Reviewer’s report

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

Version: 0 Date: 16 Aug 2018

Reviewer: Nicholas Kassebaum

Reviewer's report:

This is an interesting paper and the analysis is one that is of use in the field of health metrics. 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. 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). What are the recommendations the authors can make about what we should change as a result of their analysis?

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).

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?

Line 19-22: I have no idea what the meaning of these sentence is
Line 25: What is "timely limited"?
Line 32: should be "mandatory 'a' few years..."

Page 2 (has methods midway down)
Line minus 2: replace badly with poorly
Line 4-8: This is more appropriate for the limitation section

Methods
Line 38-42: This should be in the results section. Replace "least diagnoses" with "the fewest diagnoses listed"
Line 53-57: This is results.
Line 60: delete "i.e."

Page 3:
Line minus 1: congregated is not the right word here.
Line 3-5: please summarize what the anonymous record linkage did/ accomplished/ how it was completed

Disease categories section = refer the reader to table 1 and 2 at some point during this paragraph
line 16: change unspecific to nonspecific
line 17: change less to fewer
Line 20: In what way was the list for morbidity "more specific"? How was this defined? How were the cause categories selected? Why were some combined and others not? Is there any precedent for the classification system selected?
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?

Page 4:
Line minus 2: Logistic regression of what? Need a lot more detail on the statistical methods used in this analysis.

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

Page 5:
Line 36-41: "diseases with delimitation problems". This entire paragraph is completely unclear.
"Influence of sex and age" --> this section should just be renamed "Logistic regression results" or "Determinants of agreement between hospital and death records"

Page 6
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? 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)? 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)?

Discussion
Page 7
Line 5: Is there any evidence of a role of providers having incomplete information during
terminal hospitalization?
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? 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? Which combination of factors should prompt further review?
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.

Level of interest
Please indicate how interesting you found the manuscript:

An article of importance in its field

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an
organisation that may in any way gain or lose financially from the publication of this
manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose
financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the
manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds
or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal