

**EULAR points to consider when analysing and reporting comparative effectiveness
research using observational data in rheumatology**

Delphine S. Courvoisier¹, Kim Lauper^{1,2}, Joanna Kedra^{3,4}, Maarten De Wit⁵, Bruno Fautrel^{3,4},
Thomas Frisell⁶, Kimme Hyrich^{2,7}, Florenzo Iannone⁸, Pedro M Machado^{9,10}, Lykke Midtbøll
Ørnbjerg^{11,12}, Ziga Rotar¹³, Maria Jose Santos^{14,15}, Tanja Stamm¹⁶, Simon R Stones¹⁷, Anja
Strangfeld¹⁸, Sytske Anne Bergstra¹⁹, Robert B.M. Landewé^{20,21}, Axel Finckh¹

¹Division of Rheumatology, Geneva University Hospitals & Faculty of Medicine, University of Geneva, Geneva, Switzerland, ²Centre for Epidemiology Versus Arthritis, Manchester Academic Health Sciences Centre, The University of Manchester, Manchester, UK, ³Sorbonne Université, Institut Pierre Louis d'Epidémiologie et de Santé Publique (iPLESP), UMR S1136, Paris, France, ⁴AP-HP, Pitié Salpêtrière hospital, Rheumatology Department, Paris, France, ⁵EULAR Patient Research Partner, The Netherlands, ⁶Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden, ⁷NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester, UK, ⁸Department of Emergency and Organ Transplantation, Rheumatology Unit, University of Bari, Bari, Italy, ⁹Centre for Rheumatology & Department of Neuromuscular Diseases, University College London, London, UK, ¹⁰National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre, University College London Hospitals NHS Foundation Trust, London, UK, ¹¹DANBIO registry, Center for Rheumatology and Spine Diseases, Centre for Head and Orthopedics, Rigshospitalet, Copenhagen, Denmark, ¹²Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Centre for Head and Orthopedics, Rigshospitalet,

Copenhagen, Denmark, ¹³biorex.si, Department of Rheumatology, University Medical Centre Ljubljana, and Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia, ¹⁴Reuma.pt registry, Rheumatology Research Unit, Instituto Medicina Molecular, Faculdade Medicina de Lisboa, Lisboa, Portugal, ¹⁵Rheumatology Department, Hospital Garcia de Orta, Almada, Portugal, ¹⁶Section for Outcomes Research, Centre for Medical Statistics, Informatics and Complex Systems, Medical University of Vienna, ¹⁷EULAR Patient Research Partner, Collaboro Consulting, Manchester, UK, ¹⁸Programme area Epidemiology, Deutsches Rheuma-Forschungszentrum (DRFZ) Berlin, Berlin, Germany, ¹⁹Department of Rheumatology, LUMC, Leiden, Netherlands, ²⁰Department of Rheumatology, Zuyderland Medical Center, Heerlen, The Netherlands, ²¹Clinical Immunology & Rheumatology, Academic Medical Center (AMC), University of Amsterdam, Amsterdam, The Netherlands

Word count: 2969

Correspondence:

Delphine Courvoisier, Division of Rheumatology, Geneva University Hospitals & Faculty of Medicine, University of Geneva, Geneva, Switzerland

Email: Delphine.courvoisier@hcuge.ch

Abstract

Background

Comparing treatment effectiveness over time in observational settings is hampered by several major threats, among them confounding and attrition bias.

Objectives

To develop European Alliance of Associations for Rheumatology (EULAR) points to consider (PtC) when analysing and reporting comparative effectiveness research using observational data in rheumatology.

Methods

The PtC were developed using a three-step process according to the EULAR Standard Operating Procedures. Based on a systematic review of methods currently used in comparative effectiveness studies, the PtC were formulated through two in-person meetings of a multidisciplinary task force and a two-round online Delphi, using expert opinion and a simulation study. Finally, feedback from a larger audience was used to refine the PtC. Mean levels of agreement amongst the task force were calculated.

Results

Three overarching principles and 10 PtC were formulated, addressing, in particular, potential biases relating to attrition or confounding by indication. Building on STROBE guidelines, these PtC insist on the definition of the baseline for analysis and treatment effectiveness. They also focus on the reasons for stopping treatment as an important consideration when assessing effectiveness. Finally, the PtC recommend providing key information on missingness patterns.

Conclusion

To improve the reliability of an increasing number of real-world comparative effectiveness studies in rheumatology, special attention is required to reduce potential biases. Adherence to clear recommendations for the analysis and reporting of observational comparative effectiveness studies will improve the trustworthiness of their results.

Introduction

Observational data are increasingly used to analyse the safety and effectiveness of new therapies in different subgroups of patients [1]. For effectiveness studies, as in randomised controlled trials (RCTs), authors typically report the proportion of patients reaching a defined clinical threshold (e.g., for rheumatoid arthritis (RA): European Alliance of Associations for Rheumatology (EULAR) response rates, EULAR/American College of Rheumatology (ACR) remission or low disease activity rates) after a set time. Comparing the proportion of responders across treatments is relatively straight forward in head-to-head RCTs, since treatment groups are similar in terms of patient characteristics by means of randomisation. However, clinical trials have restrictive inclusion criteria and usually short follow-up, and thus do not provide a full picture of clinical responses for the broader patient population seen in clinical practice, especially for chronic diseases [2]. Pragmatic RCT may provide a more real-world picture of comparative effectiveness due to more liberal inclusion criteria but also have short follow-up time, at least under full randomisation [3].

While comparative effectiveness should be assessed also in observational studies and registers, the interpretation of the results is hampered by the limitations of observational studies [4], and in particular two potential limitations. The first limitation is related to confounding. For example, in RA registers, non-tumour necrosis factor inhibitors (TNFi) biologic disease-modifying anti-rheumatic drugs (DMARDs) are often prescribed to older patients, with a higher burden of disease compared to patients receiving TNFi [5, 6]. Assumed advantages of one of the treatments may channel patients with special characteristics, with the consequence that disease activity evolution can be incorrectly attributed to the use of the treatment. This issue is often referred to as confounding by indication or channelling bias. The second limitation is related to a specific type of selection bias called attrition bias. Attrition bias occurs when there are systematic differences between treatment groups in the number or in the way patients are lost from a study [7]. Indeed, when considering effectiveness after a certain time, it is necessary to determine how to take into account patients who stopped the treatment, for example, due to an adverse event or lack of effect, and patients lost to follow-up (e.g., who stopped participating in the registry). Patients who remained on the same treatment may have a better response to the treatment, thus resulting in a selection bias in favour of responders, yielding an overestimation of effectiveness. If there is differential attrition bias, such as more frequent treatment discontinuation of one of

the treatments, or discontinuation of the treatment for different reasons, the comparative effectiveness analysis will be biased.

EULAR has previously published points to consider (PtC) on how to use observational data to analyse and report safety data in biologic registers and report clinical trial extension studies [8, 9]. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines offers a starting framework on how to report studies. With respect to biologic registers, these PtC build on the STROBE guidelines [10], aiming to provide more detailed guidance on reporting complex exposure characteristics, such as time being exposed, drop-out, and change from one exposure to another, with a clear focus on effectiveness outcomes and their analyses. There is an unmet need for PtC on the analysis of effectiveness in purely observational real-world data, especially registers, addressing three key aspects of real-world effectiveness. First, baseline of treatment is often hard to ascertain since patients start and stop different treatments over time. Thus, the one-year follow-up of one treatment could happen 3 months after this treatment was stopped at month 9, and correspond to the start of another treatment. Second, visits often occur at variable time points. Third, treatment discontinuation is substantial and may be informative on treatment success, for instance when patients stop for ineffectiveness. A task force was created with the aim of developing EULAR PtC to analyse and report comparative effectiveness over time (e.g., treatment response rate after a set time) in rheumatology.

Methods

After approval by the EULAR Executive Committee, the convenors (DC, AF) and the fellow (KL) convened a multidisciplinary task force to develop the PtC, guided by the consensus process outlined in the 2014 updated EULAR Standard Operating Procedures (SOPs) [7]. The task force consisted of: 8 rheumatologists, 4 epidemiologists/rheumatologists, 2 statisticians (DSC, TF), 2 patient representatives (MDW, SRS) who were also social sciences researchers, and 2 health professionals (TS, AS).

Two 1-day face-to-face task force meetings were held. The first meeting was convened in March 2019 to clarify the focus of the task force, identify the scope of methods considered in the systematic literature review (SLR), and determine alternative sources of information on accurate analyses to assess comparative effectiveness.

The SLR was performed by the research fellow (KL), with support from two task force members (JK, SAB) and one of the convenors (DC), to identify relevant peer-reviewed publications published in key rheumatology journal (Scientific Journal Ranking >2) in a 10-year period (between January 2008 and March 2019) and see the evolution of analysis and reporting over time. Studies without full text or with less than 100 patients were excluded. The aim was to identify studies comparing treatments on various outcomes in longitudinal observational studies of real-world patients' populations. Of the 9969 abstracts screened, 305 full-text articles were assessed for eligibility; with 211 articles included, only 35% of studies mentioned attrition, and the majority did not use a method that allows adjusting simultaneously for confounding and attrition when estimating comparative effectiveness over time (for a full description of the SLR, see [11]). During the first meeting, the task force also decided to perform a statistical simulation study to assess the accuracy of various methods found in the SLR or those suggested by task force members.

A first draft of the PtC, including 13 items, were prepared by the fellow (KL) and the two convenors (DC, AF). The SLR and simulation results were presented to the task force at a second meeting in November 2019, where the task force formulated a set of overarching principles and consensus statements, based on the initial draft of the PtC. Consensus, defined as $\geq 75\%$ of participants voting ≥ 8 on a 10-points scale to the inclusion of a given item, and on exact wording was undertaken through a two-round online Delphi, with the possibility to leave comments. When no consensus was reached, the statement was reformulated and submitted to a second vote. The mean and SD of the level of agreement (LoA) of task force members, as well as the percentage of participants voting $\geq 8/10$, were then calculated.

The final manuscript was reviewed and approved by all task force members and approved by the EULAR Council (formerly EULAR Executive Committee).

Results

Three overarching principles (Table 1) and 10 PtC (Table 2) were formulated.

Overarching principles

A: Treatment effectiveness relates to how well a treatment performs in routine clinical settings

Although this overarching principle can easily be endorsed by everyone, it depends on how comprehensively it is defined by all stakeholders involved, including patients, and, potentially, carers (non-professional persons helping patients). It is critical that patients are involved in the selection of outcomes that should be measured because their perspective on outcomes that are important differs from those of researchers, health professionals and other stakeholders. Furthermore, how well a treatment performs is often a matter relative to other treatments, instead of an absolute assessment. In practice, there are many therapeutic options available, and the study is more useful if it contains "all" of these, rather than just a comparison of two or three treatments. This improves the possibility to evaluate channelling and gives a more complete picture of the effectiveness relative to the options that would actually be relevant choices in practice. This is in line with the EULAR 2018-2023 strategy, aiming at delivering a comprehensive quality of care framework in patients with RMDs (https://www.eular.org/eular_strategy_2018.cfm).

B: Observational studies have several limitations, including confounding and missing data

Observational studies often have longer follow-up than RCTs and represent 'real-life' patients as seen in a typical clinical practice, with multimorbidities, unscheduled changes in treatment, and incomplete adherence. They are also necessary to investigate some exposures that could not, technically or ethically, be randomized. Observational studies are thus invaluable companions to RCTs. However, data can be hampered by confounding since patients are not randomized.

The main issue with missing data in observational studies is more one of quantity than of quality. Indeed, observational studies often have much more missing data than RCT, in part due to lower manpower, but also due to longer follow-up. In addition, their design as non-interventional studies mirrors clinical practice. This means patients may move to another region (and be lost for follow-up) – but they could also be lost to follow-up due to the severity of their (comorbid) disease [12, 13]. Patients may decide to stop participating in the study, or they may decide to not fill in specific data. Clinicians on the other hand will also perform differently

according to specific patient characteristics or routine procedures. Therefore, missing data may be sometimes missing at random, but not always.

C: Robust and transparent epidemiological and statistical methods increase the trustworthiness of the results

Evidence-based medicine supports clinical decision-making, allowing results to ‘make sense’, thereby ensuring better adherence to treatments and advice. It may also potentially improve patients’ quality of life by helping them to be confident that they made the best possible choice. For complex observational studies, achieving this trustworthiness of results requires particular attention to robust, transparent, and detailed methods.

Table 1: EULAR-endorsed overarching principles for comparative effectiveness research with observational data in rheumatology, with levels of agreement

	LoA, mean (SD)	% votes ≥8/10
A. Treatment effectiveness relates to how well a treatment performs in routine clinical settings	9.7 (1.0)	94%
B. Observational data have several limitations, including confounding and missing data	9.7 (0.8)	94%
C. Robust and transparent epidemiological and statistical methods increase the trustworthiness of the results from observational data	9.8 (0.4)	100%

Numbers in the column ‘LoA’ indicate the mean and SD (in parentheses) of the LoA, as the mean agreement of the task force members on a 0-10 scale.

EULAR, European Alliance of Associations for Rheumatology; LoA, level of agreement.

Points to consider

PtC 1: Reporting of comparative effectiveness in observational studies must follow the STROBE guidelines

The STROBE guidelines already provide comprehensive reporting guidelines for observational studies [10]. However, they lack specific recommendations for longitudinal analyses.

PtC 2: To provide a more complete picture of effectiveness, several outcomes across multiple health domains should be compared

Effectiveness is a complex construct and cannot be assessed by a single outcome. Though several studies can each look at a different outcome, a more prudent approach is to include several outcomes, across multiple health domains, to acknowledge the variety of interests of the involved stakeholders.

PtC 3: Loss to follow-up from the study sample must be reported by treatment

The following two statements aim to address potential attrition bias, by providing necessary information about the extent of loss to follow-up and the potential differential loss to follow-up. Loss to follow-up is defined as having no additional information about a patient after a given time point. In contrast, treatment discontinuation is defined as knowing that the patient stopped a specific treatment at a given time point, whether or not there is information after that time point (e.g., start of a new treatment). Because treatments are often composed of several treatments, it may be necessary to be more specific when describing changes in therapies than simple start and stop of main treatment (e.g., start of conventional synthetic DMARD, in addition to a biologic/targeted synthetic DMARD). It is necessary to report loss to follow-up by treatment or treatment combination, in order to provide information on potential differential loss to follow-up.

PtC 4: The proportion of patients who stop and/or change therapies over time as well as the reasons for treatment discontinuation must be reported

Though the rate of treatment discontinuation may be similar across treatments, the reasons for this discontinuation could differ between treatments. Reasons for discontinuation have also changed since treat-to-target approaches have become more frequent and may call for treatment tapering, especially for patients

under combination therapy. For some RMDs, treatments may sometimes be discontinued when patients are in sustained clinical remission [14, 15], in other words *due to effectiveness*. Thus, in a worst-case scenario, one treatment could have only discontinuation for adverse events, while another could have discontinuation for remission. Consider also examining characteristics of patients who stopped or changed therapies by reason for treatment discontinuation, to determine the importance of attrition bias (e.g., age, gender, and baseline disease severity for each reason of treatment discontinuation per treatment).

PtC 5: Covariates should be chosen based on subject matter knowledge and model selection should be justified

Similar to any adjustment for confounding, the list of covariates for effectiveness at a given time point should be determined based on known potential confounders. Indeed, even recent advances in model selection may still have important issues related to being too data-driven[16], including bias in variable selection, overestimation of parameters, and inflated type I error.

PtC 6: The study baseline should be at treatment initiation and a description of how covariate measurements relate to baseline should be reported

In open cohort studies, determining baseline may become quite difficult. Efforts should be made to accurately define baseline in each study, and explicitly describe whether covariates were measured at baseline. For instance, the visit to assess disease activity could have occurred 2 weeks prior to treatment initiation, while imaging data were obtained at a visit 2 months later. In addition, registers often contain several treatment courses per patients. Consider using data from all treatment courses for the same patient, applying appropriate statistical methods to take into account non-independence.

PtC 7: The analysis should be based on all patients starting a treatment and not limited to patients remaining on treatment at a certain time point

Due to attrition, analysing only patients still on treatment at a certain time point (e.g., 1 year) would lead to bias, by considering only those patients for whom the treatment did not need to be discontinued. Complete case analysis may lead to larger bias as follow-up time and thus attrition increases.

PtC 8: When treatment discontinuation occurs before the time of outcome assessment, attrition should be taken into account in the analysis.

Attrition due to treatment discontinuation is a special case of informative censoring, whereby the patients stopping treatment differ from patients remaining on treatment, for instance by having a smaller decrease in disease activity. Several analysis methods are available to correct this selection bias. However, an increase in the response rate should be interpreted carefully since an apparent increase may represent a selection of patients for whom the treatment worked well instead of an increase of treatment effectiveness over time. In this point to consider, we encourage researchers to consider using multiple imputation techniques and/or causal inference models such as inverse probability weighting, which have been shown to be more accurate than complete case analyses [17]. When data are missing at random, that is when the missingness pattern is dependent on some other variables but can be predicted from available information [18], both methods have been shown to provide reliable estimates (see, e.g., [17, 19, 20]). Nevertheless, because of the importance of model specification of missingness, some simulation studies have shown no better results from complete case analyses than from multiple imputation and IPW [21]. Indeed, other studies showed better results from IPW or multiple imputations methods when the mechanism of either dropout or death were correctly specified [22, 23].

In this framework, members of the taskforce were presented a simulation study that examined the impact of specifying missingness of effectiveness outcome due to treatment discontinuation and attrition [24]. This study used data generated based on a collaboration of registers of biologic DMARDs including ~ 50,000 RA patients. The effectiveness measure assessed was low disease activity (LDA) rate at 1 or 2 years. The methods compared included complete case (CC), Lundex [9], inverse probability weighting (IPW) [17], and a specific multiple imputation model called Confounder-Adjusted Response Rate with Attrition Correction (CARRAC). For both IPW and multiple imputations models, the covariates to specify missingness comprised reasons for treatment discontinuation, in addition to more usual patient characteristics. The conditions tested included

having between 10% and 30% of patients stopping treatment or being lost to follow-up. These percentages were allowed to vary between treatment groups, to investigate differential attrition. Furthermore, a condition evaluated the impact of informative attrition, where CDAI at the time of response rate (1 year) influenced the chance of having discontinued treatment, thus making data 'not missing at random' (NMAR). Results showed that CC usually overestimated LDA at 1 year, and Lundex methods underestimated LDA at 1 year, whereas IPW and CARRAC were usually unbiased. Even though effectiveness estimates assessed by CC or Lundex methods were often quite biased for each treatment, the difference in LDA between two treatments were often closer to the true difference value.

PtC 9: Sensitivity analyses should be undertaken to explore the influence of assumptions related to missingness, particularly in case of attrition

Since assumptions and choices of covariates can have a strong impact on the estimates of effectiveness, sensitivity analyses considering different reasonable alternatives will help determine the robustness of the findings. For instance, using complete case analysis assumes that the effectiveness of the treatment was similar for those who remained on treatment and for those who discontinued (e.g., for lack of loss of effect). The estimate from a second analysis considering all patients who discontinued treatment as non-responders would provide the opposite viewpoint that all discontinuations are due to ineffectiveness. Thus, showing the results of both analysis gives an idea of how much effectiveness can vary based on the assumptions underlying the analyses.

PtC 10: Authors should prepare a statistical analysis plan in advance

Statistical analysis plans (SAP) protect the analyses from becoming too data-driven, influenced by what is seen in the initial descriptive results. This is particularly important for observational studies since analyses are much less clear-cut than for randomized trials. Consider including details on covariates included for adjustment, how these covariates will be included in the models (e.g., age as a continuous linear variable, or as a categorical factor), which outcomes will be considered, which analyses will be done, and which sensitivity analyses will be run.

Table 2: EULAR-endorsed points to consider when analysing and reporting comparative effectiveness

research with observational data in rheumatology, with levels of agreement

	LoA, mean (SD)	% votes ≥75% yes
1. Reporting of comparative effectiveness in observational studies must follow the STROBE guidelines	9.7 (0.7)	100%
2. To provide a more complete picture of effectiveness, several outcomes across multiple health domains should be compared	9.6 (0.5)	100%
3. Lost to follow-up from the study sample must be reported by the exposure of interest	9.7 (0.5)	94%
4. The proportion of patients who stop and/or change therapies over time as well as the reasons for treatment discontinuation must be reported	9.7 (0.6)	94%
5. Covariates should be chosen based on subject matter knowledge and model selection should be justified	9.5 (0.7)	100%
6. The study baseline should be at treatment initiation and a description of how covariate measurements relate to baseline should be included	9.5 (0.5)	100%
7. The analysis should be based on all patients starting a treatment and not limited to patients remaining on treatment at a certain time point	9.8 (0.4)	100%
8. When treatment discontinuation occurs before the time of outcome assessment, the attrition should be taken into account in the analysis. Consider using multiple imputation techniques and/or causal inference models such as inverse probability weighting (IPW)	9.3 (1.0)	100%
9. Sensitivity analyses should be undertaken to explore the influence of assumptions related to missingness, particularly in case of attrition	9.6 (0.6)	100%
10. Authors should prepare a statistical analysis plan in advance	9.6 (0.7)	100%

Numbers in the column 'LoA' indicate the mean and SD (in parentheses) of the LoA, as the mean agreement of the task force members on a 0-10 scale.

EULAR, European Alliance of Associations for Rheumatology; LoA, level of agreement; STROBE, STrengthening the Reporting of OBServational studies in Epidemiology.

Discussion

Observational studies are becoming more comprehensive and detailed. Their longer follow-up allows for a better understanding of the long-term effect of treatments. However, researchers need to be mindful of the risk of biased estimations of effectiveness. Since no solution to adjust for this risk will be perfect, guidance on which information should be reported to allow a fair assessment of potential bias is critical. Indeed, these PtC expand on the STROBE guidelines regarding the importance of describing missing data patterns. Similar to STROBE guidelines, they are relevant not only to RMDs, but to most medical fields using cohort studies to assess effectiveness, and especially to chronic disease treatments.

To our knowledge, no other non-governmental organisation representing patients, healthcare professionals and scientific societies to date has developed recommendations for comparative effectiveness studies. Yet the need for guidelines becomes increasingly evident. First, evidence accrues from numerous publications in statistics, across various medical fields, focused on missing outcome data over time and how to impute them [11-15]. Overall, these studies find that missingness is often informative (i.e., associated with either exposure, or the outcome that should have been measured), thereby making the data 'not missing at random' (NMAR). These results reinforce the message that showing missing data patterns is necessary, to inform readers about differential attrition bias, which would cause a difference in the strength of association found between treatment and the effectiveness outcome. Second, discontinuation of treatment for remission is an option, and thus previous methods such as the simple Lundex approach [9], which considered all patients who stopped treatment as non-responder, are less appropriate than before. Though this trend may be stronger in some countries than in other, depending on local practices or recommendations, evolution in standards of care will continue, as will the need for well-documented reporting and analysing of effectiveness.

Compared to previous EULAR-endorsed PtC, Oxford Centre for Evidence-Based Medicine Levels of Evidence were omitted because no clinical studies were included. Thus, as recommended by EULAR SOP, we downgraded our recommendations to 'points to consider' due to the lack of strong data-driven evidence. However, the agreement between task force members was very high. Though this taskforce represents experts from 11 countries, a limitation is that there was only one representative from Eastern Europe.

Finally, as analyses of observational data become more complex and to accommodate more intricate research questions and data collection, supporting tools should be provided to researchers. These PtC are one tool to support correct reporting of comparative effectiveness studies. Another available support is the EULAR Virtual Research Center (VRC) offering a range of resources including clinical research support. Investigators of future studies should be encouraged to implement variables to be able to adhere to these recommendations, e.g., providing reasons for treatment discontinuation. R packages, SAS procedures, or any other statistical software should be developed to easily implement state of the art analyses, with a detailed documentation clarifying the substantive choices that fall to the investigators.

Twitter accounts: DSC @delcourvoisier, KL @k_lauper

Competing interests:

DSC has received consulting fees from Abbvie, MSD, and Pfizer outside submitted work.

KL has received speaker fees from Gilead-Galapagos and grant/research support from AbbVie outside submitted work.

ZR has received consulting and speaker fees from Abbvie, Eli Lilly, Novartis, MSD, Pfizer, Roche, Sandoz outside submitted work.

PMM has received consulting/speaker's fees from Abbvie, BMS, Celgene, Eli Lilly, Janssen, MSD, Novartis, Orphazyme, Pfizer, Roche and UCB, unrelated to the work presented in this manuscript, and is supported by the National Institute for Health Research (NIHR),

University College London Hospitals (UCLH), Biomedical Research Centre (BRC). **Disclaimer:**

The views expressed here are those of the authors and do not necessarily represent the views of the (UK) National Health Service (NHS), the National Institute for Health Research (NIHR), or the (UK) Department of Health, or any other organisation.

FI has received consulting/speaker's fees from Abbvie, BMS, Celgene, Eli Lilly, Galapagos, Janssen, MSD, Novartis, Pfizer, SOBI, Roche and UCB, unrelated to the work presented in this manuscript.

TS has received grant/research support from AbbVie and Roche, has been a consultant for AbbVie and Sanofi Genzyme, and has been a paid speaker for AbbVie, Roche, Sanofi and Takeda.

AF has received grant/research support from AbbVie, BMS, Eli-Lilly, Galapagos, and Pfizer, has been a paid speaker for AbbVie, BMS, Eli-Lilly, Novartis, on Novartis.

MW operating for Stichting Tools has received fees for lectures or consultancy provided by MW from Celgene, Eli Lilly, Pfizer and UCB, over the last three years, unrelated to the work presented in this manuscript.

AS has received speaker's fees from AbbVie, BMS, Celltrion, MSD, Pfizer, and Roche, unrelated to the work presented in this manuscript.

SRS has received consulting/speaker's fees from 67 Health, Ampersand Health, Envision Pharma Group, Janssen, and On The Pulse Consultancy, and is an employee of Envision Pharma Group, unrelated to the work presented in this manuscript.

LØ has received grant/research support from Novartis, unrelated to the work presented in this manuscript

References

- 1 Cave A, Kurz X, Arlett P. Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe. *Clin Pharmacol Ther* 2019;106(1):36-9.
- 2 Kilcher G, Hummel N, Didden EM, Egger M, Reichenbach S, GetReal Work P. Rheumatoid arthritis patients treated in trial and real world settings: comparison of randomized trials with registries. *Rheumatology (Oxford)* 2018;57(2):354-69.
- 3 Hohenschurz-Schmidt D, Kleykamp BA, Draper-Rodi J, et al. Pragmatic trials of pain therapies: a systematic review of methods. *Pain* 2021.
- 4 Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002;359(9302):248-52.
- 5 Frisell T, Baecklund E, Bengtsson K, et al. Patient characteristics influence the choice of biological drug in RA, and will make non-TNFi biologics appear more harmful than TNFi biologics. *Ann Rheum Dis* 2018;77(5):650-7.
- 6 Lauper K, Nordstrom DC, Pavelka K, et al. Comparative effectiveness of tocilizumab versus TNF inhibitors as monotherapy or in combination with conventional synthetic disease-modifying antirheumatic drugs in patients with rheumatoid arthritis after the use of at least one biologic disease-modifying antirheumatic drug: analyses from the pan-European TOCERRA register collaboration. *Ann Rheum Dis* 2018;77(9):1276-82.
- 7 Nunan D, Aronson J, Bankhead C. Catalogue of bias: attrition bias. *BMJ Evid Based Med* 2018;23(1):21-2.
- 8 Buch MH, Silva-Fernandez L, Carmona L, et al. Development of EULAR recommendations for the reporting of clinical trial extension studies in rheumatology. *Ann Rheum Dis* 2015;74(6):963-9.
- 9 Dixon WG, Carmona L, Finckh A, et al. EULAR points to consider when establishing, analysing and reporting safety data of biologics registers in rheumatology. *Ann Rheum Dis* 2010;69(9):1596-602.
- 10 von Elm E, Altman DG, Egger M, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335(7624):806-8.
- 11 Lauper K, Kedra J, de Wit M, et al. Analysing and reporting of observational data: a systematic review informing the EULAR points to consider when analysing and reporting comparative effectiveness research with observational data in rheumatology. *RMD Open* 2021;7(3).
- 12 Fantini F, Gerloni V, Gattinara M, Cimaz R, Arnoldi C, Lupi E. Remission in juvenile chronic arthritis: a cohort study of 683 consecutive cases with a mean 10 year followup. *J Rheumatol* 2003;30(3):579-84.
- 13 Lunt M, Watson KD, Dixon WG, et al. No evidence of association between anti-tumor necrosis factor treatment and mortality in patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. *Arthritis Rheum* 2010;62(11):3145-53.
- 14 Horton DB, Onel KB, Beukelman T, Ringold S. Attitudes and Approaches for Withdrawing Drugs for Children with Clinically Inactive Nonsystemic JIA: A Survey of the Childhood Arthritis and Rheumatology Research Alliance. *J Rheumatol* 2017;44(3):352-60.

- 15 Smolen JS, Landewe RBM, Bijlsma JWJ, 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(6):685-99.
- 16 Sauerbrei W, Perperoglou A, Schmid M, et al. State of the art in selection of variables and functional forms in multivariable analysis-outstanding issues. *Diagn Progn Res* 2020;4:3.
- 17 Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res* 2013;22(3):278-95.
- 18 Rothman KJ, Lash TL, VanderWeele TJ, Haneuse S. *Modern epidemiology*. 2021.
- 19 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(10):1165-75.
- 20 Wen L, Terrera GM, Seaman SR. Methods for handling longitudinal outcome processes truncated by dropout and death. *Biostatistics* 2018;19(4):407-25.
- 21 Lewin A, Brondeel R, Benmarhnia T, Thomas F, Chaix B. Attrition Bias Related to Missing Outcome Data: A Longitudinal Simulation Study. *Epidemiology* 2018;29(1):87-95.
- 22 Welch CA, Sabia S, Brunner E, Kivimaki M, Shipley MJ. Does pattern mixture modelling reduce bias due to informative attrition compared to fitting a mixed effects model to the available cases or data imputed using multiple imputation?: a simulation study. *BMC Med Res Methodol* 2018;18(1):89.
- 23 Lotz M, Miyahara S, Tang G. Pattern Mixture Models Incorporating Reasons for Dropout. In: *Biometrics section: Proceedings of the Survey Research Methods Section, American Statistical Association*. 2008.
- 24 Lauper K, Mongin D, Gabay C, Finckh A, Courvoisier D. SAT0588 COMPARATIVE EFFECTIVENESS RESEARCH IN OBSERVATIONAL SETTINGS: EVALUATING TWO NEW METHODS TO ANALYSE RESPONSE RATES. *Annals of the Rheumatic Diseases* 2019;78(Suppl 2):1386.