Modeling the Effects of Epidemics on Routinely Collected Data

XIAOMING ZENG, MD, MICHAEL WAGNER, MD, PhD

Abstract The use of routinely collected data, such as absenteeism, to provide an early warning of an epidemic will depend on better understanding of the effects of epidemics on such data. We reviewed studies in behavioral medicine and health psychology in order to build a model relating known factors related to human health information and treatment seeking behavior and effects on routinely collected data. This review and modeling effort may be useful to researchers in early detection, simulation, and response policy analysis.


Introduction

The development of computerized epidemic early detection systems has stimulated interest in new approaches to public health surveillance based on analysis of routinely collected data. The key idea underlying this new paradigm is that epidemics perturb the normal patterns of over-the-counter drug purchases; work and school absenteeism; emergency room visits, and other routinely collected data. This new paradigm rests on assumptions (largely untested) that these effects occur early, and that such data can be integrated and analyzed quickly and effectively.

Such an approach to epidemic detection depends in many ways on a basic understanding of the effects of epidemics on such data sources. For example, the prioritization of data sources for initial study would be influenced by evidence that the source can provide an early signal for some subset of diseases of interest. The detection algorithms themselves will be highly dependent on their internal models of the patterns of data to be detected.

Moreover, the analysis of routinely collected data will especially crucial to the development of early detection system for bioterrorism. Since most bioterrorist diseases of concern have never occurred, the knowledge about their effects on patterns of absenteeism, over-the-counter purchases, and emergency room visits is essentially nonexistent. This leads two possible approaches to researchers: analytic extrapolations from the data pattern of similar diseases such as Influenza, and model-based analyses. In this research, we focus on the second approach.

The models that we will consider are models that serve to tie epidemics to effects on routinely collected data. There exist models from the fields of Health Psychology and Health Services Research—models of information seeking, self-treatment, health services utilization, and care seeking—that may be useful to researchers in early detection. However, this approach is a difficult one because human behavior in such a setting can be expected to be complex and variable because when an individual experiences symptoms that may be indicative of possible effect of disease, his or her reaction and behavior are varied and conditioned on multiple factors.

In this paper, we review the literature that contributes to an understanding of the behaviors—care seeking and information seeking—of people during the period between the onset of symptoms and the time that they seek medical attention. We integrate health psychological models from the literature with a model of early detection of epidemics to form a model that covers human behavior in the time period from first symptoms to medical treatment and con-
nect it to expected effects in the patterns of routinely collected data. We report a comprehensive literature review and show how the results of previous research in the literatures on health psychology and health service research apply to predicted effects on data sources of interest to public health surveillance researchers. Although our primary goal is to build a model for the early detection of bioterrorism attack, the model should also have applications to naturally occurring outbreaks.

**Methods**

We reviewed theoretic behavioral and cognitive models of patient’s responses for diseases that would cause similar symptoms to known bioterrorism agents. We combined ideas from these models with a model of early detection of bioterrorism attack from routinely collected data (see Figure 1).

We then conducted a literature review on factors influencing patients’ behaviors and the pattern of health service utilizations after onset of symptoms and. We limited the scope of the literature review to the reactions of humans to symptoms such as shortness of breath that would conceivably be a result of diseases caused by bioterrorism attacks. The study focused on human behavior in the period between the onset of initial symptom and the first visit to health care facilities. We assumed that once a patient was in a health-care environment, that the kinds of data that would be available would be symptom and sign data for which there are already excellent processing models. We excluded from our scope the effect of public announcements or warnings about epidemic conditions on such behaviors, not because such effects are unimportant (see [3] for example) but to limit the complexity of analysis. The psychological and behavioral models that we adapted from the literature included variables such as personal awareness that would be influenced by such announcements.

We searched for published studies using MEDLINE, PREMEDLINE and PsycINFO databases. The key phrases we initially used were care-seeking behavior, information seeking, health services utilization, and illness behaviors. After review of the behavior models, we used symptom recognition, illness interpretation, illness representation and treatment seeking to find related literatures for each stage in our model. To obtain studies of specific diseases or symptoms, we used searching strategies that combined diseases or symptoms (e.g., AMI, URI, Asthma, Dyspepsia) and the keywords used above. We obtained 932 citations from MEDLINE, 45 citations from PREMEDLINE and 4057 citations from PsycINFO. After manually reviewed titles and abstracts (if available), we read 203 articles. A full bibliographic list is available at <http://www.cbmi.upmc.edu/~xmzeng>.

---

**Figure 1** Model of Patient Behavior during Epidemics.
Combined Model

Behavioral modeling to explain patient behaviors such as information seeking, self-treatment, and treatment seeking after the onset of symptoms is a well-developed area of health psychology research. Andersen reviews several such models.4 We integrated these models with a data-driven epidemic detection model to produce a model that predicts patient behaviors after the onset of symptoms, and then extends to the expected effect on routinely collected data (see Figure 1).

The aspects of the model derived from health psychology research involve the following four consecutive phases:

Recognition of Symptoms

This phase starts when the patient or family of the patient recognizes symptoms caused by biological agents. The process of recognition is very individualized and is determined by not only severity of symptoms but also personal awareness factors such as level of anxiety, cultural difference, attentional differences, situational factors, stress, and emotional mode. In Figure 1, these effects are shown as arrows into the recognition phase. During this phase, infected individuals may behave no differently than usual, so the direct effect on routinely collected data may be limited. However, decisions to stay home from work or school may occur at this stage suggesting that absenteeism may be a very early indicator of disease outbreaks. This potential effect on routinely collected data is shown as an arrow to absenteeism data in Figure 1.

Interpretation of Symptoms

The next stage is the personal interpretation of the symptoms. Interpretation is influenced by prior experience, knowledge about the symptoms, and beliefs about their seriousness. The patient himself may label the symptoms casually and nonspecifically, e.g., flu, cardiac or not important. At this stage, patients may differentially perceive the threat of symptoms based on level of vulnerability or seriousness.5 The patient may have conversations with friends, neighbors and relatives. This group is referred as the lay referral network, which is an informal network of family and friends who offer their own interpretations of symptoms oftentimes well before any medical treatment is sought.6 In Figure 1, these effects are shown as arrows into the recognition phase.

During this phase, it is known that Internet information resources increasingly play the role of the lay referral network, and may represent a type of data that may be influenced markedly patients’ response to illness.7,8 Therefore it is possible that potentially health related website query log files will be early indicators of disease outbreaks from this stage. This potential effect on routinely collected data is shown as an arrow to web query data.

Cognitive Representation of Illness

Patients form conceptual hypotheses about their underlying illness at this stage. These conceptions of illness are influenced by the media, personal experience, and family and friends who have had experience with particular disorders.9 Patients hold organized conceptions—which are referred to in the literature as disease prototypes—for certain disease based on anecdotal information. Disease prototypes such as Acute Myocardial Infarction help people organize and evaluate information about physical sensations that might otherwise not be interpretable. This organized conception of disease influences how people seek and interpret new information and their treatment-seeking decisions.10 The information-seeking behavior will be more disease oriented compared to symptom-oriented information seeking during interpretation stages. Therefore, the query data in health related website may show different searching patterns.

Seeking Treatment

This stage begins when the patient makes the decision to seek either professional or non-professional treatment. Treatment-seeking behavior is psychologically influenced by perceived response efficacy, i.e., beliefs about the effectiveness of the anticipated response, and self-efficacy, i.e., belief about one’s ability to perform the recommended response and confidence in perceived symptoms.5 Many other factors (see Figure 1) will also influence treatment-seeking behavior.

If the patient decides to go to emergency room or clinic for treatment, then hospital data sources such as ICD-9 codes and electronic medical records will show large amounts of relatively specific information useful for outbreak detection. In contrast, if the patient decides not to seek treatment, he will seek other approaches or delay treatment. Therefore, data sources such as over-the-counter drug sale or self-treatment merchandise sales may provide the only information about these sick individuals for
detection. This potential effect on routinely collected data is shown as arrows to ER data, grocery store over-the-counter drug purchasing data, pharmacy data, phone call to primary care physician, etc in Figure 1.

Parameters for the Model

In this section, we summarize the empirical studies on behavior that can begin to inform us about how to parameterize the model in Figure 1.

Factors Influencing Psycho-Behavioral Steps

The psycho-behavioral steps from recognition of symptoms to final treatment seeking are known to be influenced by demographic, psychological and socioeconomic factors in addition to pathophysiological factors such as atypicality of symptoms or seriousness of illness (rounded rectangles in figure 1). The selected literature on these factors is summarized in Table 1.

Patient Behaviors with Specific Diseases

Another relevant literature concerns the treatment seeking behavior of patients with certain diagnoses or syndromes, such as chest pain, shortness of breath, heartburn, URI or dyspepsia. Table 2 summarizes this literature.

Discussion

To our knowledge, this paper is the one of the first efforts to relate information from health services research and health psychology to the problem of early detection of epidemics. Our method was to create a merged conceptual model that could predict, if parameterized, the effects of known relevant variables on routinely collected data sources. This model may have utility to three groups of researchers.

Use in Detection Research

This analytical framework provides insights and gen-
erates hypotheses about the value of various data sources for early detection of disease outbreaks. For example, in a published hypothetical Smallpox scenario, the projected time delay from the time of infection of the first case to the first emergency department visit is 12 days and to the first definitive diagnosis is 15 days. Insights about the likely behaviors of the infected persons in these “silent” days would be valuable guides to developers trying to build detection systems capable of detecting such epidemics early. In particular, this analysis suggests that absenteeism and web health query data from health-related websites may provide the earliest indication of such an epidemic. Many United States citizens use Internet as one of their information resources about health care. According to our model, we would expect that patients start querying health related information from Internet at the early stage of recognition of the symptoms. As we have found little empirical evidence to support this hypothesis, such studies should be a research priority.

**Use in Simulation of Data**

Due to the almost complete absence of real data from bioterrorism attacks, computer simulations of data are necessary for the validation of detection algorithms. If the detection algorithms are to be based on patterns of routinely collected data such as absenteeism, over-the-counter-purchases, web health queries, or time of arrival to emergency department, then the simulations will require models of the type described in this paper. This model serves to identify many needed studies.
Use for “What-if” Analyses of Communication and Response Strategies

The model also suggests a way to predict the effects of certain response options such as public advisories on the process of detection. Such interventions could influence patient behavior at many points. For example, disease-related information could alter patients’ behavior after symptom onset and shorten the delay in seeking treatment. Announcements and warnings can create either a positive or a negative feedback loop in which the detection process gets information earlier or later depending on whether the announcement has the effect of stimulating people to emit more observable behaviors such as web queries or emergency room visits, or less as would be the case should people become frightened of quarantine, or increased risk of infection. Our integrated behavior-data-detection system model is a starting point for models that can represent such complex interactions.

Needed Research

The existing literature although large, does not address well the needs of this new detection paradigm. It does not provide the quantitative parameters and fine-grained time detail (e.g., patients’ behavior every day or even every several hours after symptom onset) necessary for detection algorithms and simulations. It does not provide sufficient care seeking behavior data for epidemic diseases. The focus of current work in health psychology is on elucidating the reasons patients delay seeking hospital treatment where as researchers in early detection we are more interested in other routinely collected data sources. Our paper identifies many areas in which studies must be conducted to facilitate this new detection paradigm.

The authors thank Margaret McDonald for reviewing this paper. This work was supported in part by grants GO8 LM06625-01, and T15 LM/DE07059 from the National Library of Medicine; contract no. 290-00-0009 from the Agency for Healthcare Research and Quality; and a cooperative agreement from the Centers for Disease Control and Prevention.

Reprinted from the Proceedings of the 2001 AMIA Annual Symposium, with permission.

References