Author's response to reviews

Title: Prediction of gastrointestinal disease with over-the-counter diarrheal remedy sales records in the San Francisco Bay Area

Authors:

Michelle L Kirian (michelle.kirian@sfdph.org)
June M Weintraub (june.weintraub@sfdph.org)

Version: 5 Date: 15 July 2010

Author's response to reviews: see over
July 15, 2010

Natalie Pafitis MSc,
The BioMed Central Editorial Team
ditorial@biomedcentral.com

Dear Ms. Pafitis:

Thank you for your email of 15 July 2010 regarding our manuscript (1185628764366661) previously titled "Prediction of gastrointestinal disease with over-the-counter diarrheal remedy sales in the San Francisco Bay Area ". My co-author and I are grateful for the reviewers’ careful consideration and thoughtful suggestions to improve our manuscript.

On behalf of my co-author, I am pleased to submit the attached revised manuscript addressing the reviewers’ comments. On the following pages, I copied each comment in italics, followed by an explanation of my response.

The manuscript is now 3270 words, and has 36 references, 3 tables, and 2 figures. The abstract is 197 words.

Thank you for the opportunity to revise this manuscript.

Sincerely,

Michelle Kirian, MPH
1. In the new title, “Prediction of gastrointestinal disease with over-the-counter diarrheal remedy sales in the San Francisco Bay Area”, please use “sales data” or “sales records”.

   - The title, “Prediction of gastrointestinal disease with over-the-counter diarrheal remedy sales in the San Francisco Bay Area” was changed to,

     **Title:** Prediction of gastrointestinal disease with over-the-counter diarrheal remedy sales records in the San Francisco Bay Area

2. In the abstract, I have a minor suggestion. Please change “did not coincide with weeks without outbreaks more reliably…” to “did not coincide with outbreak weeks more reliably…”

   - “Signals generated by forecasting with the diarrheal remedy sales model did not coincide with outbreak weeks more reliably than signals chosen randomly.” was changed to,

     **(P2, Paragraph 3)** Signals generated by forecasting with the diarrheal remedy sales model did not coincide with outbreak weeks more reliably than signals chosen randomly.

3. p. 6. While I appreciate the efforts to sharpen the language, the phrase “the proportion sales of non-promotional diarrhea remedy to sales of non-promotional drugs for all categories combined…” needs fixing, and simply moving the first “sales” after “remedy” would take of the problem.

   - “Our analysis variable was the proportion sales of non-promotional diarrhea remedy to sales of non-promotional drugs for all categories combined (Diarrheal Remedy Sales).” was changed to,

     **(P6, Paragraph 1)**Our analysis variable was the proportion of non-promotional diarrhea remedy sales to sales of non-promotional drugs for all categories combined (Diarrheal Remedy Sales).

4. p. 8, another minor suggestion: “An outbreak week was any week [when] one or more outbreaks started, ended, or was ongoing.”

   - “An outbreak week was any week where one or more outbreaks started that week or prior to that week but ended that week or later.” was changed to,

     **(P7, paragraph 2)** An outbreak week was any week when one or more outbreaks started that week or prior to that week but ended that week or later.
5. Similarly, consider condensing with “…sensitivity was calculated as the number of outbreak weeks with a signal divided by the total number of outbreak weeks.”
   - “Model sensitivity was calculated as the number weeks where there was a signal and at least one outbreak divided by the total number of outbreak weeks.” was changed to,

   **(P7, Paragraph 2)** “Model sensitivity was calculated as the number of outbreak weeks with a signal divided by the total number of outbreak weeks.”

6. Wording that should definitely be changed is: “sensitivity and specificity calculations were also performed for three sets of randomly chosen dates.” Would it be correct to say “sensitivity and specificity calculations were also performed after replacing the outbreak weeks with randomly chosen dates”?
   - “To further test the ability of the Diarrheal Remedy Sales model to predict outbreaks, sensitivity and specificity calculations performed for three sets of randomly chosen dates.” was changed to,

   **(P7, Paragraph 2)** “To evaluate if model derived alerts identified outbreak weeks more reliably than randomly chosen alerts, sensitivity and specificity calculations were repeated for three sets of randomly chosen dates.”

7. p. 9 Please reorder the description of components of Figure 1 to match the caption and the order of the components.
   - “Plots of Diarrheal Remedy Sales and Gastrointestinal Cases and Outbreaks” was changed to,

   **(Figure 1, Title)** “Plots of Outbreak-Associated Gastrointestinal Cases, Individual Gastrointestinal Cases, Diarrheal Remedy Sales, and Differenced Diarrheal Remedy Sales”

8. p. 10, “the proportion [of] diarrheal remedy sales…” needs the added word.
   - “From July 2003 through December 2007, the proportion diarrheal remedy sales to” was changed to,

   **(P9, Paragraph 2)** “From July 2003 through December 2007, the proportion of diarrheal remedy sales to…”

9. The online edits should be expanded to “Four signals were generated by the Diarrheal Remedy Sales model [on the weeks of] 6/11/06, 1/29/06, 10/15/06 and 6/10/07”.
   - “Four signals were generated by the Diarrheal Remedy Sales model (6/11/06, 1/29/06, 10/15/06 and 6/10/07).” was changed to,
“Four signals were generated by the Diarrheal Remedy Sales model (on the weeks of 6/11/06, 1/29/06, 10/15/06 and 6/10/07).”

10. As noted previously, signals do not have sensitivity or specificity; only tests do. Please use language such as “The sensitivity of model-generated methods for the random signals was identical to their sensitivity to the true outbreak weeks, further supporting...”

- “Three sets of four random signals had identical sensitivity and specificity to those generated by the model, further...”

(P9, Paragraph 4) “The sensitivity and specificity of the model was identical to a random selection of three sets of four signals, further...”

11. p. 11. Similar to the above discussion, please use language such as “The agreement of model-generated signals with outbreak weeks was no better than their agreement with randomly generated weeks.” I stress this for readers who are not familiar with detection theory methods.

- It is not clear what text the reviewer is referring to; however, we believe that our wording regarding the randomly generated alerts is understandable to the reader. One such instance the reviewer may be referring to is, “Signals generated by forecasting with the diarrheal remedy sales model did not coincide with outbreak weeks more reliably than signals chosen randomly.” We would be happy to review this point further if necessary.