Reviewer’s report

Title: Addressing the deficiencies in the evidence-base for primary practice in regional Australia - Sentinel Practices Data Sourcing (SPDS) Project: A Pilot Study

Version: 2 Date: 10 June 2013

Reviewer: Tom Love

Reviewer’s report:

Major compulsory revisions

• Acknowledge and address bias arising from the definition of the denominator – the sample, and whether there is bias as a consequence of an ill population being more likely to visit the practice (implied by the difference between the sample age sex profile in Figure 4, and the ISML profile in Figure 3 – unsurprisingly 0-4 year olds are highly represented in the sample, whereas people in their twenties are low). At the very least there should be some estimate of the proportion of the local population who have consulted, and therefore the extent of such bias. If this correction was actually performed in the process of standardisation, then it should be reported fully. Otherwise it is hard to defend the results as a true estimate of prevalence, comparable with population prevalence from the Australian population overall.

• The validity of the underlying disease coding is not addressed, and must be considered before a meaningful claim can be made that this approach produces “a valid and sensitive surveillance system on chronic diseases”. The authors should note the extensive validation conducted on the UK GPRD (see for example the 2010 review of validations by Herret in BJ Clin Pharmacology, which makes recommendations for assessing and reporting validity).

• The interpretation of key results is idiosyncratic. In the abstract and discussion the authors draw attention to anxiety, hypertension and obesity, where they observe a higher than national rate, but their results actually found much lower than national rates for a number of diseases, including heart disease, osteoarthritis, osteoporosis and COPD. COPD in particular, is less than 1/3rd of the national rate. Is this real (in which case it would be interesting to have some discussion about why it is the case)? Or is this an artefact of coding problems, as noted above (in which case it rather undermines the claim to validity)? These results should be discussed more fully, and their relevance to the overall validity of the dataset considered.

Minor essential revisions

• Fully report data manipulation. The method section refers to data cleansing (as does the author contribution statement). But the nature of the data cleansing is not made clear. What were the issues in the raw data which had to be corrected?
Data cleansing is to be expected, but transparent reporting of method would require the nature of such data pre-processing to be explicit.

• Fully report the coding system. The underlying coding system for morbidity is not clearly stated. Was it ICPC, SNOMED or ICD-10 (ICD-10 is implied, but not stated clearly, in the discussion section, but this would be an unusual coding system for routine data in general practice)?

• Amend Figure 5. A polar plot of prevalence by age is an unconventional and unduly complex presentation, which actually obscures the patterns in the data. I strongly recommend a simple x-y line plot for this diagram, which is likely to be much more accessible to the reader, and will display the patterns by age much more clearly to the eye.

Level of interest: An article of limited interest

Quality of written English: Acceptable

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

Declaration of competing interests:

I have no financial or non-financial competing interests in the publication of this paper.