Author's response to reviews

Title: Validity and timeliness of syndromic influenza surveillance during the autumn/winter wave of A(H1N1) influenza 2009 Results of emergency medical dispatch, ambulance and emergency department data from three European regions

Authors:

Nicole Rosenkotter (nicole.roenkoetter@maastrichtuniversity.nl)
Alexandra Ziemann (alexandra.ziemann@maastrichtuniversity.nl)
Luis Garcia-Castrillo Riesgo (urggri@humv.es)
Jean Bernard Gillet (jeanbernard.gillet@uz.kuleuven.ac.be)
Gernot Vergeiner (gernot.vergeiner@leitstelle-tirol.at)
Thomas Krafft (Thomas.krafft@maastrichtuniversity.nl)
Helmut Brand (helmut.brand@maastrichtuniversity.nl)

Version: 2 Date: 7 July 2013

Author's response to reviews: see over
Dear Dr Kavanagh,

Dear Dr Fowlkes,

Dear Editorial Board members,

Thank you very much for giving us the opportunity for re-submission and for your valuable comments which led according to our opinion to a further improvement of the paper.

We have taken care of all comments. Since some comments touched topics that influenced the analysis (choice of correlation analysis, stability of the baseline period) we adapted our analysis and did some re-calculations. The method section is revised accordingly and the “new” results are presented and discussed now in the paper. You can find more detailed information to each individual comment in the right column of the table below.

We hope our revisions fulfil your expectations and lead to qualification for publication in BMC Public Health.

Kind regards,

Nicole Rosenkötter

**Reviewer 1 (Kim Kavanagh)**

**Reviewer’s report: Major Compulsory Revisions**

1. I do not believe that the assessment of correlation is appropriate for the data. Pearson correlation assumes normality of the data and a linear relationship and it is not apparent if that this is the case here. A more appropriate technique would be to use spearman rank correlation or the cross correlation function (CCF). The CCF is likely to be most powerful when considering the degree of similarity between the reference data and emergency data time series. Please see and refer to the approaches summarised in Dailey et al. (2007) Timeliness of

**Response to reviewer comments**

Thanks for this very helpful suggestions. We accepted your proposal and compared all correlation approaches. The Spearman rank correlation was in general close to the Pearson correlation coefficient. Given the additional information on the time lag provided by the cross correlation function we report now this measure in the paper and took the Pearson correlation coefficients out. We revised the method and result section accordingly.
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2.</strong> Dailey et al. also recommend running the reference data through the detection algorithm to see where this data would generate alarms. I think that doing so would make your conclusions on the timeliness of the system stronger.</td>
<td><strong>We prefer to stick to the applied “real-life” approach. We know the beginning of the influenza pandemic based on the report of the sentinel surveillance systems of the respective public health authorities. Thus, we know the point of time when the beginning of the pandemic was officially confirmed and we think it is relevant to compare the aberrations detected in the syndromic surveillance data sources against it.</strong></td>
</tr>
</tbody>
</table>
| **3.** The baseline period used in the CUSUM is defined for each data source but there is no assessment of the data which is observed in this period. This should be included in the paper. Was the process in control during the baseline period? In particular are the baseline periods for the Spanish and the Belgian data stable? They use a period when the first wave of the pandemic was occurring. This will surely be very different from the data in the Austrian baseline which is pre-pandemic. This should be added to the paper. | **We took your comment seriously and checked our baseline data in this regard. This led to a readjustment of baseline periods. We excluded now months with significantly higher daily case numbers (e.g. in EMD-AT the months December to March are excluded from the baseline period 2005-2008) to make the baselines more stable. This led also to a re-calibration of the CUSUM algorithm in each data set.**

All sections as well as tables and figures were adjusted accordingly. |
| **4.** In an extension to point 2, there is no outline of the criteria used to signal alarms in the data. Assuming estimation of the mean count levels from the (hopefully) stable baseline period what technique was used to calculate the upper bounds of the CUSUM? Did this differ between the different data sources? This should be added to the methods and considered in the results section. | **The method section is extended in order to provide more detail on the Poisson CUSUM algorithm. Table 4 (former table 3) gives now also the Poisson CUSUM parameters for each data set.** |
| **5.** Assessment of the sensitivity and specificity of the method is assessed against the occurrence of the outbreak in the area but it is not known if the individuals in the data truly had Influenza A H1N1 or not therefore it is not truly sensitivity and specificity. The limitations of using this type of data to measure sensitivity and specificity should be added to the discussion. | **We added a paragraph in the discussion in which we discuss limitations of clinical sentinel surveillance data.**

However, with reference to ECDC we argue that for identifying the course of the pandemic, the primary interest in our study, clinical sentinel surveillance data are the preferred data source. |
| **Minor Essential Revisions** |   |
| **6.** The last sentence in the conclusions section of the abstract is very long and confusing. Please reword. | **We shortened the sentence.** |
| **7.** In the background section reference to the Scottish of analysis of help-line data could be added with reference [10]. [10] doesn’t cover all | **We added the suggested reference.** |

8. Tyrolean sick leave data is used as a substitute for sentinel data. Do you have any idea of how representative this data is likely to be? Are all individuals in Austria required to have health insurance or would this represent a particular demographic of society? The representativeness of this reference data should be discussed in the text.

9. The start of the outbreak in the Austrian data is taken to be the official reported start. A few details of how this was decided from reference [25] would be helpful. Also how was the end date decided?

10. On page 12, “stratified application of CUSUM algorithms” is mentioned. How was this implemented? More detail is required in the text.

11. On page 13, the daily approach sensitivity and specificity paragraph is very repetitive of the paragraph preceding it. Perhaps it could be summarised to say that a similar approach was taken to with the weekly definition.

12. The reference data column in Table 1 is not clear, in particular the reporting delay illustrations. There needs to be text to explain what the shaded regions mean. This could be added to the footnote and explained in the text of the methods.

13. Table 3 should be amended. The caption should explain that the statistics show the daily counts. With this in mind what are the Z and H rows showing? I assume these are the test statistics rather than a count. Given that the test statistics have no interpretation without knowing the comparative cut-off value I think it would be wiser to report the p-values here. I would also change the row names to describe what the test is assessing rather than just the test name i.e. something like “Differences between week days evaluated by Kruskal-Wallis test”.

14. In the footnote of the Table 3 I don’t think that “A=data availability period” is very clear. Does
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>this mean all of the data or the baseline period i.e. the data before period B?</strong></td>
<td><strong>We changed data presentation of table 3 (now table 4). We illustrate now for each data set the characteristics during the individual baseline period and during the test period (week 36/2009 – 52/2009).</strong></td>
</tr>
<tr>
<td><strong>15. In Table 3, it would have been helpful to see mean counts broken down by day of the week to understand the day of the week effects in each of the data sources.</strong></td>
<td><strong>The mean counts per weekday are now included in table 4 (former table 3).</strong></td>
</tr>
<tr>
<td><strong>16. Is the day of the week effect a weekend vs. week day effect? If so it might have been better to test Mon-Fri grouped against Sat &amp;Sun grouped counts.</strong></td>
<td><strong>We analysed differences in the mean counts per weekday. Differences were found in EMD-AT and ED-ES data. The differences are now described in the first paragraph of the result section.</strong></td>
</tr>
<tr>
<td><strong>17. It would be helpful if the caption for Figure 1 could have dates along with week numbers. It would then be easier to relate to the dates in Table 1.</strong></td>
<td><strong>The dates are included in the caption of figure 1-3 (former figure 1).</strong></td>
</tr>
<tr>
<td><strong>18. In Figure 1e (ED_ES) did the official pandemic definition really not hold for one week during the period? This seems strange and should be mentioned in the description of the reference data in the text if so</strong></td>
<td><strong>The public health authority in Cantabria, Spain uses a fixed threshold to identify increased influenza activity. In week 48 the cases fall under this threshold and in week 49 this threshold was passed again. We add a description in the “methods/data sets/reference data” section</strong></td>
</tr>
<tr>
<td><strong>19. Table 4 caption -please include dates with the week numbers here.</strong></td>
<td><strong>The caption of table 4 (now table 6) is changed accordingly.</strong></td>
</tr>
<tr>
<td><strong>20. In the discussion there are several mentioned of performance and in particular “moderate performance” and “weakest performance”. These are rather vague terms and perhaps it would be better to talk specifically in terms of the validity and timeliness</strong></td>
<td><strong>The discussion is rewritten. As you recommended we focus now on the validity and timeliness measures specifically. Additionally, we have added now subheadings in the discussion, separating the discussion of ED, EP and EMD data.</strong></td>
</tr>
<tr>
<td><strong>21. When p-values are included in the text the actual p-value should be reported rather than p&lt;0.05 etc.</strong></td>
<td><strong>We changed the correlation analysis from Pearson correlation to the cross correlation function (see also the answer at your comment number 1). No p-values are given for this type of analysis. Significance was assessed based on the confidence intervals in the SPSS outputs. However, in table 4 (former table 3) p-values are reported as exact values for the Kolmogorov-Smirnov test and the Kruskal-Wallis test.</strong></td>
</tr>
<tr>
<td><strong>22. In the discussion (page 20), it is mention that the algorithm could be specified and fine-tuned to increase performance. This is very vague. Please include some specifics of what could be done to improve the model.</strong></td>
<td><strong>This is now specified in the discussion section. We also refer now to a published study in which fine-tuning of the Poisson CUSUM algorithm in ED-ES data was performed.</strong></td>
</tr>
<tr>
<td><strong>23. On page 20, the text says &quot;As has been shown in</strong></td>
<td><strong>See above (comment 22). We added now</strong></td>
</tr>
<tr>
<td>24.</td>
<td>On page 21, the CUSUM approach is justified by its ease of application. I do not think that other methods are particularly hard to apply and indeed the regression algorithm discussed in Kavanagh et al. is deliberately kept simple in order to maintain flexible application to different time series. I think that the wording of this statement should be reconsidered. Future work could consider regression approaches.</td>
</tr>
<tr>
<td>25.</td>
<td>Some English needs correcting. a. Page 6, 3rd paragraph: “syndromic influenza surveillance systems focus not on..” change to “syndromic influenza surveillance systems do not focus on..” b. Same paragraphs: “diagnostic findings made during examinations” remove plural - “diagnostic findings made during examination” c. Page 9, 1st paragraph: “As substitute documented sick leaves...” change to “As a substitute documented sick leave...” d. Page 11. “Cases to which respiratory syndrome of ILI was assigned were aggregated to the variables respiratory syndrome or ILI cases per week and per day for further analyses” change to “Cases to which respiratory syndrome of ILI was assigned were aggregated by week and by day for further analyses” e. Page 18, 3rd paragraph. Change “...study of Cowling et al who applied also the CUSUM....” to “...study of Cowling et al who also applied the CUSUM....”</td>
</tr>
</tbody>
</table>

### Discretionary Revisions

| 26. | In Table 1, for Austria and Belgium there are two rows of data but the reference data column applies to both but it looks like it is split into two rows as well. This confused me initially. Perhaps the reference data column would be better as a separate table. | We made a new, separate table for the reference data (now table 2) to improve clarity of the presentation. |
| 27. | In Table 2, the ED-BE breakdown is not provided because there are lots of different free text combinations. It may be better just to have this as one row for the total count rather than a breakdown of NAs which does not look good in the table. | Changed according to your recommendations. |
| 28. | In the text the data for Spain is introduced as Santander (page 8) but in the following text is referred to as Cantabrian. This is slightly confusing and it would be best to be consistent throughout. | We changed the text in the Method section data sets/syndromic surveillance data in order to ease the differentiation between the data sources. It is now made clearer if the data are from local-, regional-, or national level. Throughout the text and also with reference to the reference data we... |
Level of interest: An article whose findings are important to those with closely related research interests

**Reviewer 2 (Ashley L Fowlkes)**

The authors pose a very interesting solution to the existing issue of finding a unified, sensitive and specific method of conducting syndromic surveillance for influenza in European countries. The use of emergency dispatch data as reported by the calling patients and the use of emergency paramedic data represent potential sources of data, and the authors propose an investigation to establish utility. The analysis ultimately did not support the utility strongly, which is an important contribution to the literature. While first some revision and clarification of the methods from this analysis is needed, the limitations of the analysis should be shared with the surveillance community.

**Reviewers report: Major Compulsory Revisions**

1. The authors use only one country’s dispatch data (EMD) without explanation of the limited analysis. If such data is not widely available, then utility is diminished.

   We revised the text in the Method section on data sets/syndromic surveillance data to clarify that the data we used are not related to availability of these data sources in Europe.

2. The authors limited the analysis to only those weeks when influenza was circulating, which undermines the specificity analysis and prevents the establishment of a baseline rate of influenza syndrome dispatches and emergency diagnoses.

   This is now discussed in the limitations section of the discussion.

3. The authors used very broad and non-specific ICD-9 codes to identify ILI or respiratory syndrome cases, thus the lack of correlation with physician-diagnosed ILI was not surprising. If I understood correctly, this is a limitation of the data available, but clarification on this point is needed.

   This topic is addressed in the method section (subsection variables) and in different sections of the discussion.

4. I was unclear as to whether or not the CUSUM aberration detection method was the most appropriate analysis for these data. If the study objective was to establish the utility of additional data sources, other simple analyses would have been much more supportive. (1) For example, was it possible to aggregate data to the week level, then compare the weekly % of calls related to ARI to the weekly % of patients with ILI and % influenza positive in national virologic surveillance?

   According to (1):
   This would have been easier indeed, but we were interested in the potential of this data sources for earlier detection of the influenza pandemic. An aggregation to weeks would have led to a loss of time as well. Therefore, we still stick to the presentation of aberration detection per day.

   According to (2):
   Please see also our responses to comment number 3 and 4 of reviewer 1.
any utility in these emergency data, since the only analysis presented was weeks with days flagged as aberrations. By first establishing a correlation in seasonality, further investigation of aberrations would have been more logical and support for these data would have been greater, even though the results did not show strong correlation.

We updated the aberration detection analysis: We were stricter in selecting suitable baseline periods for aberration detection by excluding months with higher mean values (probably influenced by seasonal influenza). The method and result section as well as the presentation of the results in the tables and figures is adapted accordingly. However, due to limited data availability in EP-BE, ED-EB and ED-ES data the baselines of these data sources might still be influenced by the summer wave (around week 30 according to ECDC) which could have resulted in too high threshold values of the Poisson CUSUM algorithm. We added this limitation in the discussion of the paper.

5. Tables 1 and 2 were helpful in understanding the data sources and methods used in the analysis. It would be very helpful to add some additional clarification to column headings, such as Tables and Figures are insufficiently labeled. I was not able to interpret the nature of the characteristics described in Table 3, nor use the corresponding paragraph in the Results to understand data being presented in the table. Are these average numbers of days per week with data available, average number of days per week with aberrations?

Unfortunately BMC Public Health asks for captions with about 15 words. However, we revised the table captions (and food notes) of all tables to improve clarity. This lead sometimes to a violation of the BMC word count rule for captions. Additionally we revised the text in the result section on table 3 (now table 4) to improve clarity.

6. The article needs editorial review. I could not identify all language or grammatical mistakes, but have noted a few major points:
   a. In the Background, the sentences switch between past and present tense used, and sometimes they are switched within the same sentence.
   b. The abbreviation for example given (“e.g.”) should not be used in the middle of a sentence, but rather after the sentence in parentheses or between commas. The 1st sentence of the 4th paragraph should be changed to “do not focus” rather than “focus not”.

   We revised the mistakes you mentioned. Additionally, the paper was submitted to a professional language editing service before resubmission.

7. General note, the switching of referring to surveillance locations and systems by country and by state (e.g., Austria and Tyrol) is very confusing for readers unfamiliar with the geography, especially in the case of Cantabria which is never defined as Spain in the methods.

   We changed the text in the Method section (data sets/syndromic surveillance data) in order to ease the differentiation between the data sources.

8. The first paragraph of the Results section is actually still a Methods paragraph.

   The paragraph is shifted to the methods section.
<table>
<thead>
<tr>
<th>Discretionary Revisions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The first paragraph of the Discussion is vague and does not really summarize your findings</td>
<td>We reformulated the first paragraph of the discussion. It delivers now a better, more specific summary of the results.</td>
</tr>
</tbody>
</table>