Author's response to reviews

Title: Evaluation of school absenteeism data for early outbreak detection, New York City.

Authors:

Melanie Besculides (mbesculides@mathematica-mpr.com)
Richard Heffernan (rheffern@health.nyc.gov)
Farzad Mostashari (fmostash@health.nyc.gov)
Don Weiss (dweiss@health.nyc.gov)

Version: 2 Date: 9 August 2005

Author’s response to reviews: see over
Dear reviewers,

Thank you kindly for the time and effort you spent reviewing our paper. We were very fortunate to have reviewers with a such deep understanding of syndromic surveillance provide feedback on our work. We have addressed each of your comments point by point below. For the purposes of clarity, we have included your original points in italics before our reply. We have also added line numbers to the paper so that you can easily find where we added text to incorporate your comments. We hope that we have sufficiently addressed all of your suggestions in the revised paper. Many thanks.

Reviewer 1 Howard S Burkom

A) The paper’s abstract contains the statement that spatial analysis identified too many clusters to be useful. What the results actually showed was that the chosen implementation of spatial scan statistics is not practical for prospective monitoring with this data; leaping to a broad conclusion about spatial analysis is not justified.

This is a valid point and we revised the wording in the abstract (lines 48-49) and conclusion (lines 282-286) to reflect this comment.

B) These absenteeism data sets have limited usefulness for 2 reasons: (a) They are very nonspecific and only partially linked with illness. (b) The number of useful data days is severely limited by the school calendar, especially considering that about 20% of scheduled days should be ignored because they are close to holidays. From the discussion beginning Despite the removal of outliers, the authors imply that even more days should be removed to get purely calendar-related absences out of the data. The conclusion I would state is that much detailed analysis involving modeling and filtering of the data would be required to reduce the noise and, given (a) and (b), this analysis would be hard to justify for a health department with limited resources.

We agree with these limitations and expressed them more clearly in the conclusion (lines 275-277).

C) Any sweeping statement about the overall utility of absenteeism would require investigation of the hypothesis, stated in the full-text conclusion, that this data source can give earlier signals than a source based on acute clinical care, and such an investigation is impossible unless further modeling and data filtering can remove some of the noise. The recent PLoS Medicine paper of Martin Kulldorff, with some of the same coauthors as the present work, has ideas that could be brought to bear. First, with 1160 participating schools, many with chronically high absenteeism or erratic reporting could be removed from the analysis and still preserve spatial coverage. The PLoS paper recommends a similar strategy to cope with missing data in some cases, and this strategy has avoided some excess clustering in ESSENCE absenteeism data. Second, the removal of space-by-day-of-week interaction could be relevant to this data source. Indeed, many noise-canceling approaches based on detailed knowledge of the school system might be tried if there was evidence that the surveillance value of the data is worth the effort. These approaches could improve the temporal, city-wide results as well as the spatiotemporal ones.

We toned down the statement to be less sweeping (lines 275-277). We agree that additional efforts could be made to adjust the data so that it may become more useful for outbreak detection. As the reviewer suggested, we could remove schools with persistently high rates of
absenteeism from the dataset, however, given the limitations inherent to the data (reason for absence not available, lack of data for full calendar year) and the fact that superior data systems are already in place, we do not feel that additional data manipulation is worth the effort. Instead, we are working to collect and analyze better data that includes chief complaint.

D) In the abstract and full-text conclusions, avoid unjustified claims and point out that extensive modeling is required for enough noise reduction to establish the potential utility of absenteeism data in this school system.

We revised the abstract and conclusion to address this (see comments A-C)

E) p. 3: The authors correctly note that findings from one data stream can support or contradict. This remark should temper the previous statement. They might say that Since findings from one data stream can support or contradict findings from another, multiple systems may increase or decrease detection performance which is one good reason why their health department is not using algorithm results from absenteeism data.

We agree and revised the text accordingly (lines 56-59)

F) p. 5-6: The CuSum reset criterion is nonstandard and did not eliminate multi-day signals due to a spike on one day that was extreme enough to cause signals for 2 or 3 consecutive days. This is a weak condition that will let very high rates push later ones over the threshold until the outbreak has nearly fully subsided. By contrast, CDCs C3 CuSum algorithm does not add any sums from previous days that have flagged. The problem with the weaker condition is that if the system keeps alarming on a long down slope, sensitivity to secondary events is lost. Please just modify the explanation of this criterion.

We found our modification of CuSum—terminating signals when the percent absent returned to within 0.5 standard deviations of the baseline mean—to be useful. The artifact signals that nearly always occurred on the two days following an extreme spike obscured the true signals we found in this retrospective analysis and are an unnecessary distraction when conducting prospective analysis. We have revised the explanation to more accurately state that this “reduced the number of multi-day signals…” (line 123-125).

G) p. 9: The final sentence in the paragraph about school A is Still, 59 schools involved in clusters had single-day excesses of 200. I’ve read the paragraph several times and still don’t see why that sentence is there. Please clarify others may not get it either.

In this statement we were trying to convey that an excess of the magnitude observed was not unusual. We rewrote the statement to be more clear (line 205-208).

H) p. 10: The statement Comparing these data to influenza surveillance is a good way is vacuous, and the authors have a valid point to make. A more descriptive phrasing such as: Trying to correlate peaks in daily time series from these data with documented rises in influenza test positives is an evidenced-based approach for testing the effectiveness would make this point.

We agree and have incorporated the suggested wording (lines 219-221).

I) p. 3: Correct the spelling of complementary.

We have fixed the spelling (line 56).
J) Reword for clarity: We calculated the median daily absentee rates among elementary/middle schools and separately among high schools and examined whether the absentee rates differed between these two groups.

We agree and reworded this sentence to be more clear (lines 90-92).

K) p. 5: Refer to the Wilcoxon signed-rank test

We do mention the use of the Wilcoxon signed-rank test on page 5 (line 98).

L) The clause we analyzed daily data retrospectively should be explained for those unused to this methodology. Make clear that the analysis was done to reflect the results of prospective monitoring (without knowledge of future data).

We agree and added this point of clarification (lines 103-105)

M) p. 3: On my first reading of the statement Each of these syndromic systems relies on non-specific, pre-diagnostic data, I was tempted to object and replace systems with data sources. Though they are certainly different systems on the front end, the data sources may be treated as a single system for decision-making purposes. But this is a matter of bias, and it has not been proved that data fusion improves public health surveillance capability. However, in a unified system, the decision given in the papers final statement that prospective surveillance of school absenteeism will not be implemented in New York City might be different because the absentee data could be used to corroborate or add further information to indications from other data sources. For this paper, the authors should just be deliberate in their choice of systems or data sources in that statement.

We agree that absenteeism data may be useful for corroborating what is found in other syndromic systems. However, given the limitations of the system and the effort required to manipulate the data, we feel that resources are better spent on obtaining higher quality data.
Reviewer 2 David Buckeridge

A) For the time-series analysis, it is not clear what portion of the data-set (October 1, 2001 to June 25, 2004) the authors used to fit their linear regression model for forecasting. This ‘training’ interval should be distinct from any ‘testing’ interval to ensure that the analysis is a reasonable model of prospective surveillance.

We have clarified that, although retrospective, our analysis mimicked daily prospective analysis. Each day’s analysis was based on all data starting with September 15, 2001 up to and including the day of analysis (lines 104-106).

B) The authors go to extraordinary lengths to ‘minimize false positives’ in the time-series data (it appears that similar steps were not taken with the space–time data). As a first step, they removed outliers, which included nearly 20% of the days. Twenty percent of values cannot be outliers, by definition. Did the authors remove these values from both the training and testing data? The authors should provide a clear description of (1) how and why they removed these values, (2) how the removal of these values might impact their analysis, and (3) the utility of absenteeism data for prospective surveillance given that nearly 20% of values are ‘outliers’.

We agree that 20% is a significant number of days to remove from a dataset. The days were removed because the increase in absenteeism on these days was explained by factors unrelated to illness. By including days with known explanations for absence the data were even more noisy than what we report in the paper. Days were removed from both the training and testing periods.

We revised the text to more clearly state the reason for removing days (lines 111-112). We changed the term ‘outlier’ to ‘extreme, uninformative data points’ throughout the text. We also revised the conclusion to more clearly describe the limitations of using school absenteeism data for prospective surveillance (lines 275-277).

C) Another step the authors took was to modify the cumulative sum applied to forecast residuals. The specific modifications appear arbitrary. The authors should explain why they modified the cumulative sum in the manner that they did and how those modifications are likely to affect their results.

See response to Reviewer 1 comment F, above, and lines 123-125.

D) The results of the spatial analysis are striking. To obtain 3 significant clusters each day, on average, something must be amiss with the space–time model. The authors provide few details about how they processed the space–time data, but it appears that there was no attempt to model baseline variation (which the authors do not describe) in the space–time data (e.g., accounting for a region by day-of-week effect). It is not clear that this modeling would help, but the authors should describe any variation that does exist in the data and explain why they are using a trailing average as a space–time forecast model if significant variation does exist. In any event, I doubt that the observed values are Poisson distributed (this would be easy and useful to check), and if they are not, then applying a spatial scan with a Poisson model is not warranted.

We agree the results are striking. This is an interesting area for future research by a statistician, but is not within the scope of this paper.

E) Page 3, third paragraph – The final version of the report cited (6) is now published.
We have updated the citation (lines 331-334)

F) Page 5, last paragraph – Was the linear regression identical to the one used before? In the spirit of reproducibility, it would be helpful to say a few more words about the final model.

Details about the model have been added (lines 114-117).

G) Page 7, second full paragraph – What was the significance level / threshold and how was it selected? Are your results sensitive to the choice of threshold?

Aside from the modifications mentioned in lines 119-126, we followed CuSum convention as implemented in CDC’s EARS system (Hutwagner, reference 8) and have added this reference and specific mention of C1, C2 and C3 signals in lines 126-127.