Author’s response to reviews

Title: Administrative database as a source for assessment of systemic lupus erythematosus prevalence: Estonian experience

Authors:
Kati Otsa (kati.otsa@itk.ee)
Sandra Talli (sandra.talli@itk.ee)
Pille Harding (pille.harding@itk.ee)
Eevi Parsik (eevi.parsik@regionaalhaigla.ee)
Marge Esko (marge.esko@ltkh.ee)
Anti Teepere (anti.teepere@itk.ee)
Marika Tammaru (marika.tammaru@itk.ee)

Version: 1 Date: 16 Jan 2019

Author’s response to reviews:

Dear Editors,

Please find our responses to the Reviewer’s comments below.

Kind regards

Kati Otsa

Reviewer reports:

Vera Golder (Reviewer 1): This is a well written manuscript describing the accuracy of a database search strategy in epidemiological studies of SLE.

The main message from this paper is that diagnostic uncertainty or the false positive rate of incorrect ICD coding for SLE diminishes with increasing number of codes applied to the same patient (in most cases).
Major suggestions for consideration:

1. One take from the main message is that perhaps the initial code for SLE per patient should be disregarded when conducting database searches, the advantages and pitfalls of this approach should be briefly discussed.

Thank you for the comment. We cannot completely agree with the reviewer’s comprehension of the main message of the manuscript. Although, demonstration of association between the repetition of diagnosis code assignment and correctness of SLE diagnosis is one of our main results, our main aim was to convey the message of different nature of the “false positive” SLE diagnoses in administrative databases. However, a clarifying sentence “So far, the researches should keep in mind that usage of administrative data which include cases with few repetitions of M32 codes may lead to overestimation of SLE prevalence.” is added to the end of the Conclusion sections (Conclusions, page 10, line 36).

2. M32 is the umbrella ICD-10 code for all systemic lupus erythematosus, and indeed in many countries this is not sufficient or specific enough for billing purposes.

Did the authors look at any of the codes under this umbrella in subgroup analyses? e.g. M32.14 - glomerular disease due to systemic lupus erythematosus.

It would be interesting to note if the PPV values improve and false positive rates reduce with more specific coding.

Yes, the data on ICD sub-codes were available for us. However, there were cases when different codes were assigned to the same patient (eg M32.14 by a nephrologist and M32.12 by a cardiologist); some specific codes were rare. This complicated considerably the structure of our research data and made potentially comparable groups very small. Therefore, in order to achieve the reasonable width of confidence intervals, we decided not to execute the analysis by the subgroups.

3. A large limitation of this study is its generalisability to other health systems, and this should be clearly stated. The high rate of reporting of the ICD-10 codes, and therefore the low probability of missed cases, is driven by the billing set-up whereby the "fee for service" is not received without the code. Without this inbuilt financial incentive the rate of reporting would be much lower, especially in ambulatory care settings where administrative support is likely to be lacking.

Thank you for drawing our attention to this point. We agree that generalisability of the results of database research is affected by limitations of the application of case ascertainment algorithms to
administrative databases of different structure and functioning principles. We hope that the
detailed description of the used database and methods provides the readers with grounds to
decide on the applicability of our particular approach in their specific context of interest. Still,
we believe that our main message – different nature of “false positive” M32 codes in
administrative databases – could be valuable for scholars working with administrative data of
different origins. The first paragraph of Conclusions section was changed as “Administrative
database research also has its inherent limitations; the necessity for ascertainment validation of
diagnoses in administrative data using other data sources consumes time and resources and may
be complicated by the issues of data protection legislation. Generalisability of the results of
database research is affected by limitations of application of case ascertainment algorithms to
administrative databases of different structure and functioning principles. We hope that the
detailed description of the used database and methods provides the readers with grounds to
decide on applicability of our findings in their specific context of interest.” (Conclusions, page 9,
line 50-55)

Minor style editing:

When citing the work of others use either "Author et al." or "Author and colleagues", not both interchangeably

Thank you, “et al” replaced by “and colleagues” at five occasions (page 1 line 44, page 4 line 43, page 7 line 16, page 7 line 45, page 8 line 41 ).

Aisha Lateef (Reviewer 2): Interesting study which provides objective estimates of accuracy in
database driven research.

Thank you