Challenges of Electronic Medical Surveillance Systems

Jaques Reifman, Gary Gilbert, Mary Parker, David Lam
Telemedicine and Advanced Technology Research Center
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702
USA

E-mail: reifman@tatrc.org / gilbert@tatrc.org / parker@tatrc.org / lam@tatrc.org

ABSTRACT

In this paper, we discuss the technical challenges of electronic medical syndromic surveillance systems intended to provide early warning of bioterrorist attacks and naturally occurring epidemics. The discussion includes challenges associated with both civilian and military environments. In particular, we address the challenges in: (1) establishing an automated data collection infrastructure, (2) achieving timely access to quality data from disparate sources, (3) developing sensitive and specific outbreak detection algorithms, and (4) developing comprehensive and realistic simulation models for detection-algorithm development and validation. In addition, we identify unique attributes of military and North Atlantic Treaty Organization settings that may affect the development, deployment, and usage of medical surveillance systems. We conclude that considerable work and research are needed to overcome these challenges, that the information provided by these systems may lack the necessary specificity for follow-on mitigating actions, and that their cost-effectiveness and practical relevance, vis-à-vis the traditional reliance on health care providers to identify outbreaks, is still to be demonstrated.

1.0 INTRODUCTION

The 2001 anthrax attacks in the U.S. and the international outbreak of Severe Acute Respiratory Syndrome (SARS) heightened the importance of information technologies that could provide early warning of bioterrorist attacks and naturally occurring epidemics. Electronic medical surveillance systems, whose genesis precedes these recent events and which are oftentimes referred to as syndromic surveillance systems, are being widely investigated as a potential dual-use early indicator of abnormal events. These systems are not intended for disease diagnosis or longitudinal health monitoring, but rather to detect impending epidemic outbreaks and identify infected individuals early in the course of their disease through disparate data sources before a confirmed diagnosis is made. Generally, these syndromic systems target the detection of abnormal patterns in non-specific data, such as school and work absenteeism, over-the-counter pharmacy sales, nurse triage calls, and data logs of clinical symptoms and signs (in the form of acute respiratory infection or gastrointestinal illness) from encounters with primary care physicians. More recently, there has been an interest in broadening the scope of such systems to improve their timeliness by integrating public health information with risk indication and vulnerability information.

A multitude of electronic medical syndromic surveillance systems are being investigated for both military and civilian settings [1,2]. The various services and agencies of the U.S. Department of Defense (DoD) are sponsoring the development of different systems for both deployed and garrison-based forces, and many other

government agencies as well as just about every state and local government in the U.S. is investigating their own approach. While there is clearly no lack of interest and resources being targeted to the development of syndromic surveillance systems, before their potential benefits can be fully exploited—not to mention the questionable practical usefulness of these systems' outputs—a number of key technical challenges need to be addressed and overcome.

In this paper, we identify and discuss numerous technical challenges for developing and deploying medical surveillance systems in both military and civilian settings. In particular, we address the challenges in: (1) establishing an automated data collection infrastructure, (2) achieving timely access to quality data from disparate sources, (3) developing sensitive and specific outbreak detection algorithms, and (4) developing comprehensive and realistic simulation models to generate development and validation data for the detection algorithms. Moreover, we discuss the unique attributes of military and North Atlantic Treaty Organization (NATO) settings that may affect the development, deployment, and usage of these systems.

2.0 DATA ACQUISTION INFRASTRUCTURE

One of the major challenges in implementing any type of automated medical syndromic surveillance system is establishing and maintaining the information processing infrastructure to collect, store, transmit, and share data for analysis. This is especially true in a mixed military/civilian environment and in locations where military forces have been forward deployed. Military and civilian agencies use disparate information systems and data communications networks to record, store, transmit, and consolidate data relevant to epidemic surveillance. Military and civilian medical information systems are often incompatible in hardware, software, data architecture, and/or data transmission protocols. There are even significant incompatibilities among government agencies and within the same agency, the individual military services, organizations, and functional activities. Some of the most useful data are not even collected digitally or are so sensitive (e.g., intelligence data) that merging them with less sensitive private data (e.g., nurse triage, doctor visits, pharmaceutical sales, school/work absenteeism) is difficult, if not impossible.

Within the DoD Health System, a 20-year effort has been underway to create a comprehensive "cradle to grave" global health information system that captures, stores, processes, and facilitates analysis of the medical records of military members, retirees, and their families. The effort is culminating in the implementation and population of a world-wide DoD clinical data repository that takes inputs from disparate DoD medical and non-medical information systems to consolidate various types of data from outpatient and hospitalization records from military and military contracted medical facilities, including clinical symptoms and signs, chief complaints, test and laboratory orders, and pharmacy data. Only recently, through initial implementation of its Theater Medical Information Program (TMIP), has the DoD effort encompassed the forward deployed military forces in war zones like Bosnia, Kosovo, Afghanistan and Iraq. Along the way, this effort has encountered significant bandwidth and information processing and infrastructure roadblocks. Fortunately, some major innovative solutions in military information processing strategy are nearing implementation. Figure 1 provides an overall illustration of the various sources and types of data that need to be integrated.

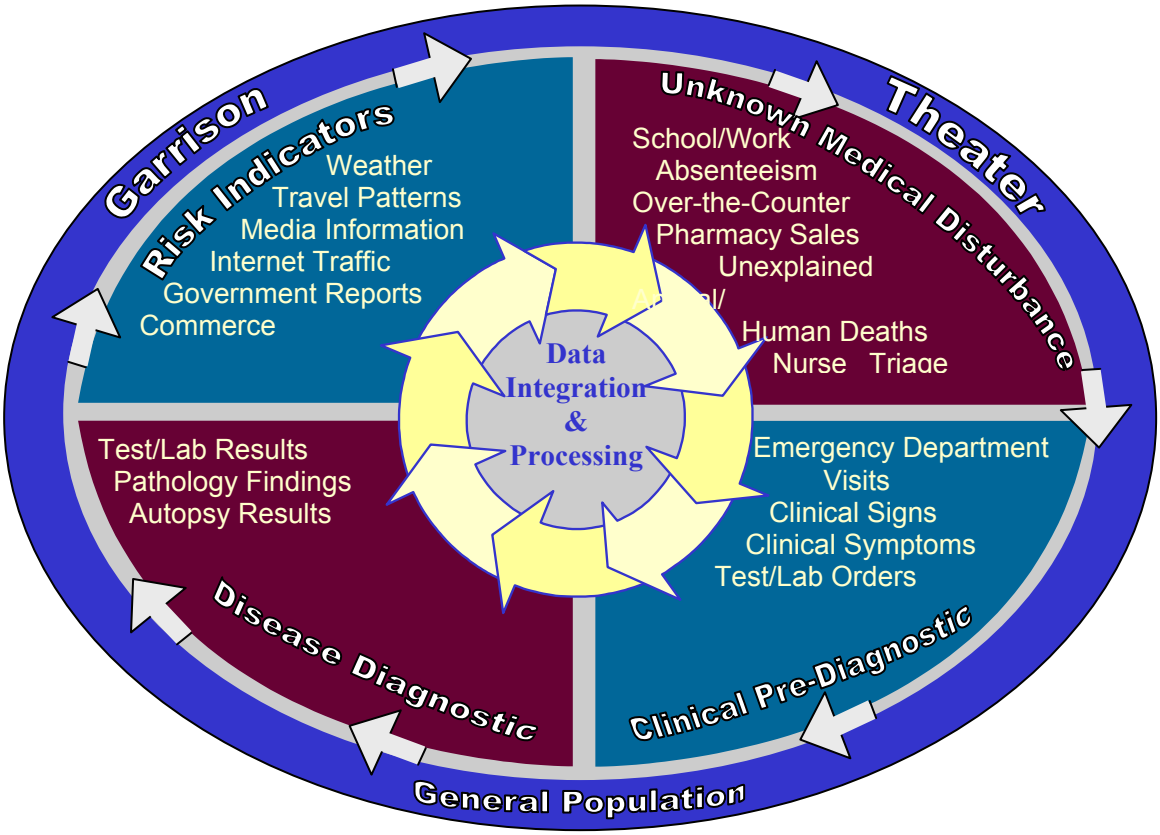

Figure 1: Sources and Types of Medical Syndromic Data that Need to Be Collected, Transmitted, Stored, and Processed for Analysis.

An emerging information architecture doctrine for the DoD, called Global Information Grid (GIG) [3], was born from concerns regarding interoperability and end-to-end integration of automated information systems. In this architecture, a distributed network-centered information architecture serves as both a collection center for information from those generating it as well as a source of information for those needing it, without a pre-planned or required communications link between information generators and information users. The emerging "net-centric" architecture for military medical operations [4] is based on a related concept (see Figure 2). Within net-centric operations, many of the medical information deposits will be of the "transmit and forget" nature (those inputting the information are not concerned with who will use it), while many of the information queries will be "blind" (those requesting and using the information will not be aware of its source). In order for a world-wide (or even regional) syndromic surveillance network to be efficiently and securely implemented within the military and between the military and civilian digital information networks that would serve as data collection sources, a net-centric GIG-type architecture is required.

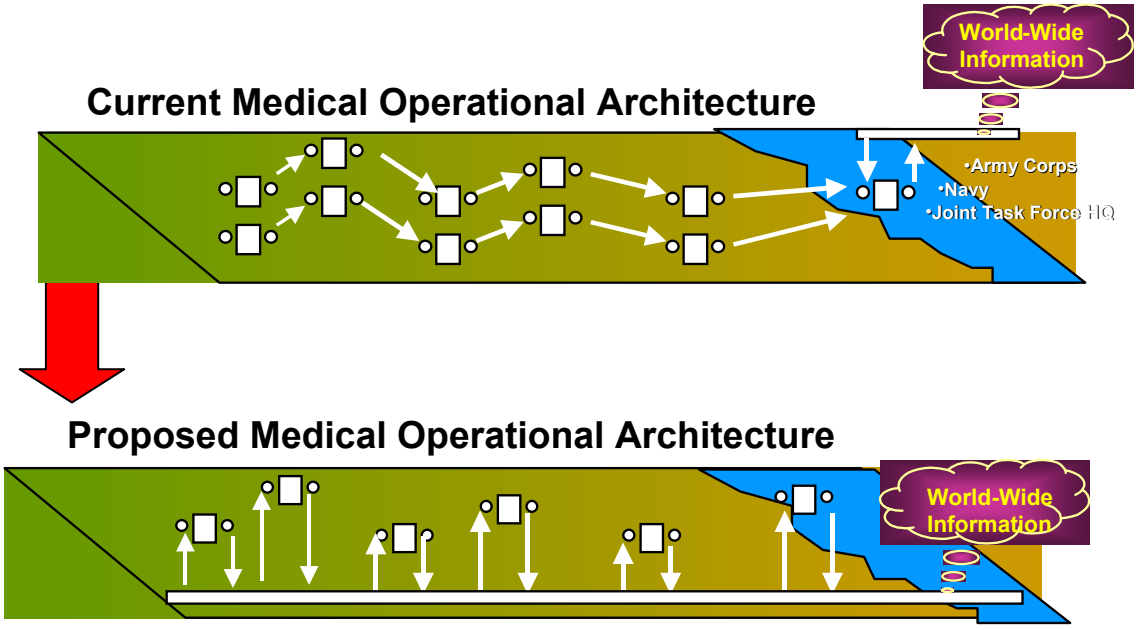

Figure 2: Proposed Network-Centric Medical Information Operational Architecture

Among NATO military members, several biosurveillance, medical information architecture standardization and systems integration and interoperability efforts are underway. This is occurring under the auspices of the Medical Information Management Systems Working Group, sponsored by the Committee of the Chiefs of Military Medical Services in NATO (COMEDS), and the Nuclear/Biological/Chemical (NBC) Medical Working Group of the NATO Agency for Standardisation. Full integration of health-care data from NATO-member government agencies, their military health services, and private health provider counterparts sufficient for efficient syndromic surveillance is a long-way off. To accelerate this process, one potential solution is the use of an "open systems" data communications network and/or open systems network enabling software that resides at the transport and network layers underneath the individual applications, session, and presentation layers, but on top of hardware device data-link and physical layers according to the Institute of Electrical and Electronic Engineers data transmission hierarchy. A standard "open systems" data architecture with standard data dictionaries, data formats, and data compression algorithms is also required if data are to be shared. Adopting open systems is key to facilitating interoperable data communications and information sharing. It is also the only economical way to deliver a solution that measurably solves the problem [5].

Sponsored by numerous U.S. government agencies, academic institutions, information technology firms, and private health care organizations and insurers, the National Forum for Health Care Quality Measurement and Reporting [6] initiated a massive project to identify a strategy and an agenda for establishing the type of national health information infrastructure that is essential to an effective medical syndromic surveillance data collection and distribution system. Initially, the Forum commissioned background papers and convened top leaders in healthcare and information technology to: 1) characterize the current state of the nation's health information infrastructure, 2) identify the primary impediments to achieving the timely flow of necessary health information across the continuum of care, along with ways to eliminate these barriers, and 3) identify actions needed to create the political will to adopt the laws, standards, business practices, and technologies necessary to create a state-of-the-art national health information infrastructure. Without such a massive effort,

the obstacles to the comprehensive near real-time health information processing required for effective syndromic surveillance may be insurmountable.

3.0 DATA SOURCES AND TIMELINESS

Selecting data sources for inclusion in an electronic medical surveillance system will be partially dictated by which phase of the detection timeline is being targeted. Figure 3 illustrates how this might be conceptualized. The far left of the spectrum represents the most timely but least specific stage along the timeline, and affords the greatest lead-time for preparation and response to an impending bio-event. The far right represents the most specific stage, the time at which a definitive diagnosis is made within a certain community. In between these two extremes are the two phases of the timeline at which most medical syndromic surveillance strategies are directed.

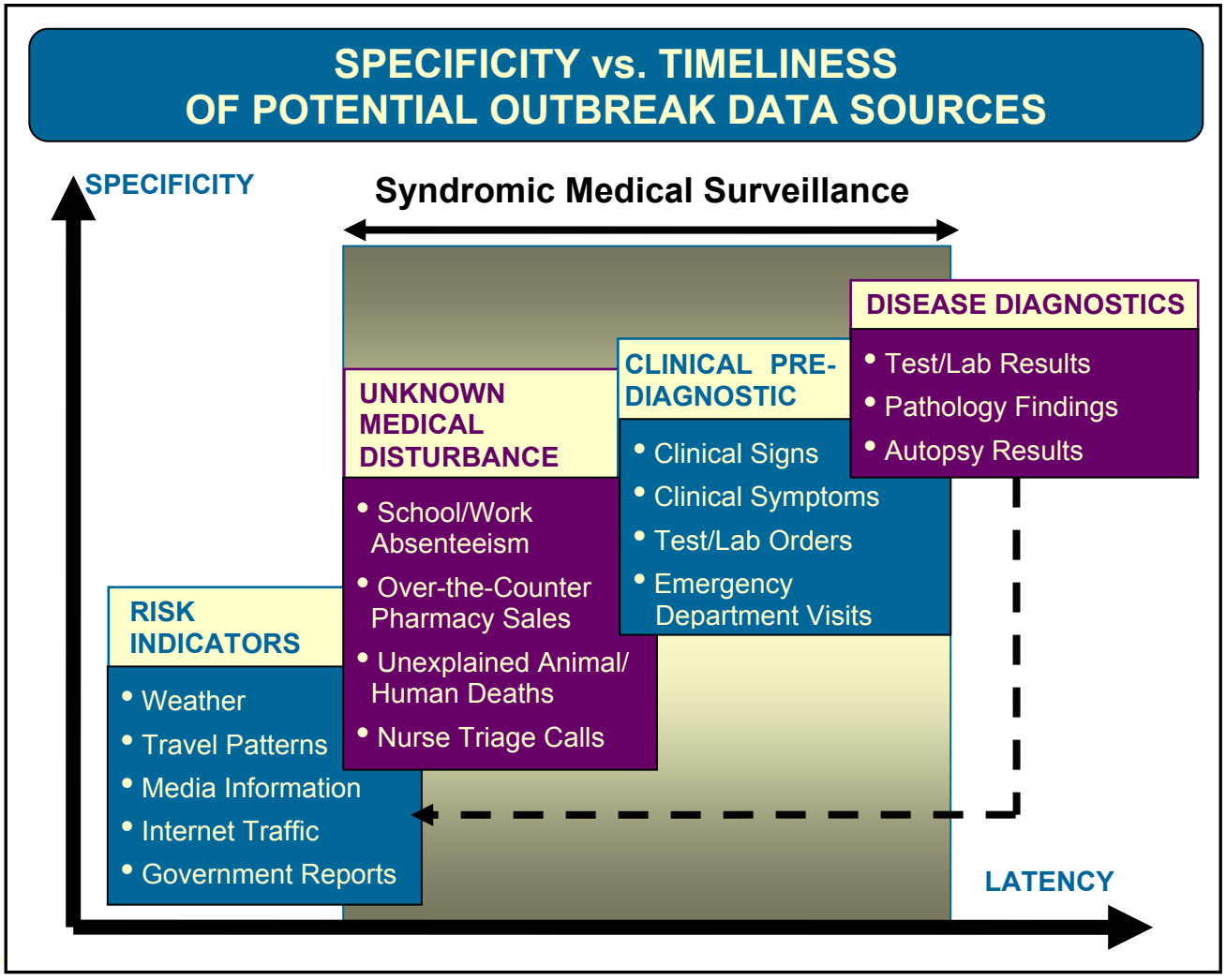

Figure 3: Conceptual Representation of Specificity versus Timeliness of Potential Data Sources for Electronic Medical Syndromic Systems

Focusing on the far left of the spectrum, the "first alerts" available may consist of risk indicators that a community is at risk for a bio-event, i.e., "conditions are favorable," prior to an actual epidemic event in that community. These markers of potential vulnerability may in some cases be extracted from data sources not directly related to healthcare. For example, local enviroclimatic information may indicate conditions are appropriate to support transmission of an insect-vectored pathogen. In the event of a known or suspected infectious disease outbreak overseas, inspection of air travel patterns might allow a reasonable prediction of which U.S. cities would first be affected due to translocation of a pathogen. An example is provided by the West Nile virus outbreak in 1999, suspected to have translocated from Israel to the U.S. [7]. Given that the overwhelming majority of air travel from Israel to the U.S. routes through the JFK airport, it may not be surprising that New York City was the initial U.S. site for ecological establishment. As noted in the SARS epidemic, information issued by foreign government health agencies may not accurately reflect the full scope and severity of an outbreak. While potentially anecdotal and non-specific, data sources such as media reports, internet traffic, and telecommunication patterns may reflect the level of anxiety within an affected community and provide clues about the true impact and extent of the event. In addition, commerce information about livestock restrictions or factory closings may also suggest a greater level of concern than that conveyed by an official government body. These indicators described above could potentially be used to assess risk for a given community prior to an actual outbreak in that community.

Moving further to the right on the timeline, unknown epidemic outbreak detection could be targeted. A bio-event could be occurring within a community, although little is known about its identity. At this stage, indirect and non-specific medical indices might be useful. Potential markers to be analyzed include such things as rates of school and work absenteeism, over-the-counter pharmacy sales, and grocery sales. If available prior to a spike in healthcare visits, such markers might be useful in the "pre-clinical" setting, as discussed by Buckeridge et al. [8]. However, the timing of such markers relative to clinical presentation has been inconsistent, suggesting this relationship is more complex, and may vary depending upon the specific pathogen and/or environment in which the epidemic occurs. For example, Musen et al. [9] noted that, during a cryptosporidiosis outbreak in Milwaukee, Wisconsin, in 1993, school absenteeism peaked four days after the peak in emergency room visits occurred, contradicting conventional notion (and Figure 3) that school absenteeism peaks before emergency department visits [8]. Gross clinical markers, such as calls to nurse triage lines and an increase in unexplained human or animal deaths, might also be informative.

The next major stage along the detection timeline would be expected, in most cases, to closely follow or be nearly coincidental with the last stage. A medical condition with more defined characteristics is detected. Markers would still be primarily pre-diagnostic where direct clinical information, predominantly in the form of emergency department visits and associated patient symptoms and signs, is exploited. Syndromic surveillance based on defined categories of illness with certain clinical characteristics, such as gastrointestinal or respiratory distress, targets this portion of the timeline. Examples include syndromic surveillance systems such as the Real-time Outbreak and Disease Surveillance (RODS) and the DoD Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) [2,10], among many others. Military disease and non-battle injury (DNBI) categories represent another similar, but less specific, way to classify medical events [11]. An illness is categorized broadly into a review of systems-type category, such as gastrointestinal or dermatologic, without qualifying characteristics or an intention to capture certain diagnoses.

The stage at the far right of the spectrum represents specific disease surveillance based on diagnostic information. While such information provides little or no response lead-time for the region from which the data originates, it is very specific and of great potential value to areas geographically remote. The performance and interpretation of diagnostic tests, however, are themselves sometimes subject to debate. For

example, the U.S. Centers for Disease Control and Prevention and the DoD do not routinely accept each other's results from environmental biosensors due to a lack of standardization between the two groups [12]. Although these sensors are not diagnostic clinical assays, the results from these sensors form the basis for recommendations regarding follow-up testing, which may impact decisions related to public health and clinical care. As described above and depicted by the dashed line in Figure 3, the presence of an epidemic in one community may prompt an analysis of risk potential elsewhere. One community's (or one segment of a population's) diagnostic information can serve as another's early indication and warning. For example, through the use of cross-correlation analysis of different time series, Sebastiani et al. [13] have shown that cases of respiratory syndromes in a pediatric emergency department can predict influenza morbidity and mortality in the general population up to three weeks in advance.

Regardless of the type of data sources included, several shared challenges to their utilization are apparent. Accuracy/validity of data, reliability/reproducibility of assessments or measurements necessary for data acquisition, accessibility of data, and protection of patient privacy are some of the key issues that must be considered in any surveillance system [2,10]. Lack of standardization in data entry and format, such as the use of free-text chief complaints from emergency departments versus the International Classification of Diseases, 9th Edition, (ICD-9) coding, as well as non-uniformity in definition of disease clusters are additional challenges currently being addressed by developers of syndromic surveillance systems. Motivated by concerns that ICD-9 codes are too fine-grained for bioterrorism detection as well as potential coding inaccuracies and inter-coder variability, several systems employ clustering of ICD-9 codes to define syndromes of interest [13]. Automated data collection within any electronic surveillance strategy is optimal, obviating the need for duplicative data entry into a system separate from that for which the data were collected. Logical data filtering and normalization are also key requirements, as integrated data are generally not readily suitable for traditional statistical analysis and may represent terabytes of information that require appropriate presentation to an analyst in order to balance the need for sensitivity with a low false positive rate.

It is important to note that the availability of accurate and reliable data is critical not only for the actual implementation and deployment of near real-time electronic medical syndromic detection systems, but also plays a key role in the development and validation of such systems.

4.0 OUTBREAK DETECTION METHODS

While a comprehensive and robust data acquisition infrastructure provides the necessary backbone for near real-time access to disparate data sources, the core of electronic medical surveillance systems lies within the data analysis algorithm. The vast majority of the proposed algorithms correctly concentrate on outbreak detection as opposed to outbreak diagnosis, as the latter is event specific, and as such, cannot identify an unmodeled, unanticipated outbreak. Outbreaks are detected as deviations between historical data representing "baseline" or "expected" values and current observed data to determine, for example, if patients are exhibiting unusual symptoms for the time of year, geography, and population. Baseline or expected values, determined for different data sources employing different mathematical formalisms, must reflect spatial variations and temporal periodicities (both daily and annual), such as increases of emergency department visits over the weekends when primary care offices are closed and during allergy season and winter months [15]. This recurrent (usual) incidence of cyclic diseases should be part of the "normal" pattern of diseases in computing the expected values, in order to recognize unusual patterns when compared with the usual.

Data-driven statistical detection approaches involving temporal, spatial, and spatio-temporal analysis have been proposed [10]. Temporal analysis accounts for the time progression of the disease/attack outbreak in a

fixed geographic location; spatial analysis examines the spatial distribution of observed cases in contrast with the background distribution for a fixed time or time interval; and spatio-temporal detection approaches consider deviations in both time and space. Knowledge-based methods that integrate surveillance data and knowledge have also been proposed, but those, perhaps due to their more intricate conceptualization and difficulty automating, are not as advanced as purely statistical methods [8]. The potential advantage of this approach is that in contrast with statistical methods that operate on low-level data, knowledge-based methods operate on data at a higher level of abstraction akin to human reasoning and could potentially represent more complex relationships, such as the modes and rates of infection transmission.

The development of systems for near real-time detection of infection outbreaks based on geographic information, both spatial and spatio-temporal approaches, faces considerable hurdles. In addition to the usual non-availability of the necessary data, for example, hospital information systems generally contain no information regarding a patient's work or school location, the information extracted from those data may be ambiguous, geographically diluted, and lack specificity. For instance, people may live, work, attend school, purchase medication, and seek medical attention in geographically dispersed locations, and exposure may occur elsewhere. Moreover, changes of demographics in a given location, e.g., construction of a large nursing home, can cause nonlinear changes in the baseline data for that location and potentially eliminate that location from analysis until updated representative baseline data can be re-established. Some of these issues, however, may not be a factor in military settings where residence, work, school, pharmacy, and health care may occur in the same restricted geographical area.

Another important consideration in the development of spatial and spatio-temporal approaches is the partitioning of a given area into a number of regions or clusters in which analysis is performed. Except for the limits imposed by the aggregation level for which data are available, predefined geographical boundaries should be avoided, as they can affect the degree to which spatial events can be detected. A desirable approach is to let the algorithm identify cluster boundaries "on-the-fly" by combining any number of close locations into the same cluster so that the most likely suspect cluster, that is, the one in which there is maximum mismatch between observed data and predicted data, is detected [16].

Due to data accessibility, lack of information ambiguity, and long experience and availability of time-series analysis methods, most detection approaches rely on temporal data analysis. A wide variety of methods exist to predict the expected value of time-series data. The simplest method, such as those based on control charts, base their predictions solely on historical data. In this approach, the expected data are simply a theoretical mean over time, which is constant for that time interval (day, week, or month of the year), and an outbreak is detected when the deviation between the expected and observed data is larger than a pre-specified threshold, typically some multiple of the standard error of the sample mean [10]. One of the limitations of such fixed seasonal models is that their estimate is constant and does not account for recent trends in the data.

More sophisticated approaches, such as regression models and classical autoregressive moving average (ARIMA) models that make estimates based on both historical data and current data, have also been proposed [10,16,17]. In this approach, outbreak detection involves comparing observed patterns with those predicted by a mathematical model. The primary benefit of ARIMA models is their ability to correct for local trends in the data so that what happened on previous days is incorporated into today's prediction. As discussed by Reis and Mandl [15], the incorporation of local information works well, for example, during a particularly severe flu season, where prolonged periods of high visit rates are adjusted to by the ARIMA model, thus reducing the occurrence of false alerts. However, in a slowly spreading outbreak, model corrections to local variations can cause ARIMA models to "adjust" to the actual outbreak, leading to over predictions and missed detection. A potential solution is to employ a hybrid detection system, incorporating both a fixed seasonal models and

ARIMA models [15].

A common issue of these methods is the tuning of the algorithms to increase detection sensitivity (i.e., to decrease the possibility of missing the detection of a true event) on the first few days after the onset of an outbreak. The challenge is due to the considerable amount of daily fluctuations or signal noise in the data and the trade-off between detection sensitivity and specificity. Signal noise impacts model accuracy, leading to prediction errors, which in turn cause miss detections and false detections. Miss detection occurs when noise in the model's predictions masks the effects of actual outbreaks, lowering the system's overall sensitivity. False detection occurs when noise spikes in the model's predictions are detected as possible outbreaks, lowering overall specificity [17]. Invariably, tuning an algorithm to improve detection sensitivity deteriorates detection specificity and vice-versa.

Tuning an outbreak detection algorithm to increase sensitivity at the early stages of an outbreak can be theoretically achieved by defining the appropriate alert thresholds and using multi-day temporal filters, in which a weighted prediction over multiple days is aggregated and compared to a threshold [17]. However, this selection is not an easy one, as the optimal threshold levels and filter weights depend on the magnitude and temporal evolution of the outbreak. Without knowing the nature of the outbreak in advance, which we never do, it is not possible to optimize these parameters. For example, to detect an outbreak in the noisiest environment, that is, to detect a small and slowly evolving outbreak, the greatest sensitivity is achieved by using a uniformly weighted multi-day temporal filter capable of smoothing out noise and picking up weak signals over many days. In contrast, to detect a large and fast evolving outbreak with low noise-to-signal ratios, the greatest sensitivity is achieved by using an exponentially weighted filter, which performs the least smoothing as it places the heaviest emphasis on the most recent days. However, this also means that less weight is attributed to the wider temporal context, decreasing the system's capability to detect small signals over many days, and limiting the ultimate sensitivity that the system can reach. As Reis et al. have demonstrated [17], different threshold and multi-day filtering strategies offer superior sensitivity for a particular stage of the outbreak. Because we never really know at what stage of the outbreak we are at any given time, it is not possible to optimize such selections. Furthermore, they have also shown that, regardless of the weighting scheme, detection sensitivity at the first stages of an outbreak drops significantly when the magnitude of the simulated outbreak is within one standard deviation of the historical mean. Hence, even if optimally tuned, it is unlikely that these systems could systematically pick early outbreaks if their magnitude is small.

5.0 OUTBREAK SIMULATION TECHNIQUES

An important step in the development of computer-based surveillance systems is validation. System validation allows for the determination of which outbreaks can be detected, how large they must be to be detected, and how early they can be detected. Due to the limited availability of outbreak data and the paucity of data available on actual germ warfare attacks, system validation is invariably performed through simulation. Two main simulation approaches, both relying on normal baseline patterns of annual recurrent incidence of cyclic diseases, are used to simulate the changes in data patterns resulting from potential covert or naturally occurring infectious disease outbreaks.

In the first approach, data representing historical normal patterns of healthcare utilization and other syndromic data are used to define a baseline upon which simulated data are superimposed to represent hypothetical outbreaks of a desired length, transmission rate, and magnitude. Invariably, due to data availability and timeliness, simulations are limited to the representation of total number of visits, chief complaints, and ICD-9

diagnostic codes collected at emergency departments [14,15,17]. For example, Reis and Mandl [15] developed a model to simulate abnormal emergency department visits of patients with respiratory illness by infusing additional visits, with varying size, shape, and duration, on top of historical data.

Such a simulation model is simple to implement and allows for parametric testing of the surveillance system and assessment of its sensitivity, specificity, and timeliness to changes in adjustable system parameters (e.g., detection thresholds). The approach has been limited to date, however, to modeling temporal variations of a single data type in a fixed, single physical location, such as the time-series analysis of acute respiratory infection of pediatric patients reported during visits to the emergency department of a given hospital. Hence, the approach has been strictly used to test univariate temporal detection algorithms. It would be difficult to broaden this simulation approach to allow for spatial and spatio-temporal data analysis and include a broader range of data types, such as the representation of different segments of the population presenting different syndromes in geographically dispersed locations, and the inclusion of over-the-counter medication and school/work absenteeism data. This is chiefly due to the lack of concurrent availability of these data types for the same exact population, the lack of linkage among the data types (e.g., most emergency departments do not record a patient work ZIP code), and, most importantly, the need for explicit mathematical models that can simulate the interdependent effects of an infectious outbreak on these data types.

The second approach allows for the modeling of such complex interdependent effects. By using numerical mathematical models, an initiating event, such as a specific disease outbreak, drives the models to represent the temporal and spatial evolution as well as the effects of the selected disease. BioWar, a scalable multi-agent network model of the impact of weaponized biological diseases in metropolitan areas, is one of a handful of such simulation models [18]. Sponsored by the Defense Advanced Research Projects Agency (DARPA), BioWar uses historical data from a small number of actual cities (e.g., Pittsburgh, San Diego, Norfolk, and Hampton) using census, school tracking, and other publicly available information as baseline upon which the simulation models are applied. The simulation models allow for the representation of the city demographic distribution and associated risk factors for the modelled disease, wind dispersion, social network interactions, and mechanisms of disease transmission and time course, and provide their simulated repercussions in the population's symptoms and signs, death rate, diagnostic tests, school and work absenteeism, and over-the-counter pharmacy sales. BioWar allows for the representation of various pathogens, such as anthrax and smallpox, as well as naturally occurring epidemics. The weaknesses of this approach are that it can only model a fixed number of cities, does not allow for inter-city modeling of disease translocation, and unanticipated outbreaks, such as SARS, cannot be readily modeled to test the robustness of the detection algorithms to unknown diseases. More importantly, there are no benchmark data to validate the underlying assumptions and accuracy of these simulation models, questioning their value as a source of data to validate the detection algorithms.

6.0 U.S. MILITARY AND NATO SETTINGS

In this section, we discuss the unique geographic, demographic, and infrastructure attributes of military and NATO settings as well as the distinct sick-call procedures of military personnel and health-care providers that may affect the development, deployment, and usage of electronic medical surveillance systems.

6.1 U.S. Military Settings

Military populations and healthcare settings present unique challenges to electronic medical surveillance. Although fixed military medical facilities in garrison may have an infrastructure similar to their civilian

counterparts, temporary facilities in deployed settings often have less robust communication capabilities. In addition, young healthy adults constitute a high percentage of the military population in an operational theater, while for many infectious diseases geriatric and pediatric age groups represent a more vulnerable population. Hence, syndromic surveillance systems deployed outside military installations could themselves serve as early indicators for syndromic systems targeted to monitor military installations. Also due to demographics, an epidemic outbreak first occurring in deployed active-duty personnel could present in a less fulminant fashion than would be expected in the general population. Another relevant difference between military and civilian healthcare is the requirement for enlisted personnel to report to sick call if planning to miss work due to illness. Non-military personnel may miss work or school without having to visit a healthcare facility. Absenteeism may, therefore, be a variable marker during the time course of an epidemic outbreak, depending upon the population in which it occurs.

In some areas of the country, a prominent military healthcare beneficiary population coexists with the civilian population. In such regions, surveillance systems designed to serve these subpopulations separately would ideally work in concert. Integration of systems like the National Retail Data Monitor [19], which monitors civilian over-the-counter pharmacy sales, with an analogous method of tracking medication prescribing patterns from military healthcare facilities, would likely provide an earlier marker of aberrant activity in that region than either data stream alone.

Similar to what is occurring within the civilian healthcare community, a number of separate medical surveillance initiatives are underway within the U.S. military healthcare system [1,2]. To promote the development of complementary efforts and to ensure that such projects are directed at individual Service and DoD-level requirements, the Assistant Secretary of Defense for Health Affairs issued a policy memorandum in November, 2003 [20]. The memorandum specifies that all information management and information technology activities related to medical surveillance within the military are to be coordinated with the Deputy Assistant Secretary of Defense for Force Health Protection and Readiness and with the Force Health Protection Council. Also published in 2003 was the DoD Health Surveillance Master Plan, a report generated by the Integrated Process Team on DoD Comprehensive Health Surveillance. This plan details a phased approach to the development of an integrated comprehensive health surveillance capability for all of DoD. Proposed initiatives include establishing a Surveillance Fusion Cell overseen by a Board of Governors, with expansion to a Surveillance Operations Center and the eventual creation of an independent Surveillance Field Operating Agency with connections to the Department of Health and Human Services, Department of Homeland Security and other U.S. government agencies.

6.2 NATO Settings

To the naïve observer, the establishment of integrated medical surveillance within NATO would appear to be relatively simple—develop a system, field it, and issue orders to use it. In reality, the issue is much more complex. NATO does not have a single medical service for its deployed troops, and certainly does not have a consolidated one for its civilian population. NATO, as an organization, does not control medical services to its personnel and populations, although (combined with Partnership for Peace) it actually has 44 disparate national military medical services. Until 1997, all operational logistics (including medical care) in NATO were by doctrine a strictly national responsibility, without any NATO-level control. In that year, new doctrine was issued (MC 319/1) specifying that logistics support was henceforth a shared responsibility between the troop-contributing nations and the NATO operational commander. This new doctrine was then extended specifically to the medical environment (MC 326/1). In accordance with current doctrine, a NATO commander can require reports on, and inspect medical assets intended to support his/her operations. As required by NATO medical doctrine [21,22], medical surveillance can be seen as a shared responsibility

between the nations and the Force Commander, but the details of how to accomplish that most effectively are still being developed. Medical units remain generally under national control, and hence, use national systems of medical records, documentation of care, and medical surveillance. The NATO mission in this context is to promote interoperability without compromising national sovereignty. The standard NATO mechanism for accomplishing this requirement for interoperability is the NATO Standardisation Program, described in detail in Allied Administrative Publication 3 [23].

In 1996, EPI-NATO [24], the first attempt at a NATO multinational DNBI medical surveillance tool for use in deployed and garrison-based forces, was developed and fielded. A modification of the British J-95 system [25], it has been used intermittently in Bosnia and Kosovo. EPI-NATO collects and tracks initial and subsequent sick calls to a medical treatment facility, identifying patients seen by symptom complex, such as climatic injury, sports injury, respiratory disease, and dermatologic problems, rather than by diagnosis. Consolidated reports of the incidence of these symptom complexes are transmitted up the chain of command to allow early visibility of outbreaks, without waiting for definitive diagnoses to be made. When appropriately used and analysed, the system functioned as designed, and allowed the early identification of at least one significant outbreak of disease in Bosnia [26]. However, in the absence of an electronic medical reporting system and a reasoning engine or algorithm to detect abnormal events, EPI-NATO suffered from delays in detection, reporting, and transmission, since each entry and detection was made manually, and forwarding of reports required specific action rather than occurring automatically. Further, the personnel with epidemiological expertise necessary to use the data optimally were not made available in adequate numbers by the nations, and data analysis was never timely or complete. Thus, EPI-NATO failed to live up to its potential and some nations simply refused to use the system. Others felt it was duplicative of their national reporting systems. Although the use of EPI-NATO was directed in policy [21], implementation problems included a lack of written implementation policy and lack of integration with existing documents on medical reporting.

With the ongoing change in NATO's missions and its realization that both military forces and civilian populations are likely targets for attack, NATO has begun to actively pursue the development of an integrated medical surveillance system. The development of an on-line military and civil disease surveillance system is one of five NBC defense initiatives approved during a recent NATO summit meeting held in Prague. Many initiatives to address this identified need are now underway. Specifications and requirements for the development and deployment of surveillance systems, capable of reporting either symptom complexes or diagnostic information to a central repository, are under development. Also, efforts by the Joint Medical Committee to evaluate various national surveillance systems to identify mechanisms for integrating them into a fully-functional NATO surveillance system suitable for both military and civilian populations are slowly moving forward.

7.0 DISCUSSION

Major obstacles to syndromic surveillance data collection are the disparate information systems and data communications networks used by the DoD and its civilian health care counterparts. Medical information infrastructures are very often incompatible in hardware, software, data architecture, and /or data transmission protocols, even among the various government agencies and the individual military services. Even if existing and planned digital information systems for both the military and civilian communities could be integrated, there remain significant bandwidth challenges especially in the forward deployed military settings. Only recently, through implementation of various components of the TMIP, has the DoD medical information infrastructure been extended to forward deployed military forces, although significant bandwidth and information processing and infrastructure issues remain to be solved. A world-wide open system with

network-centric medical information reference architecture, implemented among diverse government and civilian agencies via an approach like the DoD Global Information Grid [3], is essential to developing an effective "all-source" medical syndromic surveillance system.

Several challenges related to the selection and utilization of data sources have been identified. As data sources are chosen further along the timeline of epidemic outbreak detection, increased specificity is gained at the expense of decreased response lead time in a given community. Dissemination of specific diagnostic information from one community, however, could prompt other unaffected communities to assess their risk of acquiring the pathogen, potentially providing ample opportunity for response planning. Data management issues such as accuracy, accessibility, confidentiality, and lack of standardization also need to be addressed. Demonstration of the added value of electronic medical surveillance systems to public health will be key to proving cost-effectiveness, in particular, due to the significant resources required to procure and maintain the needed data acquisition infrastruture. In addition, defining the scope of response appropriate for a given type of alert, cognizant of the inherent resource limitations and minimal mobilization times associated with the response community, poses a significant challenge to their implementation and sustainability.

Challenges associated with near real-time outbreak detection methods abound. First, there is the common trade-off issue concerning sensitivity versus specificity of detection, which is reflected by the "appropriate" selection of alert threshold levels. Setting the thresholds too low improves timelines of detection and sensitivity (i.e., miss detection) at the expense of specificity (i.e., false detection), while setting the thresholds too high improves specificity at the expense of timeliness and sensitivity. As the optimal selection of threshold levels depends on the magnitude and time evolution of the outbreak, which are not known a priori, it cannot be realized. Second, the use of systems that employ geographic locations in their reasoning process are seriously impaired by the lack of the necessary data and ambiguity, as exposure location is not necessarily linked to where an individual lives, works, purchases medication, and seeks medical care. Next, metrics need to be generated and used to systematically and quantitatively measure the effectiveness of electronic medical surveillance systems against more traditional methods of detecting and tracking infectious disease outbreaks. Finally, due to the large amount of noise in syndromic data, the commonly accepted approach to detect outbreaks by comparing current observations with expected values may be conceptually flawed. During the initial stages of the 2003 SARS epidemic in Hong Kong, there was no detectable difference between the number of patients being registered, first believed to have contracted viral pneumonia, and the expected number of pneumonia cases for that time of the year [27]. The first useful hint that something unusual was happening occurred when a large number of health care providers themselves started to get sick. As a consequence, Hong Kong hospitals are now monitoring absenteeism among health care providers. While such an approach would be useful is detecting a re-emergence of SARS, it is unlikely that it will be valuable in detecting the "next SARS," that is, the emergence of a new, unanticipated disease for which transmission mechanisms are not known in advance.

Another key challenge is the quantitative validation of these systems. Simulation techniques currently available—to determine which outbreaks can be detected, how large they must be to be detected, and how early they can be detected—are inadequate, lacking the capability to model complex scenarios. They are either limited to model temporal variations of a single variable, e.g., the number of total daily emergency department visits, or are restricted to model fixed geographical regions and pre-specified outbreaks, such as anthrax and smallpox terrorist attacks in a given city. For instance, it is not possible to fully model the sequence of events of the anthrax attacks in the U.S. or those leading to the international SARS outbreak, preventing the validation of syndromic surveillance systems with the complexity of real-world events. Moreover, the lack of benchmark data precludes the validation of the simulation models themselves.

Differences between military and civilian populations and healthcare facilities should be kept in mind as electronic medical surveillance systems are developed and deployed. Disease presentations and the timeline for data sources used as markers of disease may vary between these groups, suggesting that a system designed for one may not be applicable to the other. From a U.S. military perspective, DoD has recently published a policy memorandum and document intended to establish a management strategy for integrating the multiple currently separate initiatives directed at health surveillance within the services. From a NATO perspective, other than EPI-NATO [24], no NATO-wide surveillance system is currently functional or deployed. NATO as an organization has formally recognized the need to develop a medical surveillance tool that will provide early warning of disease outbreaks among deployed forces and civilian population, and is progressing in that direction. Since medical data in NATO has not previously been considered anything other than national data, security of medical data has strictly been a national responsibility. A new requirement has been identified to ensure that multinational procedures are set up to meet the privacy requirements of the most stringent nation participating, and to identify required data fields and reporting mechanisms. Issues, such as standardization of current systems, differing national policies on medical data protection, and the need for integrated multinational monitoring of the civilian population, have slowed the development of such systems.

8.0 CONCLUSIONS

There is considerable activity in many agencies of the U.S. government, academia, private sector, and, more recently, in NATO to develop electronic medical surveillance systems intended to provide early warning of bioterrorist attacks and naturally occurring epidemics from monitoring a wide variety of pre-clinical and pre-diagnostic data sources. Currently in various stages of development and evaluation, these independently developed systems display striking similarities in the types of data they collect and use, the underlying detection algorithms, and in their overall system architecture. They also share enormous technical challenges, such as the timely access of quality data from disparate sources, algorithms that lack the necessary sensitivity and specificity, and inadequate means for algorithm validation.

If ongoing electronic surveillance efforts within the U.S. and its NATO partners are to succeed, a refocused basic and applied research program is needed to: 1) identify and validate more reliable sources of evidence, 2) improve methods for data collection, standardization, and dissemination through an integrated open systems global infrastructure, 3) develop more effective analytic approaches and prediction algorithms, 4) broaden outbreak simulation efforts to account for introduction of unknown agents over wider regional areas of observation, and 5) identify more reliable validation methods.

In summary, the challenges of establishing an information processing infrastructure for timely collection of medical syndromic data seem to be essentially one of implementation. Conversely, the challenges in detection algorithms are more fundamental, requiring substantial research. Ideally, what is needed is a comprehensive computer system that can take in myriad types of data from multiple dispersed data sources and—through a combination of inductive, deductive, and non-monotonic reasoning algorithms that combine infection disease domain knowledge and evidential data—synthesize the data into useful information, such as a priority list of possible initiating events to be considered by military and public health providers. Research in outbreak simulation models also needs to be expanded. Agencies funding such efforts should require developers to make them open source. In the absence of real-world data for model validation, unrestricted availability would accelerate model development and allow the scientific community at large to review the underlying model assumptions.

Apart from these vast technical challenges, the practical relevance of medical surveillance systems for early

detection of outbreaks in deployed and garrison-based military personnel and the general population is still to be demonstrated. It is unclear if such systems will be more effective than the traditional reliance on doctors, nurses, and hospitals to alert public health officials of a disease outbreak before hospitals are flooded with very sick patients. It is questionable if they can be tuned to be sensitive enough to identify abnormal patterns and set off alerts when just a few individuals are infected, while limiting the number of false positive alerts. Assuming that the system detects an abnormal event, e.g., in the number of total daily emergency department visits, what information does it offer that military and public health providers can act on? As recent outbreaks have shown, knowing only that an abnormal event is occurring without having more specific information, such as an understanding of the mechanisms of disease transmission, is not likely to lead to timely and effective actions that can reduce morbidity and mortality rates. In the short term, we need to better understand the capabilities of these systems. One possibility, is to generate and use metrics to systematically and quantitatively measure the effectiveness of deployed systems against more traditional methods of detecting and tracking infectious disease outbreaks.

9.0 REFERENCES

[1] Kun LG and Bray DA. Information infrastructure tools for bioterrorism preparedness. IEEE Engineering in Medicine and Biology, September/October 2002.

[2] Garshnek V and Burgess L. Infectious disease surveillance – A review of current system. Presented at the Bioterrorism Preparedness Summit, University of Hawaii, 19-21 October 2003. (available as part of the Annual Report 2003 to TATRC under the project Bioterrorism Preparedness for Infectious Diseases, #DAMD17-03-2-0018).

[3] Gilbert GR, May JH, Rocca MA et al. The global grid telemedicine system consult broker. Telemedicine Journal. 2004 (in press).

[4] Mease AD, Klein T., Kensinger EE, Kruse BW, Schaeberle DC, Whitlock WL, Gilbert GR. Information requirement analysis supporting net-central medical operations. Proceedings of the American Telemedicine Association Annual Meeting, Phoenix, AZ, May 2000.

[5] Holcomb GK. 2003. Cost-effective open standards-based solution for homeland security, bioterrorism defense, and healthcare delivery. White Paper, Tulane University. 2003. Available at <http://www.som.tulane.edu/tccep/documents/LouisianaHomelandSecurity.pdf>.

[6] National Forum for Health Care Quality Measurement and Reporting. Washington, D.C. Available at <http://www.qualityforum.org/prnursinghomes11-19-01.pdf>.

[7] Lanciotti RS, Roehrig JT, Deubel V et al. Origin of the West Nile virus responsible for an outbreak of encephalitis in the northern United States. Science. 1999; 286(5448):2333-7.

[8] Buckeridge DL, Graham J, O'Connor et al. Knowledge-based bioterrorism surveillance. Proceedings of the AMIA Symposium; 2002:76-80.

[9] Musen MA, O'Connor MJ, Buckeridge DL et al. A knowledge-based approach to temporal abstraction of clinical data for disease surveillance. Proceedings of the NATO Human Factors in Medicine Symposium on Operational Issues in Chemical and Biological Defense, Estoril, Portugal, May 14-17, 2001. Available at:

<http://www-smi.stanford.edu/pubs/SMI_Reports/SMI-2001-0891.pdf>.

[10] Mandl KD, Overhage JM, Wagner MM, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. Journal American Medical Informatics Association (in print).

[11] STANAG 2050 (5th Ed). Statistical Classification of Diseases, Injuries, and Causes of Death. March 1989.

[12] Parker M, Emanuel P, Wilson P et al. Proceedings from the consensus conference on the role of biosensors in the detection of agents of bioterrorism. Homeland Defense Journal, on-line edition, 28 January 2004. Available at <http://www.homelanddefensejournal.com>.

[13] Sebastiani P, Mandl KD, Szolovits P, Kohane IS, and Ramoni MF. Pediatric patients are early sentinels of influenza illness and mortality. (under review, February 2004).

[14] Lober WB, Karras BT, Wagner MM et al. Roundtable on bioterrorism detection: Information system-based surveillance. Journal of the American Medical Informatics Association. 2002; 9:105-15.

[15] Reis BY and Mandl KD. Time series modeling for syndromic surveillance. BMC Medical Informatics and Decision Making. 2003; 3:2 <http://www.biomedcentral.com/1472-6947/3/2>.

[16] Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. Journal Royal Statistical Society. 2001; 164:61-72.

[17] Reis BY, Pagano M, and Mandl KD. Using temporal context to improve biosurveillance. Proceedings of the National Academy of Sciences. 2003; 100:1961-5. Available at <http://www.pnas.org/cgi/doi/10.1073/pnas.0335026100>.

[18] Carley KM, Fridsma D, Casman E et al. BioWar: scalable multi-agent social and epidemiological simulation of bioterrorism events. Proceedings of the NAACSOS conference, Pittsburgh, Pennsylvania. 2003. Available at <http://www.casos.ece.cmu.edu/projects/BioWar/>.

[19] National Retail Data Monitor Fact Sheet. Available at <http://www.health.pitt.edu/rods>.

[20] Health Affairs policy memorandum 03-022. Medical surveillance information management strategy for Force Health Protection. 6 November 2003.

[21] AJP-4.10. Allied joint medical support. February 2002.

[22] AMEDP-6(B). NATO handbook on the medical aspects of NBC defensive operations. July 1994.

[23] AAP-3. Procedures for the development, preparation, production and the updating of NATO standardization agreements (STANAGs) and allied publications.

[24] United States European Command. Daily disease and non-battle injury medical surveillance using EPINATO. 13 February 2003. Available at <http://www.eucom.mil/Directorates/ECJ4/ECJ4-MR/prevmed/>.

[25] Wright LA, Demicheli V, Gillespie WJ et al. Morbidity surveillance in the British Army - the first 12

months. Journal Royal Army Medical Corps. 1998; 144:11-7.

[26] Lam D, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Command and Materiel Command, personal communication, 13 February 2004.

[27] Chiu L, Princess Margaret Hospital, Hong Kong, personal communication, 20 October 2003.

ACKNOWLEDGEMENTS

The first author was supported in part by the Combat Casualty Care and the Military Operational Medicine research programs of the U.S. Army Medical Research and Materiel Command, Ft. Detrick, MD.

DISCLAIMER

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or the U.S. Department of Defense.

"This paper has been approved for public release; distribution is unlimited."