Reviewer’s report

Title: Time series analysis of the relationship of readily available diarrheal remedy over-the-counter drug sales data and gastrointestinal disease in the San Francisco Bay Area

Version: 1 Date: 7 June 2010

Reviewer: Howard S Burkom

Reviewer’s report:

I have called for "major compulsory revisions" because I think the requested changes are important, but they should not be burdensome. I would like to see more detail about the technical results, but depending on these details, no more results are necessarily needed.

First, here are answers to the BMC reviewer questions regarding the manuscript “Time series analysis of the relationship of readily available diarrheal remedy over-the-counter drug sales data and gastrointestinal disease in the San Francisco Bay Area”:

1. Is the question posed by the authors well defined?
   --Judging from the background, discussion and conclusions, the authors’ main question is the “validity and representativeness of over-the-counter drug sales data for prospective outbreak detection” (bottom of p. 2), about the utility of OTC data, not about time series analysis. The reader is given few details of the time series analysis, and requested details are listed below.

2. Are the methods appropriate and well described?
   --The methods are not sufficiently described, but from what I can understand, they seem appropriate.

3. Are the data sound?
   --The data do give information about OTC sales. The authors did not collect them directly but describe them sufficiently and acknowledge some problems.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
   --Does not apply.

5. Are the discussion and conclusions well balanced and adequately supported by the data?

6. Are limitations of the work clearly stated?
   --For 5 and 6, generally yes, though a few modifications are requested below.

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
Yes, the citations are fine.

8. Do the title and abstract accurately convey what has been found?
--The title is awkward and should be more descriptive of the text, which says little about time series analysis. The abstract needs some modification.

9. Is the writing acceptable?
--Generally yes. Corrections listed below mostly minor.

*** Major Compulsory Revisions (which the author must respond to before a decision on publication can be reached):

Two types of revision seem necessary: better organization of the ideas regarding data utility, and a more detailed discussion of the analysis and results evaluation.

1. Data utility, and significance of findings: It is important to make the distinction between evaluation of utility of NRDM data and of OTC data in general. On p. 2, line 3, the authors overstate when they say they “evaluated the ability of readily available diarrhea remedy over-the-counter drug sales data to predict”. Their later statements are clearer on this point, especially the first full paragraph on page 7, beginning with “Our study was not able to assess whether improvements in over-the-counter drug sales reporting systems might improve the performance of this type of syndromic surveillance. Systems that include more complete information on retailer participation and more detailed sales data might perform better than the system that was available to us…” From my own projects and from talking to users at conferences, I am familiar with the weaknesses of the NRDM dataset. What I am asking is that the authors make this distinction more prominent in the early part of the paper and also discuss it in the Limitations section. Indeed, the Public Health Agency of Canada has reported better results with their own OTC data.

More specifically, the statement that “over-the-counter drug sales did not coincide with diarrhea due to other [beside Norovirus] etiologies” reminded me of my experience evaluating records of prescription drug sales. In that project, no one could demonstrate any correlation between GI-related prescriptions and illness outbreaks until a physician investigated the products included and removed many of them for improved specificity, and then correlations were found. Purchasers of NRDM data cannot modify the data in this way.

I agree with their statement that “the ability of over-the-counter syndromic surveillance to enhance the detection of waterborne disease outbreaks has not been adequately demonstrated”. However, they also point out that “Our study was not able to assess whether improvements in over-the-counter drug sales reporting systems might improve the performance of this type of syndromic surveillance”, and they should not elsewhere extrapolate their findings to OTC data in general. Judicious reorganization and some rewording are needed.

2. More detailed discussion of results and evaluation

The authors’ discussion seems to confuse forecasting and detection. On page 2 is the statement, “Signals generated by forecasting with the drug sales model did
not perform better than signals generated randomly”. How do signals “perform”? Do the authors mean that random signals match the reported outbreaks as well as time series anomalies based on forecasting? To me, this finding means that the time series collected are not what should be monitored. The related discussion on page 6 also needs clarification. Please explain.

Explain how sensitivity and specificity were defined to clarify the statement “The sensitivity and specificity of these alerts were calculated by week”.

I had to read the statement “Forecasting with the univariate over-the-counter drug sales model produced four signals” several times. I think that it means that there were 4 instances where the observed value exceeded the forecast value by more than the upper 95% confidence limit. Is this true? These explanations should not require much space.

*** Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

p. 2, title: the phrase “diarrheal remedy over-the-counter drug sales data” is awkward in the title and in the text. One suggestion is “over-the-counter sales records of diarrheal remedies”.

p. 3, “may include as few as one or two cases”
--The most liberal definitions of outbreaks require at least two cases—please modify the text to read “as few as two cases”.
--The statement “Data were received from three adjacent counties and transmitted in electronic formats” should be rewritten as “Data were transmitted in electronic formats from three adjacent counties”. If this is not correct, then a better explanation is needed.

“Only cases with residence in the respective county were sent by each county…”
--Wouldn’t it be more correct to say “were requested from each county…”?

--Please change “receives outbreaks records” to “receives outbreak records” or “receives records of outbreaks.”
--In the data descriptions, the general reference to “data” is too vague and should be clarified with references to data records and their fields. In particular, please replace the clause “gastrointestinal case and outbreak data were aggregated into a single dataset” with more detail.

“drug sales data are divided into 18 categories based on common use, form and whether intended for adult or pediatric populations..”
--Does this description mean that the categories are such as “adult/pills”, “child/liquid”? Please give examples. This sentence needs to be split into 2 sentences.

"Data distinguish between promotional and non-promotional sales”
-- How do the data make this distinction?

p.5, Please add the missing word in “From July 2003 through December 2007, the proportion diarrheal remedy sales …“
p. 6, As in the title, I suggest replacing “NRDM diarrheal remedy over-the-counter sales data” with “NRDM over-the-counter sales data for diarrheal remedies”

p. 7, Please correct the grammar in “the incidence of respiratory illnesses… is likely to greatly exceeded that of diarrheal illnesses”

p. 8 The Conclusions section says that “we cannot exclude the possibility that this system may[]be able to detect larger outbreaks”. No system has been described. What do the authors mean?

*** Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

p. 7 “The number of surveillance systems and lack of published reports on over-the-counter drug sales monitoring systems, and NRDM specifically, suggests publication bias may be present.”

To me, this reference to “publication bias” suggests that more successful applications exist but are not published. Is this interpretation correct? I would have expected bias in the other direction, with the successful surveillance systems more likely to publish. Having spoken to multiple users, I find it much more likely that the combination of understaffed user departments and weaknesses in the data has resulted in a reduced number of evaluations, successful experiences, even regular usage. The authors may wish to revise the statement based on experience.

**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.