PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with Ischaemic heart disease: a cohort study with nested case-control analysis
AUTHORS	Reeves, David; Springate, David; Ashcroft, Darren; Ryan, Ronan; Doran, Tim; Morris, Richard; Olier, Ivan; Kontopantelis, Evangelos

VERSION 1 - REVIEW

REVIEWER	Rosemary Tate University of Sussex UK
REVIEW RETURNED	17-Feb-2014

GENERAL COMMENTS	A well written and interesting paper. My only major comment is that it would be useful if the authors could put a bit more emphasis on the fact that not only are they using different data sources, but also that there are quite a few differences in the software. It would be helpful to outline a few of the differences between Vision and EMIS - for example, I believe EMIS is more text based, and there are a number of other differences, not necessarily that small? It is very encouraging that they appear to be compatible for research studies. Minor points p 4. 3rd para. "However, these studies did not replicate ..." I suggest this sentence is removed as it implies a limitation of the Qrisk tools, which is a bit misleading and also not really relevant here.

REVIEWER	Emily Herrett London School of Hygiene and Tropical Medicine, United Kingdon
REVIEW RETURNED	18-Feb-2014

GENERAL COMMENTS	All of the comments here are also addressed in my review. 1. While the objective of the study is clearly defined, the aim is not clearly stated in the main body of text. I think the aim was to compare the results of an analysis in CPRD with the same analysis in QResearch. 12. I am worried that the authors place too much emphasis on their results showing the validity of CPRD and QResearch data. They have a good conclusion which sums up the limitations quite nicely, but I think that the limitations about the interpretation of the data could be further discussed. This comparison cannot provide firm evidence of external validity because both CPRD and QResearch are based on the same kind of data.

15. While the manuscript is written with good English, I felt that it was verbose.

This paper by Reeves et al compares an analysis of the association between statins and ischaemic heart disease using data from one UK primary care database (CPRD) with a previously published analysis from a different UK primary care database (QResearch).

It appears that the authors were able to accurately replicate the analysis performed in QResearch data. The data were managed in a way that I would expect to see in a CPRD analysis. Given the previous paper by Kontopantelis et al in 2013 describing differences in QOF attainment by clinical system, the current paper is of some interest. If there were differences in the recording of key requirements within the clinical system, then there may be different findings at analysis. It is reassuring to find the same result in each database. I have the following comments for the authors:

1. The manuscript is too long. The introduction and methods should be trimmed down to a more manageable length. Some of the detail could be placed in supplementary material.

2. I think that the justification for the study is somewhat hidden in the introduction. I would suggest that the authors try to focus the introduction. Some parts felt less relevant than others and might fit better in the discussion. The authors touch upon the paper by Kontopantelis and colleagues as a reason for this analysis. I think that this is an important point because it justifies the study. The overall aim of the study could also be stated a little more clearly at the end of the Introduction. I think the aim was to compare the results of an analysis performed in CPRD to the same analysis performed in QResearch?

3. The authors mention the three major UK PCDs. The authors may wish to also refer to ResearchOne, a relatively new PCD with over 5 million patients.

4. The authors state that comparing different primary care datasets cannot prove validity. I agree – the coding errors of one UK primary care dataset are likely to be very similar to another dataset because of GP and patient behaviour. I think the strongest conclusion that we can make from a comparison of UK primary care datasets is that they are as good (or as bad) as each other. I thought that the final conclusion was nice – that taken with other studies, this analysis adds to the body of evidence suggesting validity. Therefore I find the title of this article and the focus on 'validity' a little misleading. I think that the limitations of the comparison need further emphasis in the paper.

5. There were some interesting demographic differences

	between the CPRD and QResearch cohorts, and the proportion of patients with CHF was more than double in QResearch compared to CPRD. There were also differences in the prevalence of hypertension. Could the authors share any possible explanations for these differences? Could this have been due to a difference in coding between the data sources? The prevalences were higher in QResearch, which tended to have an older population. Were the age-specific prevalences of these diagnoses the same between the datasets?

VERSION 1 – AUTHOR RESPONSE

Reviewer Name Rosemary Tate
Institution and Country University of Sussex UK
Please state any competing interests or state 'None declared': None declared

A well written and interesting paper. My only major comment is that it would be useful if the authors could put a bit more emphasis on the fact that not only are they using different data sources, but also that there are quite a few differences in the software. It would be helpful to outline a few of the differences between Vision and EMIS - for example, I believe EMIS is more text based, and there are a number of other differences, not necessarily that small? It is very encouraging that they appear to be compatible for research studies.

We have added text to the methods section outlining some of the main differences between the two systems, in particular in how the systems are navigated and data entered. We also reference a direct comparison between the two.

Minor points

p 4. 3rd para. "However, these studies did not replicate ..." I suggest this sentence is removed as it implies a limitation of the Qrisk tools, which is a bit misleading and also not really relevant here.

We did not intend to imply a limitation to the Qrisk tools, but rather to mean that the questions addressed by these studies were quite different, and that our question had not already been answered by these risk validation studies. The validations conducted of the Qrisk tools did not compare risk algorithms derived from the two databases, but rather investigated the validity of predictions made from applying the QResearch-based algorithms to the THIN database. We have revised the passage to clarify our intent.

We think it important to make the point as some of our colleagues were of the impression that these Qrisk studies had done the former and had therefore addressed the same research question as ourselves.

Reviewer Name Emily Herrett
Institution and Country London School of Hygiene and Tropical Medicine, United Kingdon
Please state any competing interests or state 'None declared': None declared

This paper by Reeves et al compares an analysis of the association between statins and ischaemic heart disease using data from one UK primary care database (CPRD) with a previously published analysis from a different UK primary care database (QResearch).

It appears that the authors were able to accurately replicate the analysis performed in QResearch data. The data were managed in a way that I would expect to see in a CPRD analysis. Given the previous paper by Kontopantelis et al in 2013 describing differences in QOF attainment by clinical system, the current paper is of some interest. If there were differences in the recording of key requirements within the clinical system, then there may be different findings at analysis. It is reassuring to find the same result in each database. I have the following comments for the authors:

1. The manuscript is too long. The introduction and methods should be trimmed down to a more manageable length. Some of the detail could be placed in supplementary material.

We acknowledge that the size of the paper is larger than average but the complexity of the project is an important contributing factor to that. We have trimmed some text from the introduction but we have not reduced the methods section since one of our main criticisms of PCD-based publications is the lack of detail that would allow a replication to be undertaken. Although we could place some of the methodological detail in a supplementary file – and will do so if the Editor requests – this approach we feel is akin to assigning the methods 'secondary' or minor status, which is rather in contradiction to our intentions with this paper.

2. I think that the justification for the study is somewhat hidden in the introduction. I would suggest that the authors try to focus the introduction. Some parts felt less relevant than others and might fit better in the discussion. The authors touch upon the paper by Kontopantelis and colleagues as a reason for this analysis. I think that this is an important point because it justifies the study.

We have added a line earlier in the introduction to inform readers about the general direction the paper will take (ie a comparison of databases). We think that the part referred to as more relevant to the discussion is the description of previous 'replication' studies, which could have gone in the discussion to put our results in the context of previous work. We thought very carefully about this, even before our first submission, but felt that these represented a large part of the justification for our own study – viz, that no truly independent full replications had been previously conducted – and therefore felt, as still feel, that it is important for this material to remain in the introduction.

The overall aim of the study could also be stated a little more clearly at the end of the Introduction. I think the aim was to compare the results of an analysis performed in CPRD to the same analysis performed in QResearch?

Correct, and a good suggestion. We have added a sentence at the end of the introduction to make the overall aim clearer.

3. The authors mention the three major UK PCDs. The authors may wish to also refer to ResearchOne, a relatively new PCD with over 5 million patients.

We briefly mention the three major UK PCDs as these have received the greatest attention by researchers and had the greatest impact on research outputs and through that, health policy. While we agree that ResearchOne is an interesting and exciting development –particularly through its more extensive record linkage - it is too new to have had any substantial impact on published research into health and health interventions. There are several other databases that arguably could make an equal claim to be mentioned, eg the aggregated Welsh GP data (SAIL), Scottish GP data, CiPCA in Keele, or other linked primary/secondary care disease registries (e.g. DARTS). Given this, we prefer to restrict our description to the (current) 'big three'.

4. The authors state that comparing different primary care datasets cannot prove validity. I agree –

the coding errors of one UK primary care dataset are likely to be very similar to another dataset because of GP and patient behaviour. I think the strongest conclusion that we can make from a comparison of UK primary care datasets is that they are as good (or as bad) as each other. I thought that the final conclusion was nice – that taken with other studies, this analysis adds to the body of evidence suggesting validity. Therefore I find the title of this article and the focus on 'validity' a little misleading. I think that the limitations of the comparison need further emphasis in the paper.

The type of validation that we have undertaken, which involved replicating the results from one data source using a second external independent source, is a very standard and widely accepted way of assessing validity – the main difference being that we have applied this to statistical model parameters rather than, as is more usual, to individual data-values. Therefore we do feel that the study is correctly described as a validity study. However, we have gone through the paper and 'toned down' text we thought made too strong a claim about validity. We also agree that the limitations of the comparison needed to be spelt out in more detail and have now done so in the final paragraph of the paper.

5. There were some interesting demographic differences between the CPRD and QResearch cohorts, and the proportion of patients with CHF was more than double in QResearch compared to CPRD. There were also differences in the prevalence of hypertension. Could the authors share any possible explanations for these differences? Could this have been due to a difference in coding between the data sources? The prevalences were higher in QResearch, which tended to have an older population. Were the age-specific prevalences of these diagnoses the same between the datasets?

The pattern of prevalence rates varied by condition, with more CCF but less hypertension in the Qresearch cohort, and similar levels of diabetes. Thus the pattern is complex and unlikely to be simply age-related. There are some geographical location differences between the cohorts, which may have contributed, and we have added a comment on these to the paper. It is plausible that coding differences between the EMIS and Vision systems may have also contributed, and we make this point. Unfortunately, QResearch provided no age-specific prevalence rates and so we are unable to conduct the comparison suggested.