Reviewer's report

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

Version: 6 Date: 17 July 2014

Reviewer: Heike Hansen

Reviewer's report:

Thank you for this manuscript. This article is a good addition to previous studies about prevalence estimates of chronic conditions in different data sources for the area in Switzerland. These analyses based on three large data sources.

Please find my comments below:

- Major Compulsory Revisions

1. Title: The term “Combining” is misleading to the reader. Maybe it is better if you use “Comparison” because the three data sources were not linked but compared. This also applies to the rest of the manuscript. In the results section you have already used the term “comparison” and in the discussion you used “complementation”.

2. Abstract: Background: “Prevalence estimates of chronic medical conditions and their multiples (multimorbidity) in the general population are scarce and often rather speculative.” Did you mean this for Switzerland? If yes, it should be written.

3. Introduction:

Paragraph 2: “However, measuring prevalence rates of chronic medical conditions poses challenges in countries where information systems are poorly developed, because of varying case definitions…”

This sentence is confusing. I do not think that the information systems are poorly developed. On the basis of your argumentation the systems are different.

4. Data protection: I miss more information in this section. Who are the other researchers? Please explain the “harmless acknowledgement”.

5. Definitions of chronic conditions:

SHS: The descriptions of the definitions of the chronic conditions are similar to table 1. Maybe the text can be shortened.

FIRE and MEDSTAT:
I missed a description how the selection of the criteria (e.g. ICD Code or blood values) took place. Please describe the process.

6. Analyses:
Please add the numbers (n) of the person who were excluded.

7. Analyses:
What is about Missing values in the three data sources? How did you handled this?

8. Results:
In the figure you wrote “drug prescribed” for FIRE data. In the text you wrote “drug use alone”. Drug prescription is not the same as drug use. Of course you could hardly control for the real use of drugs in the patients, but you should use the term carefully. Or you explain and discuss the term, respectively. The same problem is in SHS data for the term “Doctor told” in figures vs. “diagnosis of …” in the text. There should be a more clear term in the text e.g. doctors told diagnosis. Or you switch it to “self reported diagnosis”. All in all you have to revise and bring into line the descriptions of the categories.

9. Discussion:
You should integrate subtitles e.g. Comparison with other studies, Strength and limitations, Conclusion.

10. Discussion: “Therefore, the transfer of prevalence estimates of chronic medical conditions from primary care and hospital settings to the general population seems feasible with the appropriate care to the methodology used.”
This issue should be discussed in more detail.

11. Discussion: Line 347
You classify dyslipidemia as a symptom-based chronic condition. I think this is a clear cut-off diagnosis measured with e.g. cholesterol value. The patient noted his high cholesterol initially not based on symptoms but on the blood values. What is your definition of symptom-based chronic condition?

12. Discussion: From Line 391
You should revise the conclusion related to the first comment. And it should be written what this study adds.

- Minor Essential Revisions

1. Data protection: “All data “WERE” anonymized…”

2. Results: Please unify the number of decimal places (one vs. two)

3. Figure legends: The sentence about the “Data sources” is in each legend the same. This can be shortened.
- Discretionary Revisions
  none

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

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

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.