Knowledge Based Public Health Surveillance

David Buckeridge MD MSc, Martin O'Connor MS, Justin Graham MD MS, Michael K Choy, Zachary Pincus, Mark Musen MD PhD, Stanford Medical Informatics, Stanford University, Stanford, CA

1. Problem

- Surveillance of pre-diagnostic 'non-traditional' data sources (e.g., school absenteeism, pharmaceutical sales) is expected to enhance the timeliness of epidemic detection
- Pre-diagnostic data are not as specific as diagnostic data, so multiple sources must be followed to reduce false positive detections
- Combined analysis of multiple non-traditional data sources requires knowledge about the relationships between data sources, but knowledge of these relationships is often qualitative and uncertain

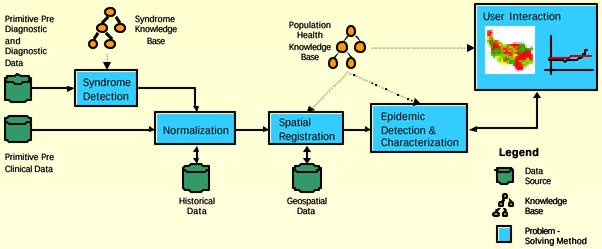
2. Knowledge Based Method

- Statistical methods perform well for focused analyses of quantitative data according to well defined models
- However, statistical models do not readily incorporate qualitative data, and can become unwieldy as the number of parameters grows
- A knowledge based approach requires explicit representation of surveillance knowledge and tasks, and enables knowledge to be applied to problem solving in a structured manner

3. Research Approach

- We are modeling the tasks involved in public health surveillance and the knowledge required to accomplish these tasks
- Based on these models, we identify or develop 'problem solving methods' (PSMs) that accomplish surveillance tasks
- We then plan to evaluate a system with different problem solving methods in terms of epidemic detection, and impact on decision-making around interventions

4. System Overview

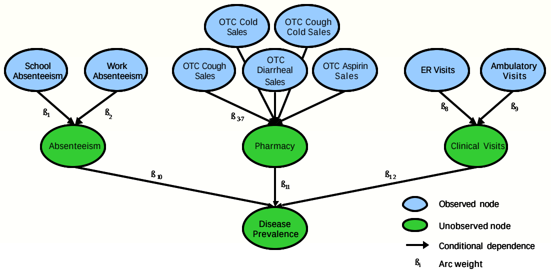


Data flow from left to right. Syndromes are detected from clinic al data according to a syndrome knowledge base (KB). Syndromes and pre-clinical data are then normalized to remove expected variation. Next, data are converted to a common spatial format. Epidemics are then detected and characterized according to a population health KB, which also guides user queries of results.

5. Current System Implementation

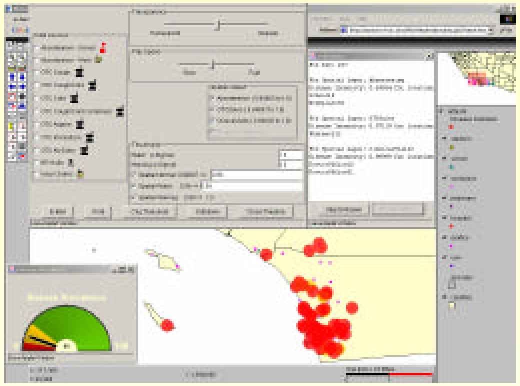
- Prototypes methods have been implemented for: Syndrome Detection using a heuristic approach, Normalization using Kalman Filtering, Epidemic Detection using a Gaussian Bayesian belief network and spatial cluster detection methods, and User Interface via a Web-Based GIS
- Based on our experience with these prototypes, we are beginning to formally model knowledge requirements, and implement additional problem solving methods (PSMs)

6. Example PSM – Bayesian Network



The diagram shows a Gaussian Bayesian belief network where the v alues of the observed nodes are standardized residuals from a forecasting method (e.g., Kalman Filtering). Arc weights are based upon knowledge of the relative importance of data sources, and can be learned from data. Temporal relationships be tween data sources can be accounted for by introducing temporal lags (e.g., Disease Prevalence at t could depend on Absenteeism at t-2, Pharmacy at t-1, and Clinical Visits at t).

7. Example PSM – Interface



8. Conclusions

- There is the potential to improve public health surveillance by using data from multiple sources, but combined analysis of non-traditional data is problematic
- Use of a knowledge based method for surveillance requires explicit description of the knowledge used in combining data, and provides a means of evaluating analytic methods for surveillance
- For more information, visit our project website – http://smi.stanford.edu/projects/biostorm

9. Acknowledgements

- Veridian Systems has been a central partner in this work. In particular, Marty Liggins implemented Kalman Filtering, and Mary Loos Sage implemented the User Interface
- Funding for some personnel has been provided by the National Library of Medicine, Palo Alto Veteran's Affairs Hospital, and the Canadian Institutes for Health Research