Author's response to reviews

Title: Modeling Emergency Department Visit Patterns for Infectious Disease Complaints: Results and Application to Disease Surveillance.

Authors:

Judith C Brillman MD (jbrilman@salud.unm.edu)
Thomas Burr PhD (tburr@lanl.gov)
David W Forslund PhD (dwf@lanl.gov)
Edward Joyce PhD (joyce@lanl.gov)
Richard Picard PhD (picard@lanl.gov)
Edith T Umland MD (eumland@salud.unm.edu)

Version: 2 Date: 21 December 2004

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

We have revised our manuscript per the reviewers’ comments. Note especially a large effort to edit for length while still responding to other requests for more description/information. We responded to each reviewer, point-by-point below, to indicate where/how changes were made.

Regards,
Tom Burr (for Judith Brillman)

Reviewer's report
Modeling Emergency Department Visit Patterns for Infectious Disease Complaints: Results Title: and Application to Disease Surveillance.
1 30 October 2004 Version: Date:
Kenneth Mandl Reviewer:
Reviewer's report:
General
The authors have responded well to previous criticisms, producing a paper that nicely describes the B-SAFER approach to outbreak detection, comparing and contrasting it with the approach taken by similar systems.

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

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
The authors should explain whether the GI outbreak described under the “Surveillance” section was confirmed with clinical or public health follow-up data or whether it was presumed based on the surveillance system findings.

This was eliminated

Discretionary Revisions (which the author can choose to ignore)

Reviewer's report
Modeling Emergency Department Visit Patterns for Infectious Disease Complaints: Results Title: and Application to Disease Surveillance.
1 6 December 2004 Version: Date:
William Lober Reviewer:
Reviewer's report:
General

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

Methods
1. The key word coding strategy should be clarified. Is negation correctly handled?
Negation discussed
The correspondence of the syndrome groups (following your expert review revisions) with those of ESSENCE or RODS could be made clear (optional).
The correspondence was clarified in the Data Stream section via:
We group daily CC counts into seven categories: respiratory, gastrointestinal (GI), undifferentiated infection (UDI), lymphatic, skin, neurological, and “other” (Table 1). The “other” category includes all visits except those in the first six categories. These grouping categories are also used by other surveillance programs, such as ESSENCE (Walter
Reed Army Institute of Research) and the Real-time Outbreak Detection System (RODS, University of Pittsburgh) [11, 13, 14]. Our grouping scheme is provided in Table 1.

Results
2. Surveillance - Figure 5 is great, but the text in the first paragraph following surveillance is confusing. Figure 5a does not appear to be daily counts, but instead the predicted counts. *Figure 5a now contains both predicted and actual counts.*

It would be nice to superimpose actual with predicted counts.

*Done.*

The "note the patterns…” sentence is unclear.

*I agree; it’s gone.*

Likewise, please make clear whether your comments about March, Jan, and Feb refer to the Forecast Error or Page’s Statistic, and the implication of these to observable events.

*Clarified. A paragraph was added which explains the relationship of these peaks to other influenza indicators.*

Discussion
3. The authors describe early detection of a respiratory outbreak (influenza) in advance of traditional mechanisms, but don’t present any data from the statewide sentinel surveillance system or from the stat lab. A detailed discussion is beyond scope, as the authors state, but some further characterization of the data to support this finding would be valuable.

*These other data streams are now briefly mentioned in the Methods section and qualitative findings from these other data streams are given in the Results section.*

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

Background
1. P1 last sentence – “might include” instead of “might be”

*done*

Results
2. Surveillance – second sentence – you mean “we focused…” or “we will focus…”? Not clear if you mean the work you did, or the results you are about to describe.

*Corrected and clarified*

Discussion
3. “sophisticated software codes” should be “sophisticated software”, or perhaps “sophisticated software code”?

*Correction to “code” made.*

4. Free text c/c paragraph “settings” instead of “setting”

*Corrected*

Discretionary Revisions (which the author can choose to ignore)

Background
1. P1 - Consider using “traditional” instead of “existing” public health surveillance, as some real time and near real-time systems have been deployed.

*done*

2. P2 – Variability may be due to more than requirement that ED’s see all patients presenting for care.

*Agree, changed to “due to large variability that arises in part because...”*

3. P4 – Its interesting to consider if syndromic surveillance is better described at active or passive.

One random set of definitions here [http://www.emro.who.int/Publications/EMHJ/0201/06.htm](http://www.emro.who.int/Publications/EMHJ/0201/06.htm)

Methods
4. Data Stream – why is the menu option for C/C rarely used?

*Clarified in text.*

5. First Order Model – could explain more about why started log scale useful in this setting (you did a fine job with what the scale is).
Done, and to clarify here:
The log scale makes the variance approximately constant, not dependent on the mean level, so analysis is simplified, and the resulting confidence limits on the original scale are wider when the expected count is higher.

The statistical distributions of most count data are skewed. That is, their histograms tend to be "stretched" to the right, which complicates model fitting. Also, a variability in counts tends to increase as a function of the count itself. By using started logs, results are more Gaussian and the usual metrics such as standard deviations are more easily interpretable. As these issues are more in the domain of statistical subtleties and not central to the main message of the paper, we comment only minimally on the related issues in the text.

A "started" log means to use the logarithm of 1 plus the count, to handle the case of 0 counts because the logarithm of 0 is negative infinity, which causes problems.

Results
6. Long term trends – specify the % population increase corresponding to the 4% visit increase, if you want to tie one to the other.
Done
7. Long term trends – are you saying that the model requires a new term in place of the C8, C9, term to accommodate non-linearity, or that adjusting C8,C9 would compensate?
Clarified in text, and:
A long term linear trend (nonzero value for C9) is easily accommodated. However, if that trend is present due, for example, to changing business practice, then it is subject to change. This would be compensated for by monitoring the residuals for patterns and adjusting C9 on an as-needed basis.

8. Use of Free Text chief complaints – this argument of c/c vs ICD-9 is reasonable, but the key word "grouping scheme" and its performance are not discussed in any detail. (see #1 above). Comparison of the categories with other groups’ work, at a minimum, would facilitate discussion here as to how the approach may be generalized. The argument could be strengthened by reference to other efforts to code and cluster c/c’s,
done
and by a discussion of why coding is not used at UH (end of second paragraph).
I’m not clear on the significance of discussing why coding is not done at UNM. I don’t think chief complaints are generally coded for purposes other than surveillance. That is why there is no standard chief complaint coding scheme.

9. The day of week section notes a striking difference between the 25% Mon increase at UH and a 3.4% increase in NYC, without commenting on possible explanations. Perhaps they are environmental? Behavioral/cultural? Statistical or other artifact?
The text has been modified to eliminate, we hope, the confusion. It's not surprising that the actual magnitude of the effect changes depending on the setting.

After editing the latest revision for length, we replaced the following 3 paragraphs:

Day-of-week patterns in EDs for potential infectious disease by individual body system classified CCs have been previously reported in the literature. Suyama [4] noted that for all infectious-disease-related ICD-9 discharge diagnoses codes, there was a Monday peak in mean number of cases (24) and a Saturday nadir (19). This is consistent with our data for most of the syndromes. However, as we have shown, there is day-of-week variability within infectious disease syndromes which is obscured by the failure to consider each body system’s pattern individually. Moshtashari also reported on a seven day a week system, New York City’s ambulance runs [37]. Data reported for respiratory related complaints demonstrated that Monday calls exceeded Saturday calls by 3.4%. This is the same day-of-week pattern that we have identified although the average relative difference is almost 25%.

ED utilization for all visits has been found to peak on Sundays and trough on Thursdays in a pediatric Emergency Department [27]. However, we did not analyze the subgroup of pediatric visits in our data. It is possible that visit patterns for the pediatric population may differ from that of adults who make up the bulk of our sample. In Israel, it is reported that ED visits peak on Sunday and fall off during the week [38]. In Israel, Sunday is the first day of the work week.
Other day-of-week patterns are described for respiratory ICD-9 discharge codes in a setting where the bulk of care is provided 5 days a week, rather than seven. In an HMO, respiratory complaints peaked on Mondays [26]. There was a smaller peak on Fridays and a mid-week nadir. It is important to understand utilization patterns on weekends, which ED data provides, as the effects of bio-terrorist or natural outbreaks are unlikely to be limited to weekdays.

with this one paragraph:
Day-of-week patterns in EDs have been previously reported in the literature [4,26,27,37,38]. Although the magnitudes of the day-of-week effects vary depending on the setting, the first day of the work week typically exhibits the greatest number of events. And, as we have shown, there is day-of-week variability within infectious disease syndromes which can be obscured by the failure to consider each body system’s pattern individually. It is also important to understand utilization patterns on weekends, which ED data provides, as the effects of bio-terrorist or natural outbreaks are unlikely to be limited to weekdays.

Also, we’re not including this in the text, because it is anecdotal, but we think: Many people get ill over the weekend. They wait until Monday to call their physicians, only to be told that the physician can’t work them in to the schedule. The patient is then referred to the ED. This pattern for acute medical illness is different than for acute traumatic injuries, where people tend to arrive shortly after the injury.

10. Surveillance - You make the comment that visual overlay of the data streams on the baseline plot is useful operationally. This is a significant generalizable claim, and could be supported with a figure. This is not compulsory, but would greatly strengthen the paper.
Actual counts added to figure 5a and Figure 6 added
11. Surveillance P2 – Why was the 2.5% false alarm rate chosen? There was no prior mention of selection of a false alarm rate or discussion of the tradeoffs (ROC for this detection algorithm? Empiric?).
The 2.5% false alarm rate was an arbitrary, reasonable choice corresponding to an upper control limit at two times the prediction error standard deviation above the baseline prediction.
12 – Surveillance – P2 – There was mention that demographic information would aid outbreak investigation. Are you suggesting that this information be collected as part of routine surveillance, or are you suggesting that these elements would be obtained from the ED log as part of an investigation?
Eliminated
Discussion
13. The language in the second paragraph is unclear, and obscures the conclusions.
We added an opening sentence to set the stage and clarify the goal of the paragraph.
It is not clear why “other” and “average” are quoted, for example. “Other” refers to the c/c category of that name, but most of the text in this paragraph appears to be presentation of the results for modeling that category, rather than a discussion of significance of the results for all categories.
Regarding "other," the manuscript has consistently quoted "other" (see the first usage in Part3 of Method) to avoid confusion with the often used English word, other. "Other" is a catch-all category
It is true that one sentence was devoted to the "other" category, because the "other" counts were available to us for the same time range and the text needed to say that.
How would the “average” year be constructed – average of all previous years? Last year only? Last n years?
Regarding "average," this was not a numerical average by day of year, but instead was a least squares fit, resulting in estimates for coefficients C1 to C11. That is why “average” was quoted -- it refers to the “one-size-fits-all” that is described in the Discussion section. We revised the last sentence to clarify, and avoided the use of “average.”
Conclusion
14. The data presented only supports the day of week effects “by infectious disease in various body
systems” and I think actually only supports the day of week effects in grouped chief complaint categories.
Correct; this was changed.

Reviewer’s report
Modeling Emergency Department Visit Patterns for Infectious Disease Complaints: Results Title:
and Application to Disease Surveillance.
1 26 November 2004 Version: Date:
Karl Ekdahl Reviewer:
Reviewer’s report:
General
This paper gives details on a model for emergency department (ED) visits to be used as a timely surveillance tools for unexpected events. The strength of the model (in its specific setting) is that it is based on a very long time series of historical data (1994 to 2002), which may of course hamper the immediate application of the model in other settings. With increasing focus on bioterror, the work is important as a good example of a timely surveillance system.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)
I do not have any major objections or comments to the article, more than that I find it too long. I believe that it could be shortened by at least 1/4, without compromising the content or the message.

We had conflicted reviews here. In our first review, we were asked to add more, and in this review we were asked to add more by one of the reviewers. The other reviewers did not have concerns about the length and this reviewer does not provide guidance as to what is not necessary. Still, we agree that cuts can be made and have done so. The revised text is less “wordy,” with minimal repetition, so we think that editing for length reduction was beneficial. Specifically: the version that the referees are commenting on was 8153 words. We first added the requested information/clarifications and increased the length to 8452 words. We then edited for reducing length and the submitted version has 7155 words (15% reduction from 8452)

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
The abbreviation ED should be spelled out the first time it is used in the manuscript.
Done

Discretionary Revisions (which the author can choose to ignore)