Reviewer's report

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

Version: 3 Date: 15 July 2010

Reviewer: Howard S Burkom

Reviewer's report:

There are no remaining major revisions. There are a few minor ones as follows, and the authors dealt with my discretionary one:

Thank you for the careful attention to my comments. My remaining criticisms below are minor, and modifications in my opinion should not require more reviewing.

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”.

In the abstract, I have a minor suggestion. Please change “did not coincide with weeks with outbreaks more reliably…” to “did not coincide with outbreak weeks more reliably…”

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.

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

Similarly, consider condensing with “…sensitivity was calculated as the number of outbreak weeks with a signal divided by the total number of outbreak weeks.”

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”?

p. 9 Please reorder the description of components of Figure 1 to match the caption and the order of the components.

p. 10, “the proportion [of] diarrheal remedy sales…” needs the added word.

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).
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...”

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.

In the authors’ response to my discretionary remark about publication bias, we agree on the interpretation of that term as positive outcome bias, and the rewording makes their logic clearer to me.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.