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.

(This paper received three reviews from its previous journal but only two reviewers agreed to published their review.)

ARTICLE DETAILS

TITLE (PROVISIONAL)	Development and validation of a clinical rule for recognition of early inflammatory arthritis
AUTHORS	ten Brinck, Robin M; van Dijk, Bastiaan T; van Steenbergen, Hanna W; le Cessie, S; Numans, Mattijs E; Hider, Samantha L; Mallen, Christian; van der Helm-van Mil, Annette

VERSION 1 – REVIEW

REVIEWER	Carlo Alberto Scirè Rheumatology Unit, Department of Medical Sciences, University of Ferrara, Ferrara, Italy Epidemiology Research Unit, Italian Society for Rheumatology, Milan, Italy
REVIEW RETURNED	04-May-2018

GENERAL COMMENTS	This is a relevant, well-designed and well-reported study developing and validating a clinical prediction rule for the accurate identification of specialist confirmed IA in GP suspected IA. The included variables seem to try to exclude widespread pain, because higher joint involvement and female gender carry lower risks in the model. I would suggest to further discuss generalisability. The inclusion criteria are poorly reproducible, because suspicion is very subjective. Furthermore, GP in the study setting are probably more skilled than the European average, because they might be strongly influenced by the EAC operating for 25 years. This point is not necessarily a weakness if a fully external validation in different setting/country will confirm the results.

REVIEWER	Alessandra Bortoluzzi University of Ferrara, Department of Medical Sciences, Italy
REVIEW RETURNED	17-Jun-2018

GENERAL COMMENTS	This is a very interesting development and validation study of a clinical model to assist general practitioners (GPs) in recognizing and referring patients suspected for inflammatory arthritis (IA). The clinical relevance of the study is high, in fact, delays and under-recognition of IAs are frequent in clinical practice, and the authors discuss this point very well. The Authors enrolled 1,288 consecutive patients in an intermediate setting between primary and secondary care, the so-called Early Arthritis Recognition Clinic (EARC). The

	Authors developed the rule in 644 patients and validated it in other 644 ones. A simple rule with seven items has been developed to assist GPs in identifying IAs. These are my comments: The authors declare that the main limitation of the study is that the setting of recognition of patients suffering from inflammatory arthritis is an “intermediary” setting and not a primary care setting. In fact, patients are referred when GPs are “unsure” of the presence of IA and data are not collected among all the patients referring to their GP for joint pain: this could result in a bias of selection of the sample, restricting the development of the rule only for patients referred to a specialist. I agree with the need to develop the rule in a primary care setting. To underline that the patients were referred in presence of a reasonable doubt for IA can enhance the significance of this clinical model that it's more properly a tool to help GPs in doubtful cases. This is clearly expressed in “Methods” and “Discussion” sections, but could the Authors mention this aspect also in the abstract? Table 1: I suppose there were no statistically significant differences in baseline characteristics between “derivation” and “validation” groups. Could the authors report the p values for that?

REVIEWER	Titilola Falasinnu Stanford School of Medicine, USA
REVIEW RETURNED	04-Jul-2018

GENERAL COMMENTS	Thank you for the opportunity to review this work. It is well written and has the potential to contribute to our understanding of risk stratification in IA. The authors of the paper used split sample validation techniques. I would recommend that the authors also split up their sample by time (e.g., develop the model using newer data validate in older data) and since there seems to be variation in the completeness of variables over time. Lines 13-46 on page 10 do not belong in the methods section - it might be more useful to put a little bit of these contextual statements in the introduction and then flesh it out in the discussion section

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

This is a relevant, well-designed and well-reported study developing and validating a clinical prediction rule for the accurate identification of specialist confirmed IA in GP suspected IA.

Answer: Thank you for taking the time to provide our manuscript with valuable suggestions and comments.

The included variables seem to try to exclude widespread pain, because higher joint involvement and female gender carry lower risks in the model.

Answer: The reviewer is completely correct. The items of the prediction rule are also discussed in the Discussion section of the manuscript on page 14.

I would suggest to further discuss generalisability.

The inclusion criteria are poorly reproducible, because suspicion is very subjective.

Furthermore, GP in the study setting are probably more skilled than the European average, because they might be strongly influenced by the EAC operating for 25 years.

This point is not necessarily a weakness if a fully external validation in different setting/country will confirm the results.

Answer: We thank the reviewer for these comments. We have expanded the discussion on generalisability of the data on pages 15, 16 and 17.

We fully agree with the reviewer that GP's doubt on the presence of IA is very subjective and that, as a consequence, inclusion was not defined by 'ticking boxes of criteria'. Theoretically a difference in experience by GPs between regions might influence the results in different settings. This reasoning is now included in the discussion section. For the interest of the reviewer, we asked the GPs in our region if they considered themselves more skilled than the European average; they indicated to be more aware on the relevance of early identification of IA but they feel not more skilled in the actual identification. Nonetheless, we completely agree that external validation in primary care is needed. To this end the discussion was expanded with a section discussing the influence of different pre-test probabilities (possibly in other regions or countries) on the performance of the model and generalizability to other settings.

Reviewer: 2

This is a very interesting development and validation study of a clinical model to assist general practitioners (GPs) in recognizing and referring patients suspected for inflammatory arthritis (IA). The clinical relevance of the study is high, in fact, delays and under-recognition of IAs are frequent in clinical practice, and the authors discuss this point very well. The Authors enrolled 1,288 consecutive patients in an intermediate setting between primary and secondary care, the so-called Early Arthritis Recognition Clinic (EARC). The Authors developed the rule in 644 patients and validated it in other 644 ones. A simple rule with seven items has been developed to assist GPs in identifying IAs.

Answer: Thank you for these kind words and the clear summary of our manuscript.

These are my comments:

The authors declare that the main limitation of the study is that the setting of recognition of patients suffering from inflammatory arthritis is an "intermediary" setting and not a primary care setting. In fact, patients are referred when GPs are "unsure" of the presence of IA and data are not collected among all the patients referring to their GP for joint pain: this could result in a bias of selection of the sample, restricting the development of the rule only for patients referred to a specialist. I agree with the need to develop the rule in a primary care setting.

Answer: Thank you for this observation. In accordance with the comment made by reviewer 1, we fully agree that external validation in GP settings is required. We have tried to make an extra emphasis on this point in Discussion on pages 15, 16 and 17.

To underline that the patients were referred in presence of a reasonable doubt for IA can enhance the significance of this clinical model that it's more properly a tool to help GPs in doubtful cases. This is clearly expressed in "Methods" and "Discussion" sections, but could the Authors mention this aspect also in the abstract?

Answers: Thank you for this valuable suggestion. We agree with the reviewer that adding this may clarify the target population of our tool. We have added this to the Objectives section of the abstract on page 2.

Table 1: I suppose there were no statistically significant differences in baseline characteristics between "derivation" and "validation" groups. Could the authors report the p values for that?

Answer: We thank the reviewer for this recommendation. There were no statistically significant differences in baseline characteristics. The p-values comparing the different characteristics between the derivation and validation dataset are now provided in Table 1 on page 25.

Reviewer: 3

Thank you for the opportunity to review this work. It is well written and has the potential to contribute to our understanding of risk stratification in IA.

Answer: Thank you kindly for taking on this opportunity and reviewing our manuscript.

The authors of the paper used split sample validation techniques. I would recommend that the authors also split up their sample by time (e.g., develop the model using newer data validate in older data) and since there seems to be variation in the completeness of variables over time.

Answer: The reviewer is correct that there is variation in the completeness of variables over time. More specifically, the questions on 'difficulty with making a fist' and 'self-reported joint swelling' were added to the questionnaire after April 1st 2012 and were missing in patients included before this time point. For this end, we have chosen to impute missing values using chained equations as stated on page 8 and 9. In order to have an equal distribution of imputed data between the derivation and validation groups, data were split by odd/even numbers.

In response to this comment of the reviewer we have now done additional analyses and split the data in two groups of equal size based on chronological time (derivation group inclusion September 1st 2010 – March 12th 2013 and validation group included March 12th 2013 – September 24th 2015). When again analysing the derivation dataset, similar odds ratios were obtained for the items 'difficulty with making a fist' as in previous analyses: 1.4 (95%CI 0.96–2.0) and for 'self-reported joint swelling': 4.6 (95%CI 2.7–7.9) for 1-3 joints; 3.1 (95%CI 1.7–5.7) for 4-10 joints; and 3.1 (95%CI 1.6–6.1) for ≥11 joints. Also, the results obtained in the validation dataset were similar. Hence the models seems to be robust despite the fact that these two variable were missing in part of the patients. Also when evaluating the performance of the multivariable model by assessing the AUC, results were similar when a different split was made. In the newer data, we obtained an AUC of 0.77 (95%CI 0.72–0.81) and 0.72 (95%CI 0.69–0.76) for the derivation and validation data set respectively. These were similar to the original values in split sample validation.

Together, we feel confident that our data, when derived in new and validated in old data, would yield similar results as the analyses included in the manuscript. Because results were similar, we have not included these additional analyses to the revised version of the manuscript; also because our manuscript already contains a lot of supplementary tables. However, if the reviewer or editor strongly feel that these additional analyses need to be presented in additional supplementary material, we are of course willing to add the results.

Lines 13-46 on page 10 do not belong in the methods section - it might be more useful to put a little bit of these contextual statements in the introduction and then flesh it out in the discussion section

Answer: In response to this comment we have revised manuscript and tried to make our explanation of the simulation in the Methods section more concise (page 10), and expand the Discussion section about the different prevalence of IA and possible influence on the performance of our simplified rule depending on the setting on page 16 and 17.

VERSION 2 – REVIEW

REVIEWER	Carlo Alberto Scirè Department of Clinical Science, University of Ferrara, Italy
REVIEW RETURNED	24-Aug-2018

GENERAL COMMENTS	I have no further comments. Many thanks for considering all my previous comments.

REVIEWER	Alessandra Bortoluzzi Rheumatology Unit, Department of Medical Sciences, University of Ferrara and Azienda Ospedaliero-Universitaria S.Anna, Cona (Ferrara), Italy
REVIEW RETURNED	24-Aug-2018

GENERAL COMMENTS	The Authors properly addressed all the comments. No further changes are required.

REVIEWER	Titilola Falasinnu Stanford University School of Medicine, Stanford, California, USA
REVIEW RETURNED	16-Aug-2018

GENERAL COMMENTS	The authors have adequately addressed my concerns.