

# Criteria for Prioritizing Statistical Anomalies Identified in BioSense

Colleen Martin, MSPH<sup>1</sup>, Jerome Tokars, MD MPH<sup>1</sup>, Gabriel Rainisch, MPH<sup>2</sup>, Jacqueline Burkholder, PhD, MS<sup>2</sup>, Kali Crosby, BSN RN<sup>2</sup>

<sup>1</sup>Centers for Disease Control and Prevention, <sup>2</sup>Constella Group LLC

## Objective

To describe a standard set of criteria for identifying potentially important anomalies and to compare the criteria with several recent public health events.

## Background

The Centers for Disease Control and Prevention (CDC) BioIntelligence Center (BIC) consists of a team of analysts that monitor, analyze, and interpret BioSense patient level clinical data from more than 370 hospitals and 1100 ambulatory care Departments of Defense (DoD) and Veterans Affairs (VA) medical facilities on a daily basis. Among a large number of data anomalies flagged by statistical algorithms each day, the BIC uses an automated process developed in concert with the BioSense Monitoring Protocols Working Group to identify a subset. These criteria are: A) anomalies for >1 patient class (outpatients, emergency department, inpatients) within the past week at the same facility, B) anomalies for  $\geq 2$  of past 3 days, C) maximum rate or count within the last 6 months (or since the facility began sending data for facilities with < 6 months of data), D) exceeds rate or count during the same time period the previous year, and E)  $\geq 1$  patient also maps to the severe illness/death syndrome. Criteria A cannot be met by DoD and VA events as these data sources send outpatient data only or by hospitals that send ED data only. Criteria D cannot be met by any facility that has not been sending data for at least 13 months. The identified subset of anomalies is fully characterized using other criteria not described here to either rule out or decide if they have potential public health importance.

## Methods

Each week, the BIC identifies about 100 anomalies that meet one or more of the BIC criteria. During January to July 2007, 6 events (Table 1) confirmed through communications with state and local health departments were examined to determine which of the BIC criteria were met on each anomaly day. The criteria were applied retrospectively for events 1 and 6 and at the time of the event for others.

## Results

The six events triggered data anomalies on 33 days. On 29 (88%) of the 33 days, at least one BIC criteria was met and would have been identified for further characterization (Table 1). No anomalies met criteria

A or E. Sixteen (57%) met criteria B, 14 (50%) met criteria C, 12 (43%) met criteria D, and 10 (30%) met more than one criteria.

## Discussion

Although data anomalies are common, it is rare that any are associated with true public health events. The BIC is primarily interested in characterizing events that are large in magnitude compared to historical data, involve multiple jurisdictions, or demonstrate unusual severity of illness. Our criteria were sensitive for identifying statistical anomalies associated with six confirmed events that involved sudden, large increases above baselines. As we continue to improve our statistical algorithms, enhance spatial analysis, and refine and automate our criteria, our specificity for identifying a subset of potentially important anomalies from the many hundreds that appear nationally in BioSense each week should increase. Simultaneously we will be decreasing the time an analyst must spend in anomaly characterization. In the future, it might be possible to enable users to filter anomalies to view in the BioSense application based on these or other criteria.

Table 1. Public health events, associated statistical anomalies, and BIC criteria met

Event	Criteria: # days met
1. GI syndrome/norovirus outbreak at DoD facility, 9 data anomalies during 1/25/07 to 2/10/07	No criteria: 1 day B: 6 days C: 2 days D: 7 days >1 criteria: 6 days
2. ILI/Fever cluster at several hospital EDs, 9 data anomalies during 1/24/07 to 2/11/07	No criteria: 2 days B: 5 days C: 5 days >1 criteria: 3 days
3. ILI/Fever cluster at several hospital EDs, 5 data anomalies during 1/30/07 to 2/8/07	B: 3 days C: 3 days >1 criteria: 1 day
4. Rash cluster in single hospital ED, 1 data anomaly on 3/15/07	C: 1 day
5. GI syndrome cluster due to synthetic data for a drill at one hospital ED, 1 data anomaly on 6/2/07	C: 1 day
6. Respiratory outbreak at DoD facility, 8 data anomalies during 6/20/07 to 7/9/07	No criteria: 1 day B: 4 days C: 2 days D: 5 days >1 criteria: 4 days