PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Improving Accuracy of Self-Reported Diagnoses of Rheumatoid Arthritis in the French Prospective E3N-EPIC cohort: A Validation Study
AUTHORS	Nguyen, Yann; Salliot, C; Gusto, Gaëlle; Descamps, Elise; Mariette, Xavier; Boutron-Ruault, Marie-Christine; Seror, Raphaèle

VERSION 1 – REVIEW

REVIEWER	Carlo Alberto Scirè University of Ferrara, Italy
REVIEW RETURNED	02-Sep-2019

GENERAL COMMENTS	The paper "Improving Accuracy of Self-Reported Diagnoses of Rheumatoid Arthritis in the French Prospective E3N-EPIC cohort" by Nguyen et al. is an interesting paper on the development and internal validation of an algorithm to accurately identify patients with Rheumatoid Arthritis in an EPIC cohort. This is not the first attempt to link the EPIC cohort with rheumatological data. Robust data form the Manchester Epidemiology Unit should be considered (for example: Lahiri M, Luben RN, Morgan C, Bunn DK, Marshall T, Lunt M, Verstappen SM, Symmons DP, Khaw KT, Wareham N, Bruce IN. Using lifestyle factors to identify individuals at higher risk of inflammatory polyarthritis (results from the European Prospective Investigation of Cancer-Norfolk and the Norfolk Arthritis Register--the EPIC-2-NOAR Study). Ann Rheum Dis. 2014 Jan;73(1):219-26. doi:10.1136/annrheumdis-2012-202481. Epub 2013 Mar 16. PubMed PMID: 23505230; PubMed Central PMCID: PMC3888611). Many studies analyzing administrative data to identify patients with RA are available and such evidence should be quoted. I would suggest to consider the following points: - The selection of patients with suspected RA in which medical data were available could introduce a selection bias toward more severe disease, inflating the accuracy. - The gold standard should be referred as 'reference standard' - Given that only subjects with self-reported diagnosis were included (positive to the test) it is unclear how sensitivity and specificity could be extrapolated to the overall population (see methods); in the results it seems that sensitivity and specificity are calculated for the IRD questionnaire and other algorithms not considering that patients with negative self-reported questionnaire were not assessed by the medical record review. This leads to a kind of verification bias, probably inflating sensitivity. The potential impact of this design should be discussed. Most of these points are already discussed as limitations. The

	authors are encouraged to defend with more arguments the strength of their methodology. While reviewing the manuscript, I would suggest to carefully check this reporting checklist: Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. J Clin Epidemiol. 2011 Aug;64(8):821-9. doi: 10.1016/j.jclinepi.2010.10.006. Epub 2010 Dec 30. Review. PubMed PMID: 21194889.

REVIEWER	Tiffany Gill The University of Adelaide, Australia
REVIEW RETURNED	09-Sep-2019

GENERAL COMMENTS	Thank you for the opportunity to review this paper. My comments are as follows: Abstract: Appropriate. Introduction: Page 4, line 34. "The diagnosis accuracy..." should be "The diagnostic accuracy..." Page 4, line 34. "vary" should be "varies" Otherwise clear and concise Method: Page 5, line 34 "besides" should be "in addition" Page 7, line 17,"..DMARD among methotrexate..." the sentence is not clear and needs to be reworked. Generally well described. Results: Well presented Discussion: Page 16, line 38 "problematic" should be "problem" Line 46 "Another limit.." should be "Another limitation..." Line 60, "option" should be "options" Otherwise clear and concise. Tables are well presented.

REVIEWER	Vibeke Videm NTNU - Norwegian University of Science and Technology, Trondheim Norway
REVIEW RETURNED	10-Sep-2019

GENERAL COMMENTS	Yann Nguyen and coworkers have studied the accuracy of self-reported rheumatoid arthritis (RA) in a large cohort of French women (n=98,995) and tested whether accuracy could be improved using additional information from a specific questionnaire about inflammatory rheumatic diseases or a medication reimbursement database. Diagnoses were verified by medical records (n=399). As expected, the authors found that self-report gave many false-positives. The positive and negative predictive values could be greatly improved by addition of information from the other sources, especially specific treatment information. The authors identified 964 RA cases in their cohort, which may be included in further RA studies. The present study confirms findings from previous publications. Of specific interest is that some of the weaknesses of the various

	sources of additional information are highlighted, e.g. in Table 4 and in the discussion. Some points for improvement: Inclusion and case identification including the subgroup numbers given in the abstract and Table 1are difficult to understand without the flow chart which is now placed in the supplement. Figure S2 should therefore be moved to the main document. To avoid too many tables/figures, the information in table 5 could be moved to the text or the supplement. As the authors mention, a main weakness with the study is that chart review was only possible for a small, non-random selection of participants with self-reported RA (~11%). The authors conclude that this probably didn't bias the results but give no good evidence. The fact that the gold standard diagnosis was only available for a small subset of patients begs the question whether all the 964 cases identified with the employed algorithms were actually correct. This lack of better validation should be more clearly mentioned. The authors state that the algorithms they have developed may be used in other cohorts. It is not clear from the manuscript exactly what these algorithms were. The relevant information is presented in table 3, but the manuscript does not contain a precise protocol of how to combine this information to reach a conclusion regarding RA/no RA for a specific participant. This may not be a problem because the algorithms have not been properly validated and therefore may not be adequate for other populations. But better presentation is necessary to make future external validation possible. The conclusions regarding use in other populations (page 17, line 7 onwards) are too strong and should be modified. The study supports that the best way to ascertain RA diagnoses still seems to be medical chart review, and that other methods will be less accurate, but more practical surrogates – and probably are reasonably good alternatives. The general conclusion that additional information characterizing joint problems and medication helps improve diagnostic accuracy of self-reported RA is the same as others have found previously. The need to send an additional questionnaire to selected participants and possible response bias shows that the employed scheme in the present study is not without potential problems. It is also important to acknowledge that there are probably differences among populations, differences in design among population-based studies, national differences regarding medication databases etc., that render general use of algorithms from one place difficult. These points should be better acknowledged.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Carlo Alberto Scirè
Institution and Country: University of Ferrara, Italy

The paper "Improving Accuracy of Self-Reported Diagnoses of Rheumatoid Arthritis in the French Prospective E3N-EPIC cohort" by Nguyen et al. is an interesting paper on the development and internal validation of an algorithm to accurately identify patients with Rheumatoid Arthritis in an EPIC cohort.

We thank the reviewer for his positive comments on our manuscript.

This is not the first attempt to link the EPIC cohort with rheumatological data. Robust data form the Manchester Epidemiology Unit should be considered (for example: Lahiri M, Luben RN, Morgan C, Bunn DK, Marshall T, Lunt M, Verstappen SM, Symmons DP, Khaw KT, Wareham N, Bruce IN. Using lifestyle factors to identify individuals at higher risk of inflammatory polyarthritis (results from the European Prospective Investigation of Cancer-Norfolk and the Norfolk Arthritis Register--the EPIC-2-NOAR Study). Ann Rheum Dis. 2014 Jan;73(1):219-26. doi:10.1136/annrheumdis-2012-202481. Epub 2013 Mar 16. PubMed PMID: 23505230; PubMed Central PMCID: PMC3888611).

We agree with the reviewer. The EPIC (European Prospective Investigation into Cancer and Nutrition) project is a large European consortium involving different cohorts across Europe. RA cases have been indeed identified in different EPIC cohorts from Denmark (Linauskas et al.), Sweden (Sundstrom et al.), Italy, Spain (Fisher et al.), and in the UK (Norfolk). The E3N cohort is the French component of the EPIC study. In each cohort, the methods to identify patients were different, depending on the available data and possibility of linkage.
Indeed, our colleagues from Norfolk robustly identified their RA cases, thanks to the linkage with the Norfolk Arthritis Register, in which all cases of RA and inflammatory polyarthritis were ascertained by examination by study research staff or medical record review rather than through self-reported questionnaires, which might provide the most robust evidence of RA ascertainment. Unfortunately, such linkage with a RA database is not available in our cohort.
Following your advice, we added a reference of this robust study (Introduction, page 4, paragraph 4):
« Other studies have ascertained self-reported RA through linkage with a medical records review, or even with clinical examination of all suspected cases [23–25]"

Many studies analyzing administrative data to identify patients with RA are available and such evidence should be quoted.

We agree and have now quoted other studies using administrative data including studies from other components of EPIC (Linauskas et al., Sundstrom et al., Fisher et al., Lahiri et al.), from the Nurse Health Study (Karlson et al.), or the Iowa Women's Health Study (Mikuls et al.). We also quoted many studies evaluating the accuracy of self-reported diagnosis including the Nord-Trondelag Health Study (Videm et al.), the Women's Health Initiative (Walitt et al.), or the Black Women's Health Study (Formica et al.).

I would suggest to consider the following points:
- The selection of patients with suspected RA in which medical data were available could introduce a selection bias toward more severe disease, inflating the accuracy.

Indeed, medical records were provided on a voluntary basis, thus not at random, and we agree that it could introduce a selection bias. We had already discussed this point in the limitations section, but have now added the following sentences to insist on this specific point.
We added in the discussion page 15 paragraph 4:
"This could have introduced a selection bias towards more severe disease, thereby inflating the accuracy. However, medical chart review confirmed the diagnosis of RA in only 41% of them, showing that both cases and non-cases provided medical charts. Also, women who provided medical charts did not differ from other women who self-reported IRD in terms of age or education level, which limits the risk of severe bias."

- The gold standard should be referred as 'reference standard'

We agree and changed the wording from gold standard to reference standard.

- Given that only subjects with self-reported diagnosis were included (positive to the test) it is unclear how sensitivity and specificity could be extrapolated to the overall population (see methods); in the results it seems that sensitivity and specificity are calculated for the IRD questionnaire and other algorithms not considering that patients with negative self-reported questionnaire were not assessed by the medical record review. This leads to a kind of verification bias, probably inflating sensitivity. The potential impact of this design should be discussed.

Women included in this validation study were both women who self-reported RA, but also women who self-reported spondyloarthritis, thus not only women who were positive to the test. We recognize that the choice of the control group is not neutral, and that our results could not be extrapolated to the overall population, but to women who self-reported inflammatory rheumatic diseases (RA or SpA). Of course, ideally, we would like to have analysed medical records from women who did not report any inflammatory rheumatic disease. This is a large cohort of mostly well-educated women who repeatedly proved very accurate when self-reporting diseases, including cancer, endometriosis, gallbladder disease etc. Follow-up is also very long, therefore the likelihood of them successively failing to report such an invalidating disease is low. Finally, in terms of investigating risk factors for RA, that the impact on statistical analyses of false negative cases, i.e. RA women who did not self-declare it, is low through dilution in the large number of non-cases in the cohort.

We have modified the following paragraph in the Discussion section page 15 paragraph 3: “Our population of non-cases were women who did not self-report RA but self-reported another IRD, which could bias our results. Ideally, we would have analysed medical records from women who did not report any IRD to determine the proportion of cases missed. Thus, reported sensitivities and NPVs should be interpreted with caution. However, our main concern was to avoid false positive cases i.e. to ascertain detected cases, rather than to avoid missing a few cases. Therefore, there may be a few undetected RA cases in the control group, but the number of these cases is likely to be small, and, given the large number of non-cases in our cohort, the risk of bias induced by the false negative cases is negligible.”

Most of these points are already discussed as limitations. The authors are encouraged to defend with more arguments the strength of their methodology.

We thank the reviewer for pointing out that most of these points have already been discussed as limitations. Nevertheless, as suggested, we added in our discussion different points to defend the strength or our study, by responding to reviewer comments.
In addition to responses to the previous comments, we added the proportion of women identified by 2 or more methods to the results section (page 13, paragraph 1) and in a comprehensive table (S4 Appendix). These data reinforce the strength of our validation process:
“In addition, 65.1% of our identified cases have been identified by at least two methods, and 16.4% and 21% have even been validated by three or four methods, respectively (S4 Appendix).”.

While reviewing the manuscript, I would suggest to carefully check this reporting checklist: Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. J Clin Epidemiol. 2011 Aug;64(8):821-9. doi: 10.1016/j.jclinepi.2010.10.006. Epub 2010 Dec 30. Review. PubMed PMID: 21194889.
We thank the reviewer for this checklist. Most items of the checklist have already been reported in our

manuscript. For some items, the information was not available (ie disease severity).

Reviewer: 2
Reviewer Name: Tiffany Gill
Institution and Country: The University of Adelaide, Australia

Thank you for the opportunity to review this paper. My comments are as follows:
Abstract: Appropriate.
Introduction: Page 4, line 34. "The diagnosis accuracy..." should be "The diagnostic accuracy..."
Page 4, line 34. "vary" should be "varies"
Otherwise clear and concise
Method: Page 5, line 34 "besides" should be "in addition"
Page 7, line 17,"..DMARD among methotrexate..." the sentence is not clear and needs to be reworked.
Generally well described.
Results: Well presented
Discussion: Page 16, line 38 "problematic" should be "problem"
Line 46 "Another limit.." should be "Another limitation..."
Line 60, "option" should be "options"
Otherwise clear and concise.
Tables are well presented.

We thank the reviewer for her kind comments. We corrected all mistakes and typo errors.
We also changed the wording, as requested, to make the sentence understandable (Page 7, paragraph 2): “women were considered as RA cases if they self-reported having RA, and had had reimbursements for any conventional synthetic or biologic DMARD used in the treatment of RA, including methotrexate, leflunomide, any sub-cutaneous tumor necrosis factor alpha (TNF-α) inhibitor, and sub-cutaneous abatacept or tocilizumab.”

Reviewer: 3
Reviewer Name: Vibeke Videm
Institution and Country: NTNU - Norwegian University of Science and Technology, Trondheim, Norway

Yann Nguyen and coworkers have studied the accuracy of self-reported rheumatoid arthritis (RA) in a large cohort of French women (n=98,995) and tested whether accuracy could be improved using additional information from a specific questionnaire about inflammatory rheumatic diseases or a medication reimbursement database. Diagnoses were verified by medical records (n=399). As expected, the authors found that self-report gave many false-positives. The positive and negative predictive values could be greatly improved by addition of information from the other sources, especially specific treatment information. The authors identified 964 RA cases in their cohort, which may be included in further RA studies.

The present study confirms findings from previous publications. Of specific interest is that some of the weaknesses of the various sources of additional information are highlighted, e.g. in Table 4 and in the discussion.

We thank the reviewer for her helpful review of our manuscript.

Some points for improvement:

Inclusion and case identification including the subgroup numbers given in the abstract and Table 1 are difficult to understand without the flow chart which is now placed in the supplement. Figure S2 should therefore be moved to the main document. To avoid too many tables/figures, the information in table 5 could be moved to the text or the supplement.

We agree with the reviewer that the flow-chart is needed in the main document. We moved it accordingly from supplementary files to the main document. We also moved Table 5 to the supplementary file (Table S2).

As the authors mention, a main weakness with the study is that chart review was only possible for a small, non-random selection of participants with self-reported RA (~11%). The authors conclude that this probably didn't bias the results but give no good evidence.

Indeed, medical records were provided on a voluntary basis, thus not at random, and we agree that it could introduce a selection bias. We moderated our claims, writing that this could introduce a selection bias inflating the accuracy, but given the comparability between women who provided their medical charts and women who did not, in terms of age or education level, the risk of severe bias is limited.
We added the following sentences in the discussion, page 15, paragraph 4:
"This could have introduced a selection bias toward more severe disease, inflating the accuracy. However, medical chart review confirmed the diagnosis of RA in only 41% of them, showing that both cases and non-cases provided medical chart. Also, women who provided their medical charts did not differ from other women who self-reported IRD in terms of age or education level, which may limit the bias."

The fact that the gold standard diagnosis was only available for a small subset of patients begs the question whether all the 964 cases identified with the employed algorithms were actually correct. This lack of better validation should be more clearly mentioned.

We agree with the reviewer that in the 964 identified RA cases there might be a few false positive cases, given that medical records were not available for every women. Nevertheless, we tried to limit their number by using different algorithms and the linkage to our medication database.
Also, the vast majority (65.1%) of our identified cases have been identified by at least 2 methods, and even 16.4% and 21% have been validated by 3 or 4 methods, respectively.
We added this information in Results, page 13, paragraph 1:
"In addition, 65.1% of our identified cases have been identified by at least 2 methods, and 16.4% and 21% have even been validated by 3 or 4 methods, respectively".
We also added a comprehensive table in the Supplementary File section (Table S3) describing the number of methods that identified each case.
We also changed the wording of the discussion to clearly mention the lack of better validation in the Discussion section, page 16, paragraph 2: "Since a proper evaluation with the reference standard (ie medical chart review) was not available for all women, there might be some false positive RA cases among them. But given the number of methods used to limit their number and their accuracy, this rate might be small. "

The authors state that the algorithms they have developed may be used in other cohorts. It is not clear from the manuscript exactly what these algorithms were. The relevant information is presented in table 3, but the manuscript does not contain a precise protocol of how to combine this information to reach a conclusion regarding RA/no RA for a specific participant. This may not be a problem because the algorithms have not been properly validated and therefore may not be adequate for other

populations. But better presentation is necessary to make future external validation possible. The conclusions regarding use in other populations (page 17, line 7 onwards) are too strong and should be modified.

We agree and provided more details on the used algorithms. The latter combined RA self-declaration, and items from the IRD questionnaire, or reimbursement of RA medications. We now specified which items from the IRD questionnaire were used for each algorithm in the Supplementary Material section (Table S1). Algorithms are also reported in table 3 and in the methods page 6, paragraph 3:
"We considered as RA cases women who confirmed having RA in the IRD specific questionnaire, and self-reported at least one of the following: 1) RA diagnosis confirmed by a rheumatologist and/or another physician (internal medicine specialist or general practitioner) 2) taking or having taken any of the RA conventional synthetic disease modifying anti-rheumatic drugs (DMARDs) or biologic DMARDs (listed in S1 Appendix, Question 34), 3) having positive RF or ACPA, or 4) at least 4 of the seven 1987-ACR criteria (listed in S1 Appendix, Questions 8,9,11,14–18)."

Also, we agree that our algorithms may not be adequate for other populations and other cohorts, since they may refer to information that have not been collected in all cohorts. Nevertheless, recorded items are quite simple and could be implemented in other cohorts, as we did here. We agree that our conclusion might be too strong, thus we moderate it.
Thus, we discussed the use of these algorithms in other cohorts as a part of the discussion, page 16, paragraph 1:
"Finally, the algorithms we devised to improve accuracy of self-reported RA diagnoses could prove useful to validate RA diagnoses in other population-based cohorts. However, they could be difficult to transpose from the French care setting to another one; thus, all data potentially available for validation (medication database, national patient registries, primary care records and/or hospital discharge databases) must be considered."

The study supports that the best way to ascertain RA diagnoses still seems to be medical chart review, and that other methods will be less accurate, but more practical surrogates – and probably are reasonably good alternatives.

We agree, it is exactly what our results suggest.
We modified our conclusion to insist on this point, (page 16):
"Even if ascertaining RA diagnoses with a complete medical chart review is probably the best option, it is often difficult and costly to be achievable in large cohorts. Therefore, it appears that obtaining other information, particularly on RA specific, treatment either from the patients themselves or from health insurance databases can be a reasonably good alternative, sparing the difficulties of obtaining complete medical charts, and the time and cost of medical chart review."

The general conclusion that additional information characterizing joint problems and medication helps improve diagnostic accuracy of self-reported RA is the same as others have found previously. The need to send an additional questionnaire to selected participants and possible response bias shows that the employed scheme in the present study is not without potential problems.

We agree and are aware that collecting additional information improves diagnostic accuracy of self-reported RA has previously been reported, as already referenced in our paper. Also, we discussed the potential response bias which is inherent to the sending of additional questionnaires.

We modified our discussion to comment on this point (page 15, paragraph 3):
"Also, our validation study relies on an additional questionnaire. Answers to this questionnaire were not obtained for all women, which might have created a response bias. However, such bias was limited by using the medication reimbursement database for women who did not answer to the IRD

questionnaire."

We modified our conclusion to specify this point, page 17:
"it appears that obtaining other information on particular on treatment either from patients themselves or from health insurance database can be a reasonably good alternative, sparing the difficulties of obtaining complete medical charts, and the time and cost of medical chart review."

It is also important to acknowledge that there are probably differences among populations, differences in design among population-based studies, national differences regarding medication databases etc., that render general use of algorithms from one place difficult. These points should be better acknowledged.

We agree that algorithms might not be easily transposed from one care setting to another. That is why we now provided more detailed information on our algorithms (Supplementary Table S1)
Also, following your comment, we added a paragraph on the difference among populations and care settings in the Discussion section (page 16, paragraph 2): "However, they could be difficult to transpose from the French care setting to another one; thus, all data potentially available for validation (medication database, national patient registries, primary care records and/or hospital discharge databases) must be considered."

VERSION 2 – REVIEW

REVIEWER	Carlo Alberto Scirè University of Ferrara and Epidemiology Unit of the Italian Society for Rheumatology.
REVIEW RETURNED	04-Oct-2019

GENERAL COMMENTS	I would like to thank the authors for their responses. I have no further comments.

REVIEWER	Vibeke Videm NTNU - Norwegian University of Science and Technology, Norway
REVIEW RETURNED	12-Oct-2019

GENERAL COMMENTS	The authors have now adequately addressed my previous concerns.