PEER REVIEW HISTORY

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

ARTICLE DETAILS

TITLE (PROVISIONAL)	COMPARATIVE EFFECTIVENESS OF BARICITINIB AND ALTERNATIVE BIOLOGICAL DMARDs IN A SWISS COHORT STUDY OF RA PATIENTS
AUTHORS	Gilbert, Benoit Thomas P.; Mongin, Denis; Aymon, Romain; Lauper, Kim; LAEDERMANN, Cédric; PERRIER, Clémentine; Mueller, Ruediger; Courvoisier, Delphine; Finckh, Axel

VERSION 1 – REVIEW

REVIEWER	Landewé, Robert
REVIEW RETURNED	N/A

GENERAL COMMENTS	Gilbert et all have analyzed 1052 treatment courses (starts) with bDMARDs and tsDMARDs in the SCQM cohort. Their focus was on baricitinib, and two co-authors are Eli-Lilly employees. Their contribution to the analysis is unclear to me. Most likely, Eli-Lilly has sponsored this study, or the SCQM-cohort, or both, but I cannot see any mentioning of sponsoring in the manuscript. This is an omission. During the last ten years I have seen many of such papers coming by, all focusing on the drug of a particular sponsor, all claiming that 'in real world' their drug was at least as good/safe as -or perhaps a bit better/safer- than the comparators/competitors. In that regard, the outcome of the analysis in this manuscript is also not rocket-science: 'Bari is at least as good as....'. It is what EULAR has already recommended in 5 versions of the EULAR RA-recommendations. From an epidemiological point of view, registries like this are not apt to analyze comparative effectiveness/safety of different DMARDs since patients starting/using these drugs are not comparable by default, leading to bias by indication. Bari, for instance, was likely a drug that was used more frequently in patients who had failed or could not tolerate one or more bDMARDs, or alternatively did not want to use MTX, or alternatively did not want to use parenteral drugs, etc etc. To some extent you can analytically adjust for such differences, and the authors have gone far to present complicated analytical techniques to convince the reader that there was appropriate adjustment- but the principles remain: inherent incomparability of drugs. What makes me a bit skeptical is that -as said- I have seen

	many papers claiming exactly the same for their b/tsDMARD of choice, and undoubtedly many more will follow. I think, in summary, that readers can better follow official guidelines, such as those from ACR and EULAR, that do not prioritize certain b/tsDMARDs, because the evidence that some drugs are better or worse than others is simply lacking. There is one additional intriguing point that deserves a bit of attention in the context of ORAL-surveillance and the outcome of the article 20 procedure rearing JAKi: patients on Bari were significantly older! If true, namely that in Switzerland, Bari (and other JAKi) are prioritized in older patients, Swiss rheumatologists may have an inconvenient problem.

REVIEWER	Nam, Jacquie
REVIEW RETURNED	N/A

GENERAL COMMENTS	This nested cohort study from the SCQM registry provides important real-world data on the use of baricitinib compared to other targeted therapies in patients with RA in clinical practice. Additional comments/ questions: - How have the authors defined ineffectiveness? It is noted that the rates of LDA and REM at 12 months did not differ between groups. Was ineffectiveness and treatment switch clinician +/- patient defined and determined? This should be clarified in the manuscript. - Higher baseline CDAI was one of the variables associated with higher drug discontinuation in the TNFi vs. BARI groups. The authors have imputed baseline CDAI and adjusted for this, but the proportion of patients with missing CDAI is very high (approximately 2/3 of patients). Similarly, at 12 months missing CDAI is high. This is noted in the limitations section in the discussion but, with the amount of missing data, can the authors confidently conclude that clinical outcomes did not differ between the BARI and TNFi groups? - The 45 day window for the 12 month CDAI seems very wide. Were there no new treatment starts within this period that could have affected this result? - As noted in the manuscript, one of the reasons patients may be less likely to switch treatment is multiple previous treatment failures. The authors have done an exploratory subanalysis in b/tsDMARD naïve patients which does not suggest this. The numbers of patients, particularly in the BARI subgroup at 12 months, however are relatively small (evidenced by the wide HR confidence intervals). This should be borne in mind when interpreting the result and should be noted in the discussion. Minor comments/ suggestions: - As BARI is also approved for severe RA, the phrase ‘mild to moderate’ could be reworded or deleted. - Probably not necessary to use ‘sharp’ when referring to the exact time point - Page 6 line 33 – ‘initiation’ rather than ‘initiations’

REVIEWER	Assadiasl, Sara

	Tehran University of Medical Sciences
REVIEW RETURNED	10-Oct-2023

GENERAL COMMENTS	The article “REAL WORLD EFFECTIVENESS OF BARICITINIB IN THE SWISS RHEUMATOID ARTHRITIS REGISTER (SCQM-RA)” has compared baricitinib, a JAK inhibitor, with TNF-inhibitors and a group of other bDMARDs, including tocilizumab, abatacept, sarilumab, and anakinra for treating a considerable number of RA patients in practice. The results have shown a superiority of baricitinib to the TNF-inhibitors in terms of discontinuation due to ineffectiveness. Currently, baricitinib and other JAK inhibitors are being used to treat a wide range of inflammatory conditions, thus there is a need to evaluate their efficacy and safety in the real world. Therefore, any data, particularly those based on daily practice can be of value in choosing the most appropriate medication for each patient. The manuscript has been thoroughly revised according to the comments of the previous reviewers. I have no further comment. Regarding the concerns about the conflict of interest in this article, I (as a researcher with no conflict of interest) should say that the outcomes do not seem to be exaggerated or biased since the majority of compared values have turned out to be equal across groups.

REVIEWER	Wang, Wei Food and Drug Administration, Division of Biostatistics, Center for Devices and Radiological Health
REVIEW RETURNED	21-Nov-2023

GENERAL COMMENTS	In this manuscript, the authors compares the effectiveness of baricitinib (BARI), a targeted synthetic DMARD (tsDMARD) with alternative biological DMARDs (bDMARDs) in rheumatoid arthritis (RA) patients, from a prospective, longitudinal cohort and the study showed that the overall drug maintenance of BARI was significantly longer compared to TNFi. I mainly provided a statistical review for a manuscript and have several comments for the author’s consideration, specifically, 1) In this study, since 1053 TC were initiated in 834 different patients. With this correlated data structure in which one patient may contribute more than one TCs, it is not appropriate to use 3-way ANOVA or Chi2 tests to compare the baseline characteristics of the study population in Table 1. The authors may select appropriate analysis methods (for example, mixed model or other methods) to account for this correlated data structure in Table 1 hypothesis testing and p value calculation. Please update the p value calculation using appropriate data analysis methods in Table 1. 2) For the main analysis, to account for the cluster or correlated survival outcome, the authors incorporated a cluster term in the analysis model as mentioned in the Statistical analysis Section. This is appropriate, but further explanation and clarification is needed to explain the cluster term used in the model analysis, and related reference is also needed for the cluster term used in the data analysis. The authors also need to provide related reference for the Fine-Gray approach to assess specific reasons for drug discontinuation (i.e. ineffectiveness, or adverse event) in a competing-risk setting. 3) In the conclusions, the authors claimed that, in this non-randomized prospective cohort study, the treatment with BARI has

	at least similar effectiveness outcomes as alternative bDMARDs. This expression is not accurate, since non-significant effectiveness outcome is not equivalent to similar effectiveness outcome. A formal equivalence test is needed to conclude similar or equivalent effective outcome between treatment groups. The authors may need to update the final conclusion to be more accurate and consistent with the final statistical analysis results

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Sara Assadiasl, Tehran University of Medical Sciences

Comments to the Author:

The article “REAL WORLD EFFECTIVENESS OF BARICITINIB IN THE SWISS RHEUMATOID ARTHRITIS REGISTER (SCQM-RA)” has compared baricitinib, a JAK inhibitor, with TNF-inhibitors and a group of other bDMARDs, including tocilizumab, abatacept, sarilumab, and anakinra for treating a considerable number of RA patients in practice. The results have shown a superiority of baricitinib to the TNF-inhibitors in terms of discontinuation due to ineffectiveness.
Currently, baricitinib and other JAK inhibitors are being used to treat a wide range of inflammatory conditions, thus there is a need to evaluate their efficacy and safety in the real world. Therefore, any data, particularly those based on daily practice can be of value in choosing the most appropriate medication for each patient.
The manuscript has been thoroughly revised according to the comments of the previous reviewers. I have no further comment. Regarding the concerns about the conflict of interest in this article, I (as a researcher with no conflict of interest) should say that the outcomes do not seem to be exaggerated or biased since the majority of compared values have turned out to be equal across groups.
 We sincerely thank the reviewer for having read our manuscript and for this encouraging comment.

Reviewer: 2
Dr. Wei Wang, Food and Drug Administration

Comments to the Author:
In this manuscript, the authors compare the effectiveness of baricitinib (BARI), a targeted synthetic DMARD (tsDMARD) with alternative biological DMARDs (bDMARDs) in rheumatoid arthritis (RA) patients, from a prospective, longitudinal cohort and the study showed that the overall drug maintenance of BARI was significantly longer compared to TNFi. I mainly provided a statistical review for a manuscript and have several comments for the author’s consideration, specifically,

 We thank the reviewer for taking the time to review our manuscript.

1) In this study, 1053 TC were initiated in 834 different patients. With this correlated data structure in which one patient may contribute more than one TCs, it is not appropriate to use 3-way ANOVA or Chi2 tests to compare the baseline characteristics of the study population in Table 1. The authors may select appropriate analysis methods (for example, mixed model or other methods) to account for this correlated data structure in Table 1 hypothesis testing and p value calculation. Please update the p value calculation using appropriate data analysis methods in Table 1.
 We acknowledge this is a confusing aspect of our study design. We updated the methods section and proposed using generalized linear mixed models. It now reads as follows:
Analyses were conducted and reported in accordance to EULAR recommendations for comparative effectiveness research.[9] Baseline characteristics were compared using generalized linear mixed

models to account for repeated treatments within the same patientsANOVA or χ2 tests as appropriate.
We updated p-values in the Table 1 accordingly (which made them slightly less significant).

2) For the main analysis, to account for the cluster or correlated survival outcome, the authors incorporated a cluster term in the analysis model as mentioned in the Statistical analysis Section. This is appropriate, but further explanation and clarification is needed to explain the cluster term used in the model analysis, and related reference is also needed for the cluster term used in the data analysis. The authors also need to provide related reference for the Fine-Gray approach to assess specific reasons for drug discontinuation (i.e. ineffectiveness, or adverse event) in a competing-risk setting.

 Thank you for pointing this out. The cluster term used is further defined in the documentation of the Coxph() function, under the title “Special terms” (https://stat.ethz.ch/R-manual/R-devel/library/survival/html/coxph.html). It computes a robust variance for the model, by applying the so-called Huber sandwich estimator. We have made that clearer in the method section and added the appropriate reference (https://www.jstor.org/stable/1912934), as follows:
The main analysis (survival analysis) accounted for clustering resulting from patients with multiple treatment courses, inducing correlation within the patient-level data. The cluster term is used to compute a robust variance for the model, by applying the so-called Huber sandwich estimator.[27] A cluster term accounted for patients with multiple TCs.
 Regarding Fine-Gray approach, we have added the relevant reference (https://www.jstor.org/stable/2670170) and one explanatory sentence in the methods, as follows:
In secondary analyses, we used the Fine-Gray approach to assess specific reasons for drug discontinuation (i.e. ineffectiveness, or adverse event) in a competing-risk setting. The Fine-Gray method takes competing risks into account when estimating the cumulative incidence function, modelling the sub-distribution hazard without treating competing events as censoring events.[28]

3) In the conclusions, the authors claimed that, in this non-randomized prospective cohort study, the treatment with BARI has at least similar effectiveness outcomes as alternative bDMARDs. This expression is not accurate, since non-significant effectiveness outcome is not equivalent to similar effectiveness outcome. A formal equivalence test is needed to conclude similar or equivalent effective outcome between treatment groups. The authors may need to update the final conclusion to be more accurate and consistent with the final statistical analysis results.

 We update the conclusions as follows:
In this non-randomized prospective cohort study, we demonstrate that treatment with drug maintenance of BARI was significantly higher than TNFi. However, we found no difference in drug maintenance when comparing BARI has at least similar effectiveness outcomes as alternative with other bDMARDs. Based on available data, the estimated 12-month response rates did not significantly differ between BARI, TNFi and OMA groups. We found no difference in treatment discontinuation for adverse event between the three groups. Overall, our results are in line with findings from randomized trials., confirm the effectiveness of BARI in daily practice and validate this agent as an alternative to bDMARDs in RA.
And also in the abstract:
Conclusions: BARI demonstrated a significantly higher drug maintenance compared to TNFi, mainly due to lower drug discontinuations for ineffectiveness., but similar We found no maintenance difference in drug-maintenance between BARI andto OMA. Clinical outcomes did not differ between the three groups. Our results suggest that BARI is an appropriated therapeutic alternative to bDMARDs in the management of RA.

VERSION 2 – REVIEW

REVIEWER	Wang, Wei Food and Drug Administration, Division of Biostatistics, Center for Devices and Radiological Health
REVIEW RETURNED	26-Dec-2023

GENERAL COMMENTS	The revised manuscript addressed all my previous comments and I have no further comments about it.