumption more plausible. In analyses that do not consider ICE information

when handling missing PRO data, the extent of the bias will depend on the strength of the underlying relation between the ICEs and the missing values, the number of ICEs that occur and the total amount of missing data.

Our findings are in line with a recent study by Haug et al.³³, who implemented an IPW method accounting for the time to death and other auxiliary variables to handle missing HRQoL data in two cancer studies. Their results indicated that, at each time point, the mean HRQoL was likely lower in the group of participants who did not report HRQoL, compared to the group who did report HRQoL. Our study reiterates the importance of incorporating not only death, but all available information about ICEs in the missing data strategy in alignment with the estimand.

Between the ICE-sensitive imputation models, the resulting estimates and standard errors (SEs) were quite similar. There were some differences in the computational time taken for the imputation: the `aregImpute` function was relatively fast, whereas the MICE 2l.norm method took longest. However, the MICE 2l.norm method implements a model with many random effects and heterogeneous within-patient variances, which the `aregImpute` function does not, so it makes sense that MICE takes more time. When the longitudinal PRO is the only incomplete variable in the data, fitting a flexible linear mixed model with ICE information and then drawing imputations from an approximation of the resulting predictive distribution is easy to implement and has a relatively low computational time.

All multiple imputation models implemented in our example were MI-FCS methods that target the joint posterior distribution of incomplete variables by specification of conditional univariate distributions. A formula-based method to specify interactions, splines and other nonlinear terms in multilevel imputation models would be a useful future addition to the ‘mice’ software. Such a method would be more user-friendly than model specification through a predictor matrix as shown in this paper. Multiple imputation by modelling the joint posterior distribution directly that takes the multilevel data structure into account has also been described in the literature^{54,55} (R software includes the `jomo` package⁵⁶). We were, however, not able to arrive at a working implementation of joint modeling MI using this software with our example data. Further research might explore joint modeling MI methods

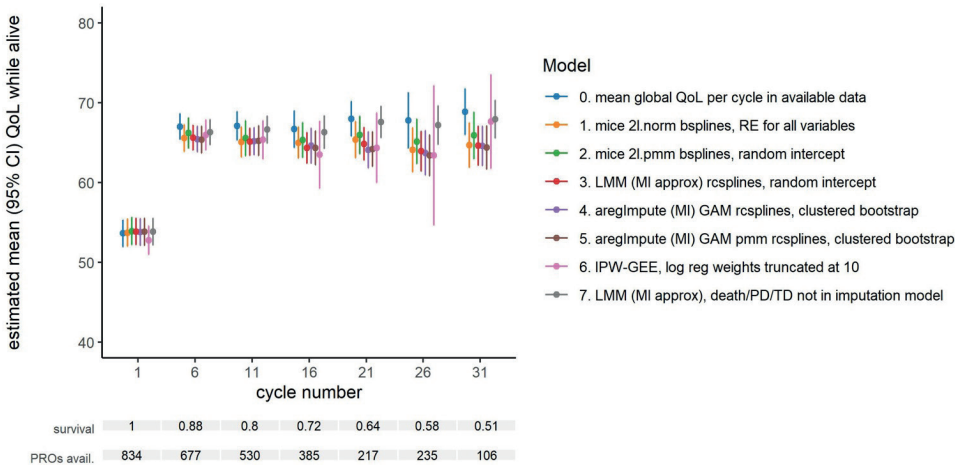


Figure 4.3 Estimated mean global QoL while alive with 95% CIs in the available data (blue) and when using various methods to address missing data (other colors).

*RE: random effects, LMM: linear mixed model, MI: multiple imputation, GAM: generalized additive model, rcsplines: restricted cubic splines, pmm: predictive mean matching, IPW-GEE: inverse probability weighted generalized estimating equations, log reg: logistic regression, PD: progression of disease, TD: treatment discontinuation

for longitudinal PRO data.

An important limitation of our trial data example is that there was very little available information in some regions of the data space (e.g., after TD). Therefore, model assumptions weigh heavily there (for example, setting time-till-TD to 0 after TD). This is emphasized by the instability of the IPW-GEE approach at later cycles. Furthermore, we can never be sure that the relation between the ICEs and the PRO in the observed data is the same in the missing data. In other words, the MAR assumption is untestable. Therefore, in a real trial analysis, we recommend performing sensitivity analyses to assess the impact of imputation model assumptions on the resulting estimates. If PRO data are intended for review of payer or regulatory submission, discussions with these stakeholders to identify methods to handle missing data that align with appropriate estimands are highly encouraged.

This study focused on the imputation of PROs on their original numerical scale. Once missing values have been imputed, various analyses can be performed, such as estimation of the mean PRO at timepoints of interest (as in our examples) or a change from baseline in the PRO. Sometimes, numerical PRO values variables are dichotomized, for example, to classify participants into responders and non-responders. The disadvantages of a PRO (non-) responder analysis in terms of bias and reduction of statistical power have been discussed in the literature^{57,58}. If dichotomization of PRO values is still desired, we recommend imputing the PROs on the original numerical scale first and then dichotomizing them in each imputed dataset before performing the final analysis⁵⁹.

5 | CONCLUSION

To account for ICE information when handling missing PRO data, the missing data model should be separated from the analysis model as ICEs may occur later than the missed measurement. Therefore, MI and IPW methods are preferred over direct likelihood-based methods to handle missing data. The choice between MI and IPW depends on whether one believes to have an appropriate model for the missing outcomes themselves or for the probability of missingness, respectively. We have shown how information about intercurrent events can be incorporated in the multiple imputation of a continuous, longitudinal PRO. Whichever method is applied, it is important that information about the timing and occurrence of ICEs is taken into account when handling missing PRO data.

6 | ACKNOWLEDGEMENTS

We would like to thank the STRATOS Initiative for their valuable methodological input during this study, and STRATOS members' involvement in SISAQOL-IMI Work Package 3.

The SISAQOL-IMI project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 945052. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

This publication reflects the views of the individual authors and should not be construed to represent official views or policies of the European Medicines Agency (EMA), the US Food and Drug Administration (FDA), US National Cancer Institute (NCI), Medicines and Healthcare products Regulatory Agency (MHRA), Institute for Quality and Efficiency in Health Care (IQWiG), Health Canada, the Norwegian Medicines Agency (NOMA), the American Society of Clinical Oncology (ASCO), the European Society for Medical Oncology (ESMO), or any other institution, organization, or entity. This publication reflects the authors' view and that neither IMI nor the EU, EFPIA are responsible for any use that may be made of the information contained therein.

REFERENCES

1. European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. The use of patient-reported outcome (PRO) measures in oncology studies. Published online 2016. Accessed April 28, 2023. https://www.ema.europa.eu/en/documents/other/appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.pdf
2. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Patient-Focused Drug Development: Methods to Identify What Is Important to Patients - Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders (Final Guidance). Published online February 2022. Accessed October 18, 2023. <https://www.fda.gov/media/131230/download>
3. U.S. Department of Health and Human Services Food and Drug Administration. Guidance for Industry - Submitting Patient-Reported Outcome Data in Cancer Clinical Trials. Published online November 2023. Accessed March 18, 2024. <https://www.fda.gov/media/173581/download>
4. U.S. Department of Health and Human Services Food and Drug Administration. Guidance for Industry - Core Patient-Reported Outcomes in Cancer Clinical Trials (Draft Guidance). Published online June 2021. Accessed August 16, 2023. <https://www.fda.gov/media/149994/download>
5. Liu L, Choi J, Musoro JZ, et al. Single-arm studies involving patient-reported outcome data in oncology: a literature review on current practice. *Lancet Oncol*. 2023;24(5):e197-e206. doi:10.1016/S1470-2045(23)00110-9
6. Teixeira MM, Borges FC, Ferreira PS, Rocha J, Sepodes B, Torre C. A review of patient-reported outcomes used for regulatory approval of oncology medicinal products in the European Union between 2017 and 2020. *Front Med*. 2022;9:968272. doi:10.3389/fmed.2022.968272
7. Pe M, Dorme L, Coens C, et al. Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review. *Lancet Oncol*. 2018;19(9):e459-e469. doi:10.1016/S1470-2045(18)30418-2
8. Fiero MH, Roydhouse JK, Vallejo J, King-Kallimanis BL, Kluetz PG, Sridhara R. US Food and Drug Administration review of statistical analysis of patient-reported outcomes in lung cancer clinical trials approved between January, 2008, and December, 2017. *Lancet Oncol*. 2019;20(10):e582-e589. doi:10.1016/S1470-2045(19)30335-3
9. European Medicines Agency. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Published online February 17, 2020. Accessed April 28, 2023. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf
10. Fiero MH, Pe M, Weinstock C, et al. Demystifying the estimand framework: a case study using patient-reported outcomes in oncology. *Lancet Oncol*. 2020;21(10):e488-e494. doi:10.1016/S1470-2045(20)30319-3
11. Thomassen D, Roychoudhury S, Amdal CD, et al. The role of the estimand framework in the analysis of patient-reported outcomes in single-arm trials: a case study in oncology. *Submitted*. Published online 2024.
12. Wolbers M, Noci A, Delmar P, Gower-Page C, Yiu S, Bartlett JW. Standard and reference-based conditional mean imputation. *Pharm Stat*. 2022;21(6):1246-1257. doi:10.1002/pst.2234
13. Noci A, Wolbers M, Abt M, et al. A Comparison of Estimand and Estimation Strategies for Clinical Trials in Early Parkinson's Disease. *Stat Biopharm Res*. 2023;15(3):491-501. doi:10.1080/19466315.2022.2116476
14. Bell J, Drury T, Mütze T, et al. Estimation methods for estimands using the treatment policy strategy; a simulation study based on the PIONEER 1 Trial. Published online February 20, 2024. Accessed April 29, 2024. <http://arxiv.org/abs/2402.12850>
15. Mallinckrodt CH, Bell J, Liu G, et al. Aligning Estimators With Estimands in Clinical Trials: Putting the ICH E9(R1) Guidelines Into Practice. *Ther Innov Regul Sci*. 2020;54(2):353-364. doi:10.1007/s43441-019-00063-9
16. Carpenter JR, Roger JH, Kenward MG. Analysis of Longitudinal Trials with Protocol Deviation: A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation. *J Biopharm Stat*. 2013;23(6):1352-1371. doi:10.1080/10543406.2013.834911
17. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. *JAMA*. 2016;315(4):407. doi:10.1001/jama.2015.19394
18. Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. *Drug Inf J*. 2008;42(4):303-319. doi:10.1177/009286150804200402
19. Bell ML, Floden L, Rabe BA, et al. Analytical approaches and estimands to take account of missing patient-reported data in longitudinal studies. *Patient Relat Outcome Meas*. 2019;Volume 10:129-140. doi:10.2147/PROM.S178963
20. Cella D, Li JZ, Cappelleri JC, et al. Quality of Life in Patients With Metastatic Renal Cell Carcinoma Treated With Sunitinib or Interferon Alfa: Results From a Phase III Randomized Trial. *J Clin Oncol*. 2008;26(22):3763-3769. doi:10.1200/JCO.2007.13.5145
21. Chase DM, Marín MR, Backes F, et al. Impact of disease progression on health-related quality of life of advanced ovarian cancer patients – Pooled analysis from the PRIMA trial. *Gynecol Oncol*. 2022;166(3):494-

502. doi:10.1016/j.ygyno.2022.06.028
22. Marschner N, Zacharias S, Lordick F, et al. Association of Disease Progression With Health-Related Quality of Life Among Adults With Breast, Lung, Pancreatic, and Colorectal Cancer. *JAMA Netw Open*. 2020;3(3):e200643. doi:10.1001/jamanetworkopen.2020.0643
23. Elmqvist MA, Jordhøy MS, Bjordal K, Kaasa S, Jannert M. Health-related quality of life during the last three months of life in patients with advanced cancer. *Support Care Cancer*. 2009;17(2):191-198. doi:10.1007/s00520-008-0477-2
24. Diehr P, Lafferty WE, Patrick DL, Downey L, Devlin SM, Standish LJ. Quality of life at the end of life. *Health Qual Life Outcomes*. 2007;5(1):51. doi:10.1186/1477-7525-5-51
25. Li Z, Tosteson TD, Bakitas MA. Joint modeling quality of life and survival using a terminal decline model in palliative care studies. *Stat Med*. 2013;32(8):1394-1406. doi:10.1002/sim.5635
26. Claessens AKM, Ramaekers BLT, Lobbezoo DJA, et al. Quality of life in a real-world cohort of advanced breast cancer patients: a study of the SONABRE Registry. *Qual Life Res*. 2020;29(12):3363-3374. doi:10.1007/s11136-020-02604-4
27. Verhulst MN, Hjerstad MJ, Van Belle S, Kaasa S, Deliens L, Pardon K. Quality of life and symptom intensity over time in people with cancer receiving palliative care: Results from the international European Palliative Care Cancer Symptom study. Montazeri A, ed. *PLOS ONE*. 2019;14(10):e0222988. doi:10.1371/journal.pone.0222988
28. Giesinger JM, Wintner LM, Oberguggenberger AS, et al. Quality of Life Trajectory in Patients with Advanced Cancer during the Last Year of Life. *J Palliat Med*. 2011;14(8):904-912. doi:10.1089/jpm.2011.0086
29. Pe M, Alanya A, Falk RS, et al. Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials-Innovative Medicines Initiative (SISAQOL-IMI): stakeholder views, objectives, and procedures. *Lancet Oncol*. 2023;24(6):e270-e283. doi:10.1016/S1470-2045(23)00157-2
30. Kurland BF, Heagerty PJ. Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics*. 2005;6(2):241-258. doi:10.1093/biostatistics/kxi006
31. Olarte Parra C, Daniel RM, Bartlett JW. Hypothetical Estimands in Clinical Trials: A Unification of Causal Inference and Missing Data Methods. *Stat Biopharm Res*. 2023;15(2):421-432. doi:10.1080/19466315.2022.2081599
32. Shardell M, Miller RR. Weighted estimating equations for longitudinal studies with death and non-monotone missing time-dependent covariates and outcomes. *Stat Med*. 2008;27(7):1008-1025. doi:10.1002/sim.2964
33. Haug N, Jänicke M, Kasenda B, Marschner N, Frank M. Quantifying bias due to missing data in quality of life surveys of advanced-stage cancer patients. *Qual Life Res*. 2024;33(4):1085-1094. doi:10.1007/s11136-023-03588-7
34. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2013;22(3):278-295. doi:10.1177/0962280210395740
35. Vansteelandt S, Carpenter J, Kenward MG. Analysis of Incomplete Data Using Inverse Probability Weighting and Doubly Robust Estimators. *Methodology*. 2010;6(1):37-48. doi:10.1027/1614-2241/a000005
36. Tilling K, Williamson EJ, Spratt M, Sterne JAC, Carpenter JR. Appropriate inclusion of interactions was needed to avoid bias in multiple imputation. *J Clin Epidemiol*. 2016;80:107-115. doi:10.1016/j.jclinepi.2016.07.004
37. Engels JM, Diehr P. Imputation of missing longitudinal data: a comparison of methods. *J Clin Epidemiol*. 2003;56(10):968-976. doi:10.1016/S0895-4356(03)00170-7
38. Huque MH, Carlin JB, Simpson JA, Lee KJ. A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Med Res Methodol*. 2018;18(1):168. doi:10.1186/s12874-018-0615-6
39. R Core Team. R: A language and environment for statistical computing. Published online 2022. <https://www.R-project.org/>
40. Van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*. 2007;16(3):219-242. doi:10.1177/0962280206074463
41. Buuren SV, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011;45(3). doi:10.18637/jss.v045.i03
42. Harrell FE. Hmisc: Harrell Miscellaneous. Published online 2022. <https://CRAN.R-project.org/package=Hmisc>
43. Rouanet A, Helmer C, Dartigues JF, Jacqmin-Gadda H. Interpretation of mixed models and marginal models with cohort attrition due to death and drop-out. *Stat Methods Med Res*. 2019;28(2):343-356. doi:10.1177/0962280217723675
44. Blackhall F, Ross Camidge D, Shaw AT, et al. Final results of the large-scale multinational trial PROFILE 1005: efficacy and safety of crizotinib in previously treated patients with advanced/metastatic ALK-positive non-small-cell lung cancer. *ESMO Open*. 2017;2(3):e000219. doi:10.1136/esmoopen-2017-000219
45. Fayers P, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A. *EORTC QLQ-C30 Scoring Manual*. Third edition. European Organisation for Research and Treatment of Cancer; 2001.
46. Fayers P, Bottomley A. Quality of life research within the EORTC—the EORTC QLQ-C30. *Eur J Cancer*. 2002;38:125-133. doi:10.1016/S0959-8049(01)00448-8
47. Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. *BMC Med Res Methodol*. 2019;19(1):46. doi:10.1186/s12874-019-0666-3
48. Von Hippel PT. How Many Imputations Do You Need? A Two-stage Calculation Using a Quadratic Rule. *Sociol Methods Res*. 2020;49(3):699-718. doi:10.1177/0049124117747303

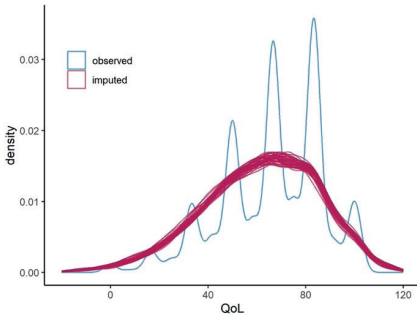
49. Bodner TE. What Improves with Increased Missing Data Imputations? *Struct Equ Model Multidiscip J*. 2008;15(4):651-675. doi:10.1080/10705510802339072
50. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30(4):377-399. doi:10.1002/sim.4067
51. Robitzsch A, Grund S. miceadds: Some Additional Multiple Imputation Functions, Especially for “mice.” Published online 2024. <https://CRAN.R-project.org/package=miceadds>
52. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using **lme4**. *J Stat Softw*. 2015;67(1). doi:10.18637/jss.v067.i01
53. Jared E. Knowles, Carl Frederick, Alex Whitworth. merTools: Tools for Analyzing Mixed Effect Regression Models. Published online February 8, 2024. <https://cran.r-project.org/web/packages/merTools/merTools.pdf>
54. Carpenter JR, Kenward MG. *Multiple Imputation and Its Application*. 1st ed. Wiley; 2013. doi:10.1002/9781119942283
55. Quartagno M, Carpenter JR. Multiple imputation for IPD meta-analysis: allowing for heterogeneity and studies with missing covariates. *Stat Med*. 2016;35(17):2938-2954. doi:10.1002/sim.6837
56. Quartagno M, Grund S, Carpenter J. jomo: A Flexible Package for Two-level Joint Modelling Multiple Imputation. *R J*. 2019;11(2):205. doi:10.32614/RJ-2019-028
57. Cappelleri JC, Chambers R. Addressing Bias in Responder Analysis of Patient-Reported Outcomes. *Ther Innov Regul Sci*. 2021;55(5):989-1000. doi:10.1007/s43441-021-00298-5
58. Cappelleri JC. Further reduction in statistical power for responder analysis of patient-reported outcomes with measurement error. *J Clin Epidemiol*. 2021;140:200-201. doi:10.1016/j.jclinepi.2021.08.017
59. Floden L, Bell ML. Imputation strategies when a continuous outcome is to be dichotomized for responder analysis: a simulation study. *BMC Med Res Methodol*. 2019;19(1):161. doi:10.1186/s12874-019-0793

APPENDIX
SUPPLEMENTARY FIGURES

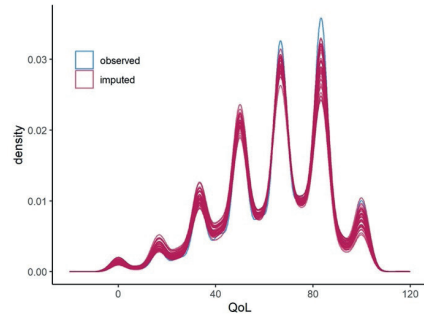
Timepoint (<i>j</i>)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
PRO measurement Y_{ij} O=available X=missing	O	O	O	O	X	O	X	X	O	X	X	X	-	-
Intercurrent events E_{ik}, S_i	PD $E_{i1} = 1$ $T_{i1} = 8$ TD $E_{i2} = 1$ $T_{i2} = 9$ Death $D_i = 1$ $S_i = 12$													
E_{ij1}	0	0	0	0	0	0	0	1	1	1	1	1		
E_{ij2}	0	0	0	0	0	0	0	0	1	1	1	1		
Time till PD (t_{ij1})	7	6	5	4	3	2	1	0	-1	-2	-3	-4		
Time till TD (t_{ij2})	8	7	6	5	4	3	2	1	0	-1	-2	-3		
Time till death (s_{ij})	11	10	9	8	7	6	5	4	3	2	1	0		
X_i , e.g., patient age at baseline	67	67	67	67	67	67	67	67	67	67	67	67		

Figure S1 Schematic representation of PRO data availability for a single example patient *i* throughout a study with 14 PRO measurement times. Corresponding variables that may be used for imputation of missing PRO data for patient *i* are shown below each planned measurement. *PD: progression of disease, TD: treatment discontinuation.

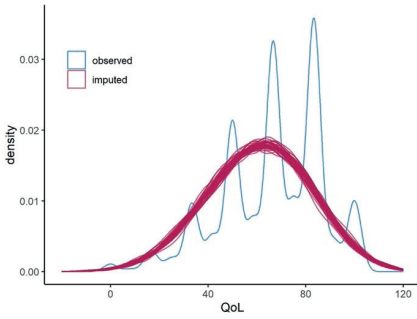
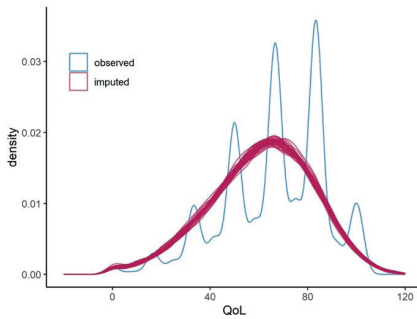
For example, to impute the missing PRO value at timepoint $j = 5$, we use, besides the observed PRO information of patient *i* over time, that the patient did not experience disease progression yet, that the remaining time until observed disease progression is 3 timepoints, that discontinuation has not yet occurred but is observed 4 timepoints later, and that death was observed 7 timepoints later. Additionally, we can use other variables that are assumed to be informative for the missing PRO value such as the patient's age at baseline.



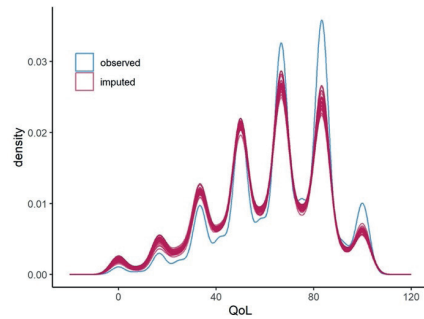
A. mice two-level normal model



B. miceadds two-level pmm

C. lme4 two-level normal model +
predictInterval

D. aregImpute normal model



E. aregImpute predictive mean matching

Figure S2 The distribution of QoL data in the available data (blue) and the imputed datasets (red) for five implemented imputation methods.