PEER REVIEW HISTORY

This paper was submitted to the ARD 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)	Long-term sustained remission in an observational study of patients with early RA –choice of remission criteria
AUTHORS	Svensson, Bjorn; Andersson, Maria; Bala, Valentina; Forslind, Kristina; Hafström, Ingiäld

VERSION 1 - REVIEW

REVIEWER	Dr. Rachel Knevel Leiden University Medical Center Department of Rheumatology Albinusdreef 2

GENERAL COMMENTS	Svensson et al. investigate which remission criteria may be best fit for use in long-term observational studies. The authors aim to study two different measures of remission in their performance of identifying patients with real remission, namely persisted remission. In their introduction the authors argue that this is important in order to study the arrest of joint destruction in observational studies. I presume the author will have good reason to study the selected remission criteria, but I cannot find this fully clarified in their manuscript. In general, I am not very convinced by the argumentation that the authors use. I found the manuscript hard to read and the reasoning not always sound. More importantly, I am not convinced that the current data support the conclusion that the ACR/EU Cr are too stringent and are therefore less appropriate for observational studies. I cannot find this supported by the data, and I do not find the explanation of the authors very clear. *There is circularity of using the same measure to define remission at each visit as to test this measure in its performance. There could be a good reason for this, but again I cannot find this in the manuscript. *Table V: The authors aim to test the ability of the different remission criterion to identify patients' "good radiographic outcome"(defined by <8 SHS progression over 8 years). Likelihood ratios are calculated

	and p-values with regard to these ratios are given. Still, the information that the LR of the DAS remission criterion is significant does not prove that this criterion performs better than the ACR/EUR criterion. To prove that, the likelihood ratios of both the ACR/EU and DAS criterion should be tested in one statistical analysis, resulting in one p-value. The high (non-significant) p-value of the ACR/EU criterion is the direct result of the low number of patients with sustained remission in this group, which has decreased the power of the likelihood analysis. It does not give any information about better or worse association of the remission criteria with good radiographic outcome. This is a classic error and could lead to incorrect interpretation of the data. Besides I would be more interested in the variance of good responders explained by different remission criteria and in the sensitivity and specificity. Looking at the data, I presume that all measures perform very poorly. * My conclusion would be that the performances of both remission criteria are bad in identifying patients with sustained remission and good radiographic response. This is very likely the result of the low true (sustained) remission in the total group of RA patients. I do not see data that suggest one of these measures is substantially better or worse than the other. Minor: *as in all observational studies, patients get lost to follow-up. The authors are very clear on the number of these patients. Could the authors discuss the implication of this on their analyses? Very likely patients in remission are more often lost to follow-up. * I like the adjustment of excluding the VAS scores from the remission criteria. *None of the tables are explained in the text. Or at least no references to the tables are present in the text. table 1 : very nice that the authors have data on the treatment. It should be better explained at which time the treatment was given. I presume baseline/time of inclusion? Could the authors state how many patients they have this information for? Now only percentages are given. * The language could be improved eg.: the phrase at page 4 line 7: "and should per protocol have a disease duration of 12 months or less" suggest that not all patients met this criteria. I am aware that with the best help still some patients get in a study without fully meeting the criteria. I am curious how many were in the current study and why it was chosen not to exclude them? *page 12 line 15 "Thus, "is not the correct word to connect the two sentences. *"criticism does not seem justified as regards patients" please correct the language *"this sentence could be improved: To what extent the criteria may

	predict little or no radiographic joint damage may be assessed by calculating ratios for the likelihood of “good radiographic outcome” of being in sustained remission." * why was good radiographic outcome defined as total SHS of less than 8 points? Over 8 years, 1 unit progression per year is not the lowest value of minor radiographic change. * Why did the authors exclude the 6 months measurement from the calculation in table III. *page 11 line 57 "joints may be swollen from reasons other than RA" I objects against this reasoning. Yes, if a patient also has gout or a recent trauma the swollen joint could be consequence of something different than RA. But in general a swollen joint in a RA patient is a clear sign of disease activity and a well trained clinician is able to distinguish different causes of arthritis.

REVIEWER	Prof. Theodore Pincus NYU Hospital for Joint Diseases

GENERAL COMMENTS	Comments to the Author Sustained remission and joint damage in an early rheumatoid arthritis cohort – choice of remission criteria. The objective of this paper was to investigate which remission criteria to use in observational and the association with radiographic damage. Comments: - This is not the first study describing the use and the impact of using different definitions of remission in patients with rheumatoid arthritis, except that one objective in this study is to look at radiographic damage which should therefore be the main objective of this study. - 839 patients were included in this study at baseline and data on 640 patients were available at 8 years follow-up. It should be stated what the reason for lost to follow-up was, because it might be that those patients with low disease activity did not want to participate anymore or that patients with severe disease activity were not able to participate anymore, influencing the results. - Sustained remission was defined as in remission at all follow-up visits. This is an interesting group of patients who are in remission at 6 months and stay in remission for 8 years. What was their baseline disease activity and did they receive any treatment? Should this group of patients not be defined as natural

	remission? - A simple T-Test or Mann-Whitney test was used to test the differences between the groups. However, three remission groups were defined. - Separate analysis were performed for men and women and it seems that gender is a onfounder. Overall, there may be other confounders which should have been considered and adjusted for in the analysis. - Table 1. Please include baseline data for the those patients in remission for patients in DAS28 remission and ACR/EU remission as well. - What treatment did patients use during follow-up? - Mean change in radiographic progression is shown in the figures. Was radiographic progression normally distributed? - Table 4 could be deleted. - One of the analysis was to assess the influence of the patient's global assessment. Why was this done as this has already been done in a number of other studies?

	- The manuscripts contains a lot of separate sentences instead of proper paragraphs.

- The manuscript received a third review at the Annals of Rheumatic Disease but the reviewer did not give permission for their comments to be published

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Comments to the Author
Svensson et al. investigate which remission criteria may be best fit for use in long-term observational studies.
The authors aim to study two different measures of remission in their performance of identifiying patients with real remission, namely persisted remission. In their introduction the authors argue that this is important in order to study the arrest of joint destruction in observational studies. I presume the author will have good reason to study the selected remission criteria, but I cannot find this fully clarified in their manuscript. In general, I am not very convinced by the argumentation that the authors use. I found the manuscript hard to read and the reasoning not always sound.
More importantly, I am not convinced that the current data supports the conclusion that the ACR/EU Cr are too stringent and are therefore less appropriate for observational studies. I cannot find this supported by the data, and I do not find the explaination of the authors very clear.

*There is circularity of using the same measure to define remission at each visit as to test this measure in its performance. There could be a good reason for this, but again I cannot find this in the manuscript.

We do not quite understand this point. We cannot see the circularity in applying different remission criteria to patients with long-standing RA and then analyze their performance including ratios for the likelihood of favorable radiographic outcome of being in sustained remission.

*Table V: The authors aim to test the ability of the different remission criterion in identifying patients "good radiographic outcome"(defined by <8 SHS progression over 8 years). Likelihood ratio's are calculated and p-values with regard to these ratios are given. Still, the information that the LR of the DAS remission criterion is significant does not proof that this criterion performs better than the ACR/EUR criterion. To proof that, the likelihood ratio's of both the ACR/EU and DAS criterion should be tested in one statistical analysis, resulting in one p-value. The high (non-significant) p-value of the ACR/EU criterion is the direct result of low number of patients with sustained remission in this group, which has decreased the power of the likelihood analysis. It does not give any information about the better or worse association of the remission criteria with good radiographic outcome. This is a classical error and could lead to incorrect interpretation of the data. Besides I would be more interested in the variance of good responders explained by de different remission criteria and in the sensitivity and specificity. Looking at the data, I presume that all measures perform very poor

We agree on the comment that low numbers may decrease the power of the likelihood analyses. The text has been modified accordingly in the revised manuscript (page 16).

The sensitivity and specificity of the criteria for being associated with radiographic progression have been calculated and added to the results section

* my conclusion would be that the performance of both remission criteria are bad in identifying patients with sustained remission and good radiographic response. This is very likely the results of the low true (sustained) remission in the total group of RA patients. I do not see data that suggests on of these measures is substantially better or worse than the other.

The groups in sustained remission by the DAS28-based criteria were considerably larger than those defined by the other criteria in spite of similar degree of joint damage, both increase in SHS and rate of radiographic progression. This fact is conceivably a clear advantage in favor of the DAS28-based criteria.

Minor:
*as in all observational study, patients get lost of follow-up. The authors are very clear on the amount of these patients. Could the authors discuss about the implication of this on their analyses? Very likely patients in remission are more often lost of follow-up.

Seventy per cent of the patents lost to follow-up had died and 40 were lost by unknown reasons.

We have reported some baseline differences between patients lost and completers but it appears unlikely that these differences could influence the main conclusions of the study.

* I like the adjustment of excluding the VAS scores from the remission criteria.

So do we but the main disadvantage is that these criteria lack the patient´s perspective.

*None of the tables are explained in the text. Or at least no reference to the tables are present in the text.

All tables are referred to in the text. We have tried to balance tables and text by giving only the main message in the text and thus avoid repeating all details.

table 1 : very nice that the authors have data on the treatment. It should be better explained to which time the treatment was given. I presume baseline,/time of inclusion at study? Could the authors state on how many patients they have this information? Now only percentages are given.

In most patients treatment was started at baseline. At baseline, one and 2 years, information on GC and DMARD treatment was available in all or almost all patients while no such info was available at 5

years in 11-13% and at 8 years 27 and 28% , respectively. This information has been added to the text in the results section.

* The language could be improved eg.: the phrase at page 4 line 7: "and should per protocol have a disease duration of 12 months or less" suggest that not all patients met this criteria. I am aware that with the best help still some patients get in a study without fully meeting the criteria. I am curious how many that was in the current study and why it was chosen not to exclude them?

One patient had a disease duration of 21 months, one 19 months, one 16, one15, two 14 and two 13 months. A delay of 2 months would not conceivably influence the results and the remaining patients are only 4. All patients were included consecutively and we decided that aspect more important than these slight protocol violences.

*page 12 line 15 "Thus, "is not the correct word to connect the two sentences.

In the previous manuscript submitted to ARD for review, there is no *thus* at the site indicated. However, there is a *thus* in page 13 line 5 but it is preceded by an *and* which connects the two sentences. Otherwise *thus* always initiates a sentence.

*"criticism does not seem justified as regards patients" please correct the language

This is the full sentence:

"So, the criticism does not seem justified as regards sustained remission in this patient material"

We believe it is clear but maybe it would be more clear like this:

So, it seems reasonable to use DAS28 Cr for the definition of sustained remission in this patient material

The sentence has been changed in the text.

*"this sentence could be improved: To what extent the criteria may predict little or no radiographic joint damage may be assessed by calculating ratios for the likelihood of "good radiographic outcome" of being in sustained remission."

We cannot identify this sentence.

* why was good radiographic outcome defined as total SHS of less than 8 points? Over 8 years, 1 unit progression per year is not the lowest value of minor radiographic change.

We have used the definition of minor radiographic change as suggested by van der Heijde(ref 12) and applied by Lillegraven et al (REF13). This definition states that an annual change of one unit is compatible with minor radiographic change.

We also tried other limits of radiographic progression like <5 SHS units (clinically relevant RP according to a definition by Omeract (2002)) and < 6 SHS units (smallest detectable change) but these limits did not show any advantage.

* Why did the authors exclude the 6 months measurement from the calculation in table III.

The reason for that is that we wanted to study long-term sustained remission. We considered it less probable to expect long-lasting remission already at 6 months after start of treatment.

*page 11 line 57 "joints may be swollen from reasons other than RA" I objects against this reasoning. Yes, if a patient also has gout or a recent trauma the swollen joint could be consequence of something different than RA. But in general a swollen joint in a RA patient is a clear sign of disease activity and a well trained clinician is able to distinguish different causes of arthritis.

This paragraph has been rewritten but still we and others (e.g. Thiele et al, ref16) believe that , in long-standing RA, tender and swollen joints may reflect some other co-existing rheumatic disorder or represent sequels of RA.

Reviewer: 2
Comments to the Author
Sustained remission and joint damage in an early rheumatoid arthritis cohort – choice of remission criteria. The objective of this paper was to investigate which remission criteria to use in observational and the association with radiographic damage.
Comments:
- This is not the first study describing the use and the impact of using different definitions of remission in patients with rheumatoid arthritis, except that one objective in this study is to look at radiographic damage which should therefore be the main objective of this study.

- 839 patients were included in this study at baseline and data on 640 patients were available at 8 years follow-up. It should be stated what the reason for lost to follow-up was, because it might be that those patients with low disease activity did not want to participate anymore or that patients with severe disease activity were not able to participate anymore, influencing the results.

- Sustained remission was defined as in remission at all follow-up visits. This is an interesting group of patients who are in remission at 6 months and stay in remission for 8 years. What was their baseline disease activity and did they receive any treatment? Should this group of patients not be defined as natural remission?

- A simple T-Test or Mann-Whitney test was used to test the differences between the groups. However, three remission groups were defined.

- Separate analysis were performed for men and women and it seems that gender is a onfounder. Overall, there may be other confounders which should have been considered and adjusted for in the analysis.

- Table 1. Please include baseline data for the those patients in remission for patients in DAS28 remission and ACR/EU remission as well.

- What treatment did patients use during follow-up?

- Mean change in radiographic progression is shown in the figures. Was radiographic progression normally distributed?

- Table 4 could be deleted.

- One of the analysis was to assess the influence of the patient's global assessment. Why was this done as this has already been done in a number of other studies?

- The manuscripts contains a lot of separate sentences instead of proper paragraphs.