Reviewer's report

Title: Utility of electronic patient records for evaluating stroke secondary prevention in primary care

Version: 1 Date: 18 November 2010

Reviewer: Susan Wells

Reviewer's report:

The authors wish to inform the design of a pragmatic randomised trial of stroke prevention by evaluating data captured during routine general practice in electronic patient records.

The title conveys this aim somewhat-perhaps a better title for the expressed aim would be "Utility of electronic patient records in primary care for stroke secondary prevention trials ". The paper is well written and very interesting but seems to digress from the original primary aim of informing a trial by critically evaluating data captured on a large GP electronic database to providing results from a before-after study.

Major compulsory revisions

The abstract is not a good reflection of the study with the summary of the methods and results particularly needing to be revised.

Methods

To truly inform a pragmatic rct (rather than to mainly provide before –after results), this study needs to supply more information or be more critical about what the data can or cannot provide regarding the eligible participants and stroke diagnosis. The participants are those who have had a stroke recorded in the electronic patient record (EPR) for the first time and the index date of stroke is given as the date of recording. We are given no information on the degree of missingness of this data (what % eligibles are actually recorded) and how likely are the participants to be representative of all eligibles. We also have no information on the proximity of this recording date to the actual stroke event date. The actual event could potentially be years before with some practices systematically going through their records and READ coding at a later date.

While the authors note that validity of medical diagnoses in GPRD has been confirmed in several studies – what is the validity of stroke diagnoses in particular and does the diagnosis include TIA?

Exclusion criteria: What is the definition of an up to standard record? And what % stroke patients were excluded on this basis?

Exposure data collection- Many relevant exposures have been collected but there is no information on the ability to extract exposures or how this was done. It may be that UK GPs are exemplary about ensuring that BP, smoking, alcohol
intake etc are all systematically coded (rather than idiosyncratically available in text fields). Furthermore, it is assumed (implicitly) that recording equates to actual management (rather than failing to document what was provided or documenting in the record in an unextractable way). Therefore the ability to extract the relevant data is very important especially for variables where there is high degree of missingness (e.g., recording of smoking status and alcohol consumption).

The rationale for having both Table 3 and 4 is unclear. They have different totals for the same variables and the summary means do not differ overly much from a clinical perspective. What does Table 3 actually represent? While there are more people who had recordings (but less who had both before and after the index stroke date)—the missing variables (either prior to or after stroke) are unlikely to be missing at random. As this data is paired/not independent before and after, Table 4 makes the most sense to me (and the ICCs calculated and appended to Table 4.) Furthermore, how were the CIs estimated for Table 4? My understanding is that they should be based on the individual participant before-after differences—for example for BP based on 14,006 differences in before—after mean systolic BP. It is not clear what has been done from the statistical analysis section of the methods.

There is no mention of Ethics approval for this study.

Discussion

The discussion was more around the before-after results than what the paper aimed to achieve. Given that the stated title was the utility of electronic patient records for evaluating stroke secondary prevention in primary care the discussion contained little critique of the methods, definition of eligibles, data extraction, missing data for each variable of interest, data validity, the impact of practice variability, the impact of ICCs on the design effect for cluster RCT in practice and the effect sizes found. For example, Table 5 is probably not representative of the practices—rather the best case scenario based on the practices who have better systems of care. Therefore has implications for a pragmatic trial.

Minor essential revisions

Just looking at the data from Table 1, with the exception of statin prescription and BMI recording, there does not appear to be any evidence for the authors' statement that risk factor management has increased from 2003-2007.

The header for Table 2 does not describe the table accurately. The footer is a much better description however it states quintile distribution whereas the table has upper and lower quartile headings.

Page 8 second paragraph I think should be Table 4 not Table 5 as documented

Discretionary revisions

There are five tables given in this paper—table 2 describing variation may be better as a figure (suggestion).
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.

Declaration of competing interests:
I declare that I have no competing interests