Author's response to reviews

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

Authors:

Ueli Zellweger (uzellweg@ifspm.uzh.ch)
Matthias Bopp (matthias.bopp@uzh.ch)
Barbara M Holzer (barbara.holzer@usz.ch)
Sima Djalali (simadjalali@usz.ch)
Vladimir Kaplan (Vladimir.Kaplan@spital-muri.ch)

Version: 9 Date: 10 October 2014

Author's response to reviews: see over
**Point by point answers to Reviewer 1**

The objective of this study was to determine the feasibility of comparing 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 manuscript improved remarkably by changing the focus from combining into comparing.

It is still unclear for me, after reading the summary, on which results the conclusion is based that definitions based on drugs use alone give the most reliable results.

We thank the reviewer for this hint. Indeed we rephrased already in the first revision the conclusion in the abstract and the end of the manuscript according to the comments of the reviewers. Unfortunately we missed to drop the sentence “that definitions based on drugs use alone give the most reliable results.” We now caught up for this.

Do the authors know for which condition the drug was prescribed? And was this indeed the indication to which the authors ascribe the drug?

We already discussed this point during the first revision. As we stated then, we are very confident about identifying diabetes mellitus based on drug use, because there is no anti-diabetic drug that could be used for another condition than diabetes. It is more difficult for dyslipidemia because use of lipid lowering drugs, especially for secondary prevention, does not imply elevated lipid levels above certain cut-off values. We already expanded on this issue in the last part of the limitation section (“Both, the definition based...underestimating bias”). For the ease of understanding we added: The identification of hypertension based on drug use is even more complicated because antihypertensive drugs can be used for heart disease or renal disease as well. However, many patients with renal disease or heart disease have hypertension. We calculated the prevalence rates for hypertension including and excluding patients who had renal disease or heart disease, and found minimal differences (not shown). Therefore, we provide prevalence estimates for hypertension based on antihypertensive drug use without excluding patients with heart failure or renal disease.

I still feel some hesitation about the results regarding obesity. If I look at the results, my conclusion would be that comparing different databases does not give a better indication of true prevalence, just information that prevalence is different using different source, and that is unclear which source is representing the truth.

We discussed also this point during the last revision (“We agree with the reviewer that obesity is clearly underrepresented in the survey data and the hospital discharge statistics and this is an important observation of the paper. We do not claim that we have solved the problem of underreporting. Furthermore, we have clearly de-emphasized the idea that combining data adds value in estimating prevalence of chronic conditions”) and stated that “chronic conditions with a less clear cut-off such as obesity were - limited to the three compared data sources - underreported in hospital data” (Discussion, section "comparison with other studies"). With the focus on comparing instead of combining we thought to have made clear that our major goal was rather about patterns and relative differences regarding age and gender than about absolute prevalence estimates. Nevertheless, in order to prevent misunderstandings we slightly modified the manuscript:

Discussion, second paragraph
Therefore, prevalence estimates of chronic medical conditions based on primary care data and hospital discharge statistics usually can be generalized to the population after a careful and appropriate harmonization of the case definitions.

Conclusion
We conclude that complementary data from different data sources might generate different prevalence estimates of chronic medical conditions in the general population. However, common age and sex patterns indicate that a careful harmonization of the definitions of chronic health conditions will provide strikingly similar age-and gender-specific distributions, even in data sources based on different settings and assessment methods.

This modification also applies to the conclusion in the abstract:
**Conclusion:** Complementary data sources can provide different prevalence estimates of chronic medical conditions in the general population. However, common age and sex patterns indicate that a careful harmonization of the definition of each chronic medical condition permits a high degree of concordance.

In addition, we dropped a sentence in the Discussion (section "comparison with other studies")
In line with others [8-10, 13], concordance between health survey and primary care data was good for clearly defined chronic conditions (diabetes and hypertension), and fair for chronic conditions with a less clear cut-off (dyslipidemia and obesity). In general, chronic conditions with a less clear cut-off were more reflected in the health survey than in the primary care data (dyslipidemia). While for diabetes comprehensive definitions performed best, for hypertension and dyslipidemia building the case definition on drug use alone was more efficient. Hospital data performed well for clearly defined diagnoses (diabetes, hypertension), but underreported severely chronic conditions with a less clear cut-off (dyslipidemia, obesity).