

Using Observational Study Data as an External Control Group for a Clinical Trial: an Empirical Comparison of Methods to Account for Longitudinal Missing Data

Vibeke Norvang

Diakonhjemmet Hospital

Espen A. Haavardsholm

Diakonhjemmet Hospital

Sara K. Tedeschi

Brigham and Women's Hospital

Houchen Lyu

Brigham and Women's Hospital

Joseph Sexton

Diakonhjemmet Hospital

Maria D. Mjaavatten

Diakonhjemmet Hospital

Tore K. Kvien

Diakonhjemmet Hospital

Daniel H. Solomon

Brigham and Women's Hospital

Kazuki Yoshida (✉ kazukiyoshida@mail.harvard.edu)

Brigham and Women's Hospital

Research Article

Keywords: clinical trials, rheumatoid arthritis, trial data

Posted Date: January 5th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1200218/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Observational data are increasingly being used to conduct external comparisons to clinical trials. In this study, we empirically examined whether different methodological approaches to longitudinal missing data affected study conclusions in this setting.

Methods: We used data from one clinical trial and one prospective observational study, both Norwegian multicenter studies including patients with recently diagnosed rheumatoid arthritis and implementing similar treatment strategies, but with different stringency. A binary disease remission status was defined at 6, 12, and 24 months in both studies. After identifying patterns of longitudinal missing outcome data, we evaluated the following five approaches to handle missingness: analyses of patients with complete follow-up data, multiple imputation (MI), inverse probability of censoring weighting (IPCW), and two combinations of MI and IPCW.

Results: We found a complex non-monotone missing data pattern in the observational study (N=328), while missing data in the trial (N=188) was monotone due to drop-out. In the observational study, only 39.0% of patients had complete outcome data, compared to 89.9% in the trial. All approaches to missing data indicated favorable outcomes of the treatment strategy in the trial and resulted in similar study conclusions. Variations in results across approaches were mainly due to variations in estimated outcomes for the observational data.

Conclusions: Five different approaches to handle longitudinal missing data resulted in similar conclusions in our example. However, the extent and complexity of missing observational data affected estimated comparative outcomes across approaches, highlighting the need for careful consideration of methods to account for missingness when using observational data as external controls to trial data.

Background

Data from observational studies and registries are increasingly being used to complement randomized controlled trials (RCTs) in clinical effectiveness research [1–3]. One recognized approach is to use observational data as an external control group to compare with clinical trial data [2–6]. The external controls may be historical or contemporaneous and may represent the natural course of disease or current standard of care [4, 5]. However, integrating trial data and observational data in one comparative effectiveness study poses methodological challenges due to the heterogeneity of data sources, not only related to the lack of randomization, but also due to differences in follow-up data [4–9].

The follow-up strategy in a clinical trial will typically include more frequent and rigorous monitoring compared with an observational study carried out in clinical practice [1, 4, 8]. This can result in differential patterns of longitudinal missing data. Addressing these differences is crucial to limit potential selection bias when using observational data sources to form external control groups for clinical trials [4, 5, 9, 10]. However, guidance in how to assess and address longitudinal missing data in this setting is scarce.

The present methodological investigation was motivated by challenges arising in a previous study in which we used data from a prospective observational study as an external contemporaneous control group to be compared with a clinical trial [11]. The objectives of this paper are to 1) describe differences in missing data patterns during follow-up in a clinical trial versus an observational study and 2) empirically examine the impact of different missing data methods on study conclusions when using observational study data to form an external control group for a clinical trial.

Methods

In the following, we explain the clinical setting and data sources that motivated our methodological investigation, the description of longitudinal missing patterns, and the empirical comparison study.

Clinical setting and data sources

Data for the present methodological investigation were provided by two Norwegian multicenter studies: the ARCTIC trial [12] and the NOR-VEAC prospective observational study [11, 13]. Both enrolled patients with recently diagnosed rheumatoid arthritis (RA) and implemented treat-to-target strategies of different stringency as patients started disease-modifying anti-rheumatic drug (DMARD) therapy.

"Treat-to-target" in RA care involves defining a disease activity treatment target when initiating a new therapy, frequent monitoring and adjusting therapy if the target is not achieved [14]. Disease activity is typically measured by a composite score calculated from clinically relevant measurements, such as the number of swollen and tender joints, the patient's global assessment of disease, and an inflammatory biomarker [15]. The preferred treatment target in RA is remission, i.e., a clinical state of no or little remaining disease activity [14, 16]. Treat-to-target is currently the recommended approach in the care for patients with RA [16, 17], however, the stringency of the strategy (i.e., how often to conduct follow-up) and how aggressive the disease activity target should be is debated.

Patients were included for the present study according to a common set of eligibility criteria (supplementary file S1). As the two original strategies of the trial showed similar outcomes in the main study [12], we pooled the groups for the purpose of the present study, giving "single-arm" trial data. Details on the respective studies providing data for this methodological investigation are available in online supplementary file S2.

Longitudinal missing data in target trial emulation

Defining and identifying missing data during follow-up requires an understanding of what "ideal" follow-up data might look like. This is not straightforward when using observational data sources as external controls for a clinical trial. We approached this issue using the target trial emulation framework. We conceptualized a hypothetical target trial [18] that compares stringent and pragmatic treat-to-target strategies for RA (supplementary file S3). To emulate baseline randomization to either the ARCTIC trial (representing the stringent treat-to-target strategy) or the NOR-VEAC observational study (representing the

pragmatic treat-to-target strategy), we used inverse probability of treatment weighting using the propensity scores calculated from baseline covariates [19, 20]. Time zero, or baseline, was set to the time of initiating DMARD therapy, which occurred at or shortly after inclusion for both studies.

The ideal data from the target trial would have all patients' outcome data at the 6, 12, 24-months follow-up visits. We separately identified and assessed three types of missingness in the trial data arm and in the observational data arm. The first type of missingness is "missing outcome data at visit". In this type of missingness, patients do have visits corresponding to the study visits that the target trial dictates. However, one or more of the components of the composite outcome score are missing. The second type of missingness is "intermittent missing visits," which can arise from the misalignment of visits occurring in the data and the study visits in the target trial. The third type of missingness is "drop-out", which arise from loss to follow-up.

Missing data approaches

We empirically evaluated five different approaches to handle these three types of missing data, as displayed in Table 1. In this section we outline the approaches and evaluation metrics

Table 1

Evaluated approaches to missing outcome data during follow-up when using data from an observational study as an external control arm to a clinical trial

Approach	Missing outcome data at visit	Intermittent missing visits	Drop-outs
Complete follow-up case analyses ¹	Exclusion	Exclusion	Exclusion
Strict censoring + IPCW ²	Censoring (set as drop-out)	Censoring (set as drop-out)	IPCW
MI + censoring + IPCW ³	MI	Censoring (set as drop-out)	IPCW
MI + IPCW ⁴	MI	MI	IPCW
MI for all ⁵	MI	MI	MI

IPCW, inverse probability of censoring weighting; MI, multiple imputation.

¹Assumptions: Patients with complete follow-up data are exchangeable with patients with missing data.

² Assumptions: The IPCW model is correctly specified when modeling the missing mechanism given previous observations in individuals with missing data due to drop-out (naturally occurring or created due to artificial censoring).

³ Assumptions: Both the MI model and the IPCW model are correctly specified. MI models missing outcome variables at visits given available information in the dataset. IPCW models the missing mechanism given the observed past for missing data due to drop-out (naturally occurring or created due to artificial censoring).

⁴ Assumptions: Both the MI model and the IPCW model are correctly specified. MI models missing outcome variables at visits and intermittent missing visits given information in the dataset. IPCW models the missing mechanism given the observed past for missing data due to naturally occurring drop-out.

⁵Assumptions: The MI model is correctly specified when modeling all missing outcome data given information in the dataset. Separate models were specified for each cohort and the imputed datasets were thereafter combined.

1. Complete follow-up case analysis

In this approach, we performed analyses in a subset of patients with complete follow-up data for the main outcome. All three types of missingness were handled by excluding patients who did not have complete follow-up data. The assumption for this approach was that patients with complete outcome data were exchangeable with patients with missing data.

2. Strict censoring plus IPCW

In the second approach, we used strict censoring and time-varying inverse probability of censoring weighting (IPCW) [21]. Subjects were censored (set as “drop-out”) at the first visit with missing outcome data or at the first intermittent missing visit, whichever occurred first. This created a monotone missing pattern and allowed the use of IPCW to account for naturally occurring or created drop-out [21]. The IPCW method assigns weights to individuals with complete follow-up data corresponding to the inverse of their estimated probability of having complete data [21]. All approaches using IPCW assume a correctly specified IPCW model to account for missing data. We specified a logistic regression model to predict the probability of missing any of the variables required to calculate the outcome data. In the calculation of the IPCWs, we used both baseline values and time-varying values of relevant covariates at available visits to predict missingness.

3. MI plus censoring plus IPCW

In the third approach, we used multiple imputation (MI) [22–24] in combination with IPCW. First, missing outcome data at completed visits were imputed using MI. Thereafter, subjects were censored at the first intermittent missing visit or the first missing visit due to drop-out, whichever occurred first. This created a monotone missing pattern and IPCW was used to account for missing outcome data due to naturally occurring or created drop-out. MI models missing outcome data given available information in the dataset. All approaches using MI assume a correctly specified MI model to account for missing data.

4. MI plus IPCW

In the fourth approach, we also used MI in combination with IPCW. First, we used MI to impute missing outcome data at completed visits and outcome data for intermittent missing visit. Thereafter, IPCW was applied to account for the remaining, naturally occurring drop-out in the imputed datasets.

5. MI for all missing

In a final approach, we used MI to impute all three types of missing outcome data: missing data at completed visits, intermittent missing visits, and drop-out. For all approaches involving MI, relevant available observations at all visits in each of the cohorts were used to inform the imputation models. Given the limited range and typically non-normal distribution of the variables to be imputed, we used multiple imputation by chained equations [24, 25]. We specified separate MI models and created 10 imputed datasets for each cohort, and these were combined into 10 final datasets. Estimates were pooled using Rubin’s rules [22].

Comparison metrics

We applied each of these five missing data approaches to conduct the comparative effectiveness study of treat-to-target strategies with different stringency. The main endpoint was binary and defined as achievement of remission or not according to the disease activity score in 28 joints (DAS28) (24). The DAS28 is a composite disease activity index with a score between 0 and 9.4. Remission is defined as a score <2.6 (25). We first compared the log odds ratio (OR) estimates, standard errors, and ORs for the treatment strategy outcomes across the five missing data approaches. We further compared the

estimated proportions of patients in remission at 6, 12, 24 months for each treatment strategy according to each treatment strategy.

Results

Cohort characteristics

A total of 188 patients from the ARCTIC trial and 328 patients from the NOR-VEAC study met the common eligibility criteria (supplementary file S1).

ARCTIC trial data

In the ARCTIC trial, 89.9% (169/188) of patients had complete follow-up data for the main outcome. Patients with incomplete follow-up data (drop-out exclusively) were younger, had less comorbidity, lower education, and more were current smokers compared to patients with complete follow-up data (supplementary file S4). Furthermore, patients with missing data also had higher disease activity at baseline.

NOR-VEAC observational study data

In the NOR-VEAC study, only 39.0% (128/328) of patients had complete follow-up data for the main outcome. Patients with incomplete follow-up data were somewhat younger and had lower education than patients with complete follow-up data; however, disease activity levels at baseline were similar (supplementary file S4).

Description of longitudinal missing data patterns

Missing data in the ARCTIC trial was monotone and almost exclusively a result of drop-out. The drop-out rate was 1.6% at 6 months, 6.4% at 12 months, and 10.1% at 24 months (Figure 1). In the NOR-VEAC observational study, the proportion of missing data was considerably higher, resulting from missing outcome data at completed visits, intermittent missing visits or drop-out, i.e., a non-monotone missing pattern (Figure 1). Drop-out in NOR-VEAC counted for 2.4% of missing data at 6 months, 9.2% at 12 months and 35.1% at 24 months. Additionally, at the 6-month visit 7.3% of patients had an intermittent missing visit and 4.0% had missing outcome data at a recorded visit, while at the 12-month visit 11.3% had an intermittent missing visit, and 4.6% had missing outcome data at a recorded visit. At 24 months, 8.5% of patients had missing outcome data at a recorded visit (Figure 1).

Results from method comparison

Impact of missing data approaches on effect estimates

Regardless of the approaches to missing data (Table 2), the odds of achieving disease remission was higher for the "stringent treat-to-target" (based on the trial data) than for the "pragmatic treat-to-target"

(based on the observational study data) in our target trial emulation.

Table 2

Differences in achievement of remission at 6, 12 and 24 months in the ARCTIC trial (Norway; 2010-2015) compared to the NOR-VEAC observational study (Norway; 2010-2018) according to different approaches to missing data¹

Approaches to missing data	Estimate	SE	Odds ratio	95% CI	p-value
6 months					
CC ²	0.756	0.288	2.13	1.21 - 3.75	0.009
Strict censoring + IPCW ³	0.423	0.223	1.53	0.99 - 2.37	0.058
MI + censoring + IPCW ⁴	0.401	0.222	1.49	0.97 - 2.31	0.072
MI + IPCW ⁵	0.401	0.220	1.49	0.97 - 2.30	0.069
MI for all ⁶	0.401	0.225	1.49	0.96 - 2.32	0.074
12 months					
CC ²	1.093	0.294	2.98	1.68 - 5.31	<0.001
Strict censoring + IPCW ³	0.879	0.259	2.41	1.45 - 4.00	<0.001
MI + censoring + IPCW ⁴	0.797	0.250	2.22	1.36 - 3.62	0.001
MI + IPCW ⁵	0.768	0.243	2.16	1.34 - 3.47	0.002
MI for all ⁶	0.687	0.248	1.99	1.22 - 3.23	0.006
24 months					
CC ²	0.429	0.288	1.54	0.87 - 2.70	0.136
Strict censoring + IPCW ³	0.407	0.284	1.50	0.86 - 2.62	0.151
MI + censoring + IPCW ⁴	0.448	0.279	1.57	0.91 - 2.70	0.108
MI + IPCW ⁵	0.401	0.268	1.49	0.88 - 2.52	0.134
MI for all ⁶	0.367	0.253	1.44	0.88 - 2.38	0.147

Approaches to missing data	Estimate	SE	Odds ratio	95% CI	p-value
SE, standard error; CI, confidence interval; CC, complete case analyses; IPCW, inverse probability of censoring weighting; MI, multiple imputation					
¹ For all approaches, inverse probability of treatment weighting using the propensity score was used to balance the two cohorts on relevant baseline covariates					
² Analyses in a subset of patients with complete follow-up data for the main outcome.					
³ Censoring of subjects with missing outcome data at a visit or intermittent missing visits, whichever occurred first. IPCW used to account for missing data due to drop-out (naturally occurring or created due to censoring)					
⁴ MI used to account for missing outcome data at visits. Censoring of subjects with intermittent missing visits. IPCW used to account for missing data due to drop-out (naturally occurring or created due to censoring)					
⁵ MI used to account for missing outcome data at a visit and intermittent missing visits. IPCW used to account for naturally occurring drop-out.					
⁶ MI used to account for all missing data.					

The complete follow-up case analysis provided higher effect estimates than the more sophisticated approaches to missing data. This tendency was most apparent at the 6-month assessment, when the OR estimate was 2.13 [95% confidence interval (CI) 1.21, 3.75] favoring the stringent treat-to-target strategy. Other approaches to missing data gave more conservative estimates ranging from OR 1.53 [95% CI 0.99, 2.37] for the strict censoring plus IPCW approach to OR 1.49 [95% CI 0.96, 2.32] for the MI for all missing approach. All three approaches involving MI to various extents yielded essentially identical results for the 6-month assessment. Statistical efficiency was evidently worse (higher standard errors) for the complete follow-up case analysis, which handled all missingness by exclusion of patients.

The 12-month assessment generally gave similar results, with the complete follow-up case approach giving the most optimistic results, whereas the other more sophisticated approaches giving more conservative and similar results. The estimates became slightly more conservative as the extent of MI use increased from the strict censoring plus IPCW approach (no MI; OR 2.41 [95% CI 1.40, 4.00]) to the MI for all missing approach (OR 1.99 [95% CI 1.22, 3.23]).

The 24-month assessment generally exhibited a similar trend to the 6- and 12-month assessments, but the discrepancies between all approaches were subtle (OR 1.44–1.57) compared with the first two time points. Since there were no intermittent missing visits at 24 months, the two combinations of MI and IPCW both used MI to impute partial missing visit data and IPCW to account for naturally occurring drop-out.

Impact of missing data approaches on response rate

The estimated proportion achieving disease remission in the emulated "stringent treat-to-target" arm (based on the trial data) was similar across the approaches to missing data (Figure 2). Since missing data in the trial were almost exclusively due to drop-out, the estimated proportion achieving remission was not affected by the different approaches to partial missing visit data or intermittent missing visit. This implies that all approaches including IPCW used this method to account for naturally occurring drop-out only. The estimated proportions achieving remission were somewhat lower at 12 and 24 months when using MI to account for drop-out (Figure 2).

In the emulated "pragmatic treat-to-target" arm (based on the observational study data), the estimated proportions achieving disease remission varied across different approaches to missing data at 6 and 12 months, but were similar at 24 months (Figure 2). At 6 and 12 months, analyses of patients with complete follow-up data and the approach using strict censoring and IPCW resulted in lower estimated proportions achieving the main outcome than the approaches handling incomplete visit data and intermittent missing visits in addition to drop-out (Figure 2). Since there were no intermittent missing visits at 24 months, the two combinations of MI and IPCW both used MI to impute partial missing visit data and IPCW to account for naturally occurring drop-out.

Discussion

We examined the impact of five alternative approaches to longitudinal missing data in an empirical example of RA treatment strategy comparison, in which we used data from a prospective observational study to form an external control group for a clinical trial. We consistently found superior outcomes of the "stringent treat-to-target" strategy (based on the trial data) compared with the "pragmatic treat-to-target" strategy (based on the observational study data), although the difference was only statistically significant at the 12-month visit. The complete follow-up case analysis tended to give higher effect estimates of the OR and wider confidence intervals compared with the other more sophisticated approaches; however, this difference decreased with longer follow-up. The four methods employing IPCW, MI, and their combinations, generally yielded similar OR estimates despite their differing modeling assumptions.

Although the literature on missing data is vast [26], our paper is unique in its focus on longitudinal missing data challenges in the emerging area of using real-world data as external controls for trial data [4]. The similarity of the patient populations and applied treatment strategies in the two studies providing data for this methodological investigation enabled us to assess missing data patterns resulting from follow-up under different study designs. We found a larger amount of missing data with more complex missing patterns in the observational study compared with the limited and monotone missing data in the trial. As a result, the differences across the alternative approaches to longitudinal missing data mainly came from the estimated disease remission proportions in the "pragmatic treat-to-target" strategy arm (based on the observational study data). Most notably, the complete follow-up case analyses gave smaller estimates for the proportions of patients reaching the desired remission outcome at 6 and 12

months in the "pragmatic treat-to-target" strategy arm. This made the estimated benefits associated with the "stringent treat-to-target" strategy (based on the trial data) appear better.

Both IPCW and MI can provide unbiased estimates under the missing at random assumption, which is weaker than the assumption for the complete follow-up case analysis. An advantage of MI is that this method efficiently uses information from individuals with partially missing data [21, 22, 24]. All available and relevant data can be included in the imputation model, including both variables related to the outcome analyses and variables associated with missingness [22, 24]. However, the MI approach is potentially sensitive to misspecification in situations where some individuals have large blocks of missing values [21]. Thus, missing data due to drop-out in the present study may make MI less appealing, especially for the 24-month time point. IPCW assumes a correctly specified model for the missingness mechanism, given observed data at previous time points [21, 27]. A correctly specified IPCW can account for missingness due to blocks of drop-out. However, IPCW can be less efficient due to the loss of information from incomplete cases [21]. Thus, the IPCW model for the trial data, with smaller amounts of missing at 6 and 12 months and a maximum of 10.1% missing at 24 months, was likely to be more efficient than the IPCW model for the observational data, with a substantial amount of missing outcome data during follow-up.

In the present empirical evaluation, censoring all patients at first missing data (strict censoring) created a monotone missing pattern in the observational data, while a monotone missing pattern already existed naturally in the trial data. Despite using IPCW to account for created or naturally occurring drop-out, the estimates from the strict censoring approach were less efficient at 12 and 24 months than approaches involving MI, reflecting the substantial loss of information due to excluded data points. This may indicate increased efficiency due to recovered information when using MI to impute all or partial missing visit data and may be preferable compared with excluding individuals at first missing value.

A limitation of this methodological investigation is the generalizability of results to other settings using an external control group. Data in the external control group of the present study was provided by a contemporaneous, prospective observational study with a patient population and follow-up strategy similar to the trial [12–14]. This is the most favorable type of external control group [4]. As a result, emulating a target trial was relatively straightforward. Other sources of observational data, such as electronic health records and insurance claims, likely pose more methodological challenges. Furthermore, as we used empirical data rather than simulations, we do not know the true underlying effect of the "stringent treat-to-target" compared to "pragmatic treat-to-target".

In conclusion, we empirically examined the impact of different approaches for missing follow-up data when using data from an observational study to form an external control arm for a clinical trial. Despite the favorable setting of having prospectively collected observational data, there were some differences in the effect estimates although the clinical conclusion was not affected. The differences mainly came from the handling of more extensive and complex missing data in the observational part of the study. When using routine observational data as external controls even more complex missingness issues are likely

expected. As the quality of a comparative effectiveness study is dependent on what we compare to, we cannot overemphasize the importance of carefully examining missing data patterns and conducting appropriate sensitivity analyses in this setting.

Abbreviations

CI	- confidence interval
DAS28	- the disease activity score in 28 joints
DMARD	- disease-modifying anti-rheumatic drug
IPCW	- inverse probability of censoring weighting
MI	- multiple imputation
OR	- odds ratio
RA	- rheumatoid arthritis
RCT	- randomized controlled trial

Declarations

Ethics approval and consent to participate

The ARCTIC trial was approved by an independent ethics committee (the Regional Committee for Medical and Health Research Ethics South-East; reference number 2010/744). Patients gave written informed consent prior to inclusion in the study. The NOR-VEAC study was approved by an independent ethics committee (the Regional Committee for Medical and Health Research Ethics South-East; reference number 2010/719). Patients gave written informed consent prior to inclusion in the study. Both studies were conducted in accordance with the guidelines and regulations applicable in Norway.

Consent for publication

Not applicable.

Availability of data and materials

The data collected in the ARCTIC trial and the NOR-VEAC observational study contain protected health information on participating patients. Therefore, the datasets generated during and/or analyzed during the current study cannot be shared.

Competing interests

E. A. Haavardsholm has received research grants to the ARCTIC trial from AbbVie, MSD, Pfizer, Roche and UCB; all other authors declare no conflict of interest.

Funding

The ARCTIC trial received grants from the Norwegian Research Council, the South-East Health Region in Norway, the Norwegian Rheumatism Association, the Norwegian Women's Public Health Association and unrestricted grant support from AbbVie, Pfizer, MSD, Roche and UCB. The NOR-VEAC study was funded by the Norwegian Foundation for Health and Rehabilitation and the South-Eastern Norway Regional Health Authority. These funding sources were not involved in the conduct of the research or preparation of this paper.

Authors' contributions

All authors were involved in drafting the article or revising it critically for important intellectual content. All authors approved the final manuscript to be submitted and agreed to be accountable for all aspects of the work. Conception and design of the study: VN and KY. Analyses and interpretation of data: VN, KY, JS, SKT, HL and DHS. Acquisition of data: EAH, MDM and TKK.

Acknowledgements

The authors would like to thank the ARCTIC investigators: Dr. Hallvard Fremstad, Dr. Tor Magne Madland, Dr. Åse Stavland Lexberg, Dr. Hilde Haukeland, Dr. Erik Rødevand, Dr. Christian Høili, Dr. Hilde Stray, Dr. Anne Noraas Bendvold, Dr. Inger Johanne Widding-Hansen and Dr. Gunnstein Bakland, and the NOR-VEAC investigators: Dr. Åse Stavland Lexberg, Dr. Cathrine Thunem, Dr. Inger Johanne Widding-Hansen, Dr Ellen Sauar Norli and Dr. Heidi Kverneggen Øvreås. This manuscript was based on work previously presented at the 2019 ACR/ARHP Annual Meeting (*Arthritis Rheumatol.* 2019; 71 (suppl 10)).

References

1. Beaulieu-Jones BK, Finlayson SG, Yuan W, Altman RB, Kohane IS, Prasad V, et al. Examining the Use of Real-World Evidence in the Regulatory Process. *Clin Pharmacol Ther.* 2020;107:843-852.
2. US Food and Drug Administration. Real-world evidence. <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>. Assessed 30 September 2021.
3. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clin Pharmacol Ther.* 2019;105:867-877.
4. Burcu M, Dreyer NA, Franklin JM, Blum MD, Critchlow CW, Perfetto EM, et al. Real-world evidence to support regulatory decision-making for medicines: Considerations for external control arms. *Pharmacoepidemiol Drug Saf.* 2020;29:1228-1235.
5. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and External Controls in Clinical Trials - A Primer for Researchers. *Clin Epidemiol.* 2020;12:457-467.

6. Schmidli H, Häring DA, Thomas M, Cassidy A, Weber S, Bretz F. Beyond Randomized Clinical Trials: Use of External Controls. *Clin Pharmacol Ther.* 2020;107:806-816.
7. Kilcher G, Hummel N, Didden EM, Egger M, Reichenbach S. Rheumatoid arthritis patients treated in trial and real world settings: comparison of randomized trials with registries. *Rheumatology (Oxford).* 2018;57:354-369.
8. Weiss NS. Generalizing from the results of randomized studies of treatment: Can non-randomized studies be of help? *Eur J Epidemiol.* 2019;34:715-718.
9. Potthoff RF. Differential losses to follow-up that are outcome-dependent can vitiate a clinical trial: Simulation results. *J Biopharm Stat.* 2018;28:633-644.
10. Haneuse S. Distinguishing Selection Bias and Confounding Bias in Comparative Effectiveness Research. *Med Care.* 2016;54:e23-9.
11. Norvang V, Brinkmann GH, Yoshida K, Lillegraven S, Aga AB, Sexton J, et al. Achievement of remission in two early rheumatoid arthritis cohorts implementing different treat-to-target strategies. *Arthritis Rheumatol.* 2020;doi:10.1002/art.41232
12. Haavardsholm EA, Aga AB, Olsen IC, Lillegraven S, Hammer HB, Uhlig T, et al. Ultrasound in management of rheumatoid arthritis: ARCTIC randomised controlled strategy trial. *BMJ.* 2016;354:i4205.
13. Brinkmann GH, Norvang V, Norli ES, Grovle L, Haugen AJ, Lexberg AS, et al. Treat to target strategy in early rheumatoid arthritis versus routine care - A comparative clinical practice study. *Semin Arthritis Rheum.* 2019;48:808-814.
14. Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Ann Rheum Dis.* 2016;75:3-15.
15. England BR, Tiong BK, Bergman MJ, Curtis JR, Kazi S, Mikuls TR, et al. 2019 Update of the American College of Rheumatology Recommended Rheumatoid Arthritis Disease Activity Measures. *Arthritis Care Res (Hoboken).* 2019;71:1540-1555.
16. Smolen JS, Landewé RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann Rheum Dis.* 2020;79:685-699.
17. Fraenkel L, Bathon JM, England BR, St.Clair EW, Arayssi T, Carandang K, et al. 2021 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. *Arthritis Rheumatol.* 2021;73:1108-1123.
18. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol.* 2016;183:758-64.
19. Rosenbaum PR, Rubin DB. The central role of propensity score in observational studies for causal effects. *Biometrika.* 1983;70:41-55.
20. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational

- studies. *Stat Med.* 2015;34:3661-79.
21. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res.* 2013;22:278-95.
 22. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* John Wiley & Sons, Inc.; 1987.
 23. van Buuren S. *Flexible Imputation of Missing Data.* Chapman and Hall/CRC; 2018.
 24. Harel O, Mitchell EM, Perkins NJ, Cole SR, Tchetgen EJT, Sun BL, et al. Multiple Imputation for Incomplete Data in Epidemiologic Studies. *Am J Epidemiol.* 2018;187:576-584.
 25. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011;30:377-99.
 26. Molenberghs G, Fitzmaurice GM, Kenward MG, Tsiatis AA, Verbeke G. *Handbook of Missing Data Methodology.* Chapman and Hall/CRC; 2015.
 27. Perkins NJ, Cole SR, Harel O, Tchetgen EJT, Sun BL, Mitchell EM, et al. Principled Approaches to Missing Data in Epidemiologic Studies. *Am J Epidemiol.* 2018;187:568-575.

Figures

Figure 1

Patterns of missing outcome data in A) the ARCTIC trial and B) the NOR-VEAC observational study after standardization of follow-up.

SJC28, swollen joint count in 28 joints; TJC28, tender joint count in 28 joints; PGA, patient's global assessment of disease; ESR, erythrocyte sedimentation rate; DAS28, Disease Activity Score in 28 joints.

Figure 2

Estimated proportion of patients achieving the main outcome according to different approaches to missing data in A) the ARCTIC trial and B) the NOR-VEAC observational study.

CC, complete (follow-up) case; cens., censoring; IPCW, inverse probability of censoring weighting; MI, multiple imputation.

Supplementary Files

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

- [20211223aSupplementaryfilesallfiles.docx](#)