Reviewer’s report

Title: Vascular disease in women: Comparison of diagnoses in hospital episode statistics and general practice records in England

Version: 2 Date: 27 August 2012

Reviewer: Markku Mahonen

Reviewer’s report:

I still think this study is important, showing that HES records can be used in epidemiologic research. However, I feel that the authors did not fully grasp my previous comments. I still think that the paper needs a substantial revision, rethinking and rewriting.

1) the name of the Journal 'BMC medical research methodology'. Methodology - 'the science of methods'. What is the method the authors are studying and propose to be used in medical research? How did they assess this method?

2) A paper on methods should be concise, precise and use defined terminology. This is not the case with this paper. For example, the authors state '...HES data are reliable....' In statistics, the term 'reliability' is very precisely defined. The authors used the word reliable as it is used in everyday language.

3) There are too many unnecessary uninformative details on specific diagnoses (for example, pulmonary embolism, subarachnoid hemorrhage). The good agreement between HES and GP's records in these diagnoses is nonsense since obviously GP must have had hospital records and HES diagnoses as the basis for his assessment. Actually, the data show good agreement with the assessment of HES coders and GPs. Obviously, both use the hospital records as the basis of their assessment. In that sense, the study is more about accuracy of the ICD codes recorded in the HES database.

3) It simply is not possible that all patients with vascular disease are hospitalized, as the data indicate, '97% of women with no HES record did not have a GP diagnosis of vascular disease. Or are these women so healthy? In that case, the generalizability of the data is poor.

What I think these data tells us:

HES records (actually, ICD codes) can be used for case ascertainment of severe vascular disease. They can be used for case classification: severe vascular disease/no severe vascular disease. Maybe even as end-point classification in the follow-up (although a separate validation study would be very important - please do it). HES records must have sensitivity which is not high but acceptable (quite a lot of cases will be missed since they are not hospitalized) - but more important have high specificity (low proportion of false positives).
If you do a validation study, the results would be truly important. If, as I assume, HES ICD codes for the most important diagnoses are valid enough, then they can be used for end point classification.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.