Author’s response to reviews

Title: Cause of death coding in Switzerland: Evaluation based on a nationwide individual linkage of mortality and hospital in-patient records

Authors:

Ueli Zellweger (ulrich.zellweger@uzh.ch)

Christoph Junker (Christoph.Junker@bfs.admin.ch)

Matthias Bopp (matthias.bopp@uzh.ch)

Version: 1 Date: 27 Dec 2018

Author’s response to reviews:

Response to Reviewers (I'd like to attach a better readable MS word version using colors and different styles, but I did not succeed to find a way to manage this - if you give me a working e-mail address, I would be pleased to send the .doc)

Reviewer #1 (Antony Stevens):

Conclusion. I found the information and opinions expressed by the authors to be extremely interesting and I am sure it will have the same effect on many readers of POPHM. In almost all cases what the authors say is perfectly clear and it is ok to leave it in the original.

We thank the reviewer for his benevolent evaluation and the big job of identifying potentially misleading phrasing, which helped to substantially improve the manuscript.

I have however recorded how I would have expressed myself in some of the text.

1. – what the authors wrote : health of general populations

comment/suggestion : health of populations  done

2. – what the authors wrote : E.g. amenable mortality, i.e. deaths due to causes that should not be fatal in the presence of effective medical care, is an indicator of national levels of personal health-care access and quality [3]

comment/suggestion : For example deaths due : : : effective medical care, known as amenable mortality, is an indicator of national levels of personal health-care access and quality [3] done

3. – what the authors wrote : also due to
There may be substantial variation in certification practices between countries as a known problem.

4. what the authors wrote: milieus

This is a quotation - we therefore have to stick to the original text.

5. what the authors wrote: For these reasons, evaluating the reliability of cause-of-death assignment and coding is a crucial issue.

6. what the authors wrote: Mostly, validation studies are restricted to letting a sample of deaths being coded by different persons [1, 7, 10, 11].

7. what the authors wrote: Agreement between original and reviewed data shows, however, rather reproducibility than accuracy of the information [12] and will not discover forms that look correct, but do not correspond to the actual facts of the case [13].

8. what the authors wrote: in timely limited samples

9. what the authors wrote: mandatory few years later

10. what the authors wrote: quality of Swiss data
11. – what the authors wrote: which attested Switzerland

comment/suggestion: which concluded that in Switzerland done

12. – what the authors wrote: pursuing the deduction of correction factors for time series

comment/suggestion: do they mean i) trying to identify correction factors with which to correct time series or ii) trying to identify ways in which time series may be corrected ???

(i) is correct and was applied to the text

13. – what the authors wrote: it may interest

comment/suggestion: it may be of interest to know which diseases done

14. – what the authors wrote: are not necessarily conforming

comment/suggestion: do not necessarily agree done

15. – what the authors wrote: entail a distorting effect

comment/suggestion: result in distortions done

16. – what the authors wrote: Concerning (???? non-) accidental deaths,

comment/suggestion: In the case of accidental deaths see 17.

17. – what the authors wrote:

comment/suggestion: Not sure what was meant by Concerning non-accidental deaths, the underlying cause of death is defined as ...“the circumstances of the accident”

The restriction to non-accidental deaths was dropped since the rules apply to all causes of death.

18. – what the authors wrote: For deceased aged between 55 and 94,

comment/suggestion: For those who die between 55 and 94 years of age, done

19. – what the authors wrote: In the average,

comment/suggestion: On average, done

20. – what the authors wrote: mostly private home

comment/suggestion: mostly private homes done

21. – what the authors wrote: get least diagnoses
comment/suggestion: get the least diagnoses  done

22. – what the authors wrote: For this since 2001 fairly complete census,

comment/suggestion: not sure what this means

We simplified to: "Since 1998, all hospitals in Switzerland have to report data on their inpatient stays to the Swiss Federal Statistical Office [26]." (after a couple of years, quasi full coverage was accomplished for the first time in 2001 - however, this detail is not relevant for this paper and can therefore be dropped.)

23. – what the authors wrote: congregated

comment/suggestion: aggregated  done

24. – what the authors wrote: , with two exceptions, however

comment/suggestion: . However there are two exceptions  done

4

25. – what the authors wrote: SPSS 25 (IBM Corp, 2017) was used to calculate

comment/suggestion: It is quite rare for the results of data processing to depend on the software that is used. This could, for example, occur in situations where the best estimate for a statistic is still an open issue. If the authors mention the software it suggests, to me, that they think that another software could have provided different results. I would not bother to mention which software was used.

The reviewer is right which his remark that results generally do not depend on the software used. Nevertheless in the last years it has become increasingly common to report which software was used. However, if the editor prefers not to mention the used software we would of course not object.

26. – what the authors wrote: From originally 74,093 deaths

comment/suggestion: Starting with 74,093 deaths  done

27. – what the authors wrote: this applied to 72,566 (97.9%)

comment/suggestion: 72,566 (97.9%) were found to be index cases  done

28. – what the authors wrote: In the sum of the 70 selected master categories

comment/suggestion: Among the 70 selected master categories  done

29. –what the authors wrote: maxima in
comment/suggestion: the highest values were obtained for... done

30. – what the authors wrote: on the individual level

comment/suggestion: at the individual level... done

31. – what the authors wrote: Compared to the Lake Geneva region

comment/suggestion: I would have been interested to know how Lake Geneva is different from the rest of Switzerland.

For the logistic regression we had to choose a reference category, for which the Lake Geneva region suggested itself as the most populated unit. Whereas there is solid evidence for the conclusion that the agreement in the Lake Geneva region is substantially lower than in the rest of Switzerland, it is not possible to derive an exact figure from our current model.

(not changed)

32. – what the authors wrote: underlying cause of death in 59%

comment/suggestion: underlying cause of death of 59%... done

33. – what the authors wrote: Discordance in less clearly defined chronic diseases may therefore be due to a large part to individual variation in diagnostic practice.

comment/suggestion: Is the variation due to the difficulty in getting the diagnosis right or is it due to the variation in diagnostic skill among doctors?

Since we have no gold standard for deciding which information – primary cause of death or main hospital discharge diagnosis – is the correct one, we cannot conclude that this is a question of insufficient diagnostic skill. In order to prevent misleading, we changed the sentence to: "... may therefore be due to a large part to increased difficulty in reaching a consensus on diagnosis."

34. – what the authors wrote: suggest a substantial part

comment/suggestion: suggest, a substantial part... done

35. – what the authors wrote: in the very old population

comment/suggestion: in older people... done

36. – what the authors wrote: The found proportions of agreement between underlying cause of death and hospital diagnoses were very similar

comment/suggestion: The proportion of agreement that was found between the underlying cause of death and hospital diagnoses was very similar... done
37. – what the authors wrote: gave slightly better results
comment/suggestion: the agreement was closer   done

38. – what the authors wrote: with our study generally performing slightly better
comment/suggestion: the figures from our study show a closer agreement  done

39. – what the authors wrote: especially larynx
comment/suggestion: such as larynx  done

40. – what the authors wrote: most obviously
comment/suggestion: most clearly  done

41. – what the authors wrote: performing worse for diabetes mellitus
comment/suggestion: with less agreement in the case of diabetes mellitus  done

42. – what the authors wrote: multiple logistic regression
comment/suggestion: I would have been interested in a formal description of the model that was fitted.

Based on a suggestion of the other reviewer we added in the methods section: "(logistic regression was performed)...to assess the influence of socio-demographic variables on the agreement of hospital diagnoses and the underlying cause of death." and in the discussion a reference to Table 3.

43. – what the authors wrote: Comparatively lower agreement for the less educated and those speaking a non-European language could point to lower health care expenditure among less advantaged individuals.
comment/suggestion: It is not obvious how lower health care expenditure can result in a lower standard of recording. Do the doctors bother less when the patients are poor? Or is it that the poor cannot provide useful descriptions of their symptoms?

The idea was indeed that doctors may be less motivated for thorough examinations when they did not expect to have to discuss results with patients and relatives. We have to admit that health care expenditure is an unfortunate phrasing. We therefore changed to: "...could point to a lower standard of/reporting (and maybe also examination) among less advantaged patients."

Reviewer #2 (Nicholas Kassebaum):
This is an interesting paper and the analysis is one that is of use in the field of health metrics.

We thank the reviewer for his thorough evaluation and the valuable suggestion helping us to improve the manuscript.

I especially like how the authors have worked to essentially do a quality evaluation of VR-coded death data based on the reliability of codes that are NOT often considered in quality measurement (i.e. looking at non-ill defined). The authors should emphasize this point/ theme more strongly.

Thanks for this suggestion. We added in the intro: "In contrast to others we did not target an overall quality measure for hospital and cause of death statistics in Switzerland (e.g., the proportion of ill-defined causes) but aimed at assessing the reliability of as many as possible specified diagnoses / causes of death."

There are several instances where the authors may not have considered the full implications of the decisions they made in analysis - not that I disagreed with the defensability of them per se, but rather that the work could be more informative if some different choices were made (e.g. matching with a classification system / cause list already in existence).

We do not agree with this. In fact, the analysis is essentially based on the ICD-10 Tabulation List for Morbidity – excluding external causes (which are difficult to translate into causes of death), nonspecified and rare causes. A combination with ICD-10 Mortality Tabulation List 2 allowed to include two additional cancer sites and several combined categories. The only category "not already in existence" is the combined category for AD and other dementia (but still showing the results for the original subcategories).

What are the recommendations the authors can make about what we should change as a result of their analysis?

(see answers in the discussion section below)

As far as the writing is concerned, it is barely acceptable. There are several sentences and phrases with unclear meaning and word usage that could be clearer. There are results in the methods section, discussion points in the intro, some results in the discussion section. I encourage the authors to carefully edit the text for clarity of language and make sure explanations are presented in neutral terms (either eliminate or at least define jargon).

We thank the reviewer for his suggestions which together with the suggestions of Reviewer 1 decisively helped to improve the quality of language.

We edited the text carefully and hope that it is now acceptable.
Some specific comments from the manuscript:

Page numbers and line numbers should be sequential.

Intro page: Could be strengthened to more clearly and articulately state the value added by this study. Why do you need to do this? What does it do that hasn't been done previously?

We already recalled in the intro that except of a meanwhile outdated study there were no such validation studies in Switzerland at all. As mentioned in the discussion also from other countries there is only one recent study of similar size. In addition – and this may be more relevant for future epidemiologic or healthcare utilization studies – we aimed at evaluating combined vs. separate disease categories where the related ICD-codes may not always be easily distinguishable. This "multi-level" approach and the search for sociodemographic determinants that could influence the agreement between hospital and cause of death data are – to our best knowledge – novel. However, we are pleased to adopt the reviewer's suggestion and added at the end of the intro: "Finally, we aimed at evaluating combined vs. separate disease categories where the related ICD-codes may not always be easily distinguishable, and at looking for sociodemographic determinants that could influence the agreement between hospital and cause of death data."

Line 19-22: I have no idea what the meaning of these sentence is with the changed phrasing this should now become more clear: "Usually validation studies are restricted to obtaining the coding of the same set of death certificates from different raters [1, 7, 10, 11]. Unfortunately an agreement between the original and the reviewed data may us tell more about reproducibility than about accuracy of the information [12]: The forms look correct, but do not provide an accurate description of the case [13]."

Line 25: What is "timely limited"?
changed to: "studies spanning short periods"

Line 32: should be "mandatory 'a' few years..."

done

Page 2 (has methods midway down)

Line minus 2: replace badly with poorly
Line 4-8: This is more appropriate for the limitation section

No – this limitation is not specific to our study and we need this information in order to underline that we took into account also secondary diagnoses and concomitant causes of death – a strength that contrasts to many other studies.

Methods

Line 38-42: This should be in the results section. Replace "least diagnoses" with "the fewest diagnoses listed"

No – these are general features of the Swiss Cause of Death Statistics, encompassing all deaths and not only our study population. Restriction to the latter would not allow the comparison with persons deceased at home or in an LTC facility.

Line 53-57: This is results.

done

Line 60: delete "i.e."

done

Page 3:

Line minus 1: congregated is not the right word here.

replaced by "aggregated"

Line 3-5: please summarize what the anonymous record linkage did/ accomplished/ how it was completed

We rephrased: "Using full date of birth, sex, and a geographical identifier, an anonymous record linkage with the Swiss National Cohort [28] could be successfully established [29], providing additional socio-demographic information (educational level, place of birth, principal language, household type) from the 2000 census."

Disease categories section = refer the reader to table 1 and 2 at some point during this paragraph

We added: "for analysis and the tables in the results section"

line 16: change unspecific to nonspecific
Line 17: change less to fewer

done

Line 20: In what way was the list for morbidity "more specific"? How was this defined?

Both lists cover the whole ICD, but the items in the morbidity list are generally more detailed (298 vs. 80 items), that means the mortality list has more combined items.

How were the cause categories selected? Why were some combined and others not? Is there any precedent for the classification system selected?

All combined categories stem from the mortality list. In view of a variable assignment of the ICD code (more or less detailed, known delimitation problems) we present the results for both, the detailed and the combined categories.

Line 25: What is the basis of the statement that dementia and AD are not distinguishable? Is this only for death? For morbidity? For both?

We do not pretend that they are not distinguishable (though in a strict practice, a proof of AD is not possible in vivo). But as the much higher agreement in the combined category shows, physicians are not very consistent in reporting.

Page 4:

Line minus 2: Logistic regression of what? Need a lot more detail on the statistical methods used in this analysis.

We added: "...to assess the influence of socio-demographic variables on the agreement of hospital diagnoses and the underlying cause of death."

Results section

Line 15-21: this is methods. this text could be exchanged with the results that are presented in the methods section

done

Line 25: Need to explain what all the numbers are in the figure. What are 5,092 and 8,888?

Line 38: What do you mean by "could be traced"? Need to be more specific
We change to: "...could be found among the principal or additional diagnoses reported for the terminal hospital stay of the same individual"

Line 47-49: All of these with low agreement are themselves non-specific or ill-defined codes. What was the basis for including them here? I thought all the non-specific codes were eliminated.

The reviewer is right that "not specified stroke" and "other cerebrovascular diseases" are in line with our definition for a non-specific disease item. They were just not removed because we wanted to keep all subcategories if other subcategories of the same combined category fulfilled the criteria.

Goes back to needing a much more comprehensive description of how the cause classification was developed. Need to also talk about classification in the discussion/ limitations section. Also on this, for each of the codes, how was the agreement within the same category/ organ system? For example, did not specified stroke still end up as a cerebrovascular or neurological disease when there was non-agreement?

This was just the motivation to compare combined and subcategories of diseases known to be prone to misclassification. The higher the difference between the proportion of agreement in the combined category and the average proportion of the subcategories, the more the latter end up in the respective combined category. An extension to organ systems is in our opinion not a surplus, since the results would either hardly change (organ systems with currently only few items) or become meaningless (in case of circulatory disease or cancer) and seriously overload the table.

Was primary hypertension associated with the disease that are known to cause/ result from it (e.g. kidneys, heart, brain, aneurysm)? This is a crucial piece of the analysis.

This sounds attractive but is impractical, since for most listed disease items there is no comprehensive consensus about precursor, concomitant or sequela diseases. We however agree that in some rare instances like primary hypertension this could enable a successor study.

Page 5:

Line 36-41: "diseases with delimitation problems". This entire paragraph is completely unclear.

We agree that the beginning "Concerning diseases with known delimitation problems" may confuse, since it prepones the interpretation. We therefore dropped this part of the sentence and restrict to a mere description of the result.

"Influence of sex and age" --> this section should just be renamed "Logistic regression results" or "Determinants of agreement between hospital and death records"
Line 17-21: There are some additional components that are potentially relevant for Switzerland that appear to have been ignored. For example, provider-to-patient language concordance. Was there any influence of whether it was a French or German speaking location where death occurred?

As can be seen from Table 3, French-speaking individuals as well as the only predominantly French-speaking region (Lake Geneva) had lower agreement. Treatment of patients in another language region is rare and can therefore not bias results.

Or being far away from home at time of death (e.g., a person with known dementia who dies at their community hospital would be expected to have high concordance, but not if they were across the country at the time of death)?

As mentioned in the limitations this does not apply: we restrict to hospital deaths, which means that the information for the death and the hospital discharge statistics stem from the same source (medical files).

What about the time elapsed between admission and death (e.g., those who die quickly after arrival are less likely to have concordant death certificates)?

For the same reason (information from the same institution) this should not have a substantial impact on agreement, either.

Discussion

Page 7

Line 5: Is there any evidence of a role of providers having incomplete information during terminal hospitalization?

Same problem: we cannot trace providers, but the potentially available information (medical records of the hospital) is the same for both, death and hospital discharge statistics.

Line 35-42: This is a rather weak interpretation of the findings of the study.

Based on what this study accomplished, what are the recommendations of the authors? Is Switzerland doing a good job? Not a good job?
Due to the scarcity of comparable studies, it is not easy to decide whether Switzerland performs better or worse than other countries. However, the limited available evidence – summarized in the first paragraph of the discussion – suggests that there may be no substantial difference to other developed countries. Also, on a cause-specific level the agreement rates are rather higher than those in the Minder-Zingg study from the 1980s.

What should they do to improve? Are there certain instances where the hospital records should be trusted more? Some where the death certificates should be trusted more?

From a mere comparison of A and B it cannot be decided whether A or B should be trusted more – this could only be achieved with a thorough case review based on full clinical documentation (see second recommendation in the next paragraph).

Which combinationm of factors should prompt further review?

We added in the Discussion a paragraph with suggestions for interventions to improve data quality in the future:

"What measures could improve the quality of the MS and the cause of death statistics in Switzerland? First, the responsible authority and data owner (Swiss Federal Statistical Office, SFSO) could do much more than at present. Currently, the SFSO publishes data reports on a yearly basis and researchers can request access to individual data, but there are no published reports about data quality. Second, the regular use of data automatically helps to improve data quality, because problems become visible and can be overcome. Especially studies combining information from different data sources may promote the evaluation of corresponding strengths and limits [33]. Third, a case review based on full clinical documentation should be carried out for the areas with lower scores in this study. For example, a sample of deaths due to diabetes could be studied in detail. The example of cancer shows that the work of cancer registries has an impact on medical documentation. An individual comparison of the cause of death statistics with the data of the cancer registries is required by law from 2020. Finally, the instruction in the scope and methods of the cause of death statistics during the basic and continuing education of physicians should be improved."

Line 38: There is no evidence presented that supports the statement that this is associated with "lower health care expenditure". Please clarify what you mean and what supports this statement.

We have to admit that health care expenditure is an unfortunate phrasing. We therefore changed to: "...could point to a lower standard of/reporting (and maybe also examination) among less advantaged patients."
Editor comments (Pauline Kim):

The only thing that I think the authors should be asked to do is to add something to the Discussion about their views on the quality of Swiss CoD data as Nick has asked for.

Are they fit for purpose?

We tried to adequately answer the questions of Nick Kassebaum (see above) and would like to refer to the first paragraph of the discussion dealing with this issue.

Is there any evidence they are of higher quality in French speaking vs german speaking populations of Switzerland?

The contrary applies: as can be seen from table 3, the Lake Geneva Region which encompasses more than 70% of the population living in French-speaking cantons, has in spite of adjusting for individual language the lowest probability of agreement (and agreement among deaths of French-speaking people is even when adjusted for area of residence lower than among those from German-speaking decedents). In order to support understanding, we added in the results "... Lake Geneva region, the only predominantly French-speaking region, agreement ..." and in the Conclusion "... regional variation (with lower agreement rates in French-speaking populations) ...".

What are the implications for interventions to improve data quality in the future? If they can address these issues in a para or two in the Discussion, I think (apart from a careful English editing) the paper should be accepted.

With pleasure we adopt this advice and added in the Discussion a paragraph with suggestions for interventions to improve data quality in the future:

"What measures could improve the quality of the MS and the cause of death statistics in Switzerland? First, the responsible authority and data owner (Swiss Federal Statistical Office, SFSO) could do much more than at present. Currently, the SFSO publishes data reports on a yearly basis and researchers can request access to individual data, but there are no published reports about data quality. Second, the regular use of data automatically helps to improve data quality, because problems become visible and can be overcome. Especially studies combining information from different data sources may promote the evaluation of corresponding strengths and limits [33]. Third, a case review based on full clinical documentation should be carried out for the areas with lower scores in this study. For example, a sample of deaths due to diabetes could be studied in detail. The example of cancer shows that the work of cancer registries has an impact on medical documentation. An individual comparison of the cause of death statistics with
the data of the cancer registries is required by law from 2020. Finally, the instruction in the scope and methods of the cause of death statistics during the basic and continuing education of physicians should be improved."