Author's response to reviews

Title: Establishing a follow-up of the Swiss MONICA participants (1984-1993): Record linkage with census and mortality data

Authors:

Matthias A Bopp (bopp@ifspm.uzh.ch)
Julia Braun (julia.braun@ifspm.uzh.ch)
David Faeh (david.faeh@ifspm.uzh.ch)
Felix Gutzwiller (felix.gutzwiller@ifspm.uzh.ch)

Version: 3 Date: 1 September 2010

Author's response to reviews: see over
Re-submission of our paper

Dear Melissa Norton

Please find below our answers to the reviewers comments. We addressed them point by point. Thanks to the review, several parts of the manuscript which were not precisely enough described, could be clarified. Our paper has therewith decently improved.
The second reviewer addressed an important point: It has to be stressed that in principle our method could be used in any country with a census-linked mortality study. We would therefore appreciate if the article could be featured, for example:

Kissing a sleeping beauty

Many cross-sectional studies and surveys miss to establish a mortality follow-up. Census-linked mortality studies may help to overcome this failure even decades later and in countries with no personal identifier available, thus enabling a longitudinal design.

Looking forward to hearing from you soon.

Sincerely yours

Matthias Bopp

Matthias Bopp PhD, MPH
University of Zurich
Institute of Social and Preventive Medicine

Review of Christa Scheidt-Nave

We wish to thank Christa Scheidt-Nave for her thorough review helping us to improve our manuscript. We adapted our paper in several respects. The changes are highlighted with the MS word track changed mode in the manuscript.

Major compulsory revisions
(1) Abstract
a. Results: The abstract should include some information that linkage success did not differ according to any key study characteristics, and that the results of survival analyses were robust to linkage step or certainty of a correct link.

Done

b. Conclusions: The conclusion that there is ‘negligible potential for selection bias’ appears too strong, because MONICA survival curves / mortality rates do differ from those of the Swiss general population, particularly in men. Instead, focus should be put on the feasibility and robustness of the data linkage procedure. In addition, the results regarding mortality patterns seem plausible but do reflect ‘healthy participant’ selection.

We agree with the reviewer that our statement was somewhat misleading. We referred to an additional selection bias introduced by the linkage process and not to the bias inherent in all health surveys with limited participation rate.
The conclusion now reads:
“Using anonymized census and death records allowed an almost complete mortality follow-up of MONICA study participants of up to 25 years. Lower mortality compared to the general population was in line with a presumable ‘healthy participant’ selection in the original MONICA study. Apart from that, the derived data set reproduced known mortality patterns and showed only negligible potential for selection bias introduced by the linkage process. Anonymous record linkage was feasible and provided robust results. It can thus provide valuable information, when no cohort study is available.”

(2) Methods
a. Linkage section, end of paragraph 1: how exactly was information on deaths which occurred before the 1990 census obtained?
Information on deaths before and after the 1990 census stems from the same source (Statistics of deaths and causes of deaths of the Swiss Federal Statistical Office). However, deaths occurred before December 1990 are not included in the standard SNC and had to be linked separately. The lack of census information for these individuals is a drawback which may affect linkage success.

We added:
“Therefore, study participants not retraceable in the SNC, i.e. in the 1990 census, had to be evaluated separately for a potential link with the official death registry. For deaths occurred before the 1990 census, linkage success is expected to be slightly lower.”

b. Linkage section, paragraph 3/List of succeeding steps, point 8: what exactly does ‘manual control and optimizing’ refer to?
We included in the manuscript:
“8) manual control and optimizing (check of remaining unlinked MONICA participants for potential partner records in the 1990 census and the 1984-90 death records; typically these records showed discordances regarding date of birth, place of residence and occupation which had prevented automated record linkage, but considering all available information and potential alternative links strongly suggested that the records referred to the same individual)”

c. Survival analysis section, paragraph 2, last sentence: ‘This procedure was also used...’ Does this sentence refer to the sensitivity analyses mentioned later on (conclusions, end of paragraph 1)?
Yes. We specified:
“Multivariable Cox regression was also used for sensitivity analysis (control for heterogeneity and errors introduced in the linkage process)”

(3) Results
a. Record linkage, end of paragraph 3: what procedures did ‘manual search’ imply?
This refers to point 8 (2b, above)
We specified „.... by manual search (cf. point 8 in the linkage section, above)”

b. Cox regression: This paragraph should include the results of sensitivity analyses (survival by linkage step, certainty of a correct link), as these are referred to in the conclusions.
We thank the reviewer for this important hint.
We added
“Sensitivity analyses using Cox models did not show any noteworthy difference of survival in function of linkage step or certainty of a correct link, when sex, age and wave was adjusted for.”
c. Age-standardized mortality rates (Table 3): this comparison seems of limited use. Since region as well as cohort (MONICA wave) appear to affect survival considerably (Table 2), excluding MONICA Ticino participants would be expected to increase the difference between observed and expected rates. *There may be a misunderstanding concerning the meaning of this table which probably relates to a lack of explanation and/or an imprecise table title. Because there are regional mortality differences, the national life tables used for the Kaplan-Meier-curves are not ideal. In order to demonstrate this, we thought that it was important to include a table that compared the MONICA study population with the general population of the same (VD/FR) region. We specified this in the results section and adapted the title of the table.*

Would it be possible to examine the effect of MONICA wave on age-standardized mortality rates and ratios by repeating the analysis with MONICA wave 1, i.e., the individuals with the longest follow-up? *Doing so we would substantially lose precision (loss of 57% of deaths) without gaining additional evidence to provide interpretable results. Nevertheless we present here – for the reviewer only – the corresponding table:*

### Table 3bis - Comparison of age-standardized mortality rates (per 100,000 person years, WHO standard population “Europe”, ages 45-79) between study and general population of the Vaud/Fribourg region: MONICA I only

<table>
<thead>
<tr>
<th></th>
<th>Linked data set MONICA/SNC (1)</th>
<th>Data of Swiss Federal Statistical Office (2)</th>
<th>Ratio (1)/(2), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males 1985-89</td>
<td>1479.7</td>
<td>1880.5</td>
<td>78.7</td>
</tr>
<tr>
<td>Males 1990-94</td>
<td>1520.9</td>
<td>1721.8</td>
<td>88.3</td>
</tr>
<tr>
<td>Males 1995-99</td>
<td>1452.3</td>
<td>1546.8</td>
<td>93.9</td>
</tr>
<tr>
<td>Males 2000-04</td>
<td>1247.9</td>
<td>1345.7</td>
<td>92.7</td>
</tr>
<tr>
<td>Males 2005-08</td>
<td>641.4</td>
<td>1185.6</td>
<td>54.1</td>
</tr>
<tr>
<td>N deaths</td>
<td>168</td>
<td>47,491</td>
<td></td>
</tr>
<tr>
<td>Females 1985-89</td>
<td>539.7</td>
<td>867.4</td>
<td>62.2</td>
</tr>
<tr>
<td>Females 1990-94</td>
<td>707.3</td>
<td>810.2</td>
<td>87.3</td>
</tr>
<tr>
<td>Females 1995-99</td>
<td>571.3</td>
<td>737.4</td>
<td>77.5</td>
</tr>
<tr>
<td>Females 2000-04</td>
<td>664.9</td>
<td>672.8</td>
<td>98.8</td>
</tr>
<tr>
<td>Females 2005-08</td>
<td>408.5</td>
<td>611.6</td>
<td>66.8</td>
</tr>
<tr>
<td>N deaths</td>
<td>86</td>
<td>28,928</td>
<td></td>
</tr>
</tbody>
</table>

MONICA: MONItoring of trends and determinants in CArdiovascular disease

(4) Discussion, end of paragraph 4: are there any alternate explanations for the observation that MONICA men but not women showed significantly lower mortality in 2000-2008 compared to the Swiss general population?

*We added to the discussion:*

“The larger difference in men than in women could relate to a lower MONICA participation rate in men. This could have led to a stronger “healthy participant” selection in men than in women.”

Do the authors see any need/perspectives for further research of this sex difference? *Since we could also trace non-participants, we have information about their survival. As one of our next papers, we have planned to assess differences in mortality between MONICA participants and non-participants. We will address gender issues in detail in that publication.*
(5) Conclusions, last two sentences: Uncovering hidden potential for public health intervention by analyses of representativeness of population studies in terms of survival - this is an interesting thought, however, it needs to be acknowledged that this would require additional information for comparison (health-related behavior, environmental data etc.).

We agree that we do not have the same information on risk factors for non-participants or the general population as for MONICA participants. However, from the SNC, important sociodemographic and environmental data is available. Again, we have planned to address this issue in the non-participants paper.

To clarify, we included:
“Most public health interventions address entire populations. However, they are often based on health surveys with limited representativeness. Better understanding of this discrepancy could uncover hidden potential for public health intervention.”

Minor essential revisions
(1) Table 1, second to last row 'start of examination' – TI is missing
Corrected

(2) Figure 1 lost to follow-up 2000-2008: does 'NA' stand for not assessed? The abbreviation needs reference in a footnote to the table.
We added: NA: not available, can only be determined when 2010 census will be linked

(3) Table 2: If the hazard ratio represents the adjusted ratio this should be denoted; covariates included and additional adjustments made need to be referenced in a footnote to the table.
Done

Discretionary revisions
(1) The authors could make the results section easier to read by changing the sequence of tables and figures and by adapting the text accordingly. Table 1 is not referred to in the text so far. The reviewer has overlooked it. Table 1 was mentioned on several occasions:

1) In the methods section, first paragraph:
“We included all three population studies from Switzerland conducted within the framework of MONICA: MONICA I (1984-86, N=3,445), MONICA II (1988-89, N=3,466), MONICA III (1992-93, N=3,252, table 1).”

2) In the results section “Kaplan-Meier plots”, second paragraph:
“In fact, participants of each MONICA wave entered the study at different dates within a time interval of few months (see table 1), causing the strongly clustered censoring points in the right part of the curves. The study entry intervals of MONICA I in Ticino and Vaud/Fribourg differed (table 1), which explains the two small clusters close to each other.”

Table 1 should be followed by Figure 1, Table 2, Figures 2 and 3.
By doing this we would also have to place the paragraph about „Cox regression“ before the part about „Kaplan-Meier plots“. In our view this would be a rather contra-intuitive sequence, because we believe that the descriptive procedure should precede the analytic procedure.

Table 3 might be omitted (see point 3c).
The table makes sense. We now better explain why.

(2) Results, Record linkage, end of paragraph 4: it might help to clarify that linkage success did not significantly differ according to any of the key study characteristics.
We specified:
“Difference in linkage success according to any of the key study characteristics – and therefore potential for selection bias – was marginal (table A1 in the appendix).”
Review of Simona Giampaoli

We thank Simona Giampaoli for the valuable comments allowing us to further improve our manuscript. The changes are highlighted with the MS word track changed mode in the manuscript.

Comments
- Background: objectives are clear and well described;

- Methods: please add some information to the methods of the MONICA at baseline in order to highlight the importance to collect mortality follow-up and assess the prediction of risk factors such as: (end of first paragraph under methods): “Weight and height, blood pressure and blood lipids were measured following standardized procedures and methods of the MONICA Project; they have been described in details in previous publications [ ]”;

  Done

- MONICA survey data cover age-range 25+, so include only comments of this age range (first paragraph under SNC data, aged 24-29 instead 10-29);

  We specified:
  The majority of unlinked records concerned individuals aged 10-29 years [1], which only marginally concerned the MONICA study population in the Vaud/Fribourg region.

- the description of the record linkage between MONICA and SNC step by step is not clear, please describe better how was performed; please include also MONICA III cohort;

  The description is complete. There are no deaths occurred before 1990 among MONICA III participants. It is therefore not included in the description.

  We added:
  “(MONICA III, which was conducted in 1992/3, is not concerned because there are no deaths occurred before 1990 among participants)”

- statistical methods are appropriate;

- Results: are well described;

- Discussion and conclusions: explain the limits of the study and discuss future opportunities:
  - possibility to link only all cause mortality

  This is a misunderstanding. Cause of death is included in the mortality record. We specified this in the methods section “Swiss National Cohort Data”:
  “Swiss census enumeration and registration of deaths occurring in Switzerland (including cause of death information) are virtually complete."

- no record linkage of hospital discharge diagnosis

  We added:
  „Finally, our procedure only allows the linkage with cause-specific mortality data but not with morbidity data. However, since hospital discharge data is available in Switzerland, there is potential for such a linkage."

- no validation of identified events
If we had discussed cause-specific mortality, we would agree that validation is an issue. However, in this paper, we only consider death from any cause (yes or no). As mentioned above, registration of deaths is virtually complete and does not need special validation.

the method offers the possibility to identify events, not to validate them, which is the expensive and time consuming part of the cohort studies. Validation is essential to answer etiological questions;

In Switzerland, the information deceased (yes or no) can be regarded as validated.

the method described is not an alternative to cohort studies.

We agree. We deleted that sentence and replaced it by:

“Anonymous record linkage was feasible and provided robust results. It can thus provide valuable information, when no cohort study is available.”

Moreover this represents a good method to use the Swiss National Cohort, but could not be applied in other countries.

We agree that some requirements must be given. However, in principle our method could be used in any country with census-linked mortality studies. Main problems are limitations by data protection and insufficient completeness. We see potential in at least a dozen European countries, including Italy (e.g. the Torino Longitudinal Study). In (mainly Nordic) countries with unique personal identification number (PIN) or mortality flagging (UK), our method is obsolete in most cases, with exception of studies which did not assess an official personal identifier.

For the future a standard consent is needed, common for all Europe in order to be able to follow the population during life reducing the confidentiality problems due to the mortality and morbidity follow-up.

We absolutely agree. However, in light of the substantial differences in data protection legislation between countries this is not realistic. Under such conditions, one could expect that only a minimal standard would be realized, insufficient to allow robust analyses and to provide valuable results.