

# A Comparison of Ambulatory Care and Emergency Department Encounters as Data Sources for Detection of Clusters of Lower Gastrointestinal Illness

W. Katherine Yih,<sup>1</sup> Allyson Abrams,<sup>1</sup> John Hsu,<sup>2</sup> Ken Kleinman,<sup>1</sup>  
Martin Kulldorff,<sup>1</sup> Richard Platt,<sup>1</sup>

<sup>1</sup>Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care,

<sup>2</sup>Kaiser Foundation Research Institute

## OBJECTIVE

We sought to compare ambulatory care (AC) and emergency department (ED) data for the detection of clusters of lower gastrointestinal illness, using AC and ED data and AC+ED data combined, from two geographically separate health plans participating in the National Bioterrorism Syndromic Surveillance Demonstration Program [1].

## BACKGROUND

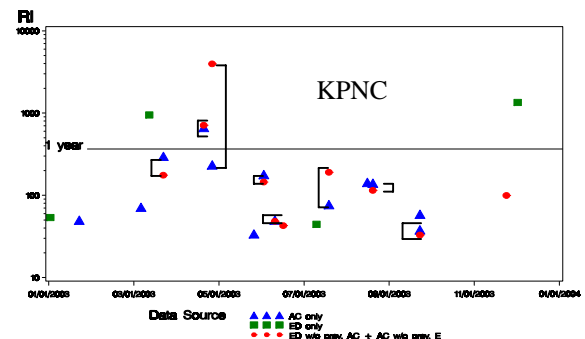
A variety of electronic health event data sources have been proposed and used for the early detection of disease outbreaks. While there is some information available about the utility of these data sources [2,3], few formal comparisons have been made among them.

## METHODS

We analyzed historical counts of new episodes of lower gastrointestinal illness (LGI) registered in both AC and ED settings at Kaiser Permanente Northern California (KPNC, population 3 million) and Harvard Pilgrim Health Care-Harvard Vanguard Medical Associates (HPHC-HVMA, population 140,000) in 2003. Encounters (extracted based on diagnostic codes from automated source databases used routinely in the provision of care) were classified as LGI if a diagnosis code in the CDC's LGI syndrome was assigned for the encounter. For each data stream—AC, ED, and AC+ED combined—repeat visits were omitted if they occurred within 6 weeks of a previous visit to the same setting for LGI. Generalized linear mixed models using 2002 KPNC data and 2000-2002 HPHC data were employed to detect unexpectedly high numbers of cases within patients' zip codes of residence [4]. For KPNC, 787 zip code areas were modeled, for HPHC-HVMA, 222. The models adjusted for day of the week, holidays, month, and seasonal effects, generating a SMART score for each cluster detected, equivalent to the recurrence interval (RI) or the expected number of days of surveillance needed for one such cluster of at least the observed magnitude to occur in the absence of any actual outbreaks. The SMART scores identify signals in individual zip code areas each day, without regard to events in adjacent areas or on consecutive days. We considered all signals that would not have occurred by chance alone more than once a month ( $RI > 30$  day).

## RESULTS

For KPNC, there were 13, 4, and 10 signals in AC, ED, and AC+ED data-streams, respectively. For HPHC, the respective numbers were 3, 7, and 14. For KPNC, 8 AC signals occurred on the same date and in the same zip code area as signals in the AC+ED data. For HPHC, 2 AC and 2 ED signals matched AC+ED signals in this way. There were no such matches between AC-only and ED-only signals for either site.



## CONCLUSIONS

We observed no overlap between signals in AC-only and ED-only data. Thirty percent of signals were identified only by combining AC and ED data. These data sources appear to contain independently useful information for syndromic surveillance. Correlation with confirmed outbreaks will be necessary to understand the public health significance of each of these data sources.

## REFERENCES

- [1] Platt R, Bocchino C, Caldwell B, Harmon R, Kleinman K, Lazarus R, Nelson AF, Nordin JD, Ritzwoller DP. Syndromic surveillance using minimum transfer of identifiable data: the example of the National Bioterrorism Syndromic Surveillance Demonstration Program. *J Urban Health* 2003;80(Suppl. 1):i25-i31.
- [2] Rodriguez C. Can emergency department chief complaint surveillance predict outbreaks of bacterial enteric illness? Presented at Syndromic Surveillance Conference, Boston, MA, November 4, 2004.
- [3] Yih WK, Abrams A, Danila RN, Green KL, Kleinman K, Kulldorff M, Miller BD, Nordin JD, Platt R. Ambulatory care diagnoses as potential indicators of outbreaks of gastrointestinal illness—Minnesota. *Morb Mort Wkly Rep* 2005 (in press).
- [4] Kleinman K, Lazarus R, Platt R. A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism (with invited commentary). *Am J Epidemiol* 2004;159:217-224.