Challenges in Using the Arden Syntax for Computer-Based Nosocomial Infection Surveillance

ROBERT A. JENDERS, MD, MS, ANUJ SHAH, MA

Abstract Context: Detection of outbreaks of infection in the hospital typically requires daily manual review of microbiology laboratory test results. This process is time-consuming, tedious, prone to error and may miss trends in infection. A standard formalism for procedural knowledge representation, the Arden Syntax, provides a vehicle for implementing algorithms for detecting such infections. Objective: To design and implement a computer-based system for detection of concerning patterns of infection or antibiotic resistance. Setting: Computer-based event monitor and central patient data repository at the Columbia-Presbyterian Medical Center (CPMC). Results: We designed a two-phase system, including initial filtering of individual patient laboratory results by Arden Syntax Medical Logic Modules (MLMs) and subsequent aggregation and analysis across patients and locations using a statistical monitor. Preliminary data for the filtration phase demonstrate a 94.8% reduction in the volume of messages that must be considered in surveillance. Conclusions: Filtering raw laboratory results using a standard formalism eases the process of aggregating data across patients and sites as well as detecting trends in infection. There is a need for augmenting such formalisms in order to enable population-based decision support.


Introduction Nosocomial Infections

The occurrence of infection in the inpatient setting is an ongoing challenge. The average incidence of nosocomial infection (NI) is 5% to 10%, while rates in academic medical centers are higher than this and may reach 28% in intensive care units. Such infections affect over 2 million patients annually in the USA, resulting in monetary costs in excess of USD 4.5 billion each year.

Surveillance is an important part of controlling these infections by facilitating detection of systemic lapses in infection control and developing appropriate interventions to repair those lapses. Surveillance typically includes three related activities: determination of predominant organisms, determination of the antibiotic resistance patterns of the organisms and identification of outbreaks of such organisms.

Unfortunately, even if microbiology laboratory data are captured in electronic format, analysis of these data to detect potentially adverse trends in organism occurrence and resistance patterns often is performed manually. To help overcome this difficulty, workers at several institutions have developed computer-based surveillance systems. Examples include GERMWATCHER at Washington University and the Computerized Infectious Disease Monitor (CIDM) at LDS Hospital. Indeed, data from the latter system demonstrate improved detection of infection while reducing personnel time by nearly 2/3. In turn, this suggests that manual surveillance is prone to error.

Knowledge Representation Standards

Despite the success of these systems, their applicability at other locations is limited, even with explicit
standard definitions for culture-based NIs. This is not only attributable to a lack of structured data in appropriate format, but also it is the result of the use of knowledge representation formalisms that are difficult to share with other institutions. Arden Syntax is a formalism for representation of procedural medical knowledge that can be used to facilitate knowledge transfer. The unit of representation in the Syntax is the Medical Logic Module (MLM), which contains enough data and logic to make a single medical decision. The Arden Syntax has been implemented at multiple sites worldwide. Further, the decision support systems sold by several major vendors use this standard.

Admittedly, even with such a standard, some site-specific changes must occur in order for a knowledge base to be transferred from one site to another. However, use of a standard formalism helps to reduce the barrier to inter-institutional knowledge transfer.

This notwithstanding, the focus of published work using the Arden Syntax has been to provide alerts and reminders related to individual patients and not to facilitate notification regarding populations of patients that would be required in NI surveillance. Indeed, examples provided in the latest version of the standard have a single-patient focus and refer only to data regarding the individual patient whose data triggered the MLM.

Goals of the Analysis

Accordingly, we sought to implement a computer-based system for surveillance of NIs. By doing this at least in part using the Arden Syntax, we planned to assess qualitatively the extent to which this formalism, which is principally focused on data regarding individual patients, may be used in surveillance of NI.

Methods

Information System Setting

At the Columbia-Presbyterian Medical Center (CPMC), a variety of patient data, including demographics, patient location information and microbiology laboratory results, are stored in a central data repository. Entry of data to this repository is monitored by a clinical decision support system (CDSS), the core of which is a clinical event monitor. The knowledge base of the clinical decision support system has been encoded as MLMs using the Arden Syntax since 1992.

The event monitor examines each patient data element that is uploaded to the repository, comparing each to a collection of triggers, each of which is linked to one or more MLMs. When a match, defined either as exact or by membership in a defined class of events, occurs, the related MLMs are triggered. These in turn retrieve patient data from the database and reach conclusions using the logic encoded in the MLM. Depending on the result, the conclusions can be sent to an authorized recipient and recorded in the repository for subsequent display (e.g., an electronic medical record).

Single-Patient Focus of the CDSS

As with the Arden Syntax generally, the CPMC CDSS accommodates only single-patient events as triggers. The unique patient identifier, the medical record number (MRN), is part of the evocation record of data that can trigger a MLM. Moreover, the MRN is an implicit and inalterable part of queries within the MLM; queries across a population are not possible.

Prior work at our institution has attempted to compensate for this by creating additional software that can analyze data across patients using the CDSS. The principal example of this was the monitor of monitors (MOM), which tracked the output of the CDSS across all patients in order to determine when an individual MLM or the entire CDSS may be malfunctioning.

Overall, then, we planned to use the single-patient-focused CDSS and its Arden Syntax knowledge base and to leverage the cross-patient analysis capability of MOM in order to perform surveillance NI. In turn, this would allow us to identify features whose inclusion in the Arden Syntax would improve its use for population-based surveillance.

Results

In order to achieve these goals, we designed a two-phase system depicted in Figure 1. The first phase employs the CDSS in its current configuration. A MLM is triggered each time a patient microbiology laboratory result is uploaded to the central repository. The MLM filters the raw data by excluding negative microbiology results, preliminary results and results already detected for a particular laboratory sample. In order to maximize sensitivity, results were not filtered by likely contaminants. For example, an organisms isolated from only 1 blood culture of 3 obtained at the same time from an individual patient
likely represents an integumentary contaminant, but we did not attempt to filter out such results.

For each event that passes this filter, the MLM writes a collection of information to a decision relation in the central repository. This collection includes the microbiological results, the probable site of acquisition of infection (estimated by elapsed days from admission, with infections early in a hospital course likely acquired outside of the hospital), the patient’s location in the hospital when the sample was obtained, the time that the sample was obtained and the time when the result was reported by the laboratory. These results are written in a structured fashion using codes that reference a controlled vocabulary. Antibiotic sensitivity patterns were not included in this initial version of the system.

As the transition to the second phase of the system, each day the interval data produced by this MLM are to be retrieved from the central repository and stored in a separate database specifically designed for NI surveillance. Using the statistical analysis previously developed for MOM, we created a distinct monitor to perform counts of specific organisms over a window of time. When a threshold count was exceeded, an alert was generated. The initial threshold was set manually, while subsequent thresholds would be determined automatically by the baseline occurrence of organisms in the microbiology data. Thresholds could be set by inpatient unit or over the institution as a whole.

Finally, we designed a common gateway interface (CGI) accessible via a World Wide Web browser, that permitted on-demand review of positive results across CPMC and by individual nursing units within the facility.

As an indication of the efficiency of filtering, we examined recent data written to the decision relation by the CDSS over a period of 12 days, restricted to blood cultures for two inpatient units, one of which is an intensive care unit. A total of 37 cultures were obtained, but these cultures resulted in the production of 192 messages by the laboratory information system, including both positive and negative results as well as preliminary and final results. In the manual system, each of these would be perused by hospital epidemiology personnel. However, of all these messages, only 10 were final, positive results, and this represents a 94.8% reduction in the volume of messages requiring perusal in the manual system.

**Discussion**

As is the case with most installations of the Arden Syntax, the CPMC knowledge base retains a single-patient focus. This is manifest particularly in the definition of events that can be detected by the CDSS and the scope of queries that can be performed within MLMs evoked by a clinical event monitor. The expense and potential disruption to the CDSS led us to create a separate, back-end process to perform cross-patient analysis for NI surveillance without altering the original functions of the CDSS.

One alternative to our method, therefore, is to modify the query capability of the CDSS to perform cross-patient queries and summation of data. An individual patient result still would trigger the MLM as it does in the current system, but each positive result would prompt a query for similar results for other patients. This, in turn, could be analyzed within the logic of a MLM to determine if a potential outbreak necessitating intervention by infection control personnel was in progress. In this situation, in order to avoid repeating the summation of data for each positive result for a particular organism, the query sub-system would have to be endowed with the capability of accessing a summation of previous counts by organized by organism and hospital inpatient unit.

Augmentation of the Arden Syntax to support such cross-patient analysis would facilitate this process. Version 2.1 of the Syntax already includes aggregation operators (COUNT, AVERAGE, STDDEV, MINIMUM, MAXIMUM) that can be used for such analysis. One possible addition to the Syntax is to facilitate the definition of the characteristics of a population whose data might be queried and over which the logic of a MLM might be applied. Arden currently

![System Architecture](image-url)
enables definitions of trigger events (evoke statement) and destinations (e.g., email or pager), although the exact specification of these definitions are site-specific and, for this reason, are segregated in curly braces. Just as events and destinations are now defined, a POPULATION statement could be added. The parameters of this statement would include the specific criteria (e.g., microbiology results) that qualify a patient for inclusion in the population of interest.

This population-based emphasis might further be clarified by adopting a change in the MLM TYPE statement. Currently the only legal values for this slot are “data_driven” and “data-driven”. A third alternative, “population-based,” might be added in order to denote that a particular MLM focuses on alerts and reminders pertinent to a group of patients as a whole.

Finally, while the reduction in message volume by filtering was very large, this was determined through the use of limited preliminary data. Further experience will determine if this degree of filtering persists.

Future Work

We plan to move our two-phase system from its current status as a proof-of-concept system to a production tool that will be used actively by CPMC hospital epidemiologist and his staff. In addition to detecting patterns of the occurrence of organisms, we will apply similar statistical analysis to patterns of antibiotic resistance. Further, we plan to expand the detection of NIs from microbiology data to other types of data. A MLM that detects nosocomial pneumonia based on natural language processing of chest radiograph reports already is in place.

Also, in order to calculate rates of infection, final negative results will have to be documented in addition to positive ones. We already have created a MLM that writes a record of negative results using structured, coded values to the decision relation. Finally, one of us (RAJ), co-chairman of the Health Level Seven committee that sponsors the Arden Syntax, will work to incorporate changes in the standard that will facilitate population-based alerts and reminders.

Summary

In order to facilitate the computer-based detection of nosocomial infections, we have created a two-phase system. The first phase uses Arden Syntax to filter microbiology laboratory data in order to retain only those results suggesting actual infection. The second phase compensates for the single-patient focus of most installations of Arden Syntax by using a statistical monitor to track results over many patients across multiple hospital inpatient units. Preliminary data suggest that the first phase provides a significant reduction in the volume of messages that must be processed. Finally, we suggest improvements in the Arden Syntax that would facilitate this type of clinical decision support.

We gratefully acknowledge the support of the New York-Presbyterian Hospital and the Agency for Healthcare Research and Quality, grant R01-HS10472-01A1.

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

References