

PROF. MAARTEN BOERS (Orcid ID : 0000-0002-6969-283X)

DR. JASVINDER A SINGH (Orcid ID : 0000-0003-3485-0006)

DR. S. LOUIS BRIDGES, JR. (Orcid ID : 0000-0003-3785-1389)

DR. JEFFREY R CURTIS (Orcid ID : 0000-0002-8907-8976)

Article type : Brief Report

A novel method to combine assessment of benefit and harm: OMERACT 3x3 methodology applied to two active comparator trials

Maarten Boers, MSc, MD, PhD,¹ Jasvinder A. Singh, MD, MPH,^{2,4,5} Stacey S. Cofield, PhD,³ S. Louis Bridges Jr, MD PhD,³ Larry W. Moreland, MD,⁶ James R. O'Dell, MD,⁷ Hongsheng Wu, PhD,⁸ Sarah Leatherman, MA,⁸ Jeffrey R. Curtis, MD, MS, MPH.³

¹Epidemiology & Biostatistics; Amsterdam Rheumatology and Immunology Center, VU University Medical Center, Amsterdam, Netherlands, ²Epidemiology and ³Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, AL, ⁴Medicine Service, Birmingham VA Medical Center, Birmingham, AL, ⁵Department of Orthopedic Surgery, Mayo Clinic College of Medicine, Rochester, MN, ⁶Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, PA, ⁷Rheumatology, University of Nebraska Medical Center, Omaha, NE, United States, ⁸Boston VA Medical Center, Boston, MA.

Contact information:

Maarten Boers, MSc, MD, PhD
Professor of Clinical Epidemiology
Department of Epidemiology and Biostatistics
PO Box 7057
1007 MB Amsterdam, Netherlands
ph: +3120 444 4474

Email: eb@vumc.nl

Conflicts of interested reported: none

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/acr.23590

This article is protected by copyright. All rights reserved.

Abstract

Introduction

The Outcome Measures in Rheumatology (OMERACT) '3x3' method analyzes the occurrence of benefit and harm simultaneously at the individual patient level. We applied this method to two recent rheumatoid arthritis (RA) trial datasets.

Methods

Treatment of Early Aggressive RA (TEAR) and the RA Comparison of Active Therapies (RACAT) randomized trial outcomes for safety were defined according to OMERACT as no events, non-serious events, and serious events. Treatment efficacy was defined as good, moderate, or no response. A good treatment response without any events was labeled an 'unqualified success', and no treatment response but at least one adverse event an 'unmitigated failure'. Chi-square or exact tests assessed the association between benefit and harm, as appropriate.

Results

In TEAR, 612 of 755 patients had response data at 48 weeks: 14% of patients experienced unqualified success and 9% unmitigated failure, with no difference between the treatment arms. Treatment response and adverse event rates were not correlated. In RACAT, 309 of 353 patients had response data at 48 weeks: 6% of patients experienced unqualified success and 11% unmitigated failure, with no differences between the treatment arms. Response and adverse event rates were negatively correlated: frequency of AE and SAE increased as response decreased ($p=0.008$).

Conclusion

We found some evidence that clinical response may be reduced by the co-occurrence of adverse events.

Significance and Innovations

- Clinical trials mostly report on benefit, and the limited information on the occurrence of harm is not related to the benefit results.
- The Outcome Measures in Rheumatology (OMERACT) '3x3' method analyzes the occurrence of benefit and harm simultaneously at the individual patient level.
- Application of this method to two recent trials in RA showed some evidence that clinical response may be reduced by the co-occurrence of adverse events.

Background

Most clinical trial reports provide far more detailed benefit assessments than harms. Frequently, the expected benefit of the experimental treatment is expressed in a responder index, e.g. the proportion of rheumatoid arthritis (RA) patients experiencing relevant benefit (e.g. American College for Rheumatology improvement criteria (ACR20),¹ or European League Against Rheumatism (EULAR) response criteria.² In contrast, adverse events typically are counted and summarized as frequencies at the treatment group level. Differences between treatments in these outcomes are sometimes summarized both as the number of patients needed to treat (NNT) or number needed to harm (NNH). However, the comparison of benefit and harm of interventions is therefore considered separately, leaving the reader unsure of how the interventions truly compare to each other.

The Outcome Measures in Rheumatology (OMERACT) Initiative has suggested trial reports can be enhanced by analyzing the occurrence of benefit and harm simultaneously, at the level of the individual patient rather than at the group treatment level.³ The underlying idea is that patients and physicians not only need to know what the chances of benefit and harm of a treatment are, but also whether these chances are correlated: in other words, it could be that the patients who are benefitting from the intervention are the same patients who are more (or less) likely to experience harms. This approach has the potential ability to inform more personalized care and improve shared decision making, in that it can be conducted in important subgroups to understand the unique benefit-risk profile of interventions across heterogeneous patient phenotypes.

Methods

The OMERACT method suggests creating 2 or 3 levels of benefit (e.g., good, moderate or no response) and likewise 2 or 3 levels of harm (e.g., no adverse events [AE], non-serious AE and serious adverse events [SAE]). The trial outcome for each patient is expressed as a pair of values expressing the level both of benefit and harm, and a contingency table (2x2 or 3x3) can be created. In the 3x3 table, patients achieving a good response without any adverse event can be labeled as 'unqualified success'⁴, and patients experiencing no response but at least one adverse event as 'unmitigated failure'.

The 6 and 12-month data data of two recent active comparator trials were re-analyzed according to this framework: the Treatment of Early Aggressive Rheumatoid Arthritis (TEAR) trial⁵ and the 'RA Comparison of Active Therapies' (RACAT) trial.⁶

The TEAR trial used a factorial design and randomized 755 early RA patients to one of 4 treatments: two groups received immediate combination therapy of either methotrexate (MTX) plus etanercept or oral triple therapy with MTX, sulfasalazine and hydroxychloroquine; and two groups received initially MTX monotherapy with the step-up to combinations at 24 weeks after insufficient response based on DAS28 score. In the primary study report, analysis at 6 but not at 12 months showed a better response in the immediate treatment groups compared to the delayed groups, and no differences in safety between the 4 treatment groups.

The RACAT trial compared addition of sulfasalazine and hydroxychloroquine or etanercept to MTX in 353 RA patients with inadequate response to MTX. Patients were randomized to one of the two treatment strategies and were switched to the alternative strategy if they showed no clinical improvement after 24 weeks. At 24 and 48 weeks (primary analysis) triple therapy was non-inferior to etanercept plus MTX.

For this post-hoc analysis, initially 3 benefit/harm categories were made as described above by applying the EULAR response criteria and examining AE reports for each patient. Patients dropping out prematurely for a patient or investigator-reported reason of “side effects or any other medical issue” were assigned a code of AE even if such an event had not otherwise been explicitly reported. In addition, the data were further collapsed into 2x2 tables categorizing patients as experiencing harm according to two scenarios: harm = SAE, or harm = any AE; and likewise, as experiencing benefit or no benefit according to two scenarios: benefit = good response; or benefit = good or moderate response. We designed a new way to present this data: each cell of the 2x2 table also contains a color-coded bar with a length proportional to the percentage in that cell. On the orange/blue (bottom left to top right) diagonal, one can see the balance between worst and best; and on the on the light grey/purple (top left to bottom right) diagonal, one can see the balance between the tradeoff categories, i.e. ‘no benefit + no harm’ vs ‘benefit + harm’. The Figure was designed in Microsoft Powerpoint and drawn by hand (line tool) where necessary. Colors were chosen to respect color blindness.⁷ The distributions were tested for significance by chi-square tests, without adjustment for multiple comparisons.

Results

TEAR trial

Of the 755 randomized patients, 693 had response data at 24 weeks and 617 at 48 weeks. At 48 weeks, 66% had experienced a non-serious AE, and 7% a serious AE (Figure 1, left top panel). Overall, 14% patients experienced an unqualified success (EULAR good response and no AE) and 9% an unmitigated failure (no EULAR response and either a non-serious AE or SAE; Figure 1). At 24 weeks, results were similar (data not shown). No clear pattern emerged to suggest that the occurrence of benefit and harm was correlated. In the 2x2 analyses, combining the counts of good and moderate responders increased the numbers of patients with unqualified success, as expected (24%), and combining those of moderate and non-response increased the numbers with unmitigated failure (36%). None of the statistical tests were significant and none found differences in the distribution of benefits and harms across the four treatment groups (Figure 1, right top panel for one of the 2x2 analyses).

RACAT trial

Of the 353 randomized patients, 321 had response data at 24 weeks, and 309 at 48 weeks. At 48 weeks, 73% experienced a non-serious, and 15% a serious AE. Overall, 6% of patients experienced an

unqualified success while 11% experienced an unmitigated failure. Benefit was negatively associated with harm: frequency of AE and SAE increased as response decreased ($p=0.008$; Figure 1, left bottom panel). In the 2x2 table analysis, this association was significant in the scenarios that contrasted SAE with no SAE (Figure 1, right bottom panel). Results for 24 weeks were similar, but the association between benefit and harm was not significant (data not shown).

Discussion

Reporting the occurrence and benefit and harm together is a new way to report trial results, and makes the trade-offs involved in choosing a certain therapy more explicit. This would aid in discussions with patients and provide for further shared decision making. We studied two active-comparator trials, and in RACAT the occurrence of treatment benefit was associated with a lower occurrence of harm, regardless of treatment allocation. In addition, RACAT clearly had more patients with AE, impacting the proportion of patients with an unqualified success.

This analysis method has limitations, and should be regarded as complementary, not a replacement of current analysis and reporting strategies. One is inherent in the way adverse events are assessed and reported: unlike benefit, the 'adverse event experience' of a patient is not easily summarized. The three categories of harm as proposed in this system are a simplification, and the largest category, 'non-serious AE' is very broad. And even with this simplification, presenting a 2-arm trial with two 3x3 tables side by side requires the reader to compare 2x9 numbers, a hard task. This can be simplified by reporting the best and worst outcomes (unqualified success and unmitigated failure), by collapsing the results into 2x2 tables, and adding a graphical representation. Graphs may ease interpretation through visual pattern recognition. The downside of course is further simplification and loss of accuracy: e.g. should we combine all AE into one category, or should we contrast SAE vs no SAE? In addition, our analysis strategy does not improve on well-known limitations of current AE definitions: the consideration of non-serious but potentially treatment adherence limiting side-effects and the likelihood of relationship to treatment. In the case of the RACAT dataset, the statistical analysis result depended on the definition of the categories.

An advantage of the method is that it is flexible, i.e. neutral to the definition of benefit and harm, preference weighting of harm (e.g. to incorporate the patient perspective), and attribution of harm to treatment. Such choices should be made in the protocol phase, and the results interpreted accordingly. For example, it could be argued that some mild AE should not be considered harm at all; and the definition of SAE is mandated by regulatory authorities, sometimes resulting in misclassification from the clinician's point of view: some AE are not grouped under the category of SAE even though they are reported as 'severe' in intensity, and some are categorized as SAE where the clinician reports them as moderate or mild in severity.

The finding of an association between less benefit and more harm in RACAT is unexpected and novel. Given its occurrence in only 1 of 4 comparisons (2 time points in 2 trials) it may also be the play of chance. The expectation would be to find a positive association, i.e., more benefit, but also more harm, but perhaps this requires a stronger contrast between therapies, as in a placebo-controlled trial. In addition, the treatments in TEAR and RACAT are known to be well-tolerated.

In conclusion, we have applied the OMERACT methodology of combined reporting of benefit and harm to two active comparator trials in RA, and suggested a new graphical summary. As a complement to existing methods, this approach is likely to provide a useful framework for simultaneously considering the risks and benefits of treatments at a patient level, both overall and for important patient subgroups such as older individuals and other patients with high comorbidity burdens, who are more likely to suffer harms.

References

1. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727-35.
2. van Gestel AM, Prevoo ML, van 't Hof MA, van Rijswijk MH, van de Putte LB, van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. *Arthritis Rheum* 1996;39:34-40.
3. Boers M, Brooks P, Fries JF, Simon LS, Strand V, Tugwell P. A first step to assess harm and benefit in clinical trials in one scale. *J Clin Epidemiol* 2010;63:627-32.
4. Mancini GB, Schulzer M. Reporting risks and benefits of therapy by use of the concepts of unqualified success and unmitigated failure: applications to highly cited trials in cardiovascular medicine. *Circulation* 1999;99:377-83.
5. Moreland LW, O'Dell JR, Paulus HE, Curtis JR, Bathon JM, St Clair EW, et al. A randomized comparative effectiveness study of oral triple therapy versus etanercept plus methotrexate in early aggressive rheumatoid arthritis: the treatment of Early Aggressive Rheumatoid Arthritis Trial. *Arthritis and Rheumatism* 2012;64:2824-35.
6. O'Dell JR, Mikuls TR, Taylor TH, Ahluwalia V, Brophy M, Warren SR, et al. Therapies for active rheumatoid arthritis after methotrexate failure. *N Engl J Med* 2013;369:307-18.
7. Okabe M, Ito K. Color Universal Design (CUD): How to make figures and presentations that are friendly to Colorblind people. 2008. Accessed 06-03-2018, at <http://jfly.iam.u-tokyo.ac.jp/color>

Legends to Figures

Figure 1.

Results of the combined assessment of benefit and harm in two randomized trials.

Top panels: TEAR trial (617 evaluable patients). Bottom panels, RACAT trial (309 evaluable patients). In each panel, benefit increases from left to right, and harm increases from top to bottom.

In the panels on the left, results of treatment groups are pooled and categorized according to the combined occurrence of benefit and harm, each in 3 categories. Results are expressed as percentage of the total group, corrected for rounding. White lines delineate the cutoffs for the 2x2 categorization in the right-hand panels.

nsAE: non-serious adverse event; SAE: Serious adverse event.

The panels on the right present the results (%) per treatment group, with combined occurrence of benefit and harm each in 2 categories: for benefit, EULAR good response (yes/no); for harm, occurrence of any serious adverse event (SAE; yes/no). Length of the diagonal bar in each cell is proportional to the percentage of patients in that cell.

The orange/blue (bottom left to top right) diagonal, shows the balance between worst and best; the light grey/purple (top left to bottom right) diagonal shows the balance between two types of tradeoff: 'no benefit + no harm' and 'benefit + harm'.

MTX: methotrexate; ETN, etanercept; triple: MTX, sulfasalazine, hydroxychloroquine.

