Using Intelligent Agents to Repurpose Administrative Data in Fostering Disease Prevention in an Outpatient Context: The Case of Pneumococcal Vaccination

Carla Coffin  
Chad Saunders  
Chandra Thomas  
Andrea Loewen  
Norman Campbell  
William Ghali  
University of Calgary

cscoffin@ucalgary.ca  
chad.saunders@haskayne.ucalgary.ca  
cmthomas@ucalgary.ca  
andrealoewen@yahoo.ca  
ncampbel@ucalgary.ca  
wghali@ucalgary.ca

Abstract

The use of intelligent agents is proposed as an economical way to repurpose administrative data in order to foster a program of disease prevention in an outpatient context. A retrospective computerized search was conducted using administrative hospital discharge data to identify patients admitted to a medical teaching unit who met the Canadian Immunization criteria for pneumococcal vaccination over a one-year period.

For identification of persons eligible for pneumococcal vaccination, administrative discharge data was shown to have a sensitivity of 83%, (confidence interval [CI] 0.73-0.92) and a specificity of 78% CI (0.64-0.91), with a positive predictive value [PPV] of 87%, CI (0.83-0.90) and a negative predictive value [NPV] of 72%, CI (0.58-0.86).

This study demonstrates that administrative data appear promising as the basis for certain clinical applications. Specifically, the reasonably high specificity and sensitivity of diagnostic codes in administrative data could be utilized to trigger appropriate pneumococcal vaccination after hospital discharge among eligible patients who might otherwise never receive this efficacious intervention. Reminder systems in a hospital setting have received mixed results although positive results have been shown in several outpatient settings but using clinical data. Therefore, before a reminder system using administrative data in an outpatient context is implemented it seemed prudent to investigate this issue further.

1. Introduction

To meet the twin demands of cost containment and quality care espoused by most health care organizations, the implementation of integrated health care information systems are required [1]. The foundation for these systems is the computer-based patient record, which is specifically designed to provide easily accessible complete and accurate data, practitioner reminders, clinical decision support systems, links to bodies of medical knowledge and other aids [2]. Unfortunately, the situation described by Dick et al. is more likely an ideal state rather than a reflection of the current situation in practice. While considerable gains have been made in many areas of health care, the vision of fully integrated systems has not been realized. To understand why this has not occurred necessitates consideration of the types of information systems that have been employed in a health care context and the conditions needed for their integration.

The remainder of this paper is organized as follows. First, the literature on the use of information systems for clinical applications is reviewed. Second, intelligent agents are proposed as a tool to address, at least in part, some of the challenges identified in the literature. Finally, an empirical investigation is presented to demonstrate the efficacy of the approach.

2. Information Systems in Health Care

In considering information systems in health care, a useful distinction can be made between administrative and clinical information systems. Administrative information systems are essentially transaction processing systems designed to handle accounting data.
for the purposes of billing and cost accounting [3]. In contrast, clinical information systems are designed for processing transactions related to medical diagnosis, treatment and follow-up [4]. Administrative information systems have a long history in health care while clinical information systems are a more recent development, although both types of systems can also serve non-transactional purposes like decision support [5]. Therefore, in order to meet the twin objectives of cost containment and quality care it is clear that both types of systems are required. Furthermore, while these systems may serve separate functions of the organization, the underlying data in each system is linked in practice, even if they are not technically integrated. That is, in order to meet the objectives of continuous quality improvement, both administrative and clinical measures are needed, the source of which is often these separate information systems. However, many of the same pressures for cost containment and improved clinical care also negate efforts to integrate the two systems on any large-scale and comprehensive manner.

A key type of administrative data is discharge information as coded using the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) [6]. Administrative computer discharge data is receiving increased attention as the basis for research on comorbidity [7], to estimate incidence of clinical conditions [8] and support quality monitoring [9] in health care. Advantages of such data include its accessibility, inexpensiveness, and encompassment of large populations and the reliability of the information systems in which it is stored. However, there still remain questions regarding data quality issues such as accuracy, completeness and possible errors in coding of certain clinical conditions [7, 8, 9, 10].

3. Impediments to Information Systems Use in Health Care

Information systems typically represent a significant investment for health care organizations and for many this cost is often prohibitive. However, even when these systems are in place, there are other factors that serve to impede their use. In fact, computer-based information systems have met with limited acceptance among physicians [1]. When these systems are not used, there is a tendency to point to the health care professionals’ refusal to change [1] or to a lack of technical expertise or experience with the systems as the source of the low utilization levels [11] or discontinuation of the information system [12]. However, drawing upon the diffusion of innovation literature, this represents a bias towards blaming the individual and not the system [13]. While certainly lack of technical knowledge or familiarity with the system can be contributing factors, a close examination of the system is likely warranted in the context of changing practice. Clinical information systems affect the structure and function of health care organizations [1] and when the implementation of information systems is perceived to interfere with traditional practice routines, they are not likely to be accepted by physicians [14]. While physicians may be reacting to potential power shifts or deskilling concerns that these systems occasion [15], it is also possible that there are legitimate professional concerns that have not been addressed with the system. Given the autonomous nature of professional practice, it is unlikely that system use can be mandated [1] so a better strategy may be to design the systems to be consistent with existing professional practice.

There is also an innovation bias evident in that it is assumed that use of information systems is warranted and desirable in all situations for all individuals [13]. Information systems in health care clearly contribute significantly to improving quality of care and cost savings, however these benefits should not be expected to be universally realized nor even attainable in many contexts and for certain groups. Information systems therefore need to be designed to address social and organizational practice as well as the technical needs of the end user.

4. Intelligent Agents and Reminder Systems

Organizations are experiencing a growing gap between their data storage infrastructure (i.e. their data warehouse) and their ability to analyze and act effectively on the information these data warehouses contain [16]. A data warehouse is an orderly and accessible repository of known facts and related data that is used as a basis for making decisions [17]. One approach to dealing with the large quantity of data is to implement data mining initiatives. Data mining is the process of discovering meaningful patterns and trends in data warehouses by using pattern recognition technologies that employ statistical and mathematical techniques [16]. Data mining can require specialized statistical and mathematical skills or training on advanced and often-expensive data mining software applications. However, intelligent agents can be used for smaller scale applications where cost and technical skills may be at a premium. Agent technologies are specialized software applications that perform a set of operations with some degree of independence or autonomy by employing some knowledge or representation of the user’s goals or desires [18].
Intelligent agents (or bots within the Internet domain) can perform many of the prerequisite tasks needed to support decision-making [16]. In many contexts, intelligent agents are used as small-scale expert systems, allowing technical novices to perform advanced technical tasks. However, in the context of many health care needs, and in particular with clinical tasks, intelligent agents are more appropriately employed in a decision support manner. That is, final decision must still remain with the health care professional, thus recognizing the complex technical and social environment in which these decisions are made. Intelligent agents, while effectively dealing with quantities of data, still require a certain quality of data to support effective decision-making.

With quality data, intelligent agents can be employed as the basis of an electronic reminder system. These systems provide reminders to health care professionals based upon triggers that are initiated when predefined criteria are met. Reminder systems are typically used to provide alerts concerning possible detrimental treatments or actions. The use of reminders with clinical data was shown to have positive effects on preventative care in an outpatient setting [19], although mixed results have been shown with preventative care measures within a hospital context using clinical data as the source of the triggers [20, 21]. The authors are unaware of any studies using administrative data to foster disease prevention in an outpatient context.

5. Repurposing Administrative Data

One approach to applying the insights outlined above is to consider solutions that deal with improving the use of existing infrastructure. The fiscal reality facing many health care organizations is that there is insufficient budget for the implementation of integrated information systems. However, these organizations may still have significant administrative information systems resources in place. As noted above, these systems can be used beyond their purely transaction processing past and be extended for use in decision support roles. These decision support roles are generally still within their prescribed intended purpose. That is, administrative data are used to support administrative decisions while clinical data are used to support clinical decisions. There appears to be a general assumption within the health care community that administrative data is not suitable to clinical applications, despite a paucity of research on the topic.

Notwithstanding these concerns, it is at least plausible that there are situations where administrative data have clinical applicability. In fact, the value-added utilization of existing administrative data for organizational decision support has been demonstrated in a disease management context [5]. However, the authors are unaware of equivalent investigations in the context of disease prevention.

Organizations seeking cost conscious solutions for quality improvement initiatives that target specific areas may wish to use intelligent agent based reminder systems that avoid changing professional practice and draw upon existing administrative data. Reminder systems in a hospital setting have received mixed results although positive results have been shown in several outpatient settings but using clinical data. Therefore, before such systems are implemented it would seem prudent to investigate this issue further. The next section describes an empirical investigation of the applicability of administrative data to disease prevention initiatives in an outpatient context.

6. Study

6.1 Background

Pneumococcal vaccination is an important preventative health care measure. *Streptococcus pneumoniae* remains a major cause of significant morbidity and mortality worldwide despite availability of an effective vaccine. In recognition of this, the 1998 and the 2002 Canadian Immunization Guidelines [22, 23] have recommended that all high-risk individuals should be vaccinated. In early 1998 the Canadian Consensus Conference on Preventing Pneumococcal Disease recommended a target vaccination rate of 80% of eligible persons by 2005. Despite this national target, the pneumococcal vaccine is underutilized and in 2001 the Canadian rates were only 15-42% of eligible persons.

While a number of innovative strategies have been proposed to improve rates of vaccination [24, 25, 26], there continues to be significant missed opportunities to provide the pneumococcal vaccine to eligible persons [27]. This study explores the potential utility of hospital discharge data as a tool for generating triggers for vaccinating people following discharge from hospital.

6.2 Sample Selection

The study population was selected from the Calgary Regional Health Authority Hospital Database at the Foothills Medical Centre, a 700-bed tertiary care facility in Calgary, Alberta, Canada. The database consists of hospital discharge abstract data routinely collected and includes up to 16 *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) coded diagnoses.
Description of administrative coding procedures in the three adult hospitals of the Calgary Health Region is previously described [7]. A retrospective computerized search was conducted to screen all admissions to the general medicine medical teaching service between January 1 and December 31 2001. The database search criteria included cases if they were younger than 65 years of age, admitted to the Medical Teaching Unit at the Foothills Medical Center, if a MTU staff physician was identified as the most responsible physician, and if the patient had diagnosis codes indicating eligibility to receive the pneumococcal vaccine. Cases were excluded if they died while in hospital, if there were multiple admissions (i.e. only 1 admission was counted) and if they were not residents of the Calgary Health region. From the total 675 patients admitted, 245 charts met the study inclusion criteria and a convenience sample of 100 cases was randomly selected from the 2-65 age group. All persons aged greater than 65 years met the age criteria for vaccination, regardless of any other comorbidities and hence would easily be identified. However, we wished to see if the administrative data was accurate in coding for the other diagnoses indicating eligibility for vaccination. Using the patient chart number captured in the administrative data records each of the 100 charts of persons aged 2-65 was reviewed to determine if the patient was eligible to receive the pneumococcal vaccine based on Canadian Task Force Guidelines (Table 1).

Table 1. Canadian immunization guidelines (adapted from [22, 23])

- All persons ≥ 65 years of age
- All persons > 2 years with asplenia, splenic dysfunction or sickle cell disease
- All persons > 2 years of age with the following conditions: chronic cardiopulmonary disease (except asthma), cirrhosis, alcoholism, chronic renal disease, nephrotic syndrome, diabetes mellitus, chronic CSF leak, HIV infection and other conditions associated with immunosuppression (Hodgkins disease, lymphoma, multiple myeloma, induced immunosuppression for organ transplantation)

Specific documents reviewed included discharge summaries, admission notes, physician daily progress notes and medication list. Each case was then labeled as eligible versus ineligible. The reviewer was blinded to the hospitals original coding of medical conditions and spent approximately 15 minutes reviewing each chart. The three reviewers were practicing physicians and investigators on this study who independently reviewed the charts. Any disagreements were easily reconciled since the criteria was usually very clear.

7. Results

Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated using the chart review data as a criterion standard and the administrative database search as the diagnostic test. 

From the total 675 patients admitted to the MTU from January-December 2001, 245 cases met study inclusion criteria. A random sample of 100 patients aged 2-65 was selected from the possible 245 cases. According to our chart review sixty-four cases were found to be eligible but only fifty-three of these were identified by the administrative database, (i.e. 11 were missed). Concomitantly the database inappropriately selected 8 patients as eligible who were deemed ineligible via chart review by the team of three expert reviewers. A re-review of the full chart was not conducted, however the reviewer summary notes were consulted to verify that these were correctly classified by the reviewer. Accepting the chart review data as criterion standard the calculated sensitivity was 83%, specificity 78%, PPV 87% and NPV 72%. Sensitivity is reported as a measure of the accuracy of capture by administrative database search when the chart reviewer identified these cases as eligible for pneumococcal vaccination. Alcohol abuse appeared to be a frequent comorbidity not identified.

Specificity was calculated to determine the accuracy of administrative database reporting of cases not eligible for vaccination when these conditions were not reported in the chart. We note that three of the eight cases inappropriately identified by the database were found to be admissions for pulmonary embolus.

Alcoholism has been identified as a risk factor for the development of pneumonia. According to a recent study by Quan et al [7], an alcohol-related diagnosis was significantly more likely to be present in pneumonia cases caused by *S. pneumoniae*, increased the likelihood of longer inpatient stays and an intensive care unit stay and was associated with higher hospital charges. Not surprisingly the Canadian Task Force Guidelines have recommended that all alcoholic patients receive the pneumococcal vaccine (see Table 1). To increase our sensitivity for identification of such high-risk individuals we incorporated previously published and validated alcohol ICD-9 diagnostic codes in our database search criteria.

Despite inclusion of these diagnostic codes, we noted that alcoholism was still the most common comorbidity missed by the administrative database search. Four cases with an alcohol-related diagnosis
were not captured by the database. The most likely explanation is possible errors in coding accuracy, an issue of concern considering that alcohol abuse is a significant comorbid factor in pneumonia.

In other validity studies of administrative data, specificity is uniformly high indicating that coders rarely code a diagnosis as present when it really is not. This study found a proven specificity of 78% for identification of persons eligible for vaccination, in that only eight cases were inappropriately captured by the database. Interestingly, three of these eight cases were admitted for treatment of pulmonary embolus. In one of these cases we noted an additional diagnosis of fatty liver disease. It is possible that this may have been misclassified as chronic liver disease, or alcoholic fatty liver disease. Similarly, this could explain inclusion of a patient with a comorbid diagnosis of “increased liver enzymes” and an unwanted side effect of using the additional alcohol-related diagnostic codes to increase our diagnostic test sensitivity. In another case admitted for urticara and angioedema the patient had been treated with high-dose steroids. Perhaps this case had been miscoded as chronic steroid use, and hence immunosuppression being identified as one of the criteria for vaccination. However, after careful review of the other diagnostic codes used in the search, the authors did not find any other comorbid diagnosis to explain inclusion by the database.

8. Discussion

The current study points to the efficacy of administrative data as a potential basis for determining eligibility for pneumococcal vaccination. However, any application of these results must recognize the technical, organizational and data quality issues that arise in the use of administrative data.

From a technical standpoint most organizations are well positioned. First, the needed data are readily available in a cost effective electronic format. This is in sharp contrast to the paper based systems that required labour intensive and thus expensive implementations. Second, the databases in which this data are stored have universal capabilities to query the data in order to answer key questions based upon predefined criteria. Finally, administrative systems, since they are essentially accounting information systems, are universally designed to facilitate the generation of reports and it is this capability that forms the basis for the solutions proposed here.

Organizational change is often the most difficult (albeit the most effective) means of implementing process improvements within organizations. Difficulties with implementation become particularly pronounced in environments where data are to be shared between professional jurisdictions where there are cultural or legal barriers preventing such flow. Even when data are shared, the time demands upon individual practitioners often negates the use of the information in a timely manner, if at all. Therefore any implementation employing administrative data needs to account for cross-boundary considerations and the nature of the action required. That is, any follow-up requirements should be cost effective and facilitate an effective use of professional time.

Concerns have been raised around the applicability of administrative data in clinical practice since it is assumed that the nature and quality of administrative data is not at a level sufficient for this innovative application. However, as this study demonstrates, administrative data can be more than adequate for certain clinical applications such as a cost-effective semi-automated reminder system. It was demonstrated that the administrative data employed in this study has 83% sensitivity and 78% specificity, which it is argued is not atypical and more than adequate for an effective vaccination reminder system. Importantly, in the current era of increased awareness of health economics, we note that such information can be obtained at minimal cost.

The authors feel that a semi-automated reminder strategy could have tremendous potential to improve rates of vaccination and decrease burden of disease on a large scale, even with possible implementation of a nation-wide monitoring system. Since the intelligent agent solution proposed exists as an autonomous software tool, once developed, it can easily be distributed across sites and database platforms in a cost shared fashion. That is, technical expertise at one site could be employed to develop the intelligent agent (e.g. bot), which could then be rolled out to other sites with minimal adaptation.

Within a particular site, a multi-tier strategy could be implemented which accounts for existing technical capabilities while maintaining consistency with existing work practices. At the most cost effective level, individual reports could be generated as part of the standard discharge process. These reports would be included with the discharge summary sent to the family doctor based upon an automatically generated report using predefined queries. The final decision to recommend vaccination is still with the family doctor. This represents the most cost effective first step since discharge reports are already being sent and the extra reporting is automatically generated so minimal additional costs are incurred. However, as noted above the time demands upon family doctors often results in these recommendations not being acted upon by the physician. Therefore, complementary strategies are suggested.

At the next level of escalation a monthly or quarterly report could be sent to public health
authorities which detail patients that could potentially benefit from pneumococcal vaccination in their area. A parallel summary report could be sent to the family doctor to serve as a reminder and also to provide an indication to the physician the scope of the problem in their area. This can heighten their awareness of the issue and facilitate planning activities such as appropriate levels of inventory of the vaccine to maintain and decisions to establish periodic vaccination clinics. These reports would be generated using the same criteria employed in the generating the individual discharge reports while being sorted by period and service region. Since some additional processing is required this complementary approach represents a moderate cost solution.

Finally, form letters could be generated and sent directly to the patient. These letters would be automatically personalized with the patients name and address and minimal details of their case while noting that they may benefit from pneumococcal vaccination with a recommendation to visit their family doctor and/or public health authority with contact information provided. This approach requires additional processing and contact with the patient and thus represents the highest cost solution, although likely still economically viable.

The strategy presented here represents a hybrid approach to integrating clinical and administrative data, while also providing the foundation for future implementation of more comprehensive integration projects. This study points to the possibility that even with the known quality problems of administrative data, the level of granularity needed for building a decision support tool, may still be sufficient for certain clinical applications. That is, technical issues and data quality may pose less of a barrier then the cultural and differing professional practice, which these systems tend to bring to the forefront.

9. Conclusions

A lack of empirical research has contributed to the assumption that administrative data is universally not applicable to clinical applications. This study demonstrates that administrative data is at minimum potentially applicable to certain clinical applications. Furthermore, data mining technologies, such as intelligent agents using administrative data can serve as the basis for a cost effective way to implement these solutions in distributable software for use in local or national clinical initiatives.

The authors feel that the next step in this program of research is to implement the proposed reminder system and evaluate its efficacy for pneumococcal and other vaccination requirements.

10. References


