



OPEN ACCESS

EPIDEMIOLOGICAL SCIENCE

Accounting for missing data caused by drug cessation in observational comparative effectiveness research: a simulation study

Denis Mongin ,¹ Kim Lauper ,^{1,2} Axel Finckh ,^{1,2} Thomas Frisell ,³ Delphine Sophie Courvoisier ^{1,2}

Handling editor Josef S Smolen

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/annrheumdis-2021-221477>).

¹Department of Medicine, University of Geneva, Geneva, Switzerland

²Division of Rheumatology, Beau Sejour Hospital, Geneva University Hospitals, Geneva, Switzerland

³Department of medicine Solna, Karolinska institute, Stockholm, Sweden

Correspondence to

Dr Denis Mongin, Department of Medicine, University of Geneva, Geneva, Switzerland; Denis.Mongin@unige.ch

Received 10 September 2021
Accepted 20 December 2021
Published Online First
13 January 2022

ABSTRACT

Objectives To assess the performance of statistical methods used to compare the effectiveness between drugs in an observational setting in the presence of attrition.

Methods In this simulation study, we compared the estimations of low disease activity (LDA) at 1 year produced by complete case analysis (CC), last observation carried forward (LOCF), LUNDEX, non-responder imputation (NRI), inverse probability weighting (IPW) and multiple imputations of the outcome. All methods were adjusted for confounders. The reasons to stop the treatments were included in the multiple imputation method (confounder-adjusted response rate with attrition correction, CARRAC) and were either included (IPW2) or not (IPW1) in the IPW method. A realistic simulation data set was generated from a real-world data collection. The amount of missing data caused by attrition and its dependence on the 'true' value of the data missing were varied to assess the robustness of each method to these changes.

Results LUNDEX and NRI strongly underestimated the absolute LDA difference between two treatments, and their estimates were highly sensitive to the amount of attrition. IPW1 and CC overestimated the absolute LDA difference between the two treatments and the overestimation increased with increasing attrition or when missingness depended on disease activity at 1 year. IPW2 and CARRAC produced unbiased estimations, but IPW2 had a greater sensitivity to the missing pattern of data and the amount of attrition than CARRAC.

Conclusions Only multiple imputation and IPW2, which considered both confounding and treatment cessation reasons, produced accurate comparative effectiveness estimates.

INTRODUCTION

In rheumatology, as in other specialties, randomised controlled trials are the gold standard when evaluating treatment efficacy. However, because of the highly selected populations, their conclusions are difficult to generalise to routine clinical practice. For this reason among others, evaluation of the effectiveness of treatments in the real-world patient population is needed.¹ Comparative effectiveness between treatments when using observational real-world data requires overcoming several difficulties.

In addition to confounding, a recurring difficulty is missing data.² The data used to define drug effectiveness can be missing at the follow-up time of

Key messages

What is already known about this subject?

- Attrition, defined as an increased selection due to participants leaving the study, is a source of bias for comparative effectiveness research, but it is often not acknowledged.
- Several imputation methods exist to account for attrition in comparative effectiveness research with observational data, but comparison of these methods is lacking

What does this study add?

- This study provides an extensive comparison of the statistical methods used to compare treatment effectiveness in an observational setting in the presence of attrition.
- Omitting to consider confounding, treatment cessation or dropouts leads to biased estimation of effectiveness. The present article provides advices, methods and code to report effectiveness properly.

How might this impact on clinical practice or future developments?

- This work suggests that comparative effectiveness results in observational data may have large over or underestimation depending on the method used.

interest, while patients are still on treatment, and the missing data may have to be imputed to avoid selection bias,³ using proper imputation methods.^{4,5} But the data of interest can also be missing because of attrition (an increasing selection due to participants leaving the study).^{6–8} Considering effectiveness among those remaining on therapy after a certain set of time (complete case analysis; CC⁹) is known to be a source of bias,^{3,10} because it excludes from the analysis patients who stopped the drug for an adverse event or lack of effect, thus resulting in a selection bias in favour of responders.¹¹ Although 'intention to treat' analysis intends to avoid this bias in controlled trials,¹² there is no consensus regarding how this should be handled in an observational study.¹³ Several statistical methods exist allowing to account for both attrition and confounding, such as inverse probability weighting (IPW)^{7,14} or multiple imputation (MI) of the outcome.^{15,16} In rheumatic diseases cohort studies, a popular method to account



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Mongin D, Lauper K, Finckh A, et al. *Ann Rheum Dis* 2022;**81**:729–736.

for potential attrition bias is the LUNDEX index.¹⁷ Analogously to non-responder imputation (NRI),¹⁸ it corrects for the attrition bias by assuming that all patients stopping their treatment are non-responders. It, therefore, multiplies the estimation of effectiveness by the estimate of each drug survival, which may result in underestimating true drug effectiveness.

Depending on how it is handled, attrition may lead to biased conclusions in comparative effectiveness research. Therefore, the characterisation and comparison of existing statistical methods are needed. The reasons to stop the treatment are often available in registers and, to our knowledge, are generally not employed in standard statistical analyses when accounting for attrition.⁸

The aim of this research is, thus, to perform an extensive simulation study to compare the ability of different statistical approaches to account for missing data caused by attrition, when studying comparative effectiveness using rheumatology observational data.

METHODS

Our study simulates a comparative effectiveness study of two treatments in patients with potentially different baseline characteristics and attrition rates. In this simulation, we compare the effectiveness of tumour necrosis factor inhibitors (TNFi) versus a biologic disease-modifying antirheumatic drugs (bDMARD) with another mode of action (OMA) in patients with rheumatoid arthritis (RA) using the clinical disease activity index (CDAI) definition of low disease activity (LDA) at 1-year follow-up as outcome. We adjust for the following characteristics at treatment start (hereafter referred to as baseline covariates): disease duration, concomitant treatment with conventional synthetic DMARD (csDMARD), the number of previous bDMARDs (prev_bDMARD), and CDAI (CDAI₀). A collection of registers is used to generate a single simulation dataset with all CDAI values at 1 year. The simulation study consists in generating, in four iterative steps, missing data of CDAI at 1 year caused by attrition to then compare how different statistical methods estimate effectiveness. The sensitivity of the results is studied by changing the amount and the pattern of the missing data.

Creation of the simulation dataset

An original data set composed of a collaboration of RA registers including TNFi and OMA, including >45 000 treatment courses¹⁹ (see table 1) was used to construct a realistic open cohort simulation data set composed of 10 000 treatment courses. The variables of this simulation data set were treatment (TNFi or OMA), CDAI value at treatment start and at 12 months (CDAI₀, CDAI₁₂, respectively), all confounders cited above and the treatment status, which indicates the last available status about the patient's treatment. It is equal to 'ongoing' if the patient is still on treatment at the time of data extraction and to the reason for treatment cessation otherwise (see online supplemental material for more details).

The simulation data set was constructed to have the same proportion of treatments (OMA and TNFi) as the whole data collection. For each treatment group, CDAI₀, disease duration, prev_bDMARD, csDMARD and the treatment status were independently randomly drawn from the entire collection. The treatment duration values in the simulation data set were then randomly sampled while matching exactly the treatment, the treatment status, and prev_bDMARD. CDAI₁₂ was then randomly sampled from the entire data set while matching exactly the treatment, the treatment status, prev_bDMARD, and the categories of CDAI₀ (see online supplemental figure 1

Table 1 Characteristics of the patients under treatment in the initial real-world register collection

	<i>Other mode of actions</i>	<i>TNF inhibitor</i>
<i>N</i>	6067	40767
<i>Disease duration (median (IQR) in year)</i>	9.8 (4.6, 17.9)	7.4 (2.9, 14.6)
<i>Treatment duration (median (IQR) in year)</i>	1.2 (0.5, 2.7)	1.7 (0.6, 4.3)
<i>Number of previous bDMARD (number (%))</i>		
0	1451 (23.9)	23 016 (56.5)
1	2019 (33.3)	12 269 (30.1)
2	1383 (22.8)	3896 (9.6)
3+	1214 (20.0)	1586 (3.9)
<i>Concomitant csDMARD (number (%))</i>		
MTX	2982 (49.2)	20 062 (49.2)
MTX +other	68 (1.1)	1080 (2.6)
None	1907 (31.4)	15 047 (36.9)
Other	1110 (18.3)	4578 (11.2)
<i>CDAI₀ (median (IQR))</i>	23.4 (16.5, 32.0)	23.0 (13.9, 33.5)
<i>CDAI₁₂ (median (IQR))</i>	10.0 (5.0, 17.3)	7.0 (2.9, 14.0)
<i>Treatment status (number (%))</i>		
Ongoing	3336 (55.0)	24 968 (61.2)
Stopped for adverse event	459 (7.6)	1876 (4.6)
Stopped for ineffectiveness	748 (12.3)	3369 (8.3)
Stopped for pregnancy	5 (0.1)	53 (0.1)
Stopped for remission	62 (1.0)	189 (0.5)
Stopped for other	313 (5.2)	1089 (2.7)
<i>Stopped for unspecified reason</i>		
Missing CDAI ₁₂ (number (%))	1680 (27.7)	9415 (23.1)
Number (%) of patients stopping after 1 year (number (%))	1017 (16.8)	6081 (14.9)
LDA at 12 month (number (%))	1811 (51.4)	15 873 (63.7)

Number of observations (N), number and proportion of patients having 0, 1, 2, 3 and more (3+) previous biological DMARD, of patient having methotrexate alone (MTX), methotrexate with other csDMARD (MTX +other), at least an csDMARD other than MTX (other) or no concomitant synthetic DMARD treatment (none), median (IQR) value of baseline CDAI (CDAI₀) and CDAI at 12 months (CDAI₁₂). CDAI, Clinical Disease Activity Index; csDMARD, conventional synthetic disease-modifying antirheumatic drugs; TNF, tumour necrosis factor.

for a graphical representation of the data generation). By doing so, we recreated the association between CDAI₁₂ and the reason to discontinue using the reasons to discontinue that happened between 12 and 36 months in the original register collection. Our simulation, thus, imposed this same association between CDAI₁₂ and discontinuations before 12 months. In the resulting data set for the 12-month follow-up, more than 55% of the treatment status are 'ongoing' (table 2).

Generation of missing data

Data can be missing completely at random when the probability of missing is independent of observed and unobserved data, missing at random (MAR) when the probability of missing depends on observed data, and missing not at random (MNAR) when the probability of missing depends on unobserved data. The patients

Table 2 Characteristics of the patients in the simulated dataset

	<i>Other mode of action</i>	<i>TNF inhibitor</i>
<i>N</i>	1295	8705
<i>Disease duration (median (IQR) in year)</i>	9.7 (4.5, 17.8)	7.3 (2.9, 14.3)
<i>Treatment duration (median (IQR) in year)</i>	1.2 (0.5, 2.7)	1.6 (0.5, 4.2)
<i>Number of previous bDMARD (number (%))</i>		
0	326 (25.2)	4973 (57.1)
1	409 (31.6)	2618 (30.1)
2	294 (22.7)	775 (8.9)
3+	266 (20.5)	339 (3.9)
<i>Concomitant csDMARD (number (%))</i>		
MTX	611 (47.2)	4330 (49.7)
MTX +other	18 (1.4)	248 (2.8)
None	427 (33.0)	3143 (36.1)
Other	239 (18.5)	984 (11.3)
<i>CDAI₀ (median (IQR))</i>	23.2 (16.4, 32.4)	23.0 (14.0, 33.3)
<i>CDAI₁₂ (median (IQR))</i>	10.2 (5.3, 17.0)	8.0 (3.4, 15.1)
<i>treatment status (number (%))</i>		
Ongoing	740 (57.1)	5355 (61.5)
Stopped for adverse event	95 (7.3)	407 (4.7)
Stopped for ineffectiveness	159 (12.3)	755 (8.7)
Stopped for pregnancy	1 (0.1)	11 (0.1)
Stopped for remission	17 (1.3)	41 (0.5)
Stopped for other	57 (4.4)	215 (2.5)
Stopped for unspecified reason	226 (17.5)	1921 (22.1)
LDA	639 (49.3)	5210 (59.9)

Number of observations (N), number and proportion of patients having 0, 1, 2, 3 and more (3+) previous biological DMARD, of patient having methotrexate alone (MTX), methotrexate with other csDMARD (MTX +other), at least an csDMARD other than MTX (other) or no concomitant synthetic DMARD treatment (none), median (IQR) value of age, body mass index (BMI), baseline CDAI (CDAI₀) and CDAI at 12 months (CDAI₁₂).

CDAI, Clinical Disease Activity Index; csDMARD, conventional synthetic disease-modifying antirheumatic drugs; LDA, low disease activity; TNF, tumour necrosis factor.

whose CDAI values at 12 months were set to missing in the simulated data set were chosen by sampling from the patients having a treatment status different from 'ongoing' with the conditional probability of having a missing CDAI₁₂ value in the original real data. These probabilities were extracted from a generalised linear model applied on the initial 'real-world' register collection estimating the probability of having a missing outcome as a function of the treatment, prev_bDMARD, csDMARD, CDAI₀, CDAI₁₂ and the treatment status. Before applying this generalised linear model, all predictors were imputed using MI by chained equation (mice) with 40 iteration, 40 samples and predictive mean matching.

Different ways of deleting data were applied, leading to different types of missing patterns for the different estimation methods considered (see table 3):

- ▶ A reference scenario (missingness condition 1) with 30% treatment cessation in both treatments and no association between effectiveness at 1 year and treatment cessation.
- ▶ To test the dependence of the observed comparative effectiveness on the amount of attrition, the proportion of

missing CDAI₁₂ was set to 10%, 15%, 20%, 25% and 30% in OMA only, and then in both treatments.

- ▶ To assess the sensitivity of the estimators to MNAR data, we changed the association between CDAI₁₂ values and missingness of CDAI₁₂ caused by attrition by modifying the OR for CDAI₁₂ yielded by the generalised linear model predicting the missing values of CDAI₁₂ (ie, the effect of CDAI₁₂ values on the odds of having a missing CDAI₁₂). It was set to 1 for the reference treatment, and to 1.07 or 1.14 (probability of having a CDAI₁₂ missing multiplied by 2 or 4 for an increase in 10 points of CDAI₁₂) for OMA only, and then in both treatments.

For treatment courses with CDAI₁₂ set to missing, the treatment durations were imputed with plausible values (treatment duration between 0 and 12 months) using MI with predictive mean matching, including CDAI₀, disease duration and treatment as covariates.

Simulation

For each condition, the simulation consisted in generating 1000 samples with missing data caused by treatment cessation and estimating the difference in LDA proportion between the two treatments with different statistical methods.

We report bias as the difference between the value estimated in the simulation sample and the true value. The true value is defined here as the LDA rate given by the treatment effect of a linear model predicting LDA on the complete simulation data set (ie, before inducing missing values) adjusting for baseline covariates. In addition, we estimate coverage as the percentage of CIs in the simulation samples which included the true value.

All code to generate simulation data, estimate measures of LDA, and generate the manuscript figures and tables is available in a repository.²⁰ All the simulation, statistical analysis and figures were made in R V.4.1.0,²¹ using the library *ipw*²² for IPW, *mice*²³ for MI with chained equation, *geepack*²⁴ for the generalised estimating equations.

Methods to estimate the LDA

The dataset analysed contained one line per patient's treatment course. All methods used a generalised linear model with Huber-White robust standard errors predicting a binary outcome as a function of the baseline covariates (referred hereafter as the adjusted model).²⁵⁻²⁷ In line with Cheung and co-authors^{26, 27}, we use a Gaussian identity link, and the coefficient for treatment provides the increase of LDA rate compared with the reference treatment. The following estimation methods were considered.

Complete case analysis

CC consists of restricting the analysis only to available data. We considered here the adjusted estimation of the LDA difference between treatments.

Last observation carried forward

All the missing CDAI₁₂ values were set to the last available value of CDAI, which could be the baseline value. The LDA rate is then estimated with the adjusted model.

LUNDEX

LDA rate ($P_{LDA-LUNDEX}$) is estimated by the proportion of patients reaching LDA (P_{LDA}) obtained by the adjusted model multiplied by the Kaplan Meyer estimates of the drug survival P_{LDA} at the time of outcome evaluation¹⁷:

Table 3 Missingness conditions for the simulation, 1 being the reference condition

Condition	Attrition for OMA (%)	Attrition for TNFi (%)	CDAI ₁₂ OR for OMA	CDAI ₁₂ OR for TNFi	Missingness for CC, LOCF, LUNDEX, NRI, IPW1	Missingness for IPW2 and CARRAC
1	20	20	1	1	MNAR	MAR
2	10	20	1	1	MNAR	MAR
3	15	20	1	1	MNAR	MAR
4	25	20	1	1	MNAR	MAR
5	30	20	1	1	MNAR	MAR
6	10	10	1	1	MNAR	MAR
7	15	15	1	1	MNAR	MAR
8	25	25	1	1	MNAR	MAR
9	30	25	1	1	MNAR	MAR
10	20	20	1.07	1	MNAR	MNAR for OMA, MAR for TNFi
11	20	20	1.14	1	MNAR	MNAR for OMA, MAR for TNFi
12	20	20	1.07	1.07	MNAR	MNAR
13	20	20	1.14	1.14	MNAR	MNAR

Attrition for OMA of TNFi indicates the percentage of missing CDAI at 12 months for the other mode of action treatment (OMA) of the reference treatment TNFi, CDAI₁₂ OR are the OR of the CDAI value at 12 months (CDAI₁₂) in the generalised linear model predicting missingness of CDAI₁₂ used to create missing values in the simulation. A CDAI₁₂ OR of 1.07 or 1.14 implies that the odds of having a CDAI₁₂ missing is multiplied by 2 or by 4 for an increase of 10 points of CDAI₁₂. The column 'Missingness for CC, LOCF, LUNDEX, NRI, IPW1' indicates if the missing data are MNAR or MAR for the methods CC, LOCF, LUNDEX, NRI, IPW not accounting for the reasons for treatment cessation in the attrition weights (IPW1). The column 'Missingness for IPW2 and CARRAC' does the same for IPW accounting for the reasons for treatment cessation in the attrition weights (IPW2) and CARRAC.

CARRAC, confounder-adjusted response rate with attrition correction; CC, complete case analysis; CDAI, Clinical Disease Activity Index; IPW, inverse probability weighting; LDA, low disease activity; LOCF, last observation carried forward; MAR, missing at random; MNAR, missing not at random; NRI, non-responder imputation; OMA, another mode of action; TNF, tumor necrosis factor.

$$P_{LDA-LUNDEX} = S_{KM} \times P_{LDA} \quad (\text{equation 1})$$

CI's were calculated by bootstrap using the quantiles of 1000 samples.

Non-responder imputation

All missing values caused by attrition were set to a non-responder value. The LDA proportion was then estimated using the adjusted model.

Inverse probability weighting 1 and 2

The inverse probability weights for treatment (ipwt) and for attrition (ipwc) were computed using a generalised linear model. Weights for treatment (ipwt) included the baseline covariates. Weights for censoring (ipwc) included the same baseline covariates in the first version of this method (IPW1), or additionally included the treatment status (IPW2). The LDA proportion was then computed using the adjusted model with weights equal to $ipwt \times ipwc$.

Multiple imputation

Missing disease activity values at 12 months were imputed using MI using chained equations (mice) with the predictive mean matching algorithm, with baseline covariates and the treatment status.

For each of the imputed samples, the LDA difference between treatments was calculated with the adjusted model. The overall estimate and its SE are then pooled from the 10 mice samples using Rubin's rule.²⁸ The overall method is hereafter named confounder-adjusted response rate with attrition correction (CARRAC).

RESULTS

In the reference scenario, CARRAC and IPW2 provided almost unbiased LDA for each treatment (figure 1), thereby estimating almost unbiased LDA difference for these two methods. CC and

IPW1 overestimated LDA in both treatments. As their overestimations were similar in OMA and in the reference treatment, the absolute LDA difference was almost unbiased. Last observation carried forward (LOCF), LUNDEX and NRI strongly underestimated LDA in both treatments. Because the underestimation was much lower in OMA, the absolute LDA difference was underestimated by these methods. The coverage was 95% or above in individual treatments only with CARRAC, but due to their smaller bias when considering the LDA difference, the two IPW methods and CC retained good coverage for the comparative effectiveness.

When increasing the association between missingness and the true value of effectiveness at 1 year for both treatments (figure 2A and online supplemental figure 2), LUNDEX, NRI, IPW1, LOCF and CC estimations were almost unaffected, while IPW2 and CARRAC started to overestimate the absolute LDA difference. When the association between missingness of CDAI₁₂ and CDAI₁₂ values existed only in one treatment (figure 2B and online supplemental figure 3), all methods overestimated the difference in LDA, leading to inadequate coverage.

When changing the amount of missingness in both treatments, CARRAC, CC and IPW1 provided unchanged LDA difference estimation when compared with the reference situation (figure 2C and online supplemental figure 4). IPW2 estimations remained approximately unbiased, but their dispersion increased significantly. The bias yielded by NRI and LUNDEX, and to a lesser extent by LOCF, was increased when the amount of missing data increased equally in both treatments.

Increasing the amount of missingness in OMA but not in TNFi decreased the absolute LDA difference between the treatments provided by CC and IPW1, while the LDA proportion difference yielded by LOCF, LUNDEX and NRI increased strongly (figure 2D and online supplemental figure 5). IPW2 stayed unbiased but started to produce more dispersed estimations when the percentage of missingness was 30%. CARRAC estimations

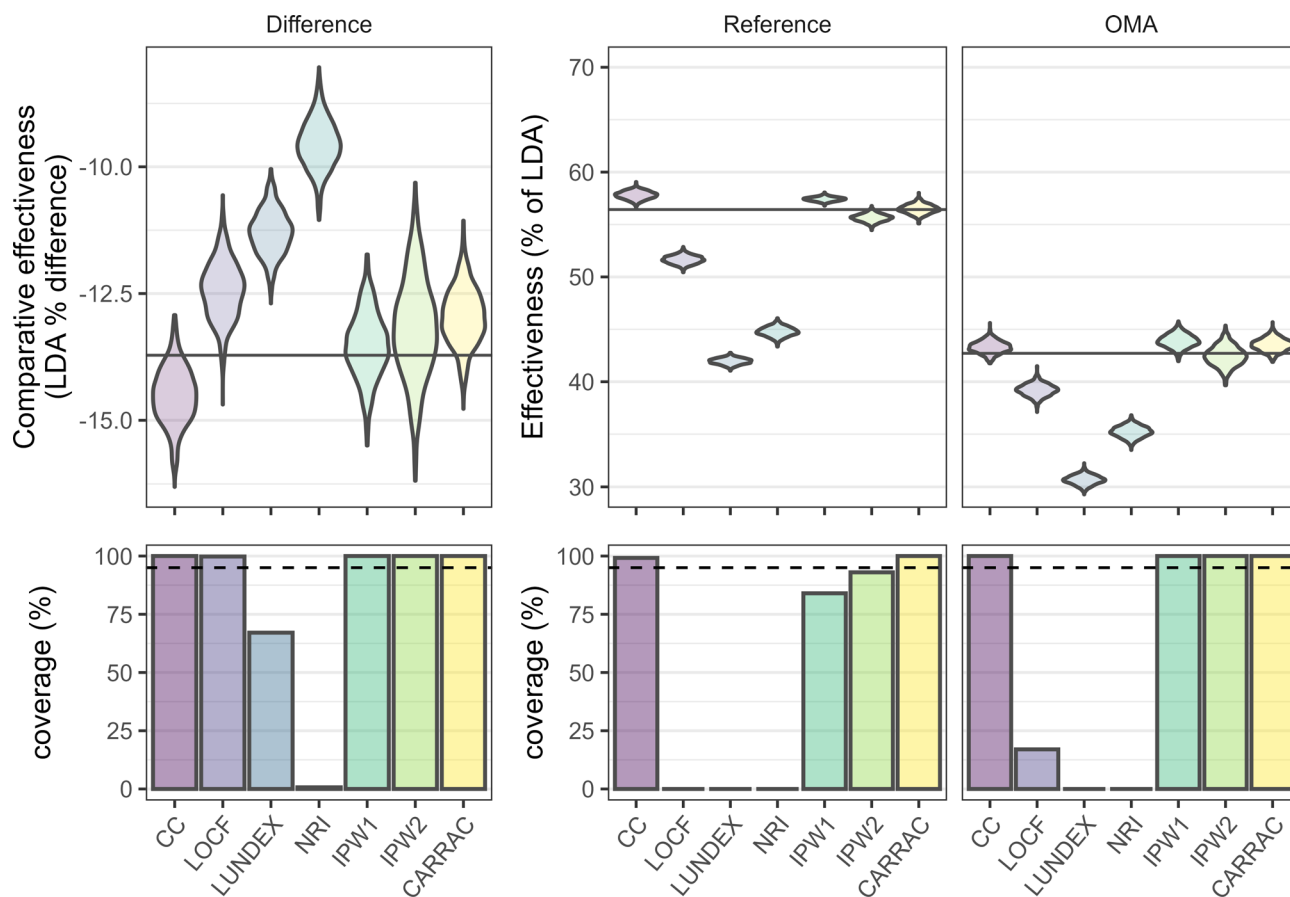


Figure 1 Distribution of the effectiveness measured by the low disease activity (LDA) proportion (upper panels) and the associated coverage (percentage of CIs in the simulation samples which include the true value, lower panels) for each treatment (reference and another mode of action (OMA)—other modes of action, middle and right panels) and for the difference between the treatments (difference, left panels). The methods analysed are complete case analysis (CC), last observation carried forward (LOCF), LUNDEX, non-responder imputation (NRI), inverse probability weighting accounting for the reasons for treatment cessation in the attrition weights (IPW2) or not (IPW1), and confounder-adjusted response rate with attrition correction by reason for drug cessation (CARRAC). The widths of the violins are fixed, so the area of the violin does not represent the number of counts. The true value is represented as a black horizontal line.

of LDA difference slightly underestimated the absolute true difference when increasing the amount of attrition but retained coverage for all missingness conditions.

DISCUSSION

Using a simulation data set generated from a real-world data collection, the present simulation study addresses the issue that some patients may stop the treatment not because it does not work but for some complex reasons including adverse events, remission, etc, which could be taken into account to accurately assess effectiveness. We, therefore, investigated several methods used to compare the response rates of two treatments in presence of confounding and attrition. We focused on attrition, by manipulating the missingness pattern to create greater treatment discontinuation in one treatment versus the other or by increasing the risk of treatment discontinuation. Because observational studies usually use adjusted models to estimate the causal effect of treatments, we used as the true estimate the difference in LDA proportion based on an adjusted model applied on the simulation data set without missing data.

We first observed that the bias in the difference of effectiveness was always lower than the strongest bias observed for the effectiveness estimated in individual treatments. This will always be the case if the estimation of effectiveness is biased in the same direction in both treatments. As expected, methods including

a model for attrition and based on MIs and IPW performed well^{29 30} as long as missingness was dependent only on known covariates (MAR data), but not on unmeasured information (MNAR data). Although an important part of the treatment cessation reasons was unknown, this information was still valuable when estimating effectiveness in the presence of attrition. Including the treatment status in the calculation of the censoring weights permitted IPW to estimate more precisely LDA for each treatment. On the other hand, it led to a higher sensibility to the amount of attrition and to the association between missingness and effectiveness value. This result highlights the importance of model specification for the missingness pattern and the difference between IPW and MI. For instance, the model used for missingness in IPW did not include an interaction term between treatment and variables predicting effectiveness at 1 year (such as baseline disease activity for instance), thus misspecifying the differential effect introduced in the data. MI using predictive mean matching, in the CARRAC method, is less sensitive to misspecification because the model defines the distribution of missing data, which has less variation between the treatments.

CC and IPW1 were biased in each individual treatment, indicating a persistent association between $CDAI_{12}$ values and their missingness. This association, reflecting the fact that patients remaining on treatment tend to have a better response than those stopping,³¹ caused the LDA estimated by CC in each

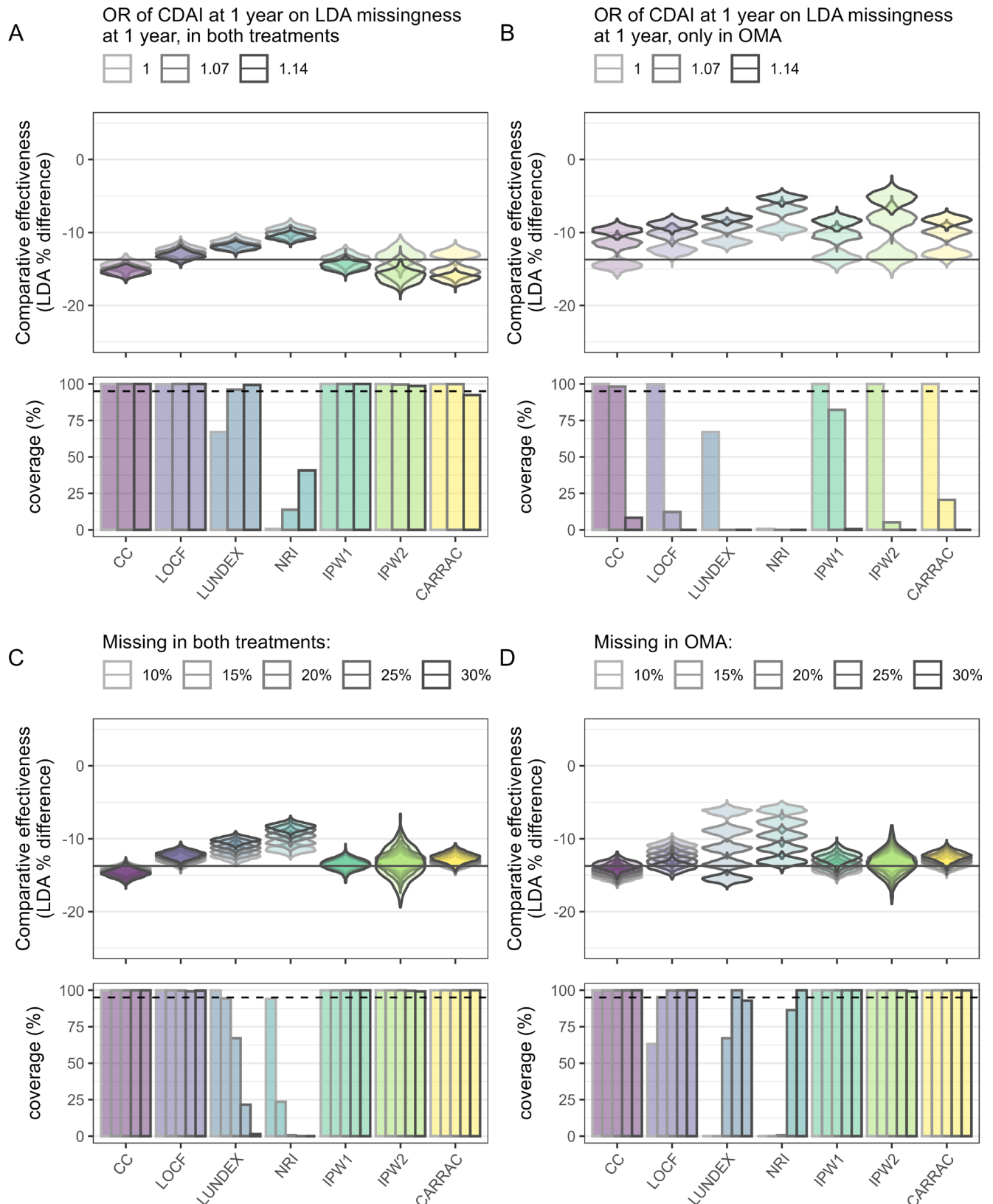


Figure 2 Low disease activity (LDA) difference between treatments and the associated coverage (percentage of CIs in the simulation samples that include the true value) of the reference situation (condition 1) and when having missing not at random in both treatments (condition 12, A), when having missing not at random in one treatment (conditions 10 and 11, B), when having a changing proportion of missing due to attrition in both treatments conditions (conditions 6–9, C) and when having a changing proportion of missing due to attrition in only one treatment (conditions 2–5, D). The methods analysed are complete case analysis (CC), last observation carried forward (LOCF), LUNDEX, non-responder imputation (NRI), inverse probability weighting accounting for the reasons for treatment cessation in the attrition weights (IPW2) or not (IPW1) and confounder-adjusted response rate with attrition correction by reason for drug cessation (CARRAC). The widths of the violins are fixed, so the area of the violin do not represent the number of counts. The true value is represented as a black horizontal line.

treatment to increase with the amount of missing data at 1 year or with the increase in association between $CDAI_{12}$ values and their missingness. IPW1 did so to a smaller extent, as its model partially accounts for the attrition. When the amount of attrition was similar in the two treatments, the biases were similar, and these two methods correctly estimated the difference in LDA. But when the missingness differed between the two treatments, these methods yielded strongly biased estimations of the differential effectiveness.

LOCF, LUNDEX, and NRI underestimated LDA in each treatment. LOCF assumes that the missing values of CDAI are identical to the last available values, although disease activity is known to globally decrease with time.^{32,33} LUNDEX and NRI considered all patients without information at 1 year as non-responders, although studies show that treatment cessation is not only due to ineffectiveness³⁴ but also due to various other reasons, such as adverse events, pregnancy, or even remission. Therefore, when increasing the proportion of missing values, the LDA proportion yielded by NRI and LUNDEX converge to 0%, which is the LDA when all data at 1 year are missing, and the one provided by LOCF converge to an intermediate value comprised between the baseline proportion of LDA and the one at 1 year.

Three main groups of methods with different ways of handling missing data caused by attrition emerged from our discussion: those including a model for attrition (IPW2 and CARRAC), those considering patients who stopped treatment as non-responders (LUNDEX and NRI) or as keeping the same disease activity in time (LOCF) and those not adjusting for attrition (IPW1 and CC). Within each group, these methods did not handle confounding the same way either, thus causing a residual difference between them.

Strength and limitations

The use of data stemming from a real register to generate a simulation data set is a strength of this study, as it allowed to test the statistical methods on close to real-world data. The large variety of methods studied and the release of the code on an open-access repository are also assets to the present work, which guarantee the accessibility of the methods. These methods are also generalisable for any disease or treatments, where treatments may be stopped for different reasons. However, as in every clinical study, the quality of the imputation will depend on model specification, so careful thought should be given to the covariates included in the model for each particular case. Nevertheless, a limitation of any simulation study is that results depend on the model used to generate the data. In the case of this study, our simulation design favours both CARRAC and IPW2, as they are the only ones making use of the treatment status used to generate the data. Thus, we may underestimate the impact of model misspecification. Another limitation is the use of only baseline disease activity, baseline confounders, and reason for stopping to estimate response rate. Though this choice corresponds to the reality of registry data, which usually have few intermediate visits, further models including repeated measurements could be informative. Finally, we used the association between the disease activity at 1 year and the reasons to stop in the future to recreate in the simulation data set the link between the reasons to stop before 1 year and the value of CDAI a patient would have had at 1 year. This procedure may underestimate the real association between reasons for cessation and disease activity. Therefore, the difference between the statistical methods presented here may be more pronounced in real applied analysis.

CONCLUSION

Correct estimation of effectiveness requires considering confounding, treatment cessation and dropouts. While CARRAC and IPW can produce proper estimates, methods omitting one of those, such as LUNDEX, NRI or CC, yield biased estimation, depending on the amount of attrition in the treatments. While the choice of methods is important, and some methods make stronger assumptions than others, model specification remains crucial. Careful justification of the model used for both missingness and confounding is necessary to obtain trustworthy and accurate results.

Twitter Denis Mongin @denis_mongin and Delphine Sophie Courvoisier @delcourvoisier

Acknowledgements The authors thank all the rheumatologists and patients who participated in the registries used to initially create the simulation dataset.

Contributors DM performed the simulation study, the data analysis, participated in the interpretation of the results, and designed and wrote the manuscript; KL participated in the study design and contributed to the interpretation of the results; AF contributed to the interpretation of the results; TF contributed to the study design and to the interpretation of the results; DSC designed the study, contributed to the implementation of the simulation study and the data analysis, participated in the interpretation of the results and is the guarantor of this work. All authors revised critically the manuscript and approved its current version.

Funding This project was supported by an unrestricted grant from AbbVie.

Competing interests None declared.

Patient and public involvement statement Because the present work is a methodological simulation study, it was not appropriate to involve patients or the public in its design, conduct or reporting.

Patient consent for publication Not applicable.

Ethics approval Because the study is a retrospective study using completely deidentified data, it was exempted from ethical approval. Approval of each local ethical committee for the collection of clinical data in each register was obtained for the register collection used to create the simulation dataset.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The simulated dataset and all the code used to generate the data, perform the simulation study and create the figures and tables have been made openly available in the following repository: <https://gitlab.com/dmongin/comparative-effectiveness>. Concerning the original register collection used to create the simulation dataset, all the data belong to the registries. Anonymised data can be shared as long as agreements are made with all participating registries.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Denis Mongin <http://orcid.org/0000-0002-4801-8395>
 Kim Lauper <http://orcid.org/0000-0002-4315-9009>
 Axel Finckh <http://orcid.org/0000-0002-1210-4347>
 Thomas Frisell <http://orcid.org/0000-0002-5735-9626>
 Delphine Sophie Courvoisier <http://orcid.org/0000-0002-1956-2607>

REFERENCES

- Guo Q, Wang Y, Xu D, et al. Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies. *Bone Res* 2018;6:1–14.
- Courvoisier D, Lauper K, Bergstra SA, et al. Op0199 points to consider when analysing and reporting comparative effectiveness research with observational data in rheumatology. *Ann Rheum Dis* 2020;79:124–5.

- 3 van der Heijden GJMG, Donders ART, Stijnen T, *et al.* Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol* 2006;59:1102–9.
- 4 Schafer JL. *Analysis of incomplete multivariate data*. CRC Press, 1997.
- 5 Mongin D, Lauper K, Turesson C, *et al.* Imputing missing data of function and disease activity in rheumatoid arthritis registers: what is the best technique? *RMD Open* 2019;5:e000994.
- 6 Nunan D, Aronson J, Bankhead C. Catalogue of bias: attrition bias. *BMJ Evid Based Med* 2018;23:21–2.
- 7 Howe CJ, Cole SR, Lau B, *et al.* Selection bias due to loss to follow up in cohort studies. *Epidemiology* 2016;27:91–7.
- 8 Lauper K, Kedra J, De Wit M, *et al.* Op0198 a Systematic Review to Inform the Euler Points to Consider When Analysing and Reporting Comparative Effectiveness Research with Observational Data in Rheumatology. *Ann Rheum Dis* 2020;79:123.2–4.
- 9 Eekhout I, de Boer RM, Twisk JWR, *et al.* Missing data: a systematic review of how they are reported and handled. *Epidemiology* 2012;23:729–32.
- 10 Markusse IM, Landewé R, Wolterbeek R, *et al.* Linear extrapolation of missing radiographic change scores in clinical trials does not spuriously overestimate group radiographic changes in rheumatoid arthritis. *Rheumatology* 2016;55:1295–300.
- 11 Horwitz RI, Viscoli CM, Berkman L, *et al.* Treatment adherence and risk of death after a myocardial infarction. *Lancet* 1990;336:542–5.
- 12 Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol* 1992;21:837–41.
- 13 Bell ML, Fiero M, Horton NJ, *et al.* Handling missing data in RCTs; a review of the top medical journals. *BMC Med Res Methodol* 2014;14:118.
- 14 Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008;168:656–64.
- 15 Floden L, Bell ML. Imputation strategies when a continuous outcome is to be dichotomized for responder analysis: a simulation study. *BMC Med Res Methodol* 2019;19:161.
- 16 Demirtas H. Practical advice on how to impute continuous data when the ultimate interest centers on Dichotomized outcomes through Pre-Specified thresholds. *Commun Stat Simul Comput* 2007;36:871–89.
- 17 Kristensen LE, Saxne T, Geborek P. The LUNDEX, a new index of drug efficacy in clinical practice: results of a five-year observational study of treatment with infliximab and etanercept among rheumatoid arthritis patients in southern Sweden. *Arthritis Rheum* 2006;54:600–6.
- 18 Moore AR, Straube S, Eccleston C, *et al.* Estimate at your peril: imputation methods for patient withdrawal can bias efficacy outcomes in chronic pain trials using responder analyses. *Pain* 2012;153:265–8.
- 19 Courvoisier DS, Chatzidionysiou K, Mongin D, *et al.* The impact of seropositivity on the effectiveness of biologic anti-rheumatic agents: results from a collaboration of 16 registries. *Rheumatology* 2021;60:820–8.
- 20 Mongin D, Lauper K, Finckh A. Data from: accounting for missing data caused by drug cessation in observational comparative effectiveness research: a simulation study. Novembre 23, 2021. Available: <https://gitlab.com/dmongin/comparative-effectiveness>
- 21 R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019. <https://www.R-project.org>
- 22 van der der WM, Geskus RB. ipw: an R package for inverse probability weighting. *Journal of Statistical Software* 2011;43:1–23.
- 23 van BS, Groothuis-Oudshoorn K. Mice: multivariate imputation by Chained equations in R. *Journal of Statistical Software* 2011;45:1–67.
- 24 Højsgaard S, Halekoh U, Yan J. The R package geepack for generalized estimating equations. *Journal of Statistical Software* 2005;15:1–11.
- 25 Frisell T, Dehlin M, Di Giuseppe D, *et al.* Comparative effectiveness of abatacept, rituximab, tocilizumab and TNFi biologics in RA: results from the nationwide Swedish register. *Rheumatology* 2019. 10.1093/rheumatology/key433. [Epub ahead of print: 21 Jan 2019].
- 26 Cheung YB. A modified least-squares regression approach to the estimation of risk difference. *Am J Epidemiol* 2007;166:1337–44.
- 27 Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004;159:702–6.
- 28 Rubin DB. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons, 2004.
- 29 Jones M, Mishra GD, Dobson A. Analytical results in longitudinal studies depended on target of inference and assumed mechanism of attrition. *J Clin Epidemiol* 2015;68:1165–75.
- 30 Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res* 2013;22:278–95.
- 31 Iannaccone CK, Fossel A, Tsao H, *et al.* Factors associated with attrition in a longitudinal rheumatoid arthritis registry. *Arthritis Care Res* 2013;65:1183–9.
- 32 Courvoisier DS, Alpizar-Rodriguez D, Gottenberg JE, *et al.* Rheumatoid arthritis patients after initiation of a new biologic agent: trajectories of disease activity in a large multinational cohort study. *EBioMedicine* 2016;11:302–6.
- 33 Aletaha D, Funovits J, Keystone EC, *et al.* Disease activity early in the course of treatment predicts response to therapy after one year in rheumatoid arthritis patients. *Arthritis Rheum* 2007;56:3226–35.
- 34 Verstappen M, van Mulligen E, de Jong PHP, *et al.* DMARD-free remission as novel treatment target in rheumatoid arthritis: a systematic literature review of achievability and sustainability. *RMD Open* 2020;6:e001220.