ABSTRACT

Gastrointestinal disease outbreak detection using multiple data streams from electronic medical records

SK Greene[1], J Huang[2], AM Abrams[1], M Reed[2], R Platt[1], and M Kulldorff[1]

[1]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA; and [2]Kaiser Permanente Division of Research, Oakland, CA, USA
E-mail: Sharon_Greene@hphc.org

Objective

To identify which syndromic and laboratory-based data streams from electronic medical records (EMRs) may be used to detect gastrointestinal (GI) disease outbreaks in a timely manner.

Introduction

GI disease outbreaks can be focal (for example, restaurant associated), generalized (for example, seasonal rotavirus increases) or intermediate (for example, widely disseminated contaminated commercial products). Health departments (HDs) are commonly notified of focal outbreaks by passive reporting, whereas generalized outbreaks in non-institutional settings are seldom reported as clusters. Intermediate outbreaks are often detected via laboratory testing, which may be subjected to backlogs and delays. Healthcare systems routinely collect in EMRs clinical data related to GI disease, such as ambulatory care diagnoses, that could be exploited for surveillance.[1] Multiple syndromic and laboratory data sources could potentially be used to prospectively detect generalized and intermediate GI disease outbreaks for situational awareness and possible epidemiological investigation.

Methods

Kaiser Permanente Northern California (KPNC) is a large integrated health care delivery system utilizing EMRs. Zip code-specific daily episode counts in 2009 were generated for 22 data streams: 9 streams of ICD-9 codes for upper and lower GI diagnoses in ambulatory care, emergency department, and inpatient settings; four streams of ICD-9 codes in association with an antibiotic prescription; one stream of text strings for hospital admissions; one stream of prescription antidiarrheal medication dispensings; one stream of stool culture tests ordered; and six streams of tests positive for GI pathogens.

We mimicked prospective surveillance of these data streams with univariate cylindrical space–time permutation scan statistic analyses,[2] using only data that would have been available as of each surveillance day. 'Alerts,' that is, cluster(s) unlikely to occur by chance alone in a 1-year period, were evaluated and will be compared with GI outbreaks from preliminary or final reports to the state HD in non-institu-tional settings affecting any of the 16 countries for which KPNC had ⩾10% population coverage. *Salmonella* and *E. coli* clusters were identified in retrospective analyses, and genetic relatedness of isolates in these clusters will be assessed.

Results

Analyses of 22 EMR data streams yielded 35 unique alerts. Seven streams did not contribute to alerts, including two streams of ICD-9 codes and an antibiotic prescription and five streams of positive tests. In total, 25 GI outbreaks (six laboratory confirmed) were reported to the state HD for the study area in 2009; the median number of recognized cases was 10 (range: 3–280). Spatio-temporal overlap, if any, between alerts in electronic data and known outbreaks will be characterized. Three clusters each of *Salmonella* and *E. coli* were identified in EMR data; the genetic relatedness of isolates in these clusters will be described.

Conclusions

We plan to identify which streams may be most useful for GI outbreak surveillance. Alerts could reflect previously unknown (particularly generalized or intermediate) outbreaks, in contrast with outbreaks reported to the state HD, which generally had small numbers of recognized cases and were highly localized. Prospectively identifying clusters of tests positive for specific GI pathogens may be a promising complement to syndromic surveillance, as isolates in clusters could be prioritized for testing, potentially reducing the time to outbreak detection.

References

1 Yih WK, Deshpande S, Fuller C, Heisey-Grove D, Hsu J, Kruskal BA, et al. Evaluating real-time syndromic surveillance signals from ambulatory care data in four states. *Public Health Rep* 2010;**125**:111–20.
2 Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med* 2005;**2**:e59.