Received: 13 August 2020 | Revised: 21 May 2021 | Accepted: 22 May 2021

DOI: 10.1111/stan.12250

ORIGINAL ARTICLE

WILEY

Information anchored reference-based sensitivity analysis for truncated normal data with application to survival analysis

Andrew Atkinson[1,2] | **Suzie Cro**[3] | **James R. Carpenter**[1,4] | **Michael G. Kenward**[5]

[1]Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK

[2]Department of Infectious Diseases, Bern University Hospital, Inselspital, University of Bern, Bern, Switzerland

[3]School of Public Health, Faculty of Medicine, Imperial College, London, UK

[4]MRC Clinical Trials Unit at UCL, London, UK

[5] Ashkirk, UK

Correspondence
Andrew Atkinson, Department of Infectious Diseases, Bern University Hospital, Friedbuehlstrasse 51, 3010 Bern, Switzerland.
Email: andrew.atkinson@insel.ch

Funding information
Medical Research Council, Grant/Award Number: MC_UU_12023/21; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: CRSK-3_190977/1

The primary analysis of time-to-event data typically makes the censoring at random assumption, that is, that—conditional on covariates in the model—the distribution of event times is the same, whether they are observed or unobserved. In such cases, we need to explore the robustness of inference to more pragmatic assumptions about patients post-censoring in *sensitivity analyses*. Reference-based multiple imputation, which avoids analysts explicitly specifying the parameters of the unobserved data distribution, has proved attractive to researchers. Building on results for longitudinal continuous data, we show that inference using a Tobit regression imputation model for reference-based sensitivity analysis with right censored log normal data is *information anchored*, meaning the proportion of information lost due to missing data under the primary analysis is held constant across the sensitivity analyses. We illustrate our theoretical results using simulation and a clinical trial case study.

KEYWORDS

censoring not at random, informative censoring, reference-based multiple imputation, Rubin's rules, sensitivity analysis, tobit regression, truncated normal data

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2021 The Authors. *Statistica Neerlandica* published by John Wiley & Sons Ltd on behalf of Netherlands Society for Statistics and Operations Research.

source: https://doi.org/10.48350/162029 | downloaded: 12.10.2024

1 | INTRODUCTION

However carefully clinical trials are designed and planned, some outcome data are often missing. This occurs when a patient is lost to follow-up, which could be, for example, due to noncompliance with the study protocol, or stopping an assigned treatment due to experiencing adverse effects.

When the outcome of interest is the time to an event, patients censored at their last-known time point can be considered as generating a type of missing data. The event occurs after they were last seen. Any analysis must make inherently untestable assumptions about the distribution of the time from the last observed visit to the event. Therefore, censored data cause unavoidable ambiguity in the analysis. Since we never know the true event time for censored data, it is important to be clear about the assumptions made about such data, and the subsequent impact of these assumptions on the conclusions. If a contextually implausible assumption concerning censored data is made, then the results from the analysis might lead to potentially misleading conclusions, which can directly influence patient care (among others Sterne et al., 2009; White & Carlin, 2010; Ibrahim, Chu, & Chen, 2012; Jakobsen, Gluud, & Winkel, 2017).

Therefore, in a trial setting the ideal is for the primary analysis to make the most contextually plausible assumptions about the censored data; the conclusion from the primary analysis is then tested using sensitivity analysis which makes different, contextually plausible assumptions concerning the censored data.

In practice, the primary analysis of time-to-event data most often makes the censoring at random (CAR) assumption, that is, that—conditional on covariates in the analysis model—the distribution of event times is the same, whether they are observed or unobserved. In a sensitivity analysis, postcensoring behavior will typically then assume data are censored not at random (CNAR), also known as informative censoring.

If the results from the primary and sensitivity analysis are broadly consistent, we can conclude that for the sensitivity analysis scenarios investigated, the outcome from the primary analysis is robust to contextually plausible departures from the primary assumption concerning the missing data mechanism. If this is not the case, and the conclusions change following the sensitivity analysis, then the investigators should report the conditions under which the conclusions may change, along with the relative likeliness of these circumstances occurring. These steps provide more confidence in the results, especially when regulators are considering new treatments for approval. Both the European Medicines Agency (EMA) and U.S. Food and Drug Administration have underlined the importance of performing sensitivity analyses in clinical trials (CHMP, 2010; NRC, 2010), with the recent EMA ICH E9 (R1) addendum providing clarity in terms of both definitions and example scenarios for such sensitivity analyses CHMP (2019).

Multiple imputation (MI) has become a popular method for such missing data analyses both for the primary analysis and subsequent sensitivity analyses, since there is software enabling accessible implementation (e.g., *proc MI* SAS, 2003, *mi impute* and *mimix* Cro, Morris, Kenward, & Carpenter, 2016; StataCorp., 2013, *mice* and *InformativeCensoring* Ruau et al., 2016; van Buuren & Groothuis-Oudshoorn, 2011).

MI involves forming a suitable imputation model based on the observed data. Typically, this will take the form of a regression model, with asymptotically normally distributed parameter estimates. Repeated draws from this conditional predictive distribution of the missing data are made to create multiple completed datasets. Each completed dataset is then analyzed using the substantive analysis model of interest to obtain multiple treatment estimates (Carpenter &

Kenward, 2013, chapter 2). Treatment and variance estimates are then averaged across multiply imputed data sets using Rubin's MI combination rules; provided the imputations are proper (e.g., Rubin 1987), statistical inferences have good frequentist properties.

Reference-based multiple imputation (RBMI), an approach originally proposed by Little and Yau (1996), has recently gained popularity as a sensitivity analysis approach due to its inherent practicality. Missing data are imputed by *reference* to an appropriately chosen group, or groups of patients from the *observed* data. For example, observed data from one arm of a clinical trial can be *re-used* to impute all unobserved data in the trial under a particular assumption, where all unobserved are expected to behave like the specified "reference" arm. Recent practical applications of RBMI are numerous, with examples in the fields of psychotherapy, endocrinology, and paediatric and geriatric care (e.g. Philipsen et al. (2015), Jans et al. (2015), Billings et al. (2018), Atri, Frölich, Ballard, et al. (2018)).

Carpenter, Roger, and Kenward (2013) defined a number of approaches for RBMI sensitivity analysis for longitudinal data with a continuous endpoint, and one of these, "Jump to Reference," has recently been extended to the time-to-event setting (Atkinson, Kenward, Clayton, and Carpenter (2019)). In summary, Jump to Reference imputes unobserved data using an imputation distribution that has parameters from the patient's randomized arm up to their last observed time point, with parameters from a specified reference arm for unobserved time points.

With RBMI methods, the model used for imputing data and analyzing data is no longer the same (i.e., *uncongenial*, Meng, 1994), and this has provoked a discussion on the operating characteristics of Rubin's variance estimate in this context (Hughes, Sterne, & Tilling, 2014; Kim, Zeng, & Taylor, 2017; Liu & Peng, 2016; Robins & Wang, 2000; Tang, 2018). In addition, with RBMI we can intuitively understand that since we reuse data, in the analysis we increase the homogeneity between the groups we are seeking to compare. If we are not careful, this dilution or mixing effect between groups decreases the variability of the resulting treatment estimate. We view this as undesirable, since this reduced variability may be seen as a "reward" for missing data.

However, recently Cro, Carpenter, and Kenward (2019) proved that, at least for continuous longitudinal data, RBMI is *information anchored*, meaning that the proportion of information lost due to missing data is held constant across the primary and sensitivity analysis. Here, we are using "information" in a rather specific sense, defined to be the inverse of the derived estimator of the sampling variance.

Furthermore, in their recent paper Atkinson et al. showed that, counterintuitively, the empirical variance decreases as the proportion of censored data increases (columns 7 and 8 of Table 1 of reference 2), whereas information anchoring was shown to hold for Rubin's variance estimator (albeit with simulated data). Building on this work, and that of Cro et al., in this article we provide theoretical results showing that information anchoring also holds for a reference-based sensitivity analysis with a Tobit imputation model assuming truncated normal data. We illustrate the approach with right censored log normal survival times, where the estimate is the difference in mean survival times by treatment.

The remainder of the paper is organized as follows. The next section introduces our example clinical trial dataset which will be used to illustrate the theoretical results. Section 3 defines information anchoring in more detail. Section 4 presents the theoretical results, which are then compared with simulated data in Section 5, along with an illustration of the approach based on the clinical trial data in Section 6. We finish with a discussion of the results and possible areas for further study.

TABLE 1 Difference between Rubin's multiple imputation (MI) variance estimator under Jump to Reference (J2R) and the information anchored variance estimate comparing theoretical estimates with empirical estimates using simulated data

n_a	Proportion of missingness (π_d)	Theoretical estimate of information anchored variance (A)	Empirical estimate of information anchored variance (C)	Theoretical estimate of MI variance under J2R (B)	Empirical estimate of MI variance under J2R (D)	Difference theoretical estimates (A-B)	Difference empirical estimates (C-D)
250	0.1	0.0081	0.0080	0.0080	0.0080	0.00009	0.00002
250	0.2	0.0081	0.0080	0.0082	0.0080	−0.00004	0.00005
250	0.3	0.0085	0.0082	0.0083	0.0081	0.00017	0.00012
250	0.4	0.0090	0.0085	0.0087	0.0084	0.00034	0.00017
250	0.5	0.0098	0.0091	0.0092	0.0089	0.00064	0.00020

Notes: Comparison of variances under J2R: A, Theoretical calculation of information anchored variance (the gold standard); B, Theoretical calculation of Rubin's MI variance; C, Empirical estimate of information anchored variance; D, Empirical estimate of MI variance under J2R.

2 | MOTIVATING DATA

The Second Randomized Intervention Treatment of Angina (RITA-2) trial (Henderson et al., 1997, 2003) randomized 1018 eligible coronary artery disease patients from the United Kingdom and Ireland to receive either Percutaneous Transluminal Coronary Angioplasty (PTCA, $n = 504$) or continued medical treatment ($n = 514$). Those patients randomized to angioplasty received the intervention in the first 3 months. The primary endpoint of the study was a composite of all cause mortality and definite nonfatal myocardial infarction. After 7 years, there were 73 deaths (14.5%) on the PTCA arm and 63 (12%) on the medical arm (difference in proportions +2.2%, 95% confidence interval [−2%, 6.4%], $p = .21$).

The study concluded that an initial policy of PTCA was associated with greater improvement in angina symptoms, with this effect being particularly present in patients with more severe angina, and that the increased risk of performing PTCA should be offset against these benefits.

RITA-2 was a pragmatic trial, so that although patients were initially randomized to PTCA or medical treatment, in the course of the follow-up patients received further procedures according to clinical need, and the trial was designed to compare a policy of beginning with medical treatment against a policy of beginning with PTCA. Subsequent nonrandom interventions (NRIs) were either PTCA, or when necessary, a coronary artery bypass graft (CABG). In the PTCA arm, 17.0% of patients had a second PTCA, while 12.7% had a CABG. By contrast, in the medical arm 27% went on to have a PTCA and 12.3% had a CABG.

The primary published analysis of our illustrative example, was an "Intention To Treat" analysis, that is, targeted a "treatment policy" estimand, with a composite outcome of time to first of death or nonfatal myocardial infarction. Consistent with this ITT approach, this analysis did not take account of nonrandomized interventions (of which, as we detail below, there were quite a number, especially in the medical therapy arm, where many patients "crossed-over" to PTCA). This suggested that our approach could be illustrated as follows: we artificially censor patients in the medical arm at the time of their first non-randomized intervention, typically PTCA. Our first analysis of the data with this artificial censoring uses within-arm imputation and therefore targets a per-protocol (i.e., hypothetical) estimand (in the language of EMA ICH E9 (R1) addendum). However, in our sensitivity analysis we impute event times for these artificially censored patients, but now using our method to assume that at these nonrandomized intervention times patients in the medical arm "jumped" to the hazard in the PTCA arm. As this is effectively what happened (because the typical nonrandomized intervention for medical arm patients was PTCA), this sensitivity analysis is targeting the treatment policy estimand. We did this to provide an empirical evaluation of the utility of our approach. This is because if our method works as intended, our sensitivity analysis is targeting the ITT estimand, and we expect it to give similar results to the original ITT analysis. Should our approach give clinically sensible results when we can cross-check it with actual data, we will have more confidence in applying it in typical applications, where such cross-checking is not possible.

In the next section we define information anchored sensitivity analysis.

3 | INFORMATION ANCHORED SENSITIVITY ANALYSIS

If we consider a two-arm clinical trial such as the RITA-2 trial, the aim of a sensitivity analysis is to assess the behavior of the treatment estimate under alternative, clinically plausible, CNAR scenarios for post-censoring data. Since in RITA-2 post-censored patients in both treatment arms

underwent PTCA, we propose using RBMI to model the post-censoring behavior based on the PTCA arm; patients "jump" from the medical arm to the PTCA arm. Therefore, we estimate the event time distribution of the PTCA arm in the usual manner, and use this to impute potential post-censoring behavior on the medical arm.

However, variance estimators for the primary analysis may behave in an unexpected manner under certain sensitivity analysis scenarios. Indeed, there are examples in which the usual variance estimator with a reference-based method *decreases* as the proportion of missing values increases (refer to Appendix A of Data S1). Such counter-intuitive properties would undermine our confidence in the approach and of course would reward trialists for losing data!

Therefore, to justify the use of RBMI it is important to assess this is not the case by quantifying the amount of statistical information available in the sensitivity analysis relative to the primary analysis. This will establish whether the sensitivity analysis is injecting new information, or taking away information, relative to the primary analysis. Cro et al. proposed the principle of *information anchoring* to address this, which we now formally define in the context of time-to-data.

Consider a clinical trial in which time-to-event data is collected from patients, denoted by Y, in order to estimate a treatment effect θ. We denote the data from those patients experiencing the event by Y_{obs}, and the data of those censored by Y_{cens}. We make a primary set of assumptions, for example, that all censored patients are "censored at random" (CAR). The estimate of θ under this primary assumption is denoted by $\widehat{\theta}_{\text{obs,CAR}}$, which, for a valid estimate under CAR, is calculated using both the observed event *and* censoring times.

Furthermore, let us assume that we are able to observe a realization of the actual event times for the censored patients, $Y_{\text{cens-realized,CAR}}$, under the primary assumption of CAR. Of course, this is a hypothetical construct, but it will help to frame the definition of information anchoring.

Taken together, the observed data, Y_{obs}, and the realization of the event times for the censored patients, $Y_{\text{cens-realized,CAR}}$, we obtain a *full* data set under the primary CAR assumption. We define $\widehat{\theta}_{\text{full},primary}$ to be the corresponding estimate of θ after fitting the primary analysis model to this full dataset.

For the sensitivity analysis, we make a different set of assumptions concerning the distribution of post-censoring data, that is, scenarios in which censoring is assumed to be informative (i.e., censored *not*at random).

Defined in the same manner as for the primary analysis, for the sensitivity analysis we have $\widehat{\theta}_{\text{obs,sensitivity}}$ and $\hat{\theta}_{\text{full,sensitivity}}$, whereby "full" is defined analogously, assuming we are able to observe the posited post-censoring behavior from our hypothetical construct denoted $Y_{\text{cens-realized,sens}}$, associated with a specific set of assumptions for the sensitivity analysis (e.g., "jump to PTCA arm").

Furthermore, we define the observed information about θ under the primary and sensitivity analyses by $I(\dots)$. Since there is less information when there is censored data, then we would expect the ratio $\frac{I(\hat{\theta}_{\text{full},primary})}{I(\hat{\theta}_{\text{obs},primary})} > 1$ and $\frac{I(\hat{\theta}_{\text{full,sensitivity}})}{I(\hat{\theta}_{obs,sensitivity})} > 1$.

Information anchored sensitivity analysis is defined as:

$$\frac{I(\hat{\theta}_{\text{full},primary})}{I(\hat{\theta}_{\text{obs},primary})} = \frac{I(\hat{\theta}_{\text{full,sensitivity}})}{I(\hat{\theta}_{\text{obs,sensitivity}})}, \tag{1}$$

so that the proportion of information lost due to missing data is constant across primary *and* sensitivity analyses. If Equation (1) above holds then we say that the sensitivity analysis is

information anchored with regards to the primary analysis Cro et al. (2019). In the next section, we show that for a two-arm clinical trial, and assuming a Tobit imputation model, RBMI for a time to event outcome is information anchored. We do this by estimating each of the components of Equation (1) above. We begin by defining the clinical trial setting and distributional assumptions, then provide variance estimates assuming fully observed behavior, before moving on to estimate the components under CAR, and finally provide the results for the specific informative censoring scenario of interest ("jump to reference").

4 | INFORMATION ANCHORED SENSITIVITY ANALYSIS FOR TIME-TO-EVENT DATA

4.1 | Clinical trial setting with time-to-event data

We consider a two-arm clinical trial in which patients are randomized either to a reference or active treatment arm. For example, the reference arm could be a placebo or standard of care treatment.

We assume the times from patient randomization to when an event occurs, typically death or treatment failure, are log normally distributed. For the primary endpoint we wish to estimate the difference in the log normal survival times between the reference and active treatment arms.

In addition, we also measure a baseline covariate per patient at the time of randomization. Furthermore, we assume that, due to randomization, the mean and variance of the baseline covariate is the same in both arms.

Patients are right censored if they deviate from the study protocol, or do not experience the event before the end of the study period. In addition, and without loss of generality, for ease of exposition we assume that patients *are only censored on the intervention arm*, and therefore those on the reference arm always experience the outcome event of interest (i.e., patients are fully observed). We assume the baseline covariate is fully observed on both arms, and further, that the log event times and baseline covariate are bivariate normally distributed.

As noted above, the treatment effect of interest is the difference in mean log time-to-event between the trial arms, and we will test if this difference is statistically significant at the 5% level using a standard t-test with shared sample variance estimate.

Throughout, we assume a single baseline covariate, but the approach generalizes to multivariable scenarios. Also note that we assume 1:1 randomization between the trial arms, so in the expectation there are equal numbers of patients in the reference and active treatment arms.

Denoting subscript r for patients on the reference arm, we therefore have baseline covariate measurements Y_{rj1} for the $j = 1, \dots, n_r$ patients, and their log event times Y_{rj2}:

$$\begin{pmatrix} Y_{rj1} \\ Y_{rj2} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_{r2} \end{pmatrix}, \begin{pmatrix} \sigma_{r11} & \sigma_{r12} \\ \sigma_{r12} & \sigma_{r22} \end{pmatrix}\right], \quad j = 1, \dots, n_r,$$

where r denotes the reference arm, and the baseline mean μ_1 is common between the reference and active arms.

Using the subscript a for the active treatment arm, we now differentiate between two sets of patients, those whose event times are observed ($\mathcal{O}$), and those whose event times are censored ($\mathcal{D}$, for deviating from protocol). Throughout, we assume that n_d of the n_a patients on the active

treatment arm are censored at some fixed time point τ, and the remaining n_o are followed until they have the event, so that $n_o + n_d = n_a$. The choice of a common censoring time τ allows the use of standard results for the truncated normal distribution in the derived expressions.

We again assume bivariate normality for the baseline covariate and event times:

$$\begin{pmatrix} Y_{aj1} \\ Y_{aj2} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_{a2} \end{pmatrix}, \begin{pmatrix} \sigma_{a11} & \sigma_{a12} \\ \sigma_{a12} & \sigma_{a22} \end{pmatrix}\right], \quad j \in \mathcal{O}$$
$$\begin{pmatrix} Y_{aj1} \\ Y_{aj2} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_{d2} \end{pmatrix}, \begin{pmatrix} \sigma_{a11} & \sigma_{d12} \\ \sigma_{d12} & \sigma_{d22} \end{pmatrix}\right], \quad j \in \mathcal{D},$$

For the primary analysis we assume censoring is at random, so that $\mu_{d2} = \mu_{a2}$. For the sensitivity analysis, we make the assumption of "Jump to Reference" (J2R) for those censored on the active arm.

We estimate the parameters of the unobserved data distribution from the *reference* arm, and use this to impute event times for censored patients on the active treatment arm, so that, post-censoring, these patients "jump to reference" (for details refer to Atkinson et al., 2019). An example of this post-censoring behavior is the "jump to PTCA" sensitivity analysis scenario noted above in the context of the RITA-2 trial. More formally, defining $S(t)$ to be the usual survival function, and assuming a patient is censored on the active arm at a time point T_i, in such a "jump to reference" scenario from this point on we use the survival function on the reference arm to impute an event time. So that,

$$S_{\text{post},i}(t|t > T_i, \text{active}) \ := S(t|t > T_i, \text{reference}),$$

where the subscript *post* denotes the post-censoring distribution. This expression is the basis for the multiple imputation process for the sensitivity analysis.

Since we are using a reference-based sensitivity analysis method, post imputation we retain the primary analysis method, the t-test, for the difference in the log normal survival times, and apply this to each of the multiply imputed datasets created under the J2R assumption for post-censoring behavior, and combine the estimates in the usual way with Rubin's rules.

Now, to confirm that the principle of information anchored sensitivity analyses holds for this scenario we require the following equality to hold, at least approximately:

$$\frac{I(\hat{\theta}_{\text{full},primary=\text{CAR}})}{I(\hat{\theta}_{\text{obs},primary=\text{CAR}})} = \frac{I(\hat{\theta}_{\text{full,sensitivity=J2R}})}{I(\hat{\theta}_{\text{obs,sensitivity=J2R}})}, \tag{2}$$

so that the proportion of information lost due to missing data is constant across primary and sensitivity analyses.

We build this result step by step, evaluating each of the components in the above equation in turn. We begin with the left-hand side of Equation (2), derive the information ratio $\frac{I(\hat{\theta}_{\text{full,CAR}})}{I(\hat{\theta}_{\text{obs,CAR}})}$. In the next subsection, we derive the upper part of this expression, when the data is fully observed for the primary analysis, $I(\hat{\theta}_{\text{full,CAR}})$. In Section 4.3, we estimate the variance following multiple imputation under CAR, $I(\hat{\theta}_{\text{obs,CAR}})$. Putting these two components together, we derive an expression for the left-hand side of the Equation (2) in Section 4.4 . The right-hand side of the equation is

then derived for the J2R sensitivity analysis scenario in Section 4.5, $\frac{I(\hat{\theta}_{\text{full,J2R}})}{I(\hat{\theta}_{\text{obs,J2R}})}$. Finally, we compare the two ratios in Section 4.6.

4.2 | Estimation of $V(\hat{\theta}_{\text{full,CAR}})$

On the active treatment arm, let us assume that we are able to observe a realization of the censored log event times under the primary assumption of CAR, $Y_{\text{cens-realized,CAR}}$, and we then combine this with the fully observed data, Y_{obs}, forming a full dataset under CAR. Fitting the primary analysis model to this dataset leads to a treatment estimate $\hat{\theta}_{\text{full,CAR}}$, the subscript making the CAR setting explicit. We start by deriving the expression for $V(\hat{\theta}_{\text{full,CAR}})$, the variance of this estimate.

Conditioning on n_d, the number of patients censored, the expected treatment effect is a weighted average of the mean time to censoring, and the mean time to event for those fully observed, compared to the mean survival time for all fully observed patients on the reference arm (μ_{r2}):

$$\hat{\theta}_{\text{full,CAR}} = \left(\frac{n_o}{n_a}\mu_{a2} + \frac{n_d}{n_a}\mu_{d2}\right) - \mu_{r2}.$$

The variance of this estimate for the fully observed data is:

$$V(\hat{\theta}_{\text{full,CAR}}) = \frac{\hat{\sigma}_{22,r}}{n_r} + \frac{\hat{\sigma}_{22,a}}{n_a} = \frac{\frac{1}{(n_r-1)}\sum_{j=1}^{n_r}(Y_{rj2} - \overline{Y}_{r2})^2}{n_r} + \frac{\frac{1}{(n_a-1)}\sum_{j=1}^{n_a}(Y_{aj2} - \overline{Y}_{a2})^2}{n_a}. \tag{3}$$

Now, we additionally assume equal numbers of patients in both arms of the study $n_r = n_a = n$, and equal variance in both arms, so that $\sigma_{r22} = \sigma_{a22} = \sigma_{22}$. The expected value of this expression under repeated draws, with the same underlying CAR assumption, is $E[V(\hat{\theta}_{\text{full,CAR}})] = \frac{2\sigma_{22}}{n}$. This establishes the expression for the top part of the left-hand side of Equation (2). In the next section, we calculate the post multiple imputation estimate for Rubin's variance under CAR.

4.3 | Rubin's variance estimate following imputation under CAR

4.3.1 | Tobit imputation model

To establish the properties of multiply imputed data on the active treatment arm, consistent with our underlying normal model, we use a Tobit model fitted to the observed data as the imputation model (Greene, 2003; Tobin, 1958). Based on estimates from this fitted model, we impute events for the censored patients, and finally derive the variance of the combined sets of observed and imputed data using Rubin's rules.

We now calculate the estimate of the variance under the CAR assumption following multiple imputation of the truncated (censored) data from the fitted Tobit model. We assume the estimates $\hat{\beta}$ from the fitted imputation model are normally distributed.

Although MI is formally Bayesian, we assume the observed data dominates the posterior distribution, and, using inferential arguments set out on pp. 56–60 of Carpenter and Kenward (2013),

we can, without any important loss of generality, assume the variance of the data is known in the imputation model.

Accordingly, the multiple imputation process is as follows:

1. Fit the Tobit model to the observed data on the active arm regressing the log normal survival times Y_{aj2} on the baseline covariate Y_{aj1}, including the appropriate likelihood terms for those censored at time τ:

$$Y_{aj2} = \hat{\beta}_0 + Y_{aj1}\hat{\beta}_1 + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \hat{\sigma}_{2.1}), \quad i = 1, \ldots, n_a,$$

   resulting in the maximum likelihood estimates $\hat{\beta}_0, \hat{\beta}_1$, and estimate of the residual variance $\hat{\sigma}_{2.1}$.
2. Approximate a draw from the Bayesian posterior distribution assuming a noninformative prior by drawing

$$\tilde{\sigma}_{2.1} = \frac{(n_o - 2)\hat{\sigma}_{2.1}}{X},$$

   where $X \sim \chi^2_{n_o-2}$. We assume the model estimates are bivariate normally distributed with mean $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)^T$ and covariate matrix:

$$\hat{\boldsymbol{V}} = \tilde{\sigma}_{2.1}\begin{bmatrix} n_o & \sum_{i=1}^{n_o}\hat{Y}_{aj1} \\ \sum_{i=1}^{n_o}\hat{Y}_{aj1} & \sum_{i=1}^{n_o}\hat{Y}^2_{aj1} \end{bmatrix}^{-1},$$

   we then take a draw from MVN($\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{V}}$), resulting in a vector of estimates $(\tilde{\beta}_0, \tilde{\beta}_1)$.
3. Impute the censored observations using the resulting regression model with one set of the estimates $(\tilde{\beta}_0, \tilde{\beta}_1)$, repeating this process until the imputed event time is greater than τ:

$$\tilde{Y}_{aj2} = \tilde{\beta}_0 + \tilde{\beta}_1\tilde{Y}_{aj1} + \tilde{\varepsilon}_i, \quad \tilde{\varepsilon}_i \sim \mathcal{N}(0, \tilde{\sigma}_{2.1}), \quad k = 1, \ldots, n_d,$$

   resulting in a complete dataset with no censored observations.
4. Repeat the above steps K times, resulting in K complete data sets.
5. Perform the substantive analysis, in this case the t-test, on each of the $k = 1, \ldots, K$ complete datasets in turn, resulting in estimates $\hat{\theta}_k$, $\hat{\sigma}^2_k$, which we combine to form overall estimates using Rubin's rules in the usual manner.

   The MI estimate of θ is $\hat{\theta}_{\text{MI,CAR}} = \frac{1}{K}\sum_{k=1}^{K}\hat{\theta}_k$, for $k = 1, \ldots K$.

   Rubin's variance estimator is defined as:

$$E[V(\hat{\theta}_{\text{MI,CAR}})] = E(\hat{W}) + \left(1 + \frac{1}{K}\right)E(\hat{B})$$

   where

$$\hat{W} = \frac{1}{K}\sum_{k=1}^{K}\hat{\sigma}^2_k, \tag{4}$$

and

$$\hat{B} = \frac{1}{(K-1)} \sum_{k=1}^{K} (\hat{\theta}_k - \hat{\theta}_{\mathrm{MI}})^2. \tag{5}$$

This is the standard MI procedure we would follow for the primary analysis assuming CAR. Now, to derive an estimate of Rubin's variance *analytically* we have to take a slightly different approach since there is no closed form solution to calculate the maximum likelihood estimators for the Tobit model.

4.3.2 | Estimation of $\boldsymbol{V}(\hat{\boldsymbol{\theta}}_{\boldsymbol{MI,CAR}})$

We begin by briefly presenting existing results for truncated normal distributions which we will subsequently use. Recall, n_d patients on the active arm are censored at a randomly defined, but fixed time point τ. Then,

$$E(\tilde{Y}_{a2j} \mid \tilde{Y}_{a2j} > \tau) = \mu_{a2} + \sqrt{\sigma_{22}} \left[\frac{\varphi\left(\frac{\tau-\mu_{a2}}{\sqrt{\sigma_{22}}}\right)}{1-\Phi\left(\frac{\tau-\mu_{a2}}{\sqrt{\sigma_{22}}}\right)} \right] = \mu_{a2} + \sqrt{\sigma_{22}}\lambda, \tag{6}$$

for $j = 1, \ldots, n_d$ patients censored at τ; φ and Φ being the density and CDF of the standard normal distribution, respectively. The fraction part (in large square brackets) of the expression above is known as the inverse Mills ratio (Greene (2003)), and,

$$\lambda = \left[\frac{\varphi\left(\frac{\tau-\mu_{a2}}{\sqrt{\sigma_{22}}}\right)}{1-\Phi\left(\frac{\tau-\mu_{a2}}{\sqrt{\sigma_{22}}}\right)} \right].$$

Note that the expression in equation (6) is just the "usual" expected value, with an additional term $\sqrt{\sigma_{22}}\lambda$, where λ is treated as a constant for specific values of μ_{a2}, σ_{22}, and τ.

Analogously, we also define:

$$E(\tilde{Y}_{a2j} \mid \tilde{Y}_{a2j} < \tau) = \mu_{a2} - \sqrt{\sigma_{22}} \left[\frac{\varphi\left(\frac{\tau-\mu_{a2}}{\sqrt{\sigma_{22}}}\right)}{\Phi\left(\frac{\tau-\mu_{a2}}{\sqrt{\sigma_{22}}}\right)} \right] = \mu_{a2} - \sqrt{\sigma_{22}}\dot{\lambda},$$

using the standard expression for the variance of the truncated normal distribution, and analogously:

$$\mathrm{VAR}(\tilde{Y}_{a2j} \mid \tilde{Y}_{a2j} > \tau) = \sigma_{a2d} = \sigma_{22} \left[1 - \frac{\varphi\left(\frac{\tau-\mu_{a2}}{\sqrt{\sigma_{22}}}\right)}{1-\Phi\left(\frac{\tau-\mu_{a2}}{\sqrt{\sigma_{22}}}\right)} \left[\frac{\varphi\left(\frac{\tau-\mu_{a2}}{\sqrt{\sigma_{22}}}\right)}{1-\Phi\left(\frac{\tau-\mu_{a2}}{\sqrt{\sigma_{22}}}\right)} - \left(\frac{\tau-\mu_{a2}}{\sqrt{\sigma_{22}}} \right) \right] \right]. \tag{7}$$

Without loss of generality, the truncation limit τ is assumed to be greater than the mean throughout the analysis.

We now derive the variance estimate under CAR. The direction we take is to write down an expression for a "typical" multiply imputed event time, and then work from there to derive an expression for Rubin's variance estimate. To do this, we combine our knowledge of the observed data, the properties of the bivariate normal distribution (see Appendix B of Data S1), and the results for the truncated normal distribution stated above.

The imputation model for the jth of n_d censored values, from the kth of K imputed datasets, is defined as:

$$\tilde{Y}_{aj2,k} = \tilde{Y}_{a2o,k} + \tilde{\beta}_k(Y_{aj1} - \overline{Y}_{a1o}) + \sqrt{\tilde{\sigma}_{22,k}}\tilde{\tilde{\lambda}} + \sqrt{\tilde{\sigma}_{22,k}}\tilde{\lambda} + \tilde{\epsilon}_{j,k}, \quad j \in \mathcal{D}, \quad k = 1, \dots, K, \tag{8}$$

where $\tilde{Y}_{a2o,k}$ is the expected value of the observed survival times for the censored patients, with

$$\tilde{\sigma}_{2.1,k}|Y_o, \hat{\sigma}_{2.1} \sim \frac{\hat{\sigma}_{2.1}(n_o - 2)}{\chi^2_{n_o-2}},$$

where $\hat{\sigma}_{2.1}$ is the estimate of the residual variance from the fitted Tobit model, or equivalently, using the properties of bivariate normality $\hat{\sigma}_{2.1} = \hat{\sigma}_{22} - \frac{\hat{\sigma}^2_{12}}{\hat{\sigma}_{11}}$;

$$\bar{\tilde{Y}}_{a2o,k}|Y_o, \tilde{\sigma}_{2.1,k} \sim \mathcal{N}(\overline{Y}_{a2o}, n_o^{-1}\tilde{\sigma}_{2.1,k}),$$

$$\tilde{\beta}_k|Y_o, \tilde{\sigma}_{2.1,k} \sim \mathcal{N}(r/q, q^{-1}\tilde{\sigma}_{2.1,k}),$$

where $r = \sum_{i=1}^{n_o}(Y_{aj1} - \overline{Y}_{a1o})(Y_{aj2} - \overline{Y}_{a2o})$ and $q = \sum_{i=1}^{n_o}(Y_{aj1} - \overline{Y}_{a1o})^2$; the coefficient $\hat{\beta}_k$ of the regression model is $\frac{\sigma_{12}}{\sigma_{11}}$;

$$\sqrt{\tilde{\sigma}_{22,k}} \mid \hat{\sigma}_{22} \sim \frac{\sqrt{\hat{\sigma}_{22}}(n_a - 2)}{\chi_{n_a-2}},$$

where $\hat{\sigma}_{22}$ is the sampling variance of the log survival times on the active treatment arm. We note at this point that we use the χ^2-distribution here as an estimate of the sampling variance of the *standard deviation* of $\hat{\sigma}_{22}$, $\sqrt{\hat{\sigma}_{22}}$; and finally,

$$\tilde{\varepsilon}_{j,k}|\ \bar{\tilde{Y}}_{a2d,k},\ \tilde{\sigma}_{22,k} \sim Tr\mathcal{N}(0,\ \tilde{\sigma}_{a2d,k}\ ,\ a = (\tau - \bar{\tilde{Y}}_{a2d,k})),$$

with $\tilde{\sigma}_{a2d,k}$ the expected value of the unobserved survival times for the censored patients. The right-hand side of the above expression denotes the truncated normal distribution with mean 0 and variance $\tilde{\sigma}_{a2d,k}$, truncated on the left-hand side at $(\tau - \bar{\tilde{Y}}_{a2d,k})$; we use this relocation so that the mean of this expression is centered at zero, with the variance as we require, *and* we ensure that multiply imputed events are greater than the original censoring time for patient j. To estimate the variance for both the observed and multiply imputed data, we have to evaluate Equation (3) in Section 4.2. Since there is no missingness on the reference arm, the first part of this expression can be calculated directly,

$$\frac{E[\hat{\sigma}^2_r]}{n} = \frac{\sigma_{22}}{n}. \tag{9}$$

For the second part of Equation (3) pertaining to the active treatment arm, we decompose the summation into observed and censored parts, substituting our new expressions for $\tilde{Y}_{aj2,k}$ and $\overline{Y}_{a2,k}$ (for the latter, refer to Appendix C of Data S1),

$$(n_a - 1)\hat{\sigma}_a^2 = E\left(\sum_{j\in o}(Y_{aj2} - \hat{\mu}_{a2,k})^2\right) + E\left(\sum_{j\in d}(\hat{Y}_{aj2,k} - \hat{\mu}_{a2,k})^2\right). \tag{10}$$

Now, to calculate the above expression following multiple imputation we need to consider *both* the components of Rubin's variance estimator, that is, the between and within imputation variance estimates (refer to Equations (4) and (5)):

Proposition 1. *Letting* $(n_a - 1) \approx n_a$ *and* $K \to \infty$,

$$\begin{aligned} E(V[\hat{\theta}_{\text{MI,CAR}}]) \approx{}& \frac{\sigma_{22}}{n} + \pi_d(1-\pi_d)(\text{VAR}(w_{k,\dot{\lambda}}) + \text{VAR}(w_{k,\lambda})) \\ &+ \frac{1}{n}\left[\frac{\sigma_{12}^2}{\sigma_{11}} + \frac{2\sigma_{2.1}}{n_o}\right] + \frac{\pi_d}{n}\sigma_{2.1} + \pi_d^2\left(\frac{\sigma_{2.1}}{n_o} + \frac{\sigma_{a2d}}{n_d} + \text{VAR}(w_{k,\lambda})\right), \end{aligned} \tag{11}$$

where $w_{k,\lambda}$ *and* $w_{k,\dot{\lambda}}$ *are normally distributed variables with mean 0, which encompass the variability of the inverse mills ratio terms in equation (8).*

To arrive at the pooled variance of the treatment difference under CAR following MI, we just add the expression above to the variance for the reference arm, $\frac{E[\hat{\sigma}_r^2]}{n} = \frac{\sigma_{22}}{n}$, and obtain,

$$\begin{aligned} E[V(\hat{\theta}_{\text{MI,CAR}})] ={}& \frac{2\sigma_{22}}{n} + \pi_d(1-\pi_d)(\text{VAR}(w_{k,\dot{\lambda}}) + \text{VAR}(w_{k,\lambda})) \\ &+ \frac{1}{n}\left[\frac{\sigma_{12}^2}{\sigma_{11}} + \frac{2\sigma_{2.1}}{n_o}\right] + \frac{\pi_d}{n}\sigma_{2.1} + \pi_d^2\left(\frac{\sigma_{2.1}}{n_o} + \frac{\sigma_{a2d}}{n_d} + \text{VAR}(w_{k,\lambda})\right). \end{aligned} \tag{12}$$

Proof. Refer to Appendices C and D. ■

4.4 | Estimation of $\frac{V(\hat{\theta}_{\text{MI,CAR}})}{V(\hat{\theta}_{\text{full,CAR}})}$

We now have the building blocks for the first result, concerning the information ratio $\frac{I(\hat{\theta}_{\text{full},primary})}{I(\hat{\theta}_{\text{obs},primary})}$ which, under the primary assumption of CAR, is rewritten as $\frac{I(\hat{\theta}_{\text{full,CAR}})}{I(\hat{\theta}_{\text{obs,CAR}})}$. In Equation (3) of Section 4.2 we defined an estimate of $I(\hat{\theta}_{\text{full,CAR}})$ for the hypothetical case of no censoring. In the previous section, the expression for $E[V(\hat{\theta}_{\text{MI,CAR}})]$ was obtained, which is an estimate of $1/I(\hat{\theta}_{\text{obs,CAR}})$. Therefore, the required ratio on the left-hand side of Equation (3) can be estimated by calculating $\frac{E[V(\hat{\theta}_{\text{MI,CAR}})]}{E[V(\hat{\theta}_{\text{full,CAR}})]}$.

Lemma 1. *The ratio of the information in the full data relative to that in the incomplete data following multiple imputation assuming CAR, using the asymptotic expressions for Rubin's variance estimator, as K tends to infinity, is bounded above by*

$$\frac{I(\hat{\theta}_{\text{full,CAR}})}{I(\hat{\theta}_{\text{MI,CAR}})} = \frac{E[V(\hat{\theta}_{\text{MI,CAR}})]}{E[V(\hat{\theta}_{\text{full,CAR}})]} \lesssim 1 + \frac{\rho^2}{2} + (1-\rho^2)\left[\frac{1}{n_o} + \pi_d + \pi_d^2 + \pi_d^3 + \pi_d^4 + \ldots\right] + \frac{\pi_d}{2} + \pi_d C\left[(1-\pi_d)\dot{\lambda}^2 + \lambda^2\right], \tag{13}$$

assuming $n = n_a = n_r, \pi_d = \frac{n_d}{n}$, *and* $\rho^2 = \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}$, *which is the correlation between the baseline measurements and the event times squared.* C *is the variance of the standard deviation of* σ_{22}, *a* χ^2*-distributed variable, so that* $C = \sigma^2\left[N - 1 - \left[\sqrt{2}\,\frac{\Gamma((N+1)/2)}{\Gamma(N/2)}\right]^2\right]$, *with* $N = n_a$ *(e.g., p. 171 of* Kenney & Keeping, 1951).

Proof. Refer to Appendix E of Data S1. ■

For the principle of information anchoring to hold, the ratio assuming CAR shown above should be, at least approximately, the same numerically as that for the sensitivity analysis following multiple imputation under the Jump to Reference assumption for censored patients.

4.5 | Estimation of $\frac{V(\hat{\theta}_{\text{MI,J2R}})}{V(\hat{\theta}_{\text{full,J2R}})}$

Under J2R, the n_d-censored patients obtain multiply imputed event times based on the reference-arm hazard. This has the effect of reducing the difference between the estimated event times on the two arms, since we now have n_d additional observations generated under the hazard from the reference arm. We referred to this phenomenon in our introductory comments as the "dilution" or "mixing" effect.

Lemma 2. *The ratio of the information in the full data relative to the incomplete data under the assumption of J2R is:*

$$\frac{I(\hat{\theta}_{\text{full,sensitivity}})}{I(\hat{\theta}_{\text{obs,sensitivity}})} = \frac{E[V(\hat{\theta}_{\text{MI,J2R}})]}{E[V(\hat{\theta}_{\text{full,J2R}})]}, \tag{14}$$

with

$$\begin{aligned} E[V(\hat{\theta}_{\text{MI,J2R}})] = &\left[\frac{\sigma_{22}}{n} + (1-\pi_d)\sigma_{a2o} + \pi_d\sigma_{a2d}\right] \\ &+ \pi_d(1-\pi_d)\text{VAR}(w_{k,\lambda_{r2}}) + 2\pi_d(1-\pi_d)\lambda_{r2}\sqrt{\sigma_{22}}\Delta_c + \pi_d(1-\pi_d)\Delta_c^2 \\ &+ \pi_d(1-\pi_d)\sigma_{22}\lambda_{r2}^2 + \frac{(1-\pi_d)^2}{n}\rho^2\sigma_{22} + \frac{3\pi_d(1-\pi_d)^2}{n^2}\sigma_{22}(1-\rho^2) \\ &+ \pi_d^2\left(\frac{\sigma_{a2d}}{n_d} + \text{VAR}(w_{k,\lambda}) + \left[\frac{1}{n} + \frac{(1+\pi_d)}{n^2\pi_d}\right]\sigma_{22}(1-\rho^2)\right), \end{aligned} \tag{15}$$

and,

$$E[V(\hat{\theta}_{\text{full,J2R}})] = \frac{\sigma_{22}}{n} + \frac{(1-\pi_d)\sigma_{a2o} + \pi_d\sigma_{a2d} + \pi_d(1-\pi_d)\Delta_c^2}{n}, \tag{16}$$

where $\Delta_c = \mu_{a2d} - \mu_{a2o}$, *with the inverse mills ratio calculated assuming* $N(\mu_{r2}, \sigma_{22})$.

Proof. For the derivation of Equations (15) and (16) refer to Appendices F and G. The proof of Lemma 2 we delay until we have stated the theorem in full. ■

Now, $E[V(\hat{\theta}_{\text{MI,J2R}})]$ in Equation (15) is a rather complicated expression, but if we focus on terms of $\text{O}\left(\frac{1}{n}\right)$ or larger, it simplifies to an expression quite similar to that which was derived for the CAR case. In fact, both expressions are dominated by the first term enclosed in brackets, but the expression under J2R starts with a term in $\frac{\sigma_{22}}{n}$ rather than $\frac{2\sigma_{22}}{n}$, which was the first term of the analogous expression under the CAR assumption (see Proposition 1).

Again, this makes sense because n_d censored observations have been replaced with new event times of a similar magnitude to those on the reference arm (in terms of the hazard). Therefore, and in line with what might be expected, the variability in the difference between the event times of both arms is somewhat reduced due the hazard dilution effect.

4.6 | Information anchoring under Jump to Reference

Equations (15) and (16) provide the building blocks for the main result concerning information anchoring.

Theorem 1. *For bivariate log normally distributed right censored event times, the variance estimate,* $E[V(\hat{\theta}_{\text{MI,J2R}})]$*, is approximately information anchored.*

Proof (outline). We hypothesize that despite using the J2R approach for sensitivity analysis, the variance inflation due to missing data following MI is the same as that under CAR. We compare the expression for the estimated variance under J2R in equation (15), $E[V(\hat{\theta}_{\text{MI,J2R}})]$, with the *predicted*variance under J2R, $E[\hat{V}_{\text{anchored}}]$. The latter we calculate using the other three terms in the equality in Lemma 2, which we recall, relates the ratios for information anchoring to hold,

$$\frac{E[V(\hat{\theta}_{\text{MI,J2R}})]}{E[V(\hat{\theta}_{\text{full,J2R}})]} \approx \frac{E[V(\hat{\theta}_{\text{MI,CAR}})]}{E[V(\hat{\theta}_{\text{full,CAR}})]} \;\Rightarrow\; E[V(\hat{\theta}_{\text{MI,J2R}})] \approx E[V(\hat{\theta}_{\text{full,J2R}})] \times \frac{E[V(\hat{\theta}_{\text{MI,CAR}})]}{E[V(\hat{\theta}_{\text{full,CAR}})]}.$$

Therefore, using the expressions for the three terms on the right hand side, which we know from earlier calculations, we can obtain the predicted *information anchored* variance,

$$E[V_{\text{anchored}}] \approx E[V(\hat{\theta}_{\text{full,J2R}})] \times \frac{E[V(\hat{\theta}_{\text{MI,CAR}})]}{E[V(\hat{\theta}_{\text{full,CAR}})]}. \tag{17}$$

Now, if we subtract the predicted term $E[\hat{V}_{\text{anchored}}]$ above from the expression for the newly derived expression for $E[\hat{V}(\hat{\theta}_{\text{MI,J2R}})]$ in Equation (15), we will obtain an estimate of the difference, which, if information anchoring holds, should be rather small numerically.

We obtain the following expression:

$$\begin{aligned} E\left[V\left(\hat{\theta}_{\text{MI,J2R}}\right)\right] - E[V_{\text{anchored}}] &\lessapprox 2\pi_d(1-\pi_d)\sqrt{\sigma_{22}}\lambda_{r2}\Delta_c \\ &+ \frac{\sigma_{12}^2}{2\sigma_{11}n} + \pi_d(1-\pi_d)\lambda_{r2}^2 + \pi_d(1-\pi_d)\text{VAR}(w_{k,\lambda_{r2}}) + \pi_d^2\text{VAR}(w_{k,\lambda}) \end{aligned}$$

$$- \sigma_{a2o}\left[\frac{\rho^2}{2}(1-\pi_d)+\frac{3}{2}\pi_d+\left[\frac{\pi_d(1-\pi_d)}{2}\right][(1-\pi_d)\dot{\lambda}^2+\lambda^2]\right]$$
$$- \sigma_{a2d}\left[\frac{\rho^2\pi_d}{2}\right]-\Delta_c^2\left[\frac{\rho^2\pi_d}{2}\right], \tag{18}$$

where we only consider terms greater than or equal to $O\left(\frac{1}{n}\right)$, and assume both $\left(1-\frac{1}{n}\right)\approx 1$ and $C \approx 0.5$ for large n. Workings are shown in Appendix H of Data S1.

This is an approximate bound on $E[V(\hat{\theta}_{\text{MI,J2R}})] - E[V_{\text{anchored}}]$, and we now investigate the approximate size of the difference in slightly more detail.

The upper bound on the difference in equation (18) is dominated in absolute magnitude by the first two terms, and the negative ones in σ_{a2o} and Δ_c^2. Focussing only on these terms, we see that the difference depends on the number of patients on each arm (n), the censoring level (π_d), the variance of the data at time 2 (σ_{22}), the variance of those observed (σ_{a2o}), the correlation between measurements at times 1 and 2 (ρ^2), the difference between the mean of those observed and those deviating on the active arm at time 2 (Δ_c), and the inverse Mills ratio relating to the censoring point τ.

Using the same argument as Cro et al. (2016), we can apply standard trial size calculation arguments to provide a realistic upper bound on Δ_c, assuming, for example, 80% power and 5% significance:

$$\Delta_c \lessapprox (\mu_{a2}-\mu_{r2}) = \sqrt{\frac{15.68\sigma_{22}}{n}}.$$

The first term in Equation (18) becomes,

$$\frac{2\sqrt{15.68}\pi_d(1-\pi_d)\sigma_{22}\lambda_{r2}}{\sqrt{n}},$$

which is of order $O(\pi_d/\sqrt{n})$, and the final term in Equation (18) is now,

$$\frac{15.68\sigma_{22}\rho^2\pi_d}{2n},$$

which is of $O(\pi_d/n)$.

Since $\pi_d \ll 0.5$ can be expected for many sensitivity analyses, the whole expression is of the order of approximately 10% of the total variance σ_{22}.

Therefore, we conclude that the upper bound on the difference is small in comparison to the absolute information anchored variance, and the principles of information anchoring have been approximately upheld following MI under J2R, confirming the proposition in Lemma 2 and Theorem 1 above. ■

Corollary 1. *If the primary analysis model is a linear regression of the outcome, adjusted for baseline, then Theorem 1 still holds.*

Proof. The imputation model remains the same when the substantive model is a linear model instead of the t-test. Following Cro et al. (2019), we replace the unconditional variance estimate from the t-test, σ_{22}, with the variance conditional on the baseline covariates, $\sigma_{2.1} = \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}$.

This final result concludes the presentation of the theoretical results for information anchoring for censored data. In the following sections, we validate these results using a simulation study, and then illustrate their application using simulated data closely modeled on the RITA-2 trial data. ■

5 | SIMULATION STUDY

We present the results of a simulation study which demonstrates the information anchoring results derived in the previous section.

We began by assuming normally distributed baseline and time-to-event outcome data without any censoring, and then increased the level of censoring on the active treatment arm. We then compared the empirical results from multiply imputing events using the Tobit model ("simulation"), with those predicted using the theoretical results presented in the previous section ("theory").

The simulation study includes patients with a baseline measurement and event or censoring times generated from a bivariate normal distribution with means and covariances as follows:

$$\mu_{\text{reference}} = [2, 1.9], \quad \mu_{\text{active}} = [2, \mu_{a2}],$$

$$\Sigma_{\text{reference}} = \Sigma_{\text{active}} = \begin{pmatrix} 0.4 & 0.2 \\ 0.2 & 0.6 \end{pmatrix},$$

with a sample size $n = n_r = n_a = 250$ in each arm. We imputed new event times for those censored at a fixed time point τ using the CAR and J2R MI methodology presented earlier with 50 imputed datasets (refer to Appendix I of Data S1 for example R code). We used mean baseline measurements $\mu_{r1} = \mu_{a1} = 2.0$ on the reference and active treatment arms, respectively, and common variance $\sigma_{11} = 0.4$. There is no censoring on the reference arm with mean log normal survival time of $\mu_{r2} = 2.0$, with μ_{a2} unknown due to censoring on the active treatment arm. Again, in line with our theoretical framework, the survival times have a known common variance $\sigma_{22} = 0.6$. Censoring was varied between 10% and 50% on the active treatment arm, with no censoring on the reference arm. We simulated 500 datasets.

Table 1 presents the results. In the following, we have dropped the $\hat{\theta}$ from the expressions to ease readability.

Column (A) shows the theoretical estimate of the information anchored variance following MI under the assumption of J2R. In other words, we calculate each of the three elements on the right-hand side of Equation (19) below with a priori known summary statistics for the simulated data:

$$E[\widehat{V}_{\text{MI,J2R}}] = \frac{E[\widehat{V}_{\text{MI,CAR}}]}{E[\widehat{V}_{\text{full,CAR}}]} \times E[\widehat{V}_{\text{full,J2R}}]. \tag{19}$$

Column (B) shows the same theoretical estimate of Rubin's variance under the J2R assumption, $E[\widehat{V}_{\text{MI,J2R}}]$, but this time calculated using the expression for that term *directly* (i.e., the left-hand side of Equation (19) above), again using a priori values.

Column (C) uses the same calculation for information anchoring as defined in column (A) (i.e., the right-hand side of Equation (19)), but estimates each of the quantities following MI of

simulated data (denoted by the wide hat), *not* using the theoretical expressions derived earlier:

$$\widehat{E[V_{\text{MI,J2R}}]} = \widehat{E[V_{\text{anchored}}]} = \frac{\widehat{E[V_{\text{MI,CAR}}]}}{\widehat{E[V_{\text{full,CAR}}]}} \times \widehat{E[V_{\text{full,J2R}}]}. \tag{20}$$

Column (D) is the analogue of Column B, but calculated using simulated data to estimate of Rubin's variance under the J2R assumption, denoted $\widehat{E[V_{\text{MI,J2R}}]}$.

Columns (A) and (B) is the difference between the theoretically derived expressions ($E[\hat{V}_{\text{MI,J2R}}] - E[\hat{V}_{\text{anchored}}]$), whereas columns (C) and (D) are the difference using simulated data ($\widehat{E[V_{\text{MI,J2R}}]} - \widehat{E[V_{\text{anchored}}]}$).

Following our theoretical expressions and assuming information anchoring holds, we would expect Column A ≈ Column B and Column C ≈ Column D.

The results in Table 1 show excellent agreement when we compare the *predicted* difference using theoretically calculated values (column "Difference theory (A-B)"), with those from simulated data (column "Difference simulation (C-D)).

The discrepancies between the true information anchored variance, and the MI estimate of it, increase as the censoring level increases as we move down the table (column "Difference simulation (C-D)"), from 0.00002 at 10% censoring to 0.0002 at 50% censoring, which are approximately of the order of magnitude of the Monte Carlo simulation error (0.00016).

These results are in line with those for continuous longitudinal data presented in Cro et al. (2019), which were also based on asymptotic arguments.

We conclude that the simulation results are consistent with our expectations and our information anchoring arguments hold.

In the next section, we investigate whether information anchoring principles hold in a clinical trial setting, inspired by data from the RITA-2 trial.

6 | ILLUSTRATIVE EXAMPLE BASED ON THE RITA-2 TRIAL

Here, we use a *simulated* dataset very closely modeled on that from the RITA-2 trial. Our theoretically derived expressions assume a common right censoring threshold, τ, for all patients. However, for the RITA-2 data this assumption does not hold, and therefore, instead of using the original data set directly, we have simulated data based on the descriptive statistics of the original data, and chosen τ to reproduce the level of censoring of the original data.

We generated bivariate normally distributed data according to the properties RITA-2 trial data, and chose τ to result in a censoring level of 27% on the medical arm, again as in the RITA-2 data. The baseline mean we define to be the same on both arms ($\mu_{\text{medical}} = \mu_{\text{PTCA}} = 0.94$). The fully observed log normal survival times on the PTCA arm have mean $\mu_{\text{PTCA}} = 1.75$. The "true" mean of the log time on the medical arm, μ_{r2}, is unknown because of the censoring:

$$\hat{\mu}_{\text{medical}} = [0.94, \hat{\mu}_{r2}], \quad \hat{\mu}_{\text{PTCA}} = [0.94, 1.75],$$

$$\Sigma_{\text{reference}} = \Sigma_{\text{active}} = \begin{pmatrix} 0.15 & -0.04 \\ -0.04 & 0.22 \end{pmatrix},$$

TABLE 2 Top part of table: Mean and *SD* (in brackets) for the medical and percutaneous transluminal coronary angioplasty (PTCA) arms of the second randomized intervention treatment of Angina (RITA-2) dataset

Treatment	*N*	Primary analysis (CAR) [1]	Intention to treat [2]	Sensitivity analysis (Jump to PTCA) multiple imputation ($K = 50$) [3]
Medical (reference)	504	$\mu_{a2o} = 1.60\ (0.34)$	$\mu_{r2} = 1.81\ (0.42)$	$\mu_{r2} = 1.81\ (0.42)$
PTCA (active)	514	$\mu_{r2} = 1.75\ (0.47)$	$\mu_{r2} = 1.75\ (0.47)$	$\mu_{r2} = 1.75\ (0.46)$
Difference in group means (t-test)	1,018	−0.15 [−0.21, −0.10] $p < .001$	0.06 [0.001, 0.12] $p = .05$	0.06 [−0.004, 0.11] $p = .07$
Rubin's variance estimator (calculated)		$E[\hat{V}_{\mathrm{MI,CAR}}] = 0.0042$	-	$E[\hat{V}_{\mathrm{MI,J2R}}] = 0.0043$
Rubin's variance estimator (following MI)		$E[\widehat{V_{\mathrm{MI,CAR}}}] = 0.0042$	-	$E[\widehat{V_{\mathrm{MI,J2R}}}] = 0.0041$

Notes: Bottom part of table: Comparison of endpoint from the primary analysis assuming CAR [1], the intention to treat analysis of the original trial [2] and the sensitivity analysis assuming "Jump to PTCA" [3].

that is, $\sigma_{a1} = \sigma_{r1} = \sigma_{11} = 0.15$, $\sigma_{a2} = \sigma_{r2} = \sigma_{22} = 0.22$, and $\sigma_{12} = -0.04$. For the medical arm, we also know the mean time of the *observed* events, $\mu_{a2o} = 1.60$, with associated variability, $\sigma_{a2o} = 0.11$.

The end-point is the difference in mean log survival time for patients on the medical arm (standard of care, the reference arm) and those receiving PTCA (the active treatment arm). Furthermore, for the primary analysis we assume there is no censoring on the PTCA arm, and patients are censored at their first NRI on the medical arm.

With this simulated dataset, and consistent with the original trial, we assume that medical arm patients censored because of a nonrandomized intervention "Jump to the PTCA arm." This allows us to compare the observed results (from analyzing the trial data) with (i) multiply imputed data and (ii) our theoretical results.

Table 2 summarizes the parameter estimates and *SD*s, along with the results from the primary and sensitivity analyses.

Column [1] provides the descriptive statistics and treatment difference for the primary analysis assuming CAR. Column [2] summarizes the results of the original trial when patients were followed up irrespective of nonrandom interventions (i.e., "intention to treat"). Column [3] presents the results when censored patients are multiply imputed under "Jump to PTCA," again calculated using both the theoretically derived results and following multiple imputation.

Under the primary analysis assumption of CAR, there is a significant treatment difference (−0.15 [−0.21, −0.10], p<.001 [column [1], row 3]), whereas the difference is only marginally significant following MI using "Jump to PTCA arm" (0.06 [−0.004, 0.11], $p = .07$ [column [3], row 3]). The results from intention to treat analysis were similar to those following multiple imputation (0.06 [0.001, 0.12], $p = .05$ [column [2], row 3]).

Consistent with our theory and the simulation results, there is little practical difference between the *theoretically* predicted estimates using our calculated quantities, and those following

multiple imputation. Furthermore, the estimates of Rubin's variance in the primary and sensitivity analysis are also very similar (final row of Table 2), which provides additional validation of our theoretical results. Plugging these and the other relevant values back into the information anchoring equation, indicated that the sensitivity analysis was approximately information anchored, as expected.

7 | DISCUSSION

Given the increasingly prominent role of sensitivity analysis in the analysis of clinical trials, exemplified by the ICH E9 addendum published in 2019 CHMP (2019), it is important to provide methods which are not only easy to implement and use, but which are also clinically plausible and contextually relevant to the trial team and other stakeholders. This also responds to the FDA mandated report by the U.S. National Research Council in 2010, which highlighted the lack of such sensitivity analysis methods NRC (2010). To address this, reference based multiple imputation methods have recently been extended for time to event data (Atkinson, 2019; Atkinson et al., 2019).

However, there has been considerable discussion and numerous publications concerning the perceived conservativeness of Rubin's variance estimator when the imputation and analysis models are not congenial (Hughes et al., 2014; Kim, 2004; Nielson, 2003; Robins, Hernan, & Brumback, 2000), a situation which occurs with RBMI (Liu and Peng (2016)). Following the arguments set out in Cro et al. (2019), we believe it is important that sensitivity analyses are information anchored, so that as we move to the sensitivity analysis we do not inflate the statistical information about the treatment effect. As the theory and the simulations presented here show, this is achieved when we use Rubin's variance estimator. By contrast, in this setting, using the bootstrap (e.g., Liu and Peng) means that the information about the treatment effect is being increased as we move to the sensitivity analysis (so we are rewarded in effect for missing data). We argue the conservativeness of Rubin's variance estimator is of less concern compared to the requirement to ensure that the sensitivity analysis does not (implicitly or explicitly) increase the statistical information about the treatment effect. Therefore, we believe Rubin's variance estimator is more appropriate to use in sensitivity analyses of this type.

The information anchoring principle states that the statistical information should be held constant across the primary and sensitivity analyses, and that it should certainly not be increased, because (Cro et al. (2019)) "an information positive sensitivity analysis is rarely justifiable, implying … that the more data are missing, the more certain we are about the treatment effect under the sensitivity analysis," and, "while information negative sensitivity analyses provides an incentive for minimizing the missing data, there is no natural consensus about the appropriate loss of information." We showed that if we wish to follow the information anchoring principle, reference-based sensitivity analysis implemented using multiple imputation is statistically appropriate for inference trials with censored outcome data that can be modelled with the truncated normal distribution. A natural example of this is time-to-event data, where the survival times follow a log-normal distribution. We built our theoretical results based on using Tobit regression as imputation model, and showed that, assuming log normal survival times and right censored data, information anchoring holds. We validated these results by simulation, with the results closely mirroring those obtained in the continuous longitudinal setting. As a reviewer pointed out, whilst we assumed no difference in the baseline

covariate between the two arms, the framework would allow an adjusted analysis of variance analysis.

A limitation of the approach is that we assume a fixed censoring point τ, primarily to enable us to use standard results for the theoretical work. Having a fixed censoring point is unusual in individual patient randomized trials, but can occur in cluster randomized trials where interventions start on a specific day, and is most often the case, by design, for emulated trials using observational data (Atkinson et al. (2020)). Notwithstanding these examples, most clinical trials have rolling recruitment and stop at a fixed calendar time, which means that the censoring time is variable. Therefore, as proposed by a reviewer, we performed additional simulations with random censoring, and compared the results with those from the theory based on a fixed censoring point (refer to Supplementary Material Appendix J of Data S1). The results suggest that, at least empirically, our theory also holds with random censoring, although this remains to be formally confirmed.

A further potential limitation of the approach presented here is using the *t*-test, or analysis of covariance, to determine the mean log difference in the event times. The most common choice for the primary analysis model would be the Cox Proportional Hazards (CPH) model, with the hazard ratio over the total follow-up period defined as treatment effect for the trial. However, it is important to recognize that the CPH model inherently assumes that the hazards are proportional, even though this is increasingly being challenged in many clinical settings (Royston & Parmar, 2011, 2013). The restricted mean survival time (RMST) is now frequently used instead of the hazard ratio as a preferable clinical endpoint. Although not always equivalent to the RMST, there are clear parallels between using the RMST and the mean log time to event used to calculate the endpoint in our theoretical framework. Nevertheless, an extension of the information anchoring theory to the proportional hazards setting would be beneficial.

We have focused solely on the "Jump to Reference" assumption (from one of a number of proposals suggested in Carpenter et al., 2013), since this was the approach which, following mapping to the time to event setting, had a clear clinical context when assuming proportional hazards (refer to Atkinson et al. (2019)).

In addition, our simulation results and those from a previous study (cf, figure 3 of Atkinson et al., 2019) indicate that information anchoring holds for censoring levels of approximately 50% or less. We believe this is sufficient reassurance for practical use since we are not primarily intending this approach to be applied to administrative censoring (which we know is completely at random and may reach high proportions), but to censoring (for various reasons, typically treatment changes) in the course of the follow-up.

The information anchoring approximation is most accurate when randomization is 1:1. When this is not the case, the theory presented here suggests how to modify the procedure to retain information anchoring. For simplicity the theory presented here focuses on the case in which there is censoring in one arm. The approach could be extended for deviations in both arms in an analogous manner to that set out by Cro et al.

Approaches varying an explicitly defined sensitivity parameter, such as δ-methods, are often presented as an alternative to RBMI for performing sensitivity analysis. Such methods introduce a δ-size change to each patient's hazard at the censoring time point, which is varied to reflect an improvement or a deterioration in their post-censoring condition (Lipkovich, Ratitch, & O'Kelly, 2016; Lu, Li, & Koch, 2015). While attractive due to their simplicity, determining a meaningful range of values for δ has often proved to be a practical drawback. The process often requires iterative discussions within the trial team, and these can become difficult to conclude satisfactorily, especially when considering delta multipliers of a hazard or odds ratio, and how

this may vary by treatment arm. Nevertheless, Cro et al. showed that the information anchoring principle holds for specific settings with delta methods, but this has yet to be demonstrated for time to event data. This is certainly an area for potential further study.

In summary, the attraction of RBMI sensitivity analysis methods is that they are accessible, that is, they are both simple to understand and straightforward to implement. Taking such an approach avoids the alternative in which we would have to explicitly model the event and censoring process, which is often rather complex to achieve in practice, even for experts in the field. We have demonstrated that RBMI in the time to event setting, albeit under specific distributional assumptions, is information anchored, providing both regulators and industry with confidence for using this approach for sensitivity analysis for time to event data.

ACKNOWLEDGEMENTS

We would like to thank the RITA-2 Investigators for use of the RITA-2 data supplied by T. Clayton from the London School of Hygiene and Tropical Medicine. In addition, we would like to thank the anonymous reviewers for their helpful comments and suggestions.

ORCID

Andrew Atkinson https://orcid.org/0000-0001-5834-8315

REFERENCES

Atkinson, A. (2019) *Reference based sensitivity analysis for time-to-event data*. (Ph.D. thesis). London School of Hygiene & Tropical Medicine.

Atkinson A., Kenward M. G., Clayton T., Carpenter J. R. (2019). Reference-based sensitivity analysis for time-to-event data. *Pharmaceutical Statistics*, *18*, (6), 645–658. http://dx.doi.org/10.1002/pst.1954.

Atkinson A., Zwahlen M., Barger D., d'Arminio Monforte A., De Wit S., Ghosn J, Chêne G., (2020). Withholding Primary Pneumocystis Pneumonia Prophylaxis in Virologically Suppressed Patients With Human Immunodeficiency Virus: An Emulation of a Pragmatic Trial in COHERE. *Clinical Infectious Diseases*, http://dx.doi.org/10.1093/cid/ciaa615.

Atri A., Frölich L., Ballard C., Tariot P. N., Molinuevo J. L., Boneva N, Cummings J. L. (2018). Effect of Idalopirdine as Adjunct to Cholinesterase Inhibitors on Change in Cognition in Patients With Alzheimer Disease. *JAMA*, *319*, (2), 130http://dx.doi.org/10.1001/jama.2017.20373.

Billings L. K., Doshi A., Gouet D., Oviedo A., Rodbard H. W., Tentolouris N, Jodar E. (2018). Efficacy and safety of ideglira versus basal-bolus insulin therapy in patients with type 2 diabetes uncontrolled on metformin and basal insulin: the dual vii randomized clinical trial. *Diabetes Care*, *41*, (5), 1009–1016. http://dx.doi.org/10.2337/dc17-1114.

Carpenter, J., & Kenward, M. G. (2013). *Multiple imputation and its applications*. Hoboken, NJ: Wiley.

Carpenter, J., Roger, J., & Kenward, M. G. (2013). Analysis of longitudinal trials with protocol deviation: A framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharm Stat*, *23*(6), 1352–1371.

CHMP. (2010). Committee for medicinal products for human use guidelines on missing data in confirmatory clinical trials. *European Medicines Agency* Retrieved January 15, 2014, from. http://www.ema.europa.eu

CHMP (2019) *Committee for human medicinal products*. ich e9(r1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. November.20.

Cro, S., Carpenter, J. R., & Kenward, M. G. (2019). Information anchored sensitivity analysis. *Journal of the Royal Statistical Society Series A*, *182*(2), 623–645.

Cro, S., Morris, T., Kenward, M. G., & Carpenter, J. (2016). Reference-based sensitivity analysis via multiple imputation for longitudinal trials with protocol deviation. *The Stata Journal*, *16*(2), 443–463.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Hoboken, NJ: Prentice Hall.

Henderson, R. A., Pocock, S. J., Clayton, T. C., Knight, R., Fox, K. A., Julian, D. G., & Chamberlain, D. A. (1997). Coronary angioplasty versus medical therapy for angina: The second randomised intervention treatment of angina (rita-2) trial. *Lancet*, *350*, 461–468.

Henderson, R. A., Pocock, S. J., Clayton, T. C., Knight, R., Fox, K. A., Julian, D. G., & Chamberlain, D. A. (2003). Seven-year outcome in the rita-2 trial: Coronary angioplasty versus medical therapy. *Journal of the American College of Cardiology*, *42*(7), 1162–1170.

Hughes RA, Sterne JAC, Tilling K (2016). Comparison of imputation variance estimators. *Statistical Methods in Medical Research*, *25*, (6), 2541–2557. http://dx.doi.org/10.1177/0962280214526216.

Ibrahim, J., Chu, H., & Chen, M.-H. (2012). Missing data in clinical studies: Issues and methods. *Journal of Clinical Oncology*, *30*(26), 3297–3303.

Jakobsen, J., Gluud, C., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide. *BMC Medical Research Methodology*, *17*(1), 162.

Jans, T., Jacob, C., Warnke, A., Zwanzger, U., Groß-Lesch, S., Matthies, S., … Philipsen, A. (2015). Does intensive multimodal treatment for maternal ADHD improve the efficacy of parent training for children with adhd? a randomized controlled multicenter trial. *Journal of Child Psychology and Psychiatry*, *56*(12), 1298–1313.

Kenney, J. F., & Keeping, E. S. (1951). *The distribution of the standard deviation, in section 7.8 of Mathematical Statistics, Part 2*. Princeton, NJ: D. Van Norstrand.

Kim, J. (2004). Finite sample properties of multiple imputation estimators. *The Annals of Statistics*, *32*(2), 766–783.

Kim, S., Zeng, D., & Taylor, M. (2017). Joint partially linear model for longitudinal data with informative drop-out. *Biometrics*, *73*(1), 72–82.

Lipkovich, I., Ratitch, B., & O'Kelly, M. (2016). Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints. *Pharmaceutical Statistics*, *15*, 216–229.

Little, R., & Yau, L. (1996). Intent-to-treat analysis for longitudinal studies with dropouts. *Biometrics*, *52*, 471–483.

Liu, G., & Peng, L. (2016). On analysis of longitudinal clinical trials with missing data using reference-based imputation. *Journal of Biopharmaceutical Statistics*, *26*(5), 924–936.

Lu, K., Li, D., & Koch, G. (2015). Comparison between two controlled multiple imputation methods for sensitivity analyses of time-to-event data with possibly informative censoring. *Statistics in Biopharmaceutical Research*, *7*(3), 199–213.

Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Sciences*, *9*(4), 538–573.

Nielson, S. (2003). Proper and improper multiple imputation. *International Statistical Review*, *71*, 593–627.

NRC. (2010). *National research council report: The prevention and treatment of missing data in clinical trials*. Washington DC: The National Academic Press: Panel on Handling Missing Data in Clinical Trials, Committee on National Statistics, Division of Behavioral and Social Sciences and Education.

Philipsen, A., Jans, T., Graf, E., Matthies, S., Borel, P., Colla, M., … Sobanski, E. (2015). Effects of group psychotherapy, individual counseling, methylphenidate, and placebo in the treatment of adult attention-deficit/hyperactivity disorder: A randomized clinical trial. *JAMA Psychiatry*, *72*(12), 1199–1210.

Robins, J., Hernan, M., & Brumback, B. (2000). Marginal structal models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550–560.

Robins, J., & Wang, N. (2000). Inference for imputation estimators. *Biometrika*, *87*(1), 112–124.

Royston, P. and Parmar, M. (2013) Survival analysis - Coping with non-proportional hazards in randomized trials. Retrieved August 31, 2013, from http://www.methodologyhubs.mrc.ac.uk

Royston, P., & Parmar, M. K. B. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine*, *30*(19), 2409–2421.

Ruau, D., Burkoff, N., Bartlett, J., Jackson, D., Jones, E., Law, M., … Metcalfe, P. (2016) *InformativeCensoring: Multiple Imputation for Informative Censoring*. R package version 0.3.4.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

SAS (2003) *SAS/STAT Software, Version 9.4*. Cary, NC.

StataCorp. (2013). *Stata statistical software*. College Station, TX: StataCorp LP.

Sterne, J., White, I., Carlin, J., Spratt, M., Royston, P., Kenward, M. G., … Carpenter, J. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ*, *339*, 157–160.

Tang Y (2018). Controlled pattern imputation for sensitivity analysis of longitudinal binary and ordinal outcomes with nonignorable dropout. *Statistics in Medicine*, *37*, (9), 1467–1481. http://dx.doi.org/10.1002/sim.7583.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, *26*, 24–36.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

White, I., & Carlin, J. (2010). Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in Medicine*, *29*, 2920–2931.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Atkinson, A., Cro, S., Carpenter, J. R., & Kenward, M. G. (2021). Information anchored reference-based sensitivity analysis for truncated normal data with application to survival analysis. *Statistica Neerlandica*, *75*(4), 500–523. https://doi.org/10.1111/stan.12250