Using Biosurveillance Whole-System Facsimiles To Compare Aberrancy-Detection Methods: Should BioSense Use SatScan?

Sylvain DeLisle MD, MBA[1,2], Fang Tian MPH[3], Hongzhang Zheng, MSc[2], Paul Sun B.Sc[1,2], Brett R. South MS[4], Holly Gaff, Ph. D.[2], Matthew Samore MD[4], Trish M. Perl, MD, M.Sc.[3]

[1]VA Maryland Health Care System, Baltimore, MD, [2]University of Maryland, Baltimore, MD, [3]Johns Hopkins Medical Institutions, Baltimore, MD, [4]University of Utah, Salt Lake City, UT, USA

BACKGROUND AND OBJECTIVE

A "whole-system facsimile" recreates a complex automated biosurveillance system running prospectively on real historical datasets. We systematized this approach to compare the performance of otherwise identical surveillance systems that used alternative statistical outbreak detection approaches, those used by CDC's BioSense syndromic system or a popular scan statistics (SatScan).

METHODS

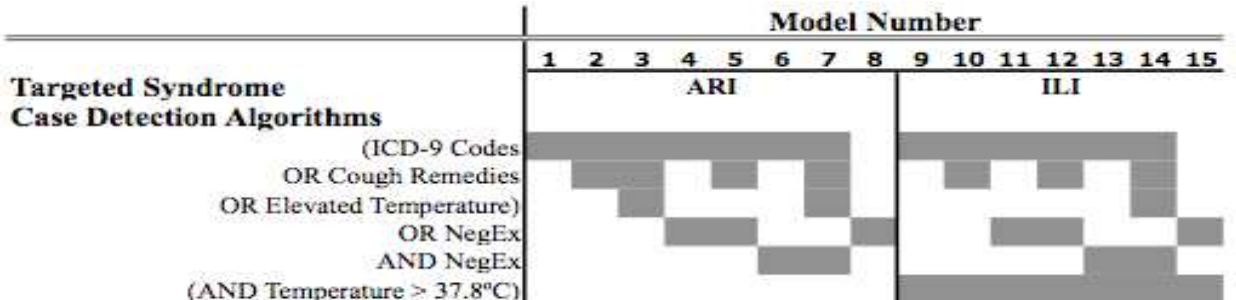
Background casecount time series were constructed by applying previously validated single-case detection algorithms (CDAs) to data mined from the Baltimore VA electronic medical record (EMR). CDAs targeted either broadly defined acute respiratory infections (ARI), or an influenza-like illness (ILI), as defined by CDC. We retained the best representatives of CDAs that included combinations of coded data (ICD-9 codes, medications, vital signs) and/or the results of computerized free-text analysis (NegEx software) of the full clinical notes for non-negated respiratory symptoms that were part of our case definitions. We injected factitious influenza cases to CDA-specific backgrounds using an age-structured metapopulation influenza epidemic model for Baltimore. We then used either SatScan time-only or BioSense's modified CUSUM statistics daily for 50 days to detect the injected outbreak. To distinguish true- from background-positive alarms, the daily statistics were performed on paired background+injection vs. background-only time series. For each CDA, the above injection-prospective-surveillance cycles were repeated once weekly for the whole study year. We computed two whole-system benchmarks: 1) the average "Detection Delay", from the time of an injection to the first true-positive statistical signal; 2) the "Workload", defined as the total number of cases included in all background-positive alarms for the entire study year.

RESULTS

SatScan alert threshold was first fixed at "$\leq$0.001". We then adjusted BioSense's alert threshold so that the average whole-system Workload best matched that obtained with SatScan. For ARI (Models 1-8 in the Table), the BioSense statistical approach led to an average detection delay of 37.1 days vs. 40.4 days for SatScan ($p < 0.0001$). For the ILI target (Models 9-15), Detection Delays were similar when all CDAs were considered together (average 38.0 vs. 39.1 BioSense vs. SatScan). For both disease targets however, a detection time advantage for BioSense over SatScan was found with the most time-effective CDAs (Models 6 and 13).

CONCLUSIONS

When coupled with the most time-effective CDAs, BioSense's modified CUSUM aberrancy-detection method provided a shorter detection delay than time-only SatScan. Experiments using whole-system facsimiles represent a useful approach to methodically evolve automated biosurveillance systems.



Whole-System Performance

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Workload (SatScan)	267	316	544	457	491	162	155	377	22	8	0	0	91	53	8
Workload (BioSense)	247	316	476	444	447	159	149	359	18	8	0	0	73	46	1
Detection Delay (SatScan)	40	40	40	40	40	40	39	40	40	38	39	40	39	39	38
Detection DeLay (BioSense)	35	35	37	37	36	35	36	38	35	37	41	42	34	37	39