Author's response to reviews

Title: Estimates of Home and Leisure Injuries treated in Emergency Departments in the adult population living in metropolitan France: a model-assisted approach

Authors:

Christophe Bonaldi (c.bonaldi@invs.sante.fr)
Cécile Ricard (ricard@ch-annecy.fr)
Javier Nicolau (j.nicolau@invs.sante.fr)
Maryline Bouilly (maryline.bouilly@gmail.com)
Bertrand Thélot (b.thelot@invs.sante.fr)

Version: 2
Date: 22 November 2013

Author's response to reviews: see over
Dear Editor,

The authors and I are pleased to read the positive evaluations regarding our manuscript (MS: 1903522451052679) and we thank you for the opportunity to respond to the reviewers' comments. The original manuscript has been modified accordingly. The changes made in the revised version are appearing in bold and/or yellow font.

All authors have approved the submission of the revised manuscript.

Yours sincerely,

Christophe Bonaldi

---

**Answers to referee Pr Juanita Haagsma**

We thank the referee for the nice evaluation of our paper and we made the suggested changes to improve the manuscript.

**Referee’s comment**

1) I was wondering whether an ethical approval was necessary to perform this study and I think the authors should state this in the methods section of the paper.

**Authors’ answer:**
Epac network has indeed received a formal authorization to collect personal data and the French Institute for Health Surveillance has a permanent access to the PMSI data allowed by the French Commission on Individual Freedom and Data Storage. These important details have been added in the revised “Materials and methods” section.

2) Secondly, I would appreciate if the authors added information in the methods section on the approximate coverage is of the EPAC surveillance network in France. What is the percentage of population that is covered by this network?

**Authors’ answer:**
The information on the coverage of the EPAC network was only available in the Supplementary Data, Table SD1. We added the percentage of adults’ population covered by EPAC in 2004 and 2008 in the “Injury data” section as suggested.

3) I suggest to include the first supplementary figure (the geographical spread of EPAC hospitals in France) in the main article so the readers who are not from France have an idea of the spread at glance.

**Authors’ answer:**
We have included the map in the main manuscript as Figure 1.

---

**Answers to referee Dr Leilei Duan**

We are pleased the reviewer found our manuscript interesting. We hope that this revised version addresses adequately his main concerns.
1. The model-assisted approach mentioned in the paper could be a good reference and example for other researchers to conduct similar studies to estimate the incidence rate of injuries in their districts where they do not grasp the whole picture of injury population.

2. In the 2nd line of the Introduction part, "Each year, a fluctuating number of volunteer hospitals in about ten French cities uses a standardized protocol to record all HLIs treated in their EDs ", might contrast with the content in the 10th line of the first paragraph, "In addition, the analysis of the hospital distribution using yearly admission statistics shows than metropolitan hospitals in the same range as the EPAC participating sample account for 70% of the total number of hospital stays recorded in France (data not shown, see Supplementary Data, figure SD3). We therefore do not think that there is any reason to suspect any major lack of representativeness in our data with regard to the ratio of HLIs to hospital stays." The yearly number of surveillance hospitals are changing, so how do authors draw such a conclusion and confirm the representativeness of hospitals?

Authors’ answer:
We agree we cannot claim that the limited sample of the participating hospitals used in the study is representative of all the French hospitals. To confirm this representativeness was not our aim in the manuscript discussion. We do not actually have any means to verify this assumption, but we only tried to check the sample was not too atypical from the French hospitals size (number of stays) distribution point of view. It is just an indication and not a proof of representativeness. However, results of the cross-validation procedure are somewhat reassuring with robust estimates when the sample changes. To clarify the point, the concerned sentence has been nuanced in the revised manuscript.

3. I would suggest that authors could make a comparison between the model-assisted method in this study and other published estimating methods which are used not only in developed countries but also developing countries

Authors’ answer:
We limited the comparison to the European closely related systems but, in spite of the attempt to standardize the data collection into the IDB framework, comparisons stay unreliable. So it seems to us it was not sensible to pursue further the comparison with extra-European systems. It is nevertheless true that a systematic review of all injury surveillance system in the world - sampling scheme, extrapolation procedures and outcomes – would be a very exciting work…

Answers to referee Pr Marie Ng

We thank the referee for her accurate remarks which helped our manuscript to be improved. We tried to make the appropriate changes.

1. My understanding is that the quantity of interest is the number of HLI (lijk). In the statistical modeling section, the authors noted that the estimation of lijk is based on estimating the ratio, aij, which links the number of cases and number of stay Stijk. To estimate aij, the authors utilized GLMM. It is not immediately clear why it is necessary to estimate first aij when the GLMM model provide direct estimates for lijk. Moreover, with Stijk included as an offset, the relationship between number of cases and number of
stays is already captured. It seems unnecessary to convert the estimates from GLMM back to \( a_{ij} \). then multiply it with stays again to obtain cases. Please clarify.

Authors’ answer:
The quantity of interest is the number of HLI at the population level, \( I_{ij} \), (with the same notations) and not \( I_{ijk} \) (the number of cases at the hospital level). Our goal was then to find a multiplier (\( \alpha_{ij} \)) between \( I_{ij} \) and the quantity \( S_{ij} \), known for the whole population of metropolitan hospitals (simply a rule of three). In other words:

\[
I_{ij} = \alpha_{ij}S_{ij}.
\]

Which is equivalent to:

\[
\sum_{k} I_{ijk}(t) = \alpha_{ij} \sum_{k} S_{ijk} \iff I_{ijk} = \alpha_{ij}S_{ijk} + \varepsilon_{ijk} \quad \text{with} \quad \sum_{k} \varepsilon_{ijk} = 0.
\]

The second member of this equivalence is then the equation 1 in the main text (the working model) which is the general context of the separate ratio estimator of a total population size method (see Levy and Lemeshow, 2008). Because we know \( I_{ijk} \) only for the Epac network sample, the idea was then to estimate \( \alpha_{ij} \) as the marginal expected value of \( \alpha_{ijk} = I_{ijk}/S_{ijk} \) using the Poisson GLMM and then the population level prediction. In the log-link Poisson GLMM context the marginal expectation of the count (or ratio which is of course strictly equivalent with the included offset) are not directly computed (see further answer to 7\text{th} comment).

A further advantage of this presentation of the method is that it is well suited with the “multiplier” approach to extrapolate national number generally used in ED-surveillance system.

2. What is the rationale behind considering orthogonal polynomial contrast for time?

Authors’ answer:
We were interested in allowing general patterns in the dependencies of the ratio on time. Because we had up to 5 equally-spaced measures (in increments of one year), we have thought to use the “natural” ordering of the time. So time has been transformed into an ordered factor along with the natural orthogonal polynomial contrasts for this type of factor. At the end, all the orthogonal polynomial terms were not significant and we simply reverted to using time as a continuous variable in the model (this detail has been added in the manuscript). We hope the further details in the main text will bring clarification.

3. It was mentioned that delta method was used to estimate the aggregated estimates of total number of cases, could the authors perhaps provide some details in the appendix of what exactly was being computed?

Authors’ answer:
We introduced an appendix with the detail computing in the revised manuscript.
4. It was noted that cross-validation was carried out to test the robustness of the model in handling small sample size of hospitals, were the results presented anywhere?

Authors’ answer:
CV results were provided at the end of the “Results/estimates” section and reported in Figure 4 (Figure 5 in the revised manuscript).

5. According to the text, in equation (2), \( \gamma 3j \) is meant to capture the interaction between age and sex. However the subscript \( j \) is defined for age not sex. Was that a typo? To capture the interaction between age and sex, the authors could have used a dummy variable on sex, the notation with subscript, \( \gamma 3j \), according to statistical convention, generally implies the interaction term is treated as a random effect. But it doesn’t seem to be so in this case. Similarly, the sex effect \( \gamma 0i \) could have been captured using a fixed effect dummy.

Authors’ answer:
We thank the referee for highlighting this mistake on the subscript: the interaction coefficient with the third spline component is \( \gamma 3i \), with \( i \) the subscript defined for sex. Moreover, we have of course used a dummy variable on sex, introduce as fixed effect. The subscript 3 was introduced to indicate the interaction with the 3\(^{rd}\) spline components (0 for the main effect) but we agree with the referee the notation could be confused according to statistical convention for hierarchical model. We have consequently modified equation (2) and main text accordingly with less ambiguous notations.

6. How is the AR(1) structure being implemented? Is that a latent model for the hyperparameter \( \sigma 2 \)?

Authors’ answer:
In the context of mixed-models, the GEE approach to correlation has been generalized to model dependence among the within group errors (level 1 residuals in our study). Introducing a correlation structure is technically enough easy with the “MASS” package used to make this work. A specific correlation structure is introduced providing a “correlation” argument (from a set of correlation structure object) as defined in “lme” package. For exhaustive details, please see the reference Pinheiro and Bates’ book (2000).

7. In equation for estimating \( aij \). on page 9, the authors included \( \sigma 122 + \sigma 222 \). It is not immediately clear why the variance of the random effects were being summed? Since the authors were trying to estimate \( (i|jk|bk,bi|jk) \) shouldn’t they be summing \( bk \) and \( bi|jk \)? And if the goal is to obtain the marginal expected value regardless of \( bk \) and \( bi|jk \), the expected value should not be conditional. And the mean over \( bk \) and \( bi|jk \) is simply zero according to the assumed distribution of random effects.

Authors’ answer:
As stated above (answer to the first comment), we indeed aimed to provide the marginal expected value (population level prediction). But in the general framework of a non-linear response, the marginal expectation, given covariates and averaged over \( b_k \) and \( b_{ijk} \), is not simply provided putting
$b_k$ and $b_{ik}$ through 0 (only true in the linear mixed model framework). In the Poisson model with a log link function we can verify that the marginal regression parameters and the conditional parameters are equal except for the intercept which is a function of the random effects variance. Intuitively, it is just the consequence of the scale change with “the exponential of the mean” which is not equal to “the mean of the exponential”. For instance, consider this simple following simulation: generate 1,000 $x \sim N(0, sd = \sigma_1)$, compute $a = \exp(\text{mean}(x))$ (the “conditional”), $b = \text{mean}(\exp(x))$ (the target “marginal”), the ratio $r = b/a$ and repeat this design, say 10,000 times. You can check that the mean of the ratio $r$ on this 10000 replicates is equal to $\exp(\sigma_1^2/2)$. We introduced an appendix with more formal details for computing the equation (3) in the revised manuscript.

8. It would be more appropriate to rearrange the text to put the description of final models in the statistical modeling section rather than result section.

Authors’ answer:
After consultation with other co-authors, it seems difficult to us to replace the “final model” section in the “statistical modeling” one. The final model fitted along with the final choices of model-building are totally dependent on the data used for this study while the “statistical modeling” section gives general and reproducible statistical strategies. In addition, detailing the final choices which led to this final model is certainly important for the good understanding of the reader. Consequently, if the referee agrees, we suggest keeping this section in the result part.

References

Main changes that have been made are the following:

1. **Title page:**
   - **Word count of the main text:** 4,304
   - **Number of figures:** 5 (the numbering of figures has been updated accordingly in the text)

2. **Abstract:** no change

3. **Introduction:** no change

4. **Materials and methods**
   - **Injury data section**
     - P.5, line 64-69: we added ethical approval details; “The personal data collection received formal approval from the French Consultative Committee for the Data Processing in Health Research (“Comité Consultatif sur le Traitement de l’Information en matière de Recherche dans le domaine de la Santé”, approval n°13547) and the French Commission on Individual Freedom and Data Storage (“Commission Nationale de l’Informatique et des Libertés”, CNIL)”
     - P.5, line 75-76: map included in the main manuscript
     - P.5, line 80-82: coverage by the network of the adults ‘population added; “Finally, the adults’ population coverage estimates of the EPAC network fluctuated from 1.5% in 2004 (with 4 hospitals) to 2.7% in 2008 (with 8 hospitals, see Supplementary Data, Table SD1)”

   - **Auxiliary data section**
     - P.6, line 97: information on the ethical approval; “The InVS is allowing accessing permanently PMSI data by the CNIL (authorization n° 902167)”

   - **Statistical modelling section**
     - P.7, line 128: we added further details to clarify the transformation of time in an ordered factor; “To investigate general patterns in the dependencies of the ratio on time (up to 5 equally-spaced measures in increments of one year), time variable has been transformed into an ordered factor with the natural orthogonal polynomial contrasts for this type of factor”
     - P.8, line 146: reference to the added appendix with details on Delta Method computing

5. **Results**
   - **Final Model section**
     - P.9, line 173 and P.10, line 200: equation (2) and (3) modified with less ambiguous notations.

6. **Discussion**
   - **Limits and strength section**
     - P.13, line 282: clarification of the sentence about the representativeness of the sample; “Consequently, although we have no means to conclude absolutely on the representativeness, we do not think there is any reason to suspect a major lack of representativeness in our data regarding the ratio of HLIs to hospital stays”

We added an appendix at the end of the manuscript to detail the computing of variance with the Delta Method and the marginal effects estimates with a GLMM.