Reviewer's report

Title: Combining complementary data sources improves prevalence estimates
validity of chronic medical conditions

Version: 6  Date: 21 July 2014

Reviewer: Joke Korevaar

Reviewer's report:

The objective of this study was to determine the feasibility of combining
information from complementary data sources.

This is an interesting topic as the data source plays an important role in
presented prevalence rates.

Major comments:

The authors studied prevalence of 4 conditions from three different databases
resulting in 12 comparisons. The authors conclude that combining provide more
valid estimates. Yet, I’m not convinced that this is indeed the conclusion that can
be drawn based upon the presented results. The pattern of all three data sources
resembles each other (per condition). Yet, the absolute values differ, and the
absolute values differ between conditions. I miss information why the FIRE
database should not be complete for the conditions studied. High BP is often
treated by the GP, and if not treated by the GP, the GP should know that the
patient suffers from high BP. So, why should this be missing in the database? Is
it a lack of communication between medical specialists and GPs, or what?

The same holds true for DM and dyslipidemia. So why should one add results
from SHS or from the hospital? In addition, I don’t think that it is remarkable that
prevalence based on hospital data is lower. Not all patients with hypertension are
treated in the hospital (not even for ambulatory care). The manuscript would
become of more interest if the authors could explain why it is to be expected that
the FIRE database is not complete. And it would become of more value if the
authors could indicate the under- or overestimation that would have occurred if
just one database was used. An if they could give what the new estimates (bases
on all sources) are, so what we gain by using three sources of data.

Next, the authors conclude that definitions based on drugs use alone give the
most reliable results. Do the authors know for which condition the drug was
prescribed? And was this indeed the indication to which the authors ascribe the
drug?

There is a difference in the results for obesity compared to the other conditions,
this condition is lower in self-reported data en far lower in hospital data, as one
could expect. My questions is what the added value of these latter two sources to
the FIRE data is for this condition. If I see the results, I think the overall
conclusion does not fit for this condition.

This manuscript clearly shows that the use source plays a role in the observed prevalence rates, yet, I cannot not see that this problem is solved (or diminished) by combining several sources.

**Level of interest:** An article of limited interest

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

'I declare that I have no competing interests'