An Antimicrobial Prescription Surveillance System That Learns from Experience

Mathieu Beaudoin, Froduald Kabanza, Vincent Nault, Louis Valiquette

■ *Inappropriate prescribing of antimicrobials is a major clinical concern that affects as many as 50 percent of prescriptions. One of the difficulties of antimicrobial prescribing lies in the necessity to sequentially adjust the treatment of a patient as new clinical data become available. The lack of specialized health-care resources and the overwhelming amount of information to process make manual surveillance unsustainable. To solve this problem, we have developed and deployed an automated antimicrobial prescription surveillance system that assists hospital pharmacists in identifying and reporting inappropriate prescriptions. Since its deployment, the system has improved antimicrobial prescribing and decreased antimicrobial use. However, the highly sensitive knowledge base used by the system leads to many false alerts. As a remedy, we are developing a machine-learning algorithm that combines instance-based learning and rule induction techniques to discover new rules for detecting inappropriate prescriptions from previous false alerts. In this article, we describe the system, point to results and lessons learned so far, and provide insight into the machine-learning capability.*

Inappropriate prescribing of antimicrobials is a major clinical problem and health concern, as well as a financial burden, in hospitals worldwide. It has been reported that as many as 50 percent of antimicrobial prescriptions are suboptimal or inappropriate (Dellit et al. 2007). Antimicrobial stewardship programs have been shown to reduce avoidable adverse effects (toxicity, antimicrobial resistance, *Clostridium difficile,* and others [Dellit et al. 2007; Valiquette et al. 2007]), improve patient health, and reduce unnecessary costs. However, these programs require the review of an overwhelming amount of clinical data by dedicated experts, which proves to be an obstacle in the current context of limited health-care resources. Therefore, hospitals are increasingly relying on automated decision support systems to review hospitalwide antimicrobial prescriptions.

The difficulties of antimicrobial prescribing lie in selecting the right antimicrobial therapy for the suspected pathogen and adjusting the dose and dosing frequency with an appropriate route of administration to ensure effective levels of

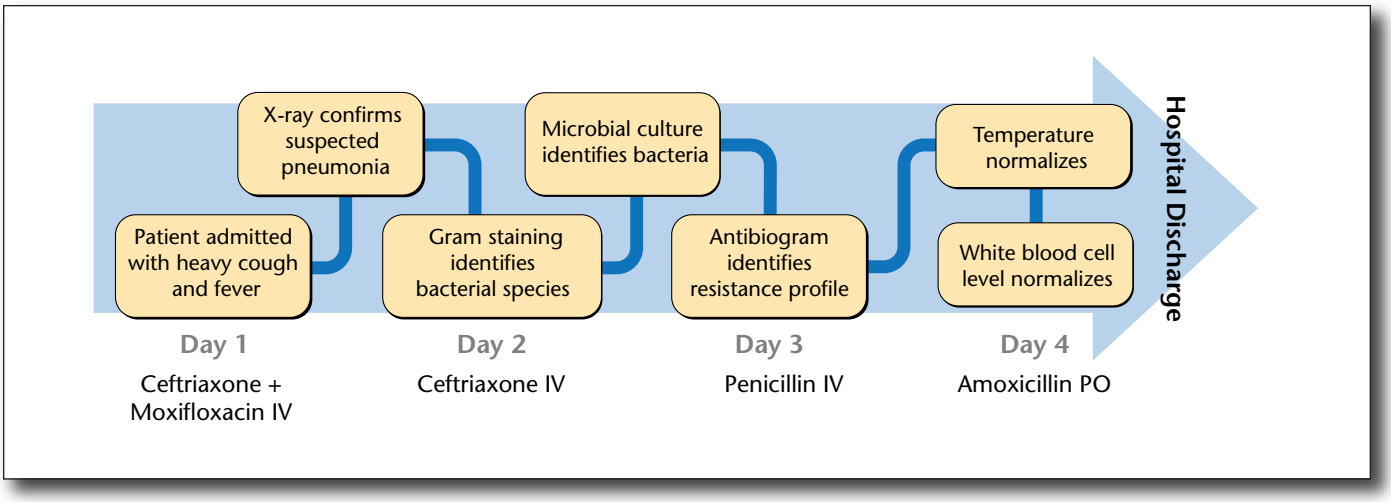

Figure 1. Example of Sequential Antibiotic Therapy.

medication are administered at the site of the infection. Inappropriate prescribing can range from selecting an antimicrobial to which the causative pathogen is resistant, making the treatment ineffective and endangering the life of the patient, to prescribing a given antimicrobial while another significantly less expensive but equally effective and safe alternative is available. Moreover, a selected antimicrobial therapy will be valid over a finite period of time; after selecting an initial empiric treatment, the physician must review his or her earlier prescription to account for newly available information.

As an example, consider the scenario of sequential therapy depicted in figure 1. A patient is admitted at the emergency room with a heavy cough and fever. The physician suspects a severe community-acquired pneumonia and confirms its diagnosis using a chest X ray. He begins an empiric broad spectrum therapy of ceftriaxone and moxifloxacin administered intravenously to cover a wide number of bacteria types associated with pneumonia. The physician requests sputum samples to be tested for common bacterial causes of pneumonia. A Gram stain is performed first and identifies the presence of Gram positive diplococci (suggestive of *Streptococcus pneumoniae*) allowing to refine the antimicrobial therapy. A microbial culture confirms within 48 hours the presence of *Streptococcus pneumoniae*. The antibiotic sensitivity profile of *Streptococcus pneumoniae* follows shortly demonstrating high susceptibility to penicillin. The prescribing physician is now able to select the most appropriate antimicrobial for this specific strain. By the fourth day, the clinical status of the patient improves; the prescribing physician can perform an early switch from an intravenous to oral antimicrobial therapy. This allows the patient to be discharged earlier with a less costly, but equally effective, oral treatment that he or she will pursue at home. Errors could occur at any point of this sequential therapy, including unduly delaying or even forgetting to make therapy adjustments (for example, keeping the patient under intravenous therapy after his or her clinical status improved sufficiently and allowed for an early switch to oral therapy).

For the most part, prescription monitoring systems use a knowledge base (KB) of rules acquired from published guidelines and experts to detect inappropriate prescriptions and prevent potential adverse events. Local and commercial solutions are generally characterized by highly sensitive rules with poor precision that trigger a high rate of clinically unhelpful alerts (Hsieh et al. 2004; Reichley et al. 2005). This high rate of false alerts impedes their use. The problem comes from the inability to create a complete and precise KB and the tendency to otherwise use a "totally inclusive" KB. It is difficult to model all variables that a prescribing physician will take into account, let alone model the decision-making process. Antimicrobial prescribing is a subjective process where physicians continually rely on their experience to select an effective treatment and prevent adverse events. In addition to published guidelines, hospitals have their own local practices (Reichley et al. 2005) that must be covered by these rules.

We have developed a new antimicrobial prescription monitoring system called APSS — antimicrobial prescription surveillance system. APSS was integrated successfully into a Canadian academic medical center. Like other antimicrobial prescription monitoring systems, APSS uses a KB that suffers from a high rate of false alerts. However, unlike any other previous system, APSS is able to learn new prescription surveillance rules. This learning capability is designed to allow APSS to self-reconfigure to local practices after deployment and to self-improve its KB over the long term supervised by user feedback. Although the application of machine learning to clinical temporal data is not new, to the best of our knowledge, this is its first application to the monitoring of drug prescriptions.

Taking the temporal nature of antimicrobial prescribing into account, APSS uses a supervised learning algorithm for discovering rules that classify temporal sequences. We approach the problem as a binary classification task where rules are used to identify "bad" temporal sequences that contain an inappropriate prescription. The algorithm we use is a combination of rule induction and instance-based learning methods that uses nearest neighbor classification with a distance function on both temporal and nontemporal parameters.

In the next section, we give an overview of APSS and discuss its development, deployment, and evaluation thus far. We then describe the ongoing development of the machine-learning extension. We follow with a presentation of preliminary results for this learning capability and conclude with ongoing and future work.

Antimicrobial Prescription Surveillance System

APSS is currently deployed at the Centre Hospitalier Universitaire de Sherbrooke, a 713-bed Canadian academic medical centre. It assists hospital pharmacists in their antimicrobial stewardship activities by identifying mismatches between prescribed antimicrobials and published and local guidelines. APSS monitors the clinical information of a patient; as new information becomes available, it verifies that the ongoing treatment remains appropriate according to contraindications conveyed by rules in its KB. The KB contains rules that account for approximately 50,000 contraindications related to inappropriate drug-drug interactions, drug-bug or drug-laboratory mismatches, cheaper alternatives, maximum daily dose, maximum and minimum dose and frequency, maximum duration, and route of administration.

APSS assists in the postprescription revision process, as illustrated in figure 2. The prescriber chooses an antimicrobial therapy after an initial assessment of the patient. The pharmacy department performs a posteriori computerized order entry into the electronic health-record system of the Centre Hospitalier Universitaire de Sherbrooke, QuadraMed Computerized-Patient Record. New and modified prescriptions are automatically sent to APSS along with all patient clinical data. APSS reviews these prescriptions using the most recent clinical information of a patient and produces documented alerts for potentially inappropriate prescriptions. The pharmacist first reviews these alerts and then contacts the prescriber by phone to recommend a prescription modification or discontinuation if deemed appropriate. If the prescriber accepts the recommendation, a new order is sent to the pharmacy and the cycle continues. The pharmacist records the validation of every revised alert.

The project began in 2005 with the objective of finding a solution to facilitate manual antimicrobial optimization. An intensive manual antimicrobial optimization program was implemented in 2004–2005 to control an outbreak of *C. difficile* infections at the Centre Hospitalier Universitaire de Sherbrooke by decreasing the use of high-risk antibiotics. Antimicrobial optimization decreased overall use of antimicrobials during this period (Valiquette et al. 2007) and the outbreak subsided. However, these measures required important resources that could not be sustained subsequently. After manual surveillance ended, overall antimicrobial consumption eventually returned and surpassed preoutbreak levels.

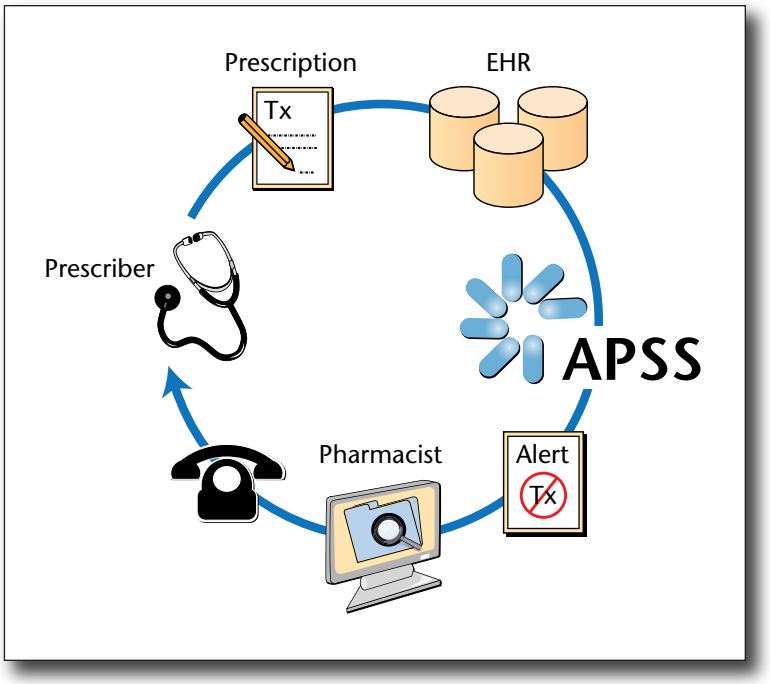

Figure 2. Asynchronous Surveillance Using APSS.

The solution put forward was to use an automated prescription monitoring system to facilitate and enhance the antimicrobial stewardship activities. We selected an asynchronous revision process because it dovetails nicely with the pen and paper prescribing practices that are still the norm at the Centre Hospitalier Universitaire de Sherbrooke. Synchronous validation at the time of computerized order entry was ruled out because it was perceived to hinder the workflow of prescribing physicians. Introducing APSS to the antimicrobial stewardship program was simple; we provided a tool that assists the hospital pharmacists already assigned to antimicrobial revision, resulting in a transparent integration into the existing prescribing practices.

We set about to develop APSS in 2007 and began with the complex task of developing its KB. Our multidisciplinary team consisted of one professor-researcher assisted by a doctoral candidate in computer science and a programmer-analyst, one infectious

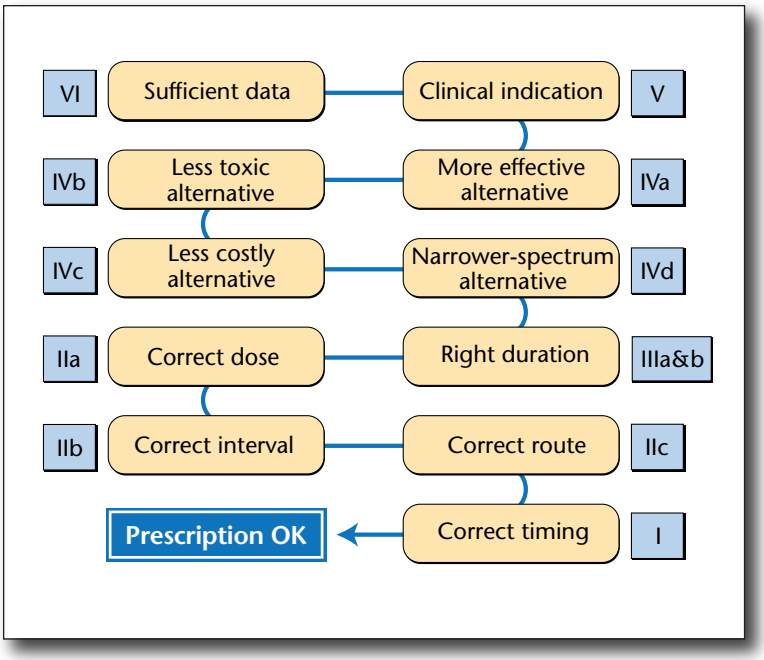

Figure 3. Standardized Model to Evaluate Antibiotic Prescriptions.

(Gyssens et al. 1992.)

diseases physician-researcher assisted by a doctoral candidate in clinical science, and one hospital pharmacist. We extracted rules from published and local guidelines and tested them retrospectively to validate the potential for prospective interventions.

The approach used by APSS to revise antimicrobial prescriptions is inspired from a standardized model for evaluating antibiotic prescriptions (Gyssens et al. 1992) and depicted in figure 3. Following this model, a prescribed antimicrobial therapy is considered valid unless it violates specified contraindications. The revision process consists of evaluating different parameters of importance for contraindications to the selected therapy. Prescriptions can be inappropriate according to different parameters, in which case multiple alerts are triggered. This revision process does not only occur at the time of prescription, but continues as additional clinical information becomes available. The antimicrobial therapy is in principle revised as additional clinical information comes in. Like it is illustrated in figure 1, the physician may diagnose a bacterial infection after an initial clinical evaluation of the patient, prescribe a particular antimicrobial therapy and, in the meantime, order microbiology tests. The results of these tests will inform on the specific species, type and sensitivity profile of the infecting bacteria. Assuming the profile indicates that the bacteria is resistant to the previously prescribed antimicrobial therapy, that is, the previously prescribed therapy is ineffective against this particular bacteria, then the prescription should be revised for a more effective alternative.

As illustrated in figure 3, prescription revision according to the Gyssens model begins by verifying whether there is sufficient available information for evaluating a prescription (Step VI). The revision process halts whenever critical information is missing. It then looks for clinical contraindications to the prescribed antimicrobial therapy (Step V). As new clinical information becomes available, the antimicrobial therapy evolves from the initial empiric broad spectrum therapy for a suspected infection to a specific narrow spectrum therapy targeted toward a specific pathogen (Step IVd). This new information may also indicate a more effective alternative (Step IVa), an equally effective but less toxic alternative (Step IVb), or an equally effective but less costly alternative (Step IVc). The total duration of the antimicrobial therapy is also monitored. A too short therapeutic course (Step IIIb) may lead to treatment failure whereas inadequately prolonged therapy (Step IIIa) increases the risk of adverse drug events and may be associated with increased antimicrobial resistance. Doses are validated (Step IIa) along with corresponding dosing frequency (Step IIb) to ensure effective serum concentration. The route of administration (Step IIc) must ensure delivery of sufficient concentration of antimicrobial to the targeted site. The timing (Step I) must also be questioned, since therapy administered too early or too late could prove useless, if not harmful.

The design of the KB for APSS follows the Gyssens model. Each step in this process corresponds to a set of contraindication rules to be checked. We characterized and segmented the antimicrobial domain knowledge according to these steps, regrouping rules to be evaluated together (for example, adequacy of dose and dosing interval). We also created higher level attributes, such as patient "type," derived from multiple attributes that represent clinical concepts used by experts when reasoning about antimicrobial therapies. Rules were created by category as a conjunction of propositions on relevant prescription parameters, patient data, patient type, and others The rule segmentation proved to be helpful in maintaining the KB. It also contributes to the efficiency of the system in that the firing of rules in each step concerns a small set of rules, as opposed to firing all the rules in one step.

The expert system for APSS (the KB and the algorithm for checking contraindication rules) was developed in C#.NET, using SQL-compatible database tables for the KB. The KB can be visualized and edited by using the knowledge management tool. The KB is for the most part maintained by clinical pharmacists. The maintenance in particular involves making changes that are required to reflect updates in published guidelines and local prescribing practices. Changes that involve new rule structures (for example, new parameters that are not accounted for in the KB) require the assistance of a computer programmer. Since the deployment of APSS in August 2010, the

interventions of a programmer have been quite limited. One instance occurred during a recent minor outbreak of *C. difficile* infections. The committee on health-care-associated infections requested the addition of temporary rules to monitor every high-risk antimicrobial.

APSS can communicate with the electronic health-record system through a data communication interface that we have developed. In collaboration with the Information Technology (IT) Department, we identified the required variables and normalized their values. We developed the exportation and importation interfaces using Health Level Seven (HL7) standards. We finally deployed the databases and knowledge management module of APSS. Support from the Centre Hospitalier Universitaire de Sherbrooke's decision makers and IT Department management was required to ensure advancement of these steps.

In August 2010, we finally deployed APSS at the Centre Hospitalier Universitaire de Sherbrooke. APSS was met by the prescribers and decision makers with unanimous appreciation and recognition. APSS increased the amount of prescriptions revised daily by the pharmacist. It also increased the impact of the antimicrobial stewardship program by enabling pharmacists to target patients who need it most. APSS enabled to extend antimicrobial surveillance from high-risk wards (for example, intensive care) to every bed of the two physical sites of the Centre Hospitalier Universitaire de Sherbrooke.

During a 53-week period, a clinical pharmacist used APSS an average of 15 hours per week. APSS evaluated 37,770 antimicrobial prescriptions. As summarized in figure 4, alerts were triggered for 10,837 (29 percent) of these prescriptions. However, alerts for 6,673 prescriptions could not be reviewed because they occurred outside of the allocated time. Of the 4,164 reviewed alerts, 2,820 (68 percent) were overridden by a pharmacist; 1,754 (42 percent) were considered clinically irrelevant while another 1,066 (26 percent) did not have sufficient clinical impact to justify modifying the current treatment (for example, alert triggered at the end of a prescription). The pharmacist contacted prescribers 1,344 times and they accepted 1,222 (91 percent) recommendations. This evaluation period was associated with a reduction of 13.5 percent in antimicrobial consumption and Canadian dollars (CAD) 305,000 (15 percent) in antimicrobial expenditures.

In September 2011, full-time surveillance began divided among a team of five pharmacists. As of September 2012, the override rate of alerts had subsided to 50 percent, with an acceptance rate by prescribers stable at 91 percent. APSS had contributed to 3,156 interventions, which were associated with reductions of 22 percent in intravenous antimicrobial consumption and CAD 688,000 in direct antimicrobial expenses. We are currently evaluating the impacts of APSS on patient health.

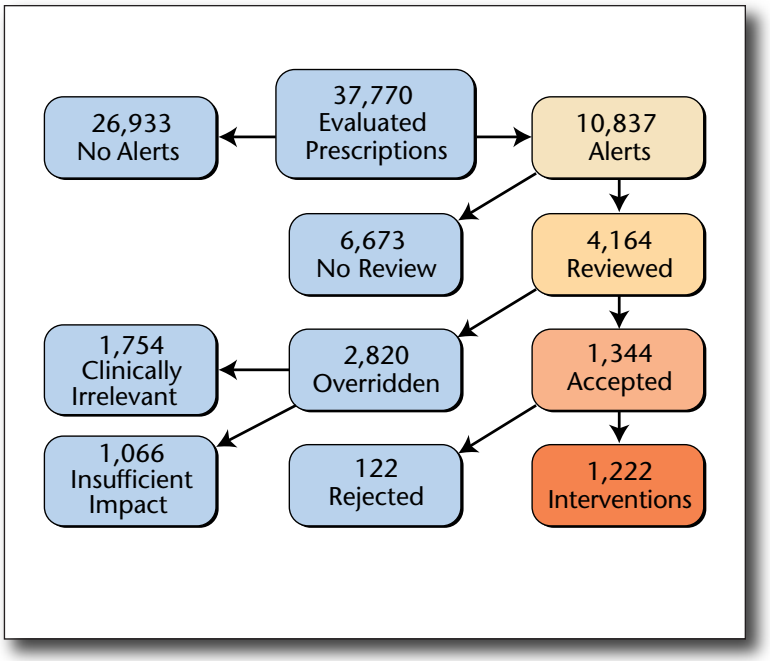

Figure 4. Alerts Triggered During a 53-Week Period.

APSS Learning Capability

The learning module of APSS aims to discover new rules for classifying inappropriate prescriptions supervised by user feedback, such as the rejection of false alerts or the identification of unflagged inappropriate prescriptions. From the beginning of the project we started investigating a mechanism to improve the KB of APSS from experience. The alerts were all documented, along with feedback from pharmacists and prescribers. We began implementing the learning module in 2011. It is not deployed and is still under testing. Here, we report on its preliminary results.

The learning module interacts with the other modules of APSS to discover and test new rules using patient data and revised alerts (see figure 5). The import module is responsible for acquiring and normalizing clinical data from the electronic health-record system and storing it into the databases of APSS. The review module uses the KB to review antimicrobial prescriptions and stores user feedback for revised alerts. The evaluation module produces custom reports on antimicrobial consumption, alerts, and others.

When deployed, we expect the learning module to enable users (such as pharmacists) to discover rules for specific antimicrobials and alert types. The user will use the evaluation module to identify commonly overridden alerts and patients who received specific antimicrobials and did or did not present the alerts of interest. This data will be used to create data sets for training and testing rule sets. Rules with sufficient recall (sensitivity) and accuracy will be sug-

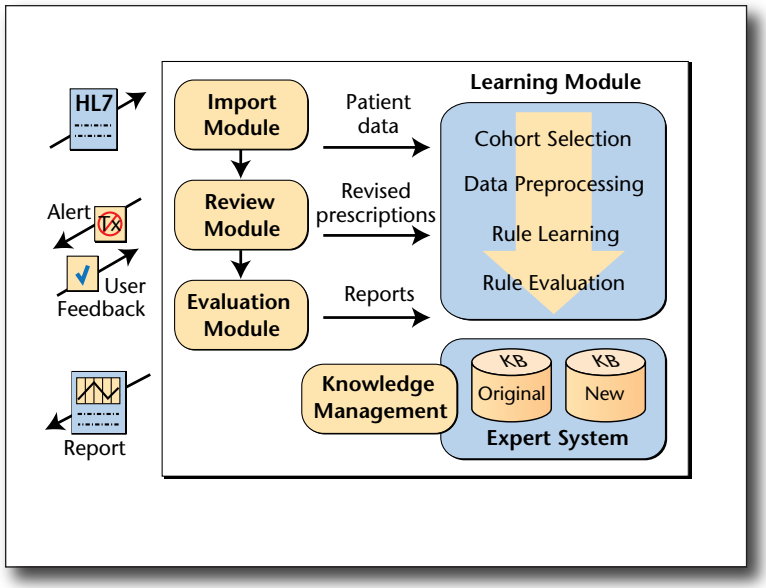

Figure 5. Overview of the Learning Module of APSS.

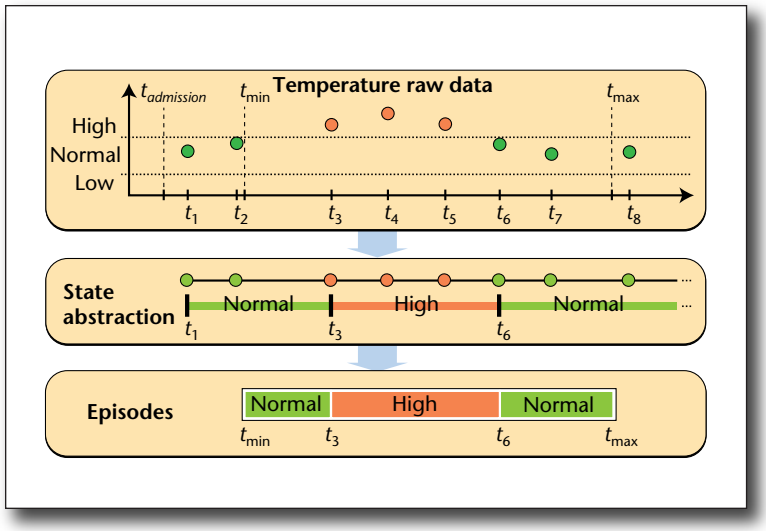

Figure 6. Example of State Abstraction.

gested for review. The user will then be able to add clinically relevant rules to the KB.

In addition to improving APSS in the long run in every hospital where it is deployed, this learning capability is expected to help configure APSS whenever it is deployed at a new hospital that has prescription practices differing from the Centre Hospitalier Universitaire de Sherbrooke where APSS is currently deployed. Another expected use of this learning module is data mining to discover unforeseen yet clinically relevant patterns of inappropriate prescribing that may be addressed by the stewardship program with targeted in-service training. One such pattern was discovered during our experimentation and is discussed in the Preliminary Results section.

Cohort Selection and Data Preprocessing

For the experiments discussed later in this article, we considered the following patient cohort: every adult inpatient (18 years of age and older) receiving at least one monitored antimicrobial admitted between January 1, 2012 and June 30, 2012. A cohort of 7,740 hospitalizations was created, consisting of 5,756 patients who received 19,172 antimicrobial prescriptions. Alerts were triggered for 7,027 prescriptions. We considered the following attributes: *gender, age,* body mass index *(BMI),* patient location *(ward),* temperature *(temp),* white blood cell count *(WCC),* neutrophil count *(neut),* creatinine clearance *(CrCl),* respiratory rate *(resp), pulse,* and blood pressure *(BP).* An attribute was also created for each medication, where prescriptions were described using their *name, dose, frequency,* and *route* of administration, as well as their revised alerts.

We used temporal abstraction (Shahar 1997) to extract a uniform and meaningful data representation from the raw clinical data of APSS. This raw data contains qualitative and quantitative attributes sampled with both time points (for example, *temp*) and time intervals (for example, *drug order).* Figure 6 illustrates the process of state abstractions for the raw temp time series. Quantitative thresholds are used to identify qualitative states that hold over a period of time, which we call episodes. We extracted a single sequence for each hospitalization. Within a hospitalization, the observation period was restricted to the ongoing antimicrobial of interest. We considered only data between the first (t_{min}) and last (t_{max}) administered dose. It ensured a common time zero (t_{min}) between sequences. We used a temporal granularity of 1 hour.

Selecting the Learning Algorithm

There are various applications of data mining and machine-learning algorithms to clinical temporal data. Association rule discovery has been used to gain insight into the causes of clinical events of interest (Bellazzi et al. 2005; Concaro et al. 2009); however, it is geared toward discovering rules for frequent patterns and performs poorly when addressing infrequent patterns (Zaki, Lesh, and Ogihara 2000), such as inappropriate prescriptions. It uses an Apriori-like strategy (Agrawal and Srikant 1994) with breadth-first search and candidate pruning based on support and confidence. The problem when looking for infrequent patterns is the necessity to lower support thresholds. It inefficiently prunes the candidate space and potentially leads to an intractable search space. It also produces an overwhelming amount of uninteresting patterns from which it is difficult to distinguish interesting ones (Zaki, Lesh, and Ogihara 2000).

Another method that was used to identify clinical events of interest is case-based reasoning. Whereas instance-based learning (Aha, Kibler, and Albert

1991) accumulates observed instances and classifies new instances using the nearest known ones, case-based reasoning (Aamodt and Plaza 1994) uses background information to create meaningful cases that are reused or adapted, in full or in part, to solve new problems. For example, case-based reasoning has been used to identify potential adverse drug events (Hartge, Wetter, and Haefeli 2006) and hemodialysis treatment failures (Montani, Portinale, and Leonardi 2006). While they are known to perform well with few instances, these algorithms are burdened with irrelevant attributes (Domingos 1996) and accumulate large quantities of cases. This is a problem when looking for a small set of highly accurate rules aimed at a human user.

A complementary approach to instance-based learning is rule induction. Rule induction disposes easily of irrelevant features, separates classes with good accuracy, and extracts a small set of rules that can lead to better predictions (Domingos 1996). However, it tends to be affected by a skewed distribution of classes and produces rules that favor the overrepresented classes (Chawla, Japkowicz, and Kotcz 2004).

Following Domingos (1996), we have chosen an algorithm that combines instance-based learning and rule induction. However, unlike Domingos (1996), which learns classification rules for a labeled set of nontemporal feature-value data, our algorithm was designed to learn classification rules for a labeled set of episode sequences in addition to nontemporal feature-value data.

Formulating the Learning Problem

Let us consider the attribute space A as the finite set of attributes for our domain and the feature space F as the finite set of qualitative values observed for these attributes. An *episode e* is defined as $<a, f, ts, te>$, where $(a = f)$ describes a symbolic state with $a \in A$ and $f \in F$ holding over the time interval $[ts, te[$. We refer to the attribute, feature, start, and end times of an episode as $e.a$, $e.f$, $e.ts$, and $e.te$, respectively. An example of episode from figure 6 is $<temp, normal, t_{min}, t_3>$.

A *sequence s* is defined by $\{e_1, \ldots, e_n \mid \forall i = 1, \ldots, n - 1 : e_i.ts \leq e_{i+1}.ts\}$, where $n = |s|$ represents the size of the sequence. We refer to the subsequence of s for the ith attribute $a_i \in A$ as $att_i(s)$ defined by $\{e_1, \ldots, e_m \mid \forall e \in att_i(s) : e \in s; e.a = a_i ; \forall j = 1, \ldots, m - 1 : e_j.te \leq e_{j+1}.ts\}$, where $m = |att_i(s)|$. A hospitalization is described as a *labeled sequence ls* defined as $\{id, s, l\}$, where *id* is a unique identifier, s is a sequence, and l is a class label that belongs to the finite set of class labels L. We focus on a binary-class problem where $L = \{appropriate, inappropriate\}$. We used revised alerts of APSS to label every sequence, where *inappropriate* indicates a true positive (alert that has been validated by a user) and *appropriate* indicates a negative (no alert) or false positive (alert that was rejected by a user).

We can now formally state the supervised machine-learning problem that concerns us. Given a finite training set of labeled sequences *TS,* discover a rule set R for inappropriate sequences. We only have two classes (appropriate and inappropriate). Learned classification rules identify inappropriate instances. The antecedent of a learned rule is a conjunction of propositions over time intervals, represented as episodes, whose satisfaction implies membership to the inappropriate class as the consequent.

```
initialize ruleset from inappropriate sequences
initialize distanceMatrix with pairwise distances
do
  for each rule r in ruleset do
    s = most similar appropriate sequence to r
    r' = modify conditions of r according to s
    update distanceMatrix according to r'
    if J-measure(r') > J-measure(r) then
      replace r by r' in ruleset
  end for each
while ruleset was improved
```

Figure 7. Pseudocode for the TIM Algorithm.

Temporal Induction of Classification Models

Our supervised learning algorithm, called temporal induction of classification models (TIM), combines instance-based learning and rule induction. Its main operations are depicted in figure 7. The rule set is initialized using all of the inappropriate sequences of the training set as maximally specific rules. The pairwise distances between the rules and sequences of the training set are computed and stored in a multidimensional distance matrix to reduce computation times. The rule refinement process aims to increase interclass distance by modifying every rule in parallel according to a rule's most similar appropriate sequence. These locally promising modifications are performed until they no longer improve a rule. Nearest neighbor classification is performed and rules are evaluated using the *J-measure* (Smyth and Goodman 1991), which quantifies the average information content of a rule.

We selected the *J-measure* for its ability to account for both simplicity and *goodness-of-fit,* measuring the probability of occurrence of a rule and its *cross-entropy* (Smyth and Goodman 1991). As a working hypothesis, a rule with high information content (that is, high probability and cross-entropy) is also likely to have a high predictive accuracy.

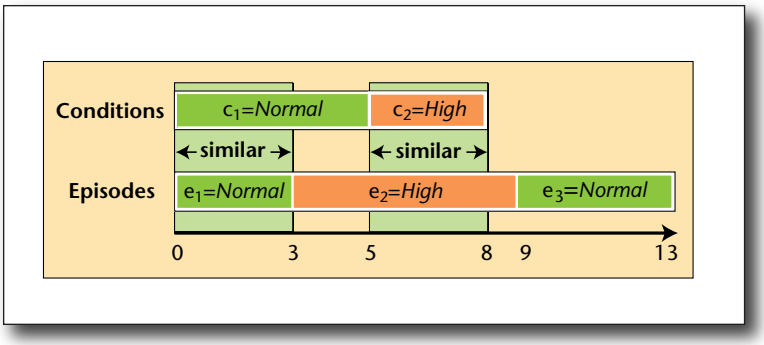

Figure 8. Example of Conditions and Episodes.

Classification

Classification is performed according to a distance function that measures the similarity between the conditions of a rule, represented by episodes, and the sequences of the data set. We use a nonsymmetric distance function where similarity is proportional to the number of conditions that a sequence satisfies. We consider that a sequence fully satisfies a condition if its episodes subsume the condition over its entire time interval, that is, a sequence is perfectly similar to a rule if it subsumes all of its conditions. Given a rule $r \in R$ with N_r attributes and a sequence $s \in TS$, the global *distance*(r, s) function is defined by equation 1. All attributes contribute equally to a normalized coefficient between [0, 1] that does not arbitrarily favor shorter rules, where 0 denotes perfect similarity. A sequence is covered by a rule if their distance is below a parameterized minimal distance threshold D_{min}, in which case it becomes labeled *inappropriate*.

$$distance(r,s) = \frac{\sum_{i=1}^{N_r} D_a(att_i(r), att_i(s))}{N_r}$$

The $D_a \in [0, 1]$ function measures the distance between the subsequences $att_i(r)$ and $att_i(s)$ for the ith attribute of r. If $att_i(s) = null$, $D_a = 1$, otherwise we use equation 2, which measures the distance between the conditions $c_j \in att_i(r)$ and episodes $e_k \in att_i(s)$.

$$D_a(att_i(r), att_i(s)) = \frac{|att_i(r)| - \left(\sum_{j=1}^{|att_i(r)|} \sum_{k=1}^{|att_i(s)|} S_F(c_j, e_k) \times S_T(c_j, e_k) \right)}{|att_i(r)|}$$

The similarity between the conditions of a rule and the episodes of a sequence considers both temporal and nontemporal parameters. The *feature similarity* function S_F measures the similarity between the symbolic features of c_j and e_k using the *overlap metric* where $S_F(c_j, e_k) = 1$ if $(c_j.f = e_k.f)$ and 0 otherwise. The *temporal similarity* function S_T is proportional to the temporal overlapping of e_k over c_j, as measured by equation 3. S_T returns a coefficient between [0, 1], where 1 implies $[c_j.ts, c_j.te[\subseteq [e_k.ts, e_k.te[$.

$$S_T(c_j, e_k) = \frac{[c_j.ts, c_j.te[\cap [e_k.ts, e_k.te[}{[c_j.ts, c_j.te[}$$

Consider the attribute-specific subsequences of figure 8. The antecedent of a rule $att_i(r)$ with conditions c_1 and c_2 overlaps the subsequence $att_i(s)$ with episodes e_1, e_2, and e_3. Conditions c_1 and c_2 are interpreted as follows: for attribute att_i (for example, *temp*), value *Normal* should be observed during time interval [0, 5[and value *High* should be observed during time interval [5, 8[. Episode e_1 partially satisfies condition c_1, whereas episode e_2 fully satisfies condition c_2. Their distance reflects this partial similarity with a coefficient of 0.2, computed as follows:

$$\begin{aligned} &D_a(att_i(r), att_i(s)) \\ &= \frac{2 - \left(\sum_{j=1}^{2} \sum_{k=1}^{3} (S_F(c_j, e_k) \times S_T(c_j, e_k)) \right)}{2} \\ &= \frac{2 - ((1 \times 0.6) + (0 \times 0.4) + (1 \times 0) + (0 \times 0) + (1 \times 1) + (0 \times 0))}{2} \\ &= 0.2 \end{aligned}$$

Refinement of the Rule Set

The intuition behind this rule refinement process is that increasing interclass distance creates more accurate rules. Each iteration provides a set of locally promising modifications selected according to a rule's most similar appropriate sequence. Rules are modified by removing the temporal overlapping between a similar condition c and episode e, resulting in a modified condition c' being either entirely removed or subsumed by c.

For example, modifying the conditions of figure 8 according to episodes e_1, e_2, and e_3 reduces the time interval of c_1 from [0, 5[to [3, 5[and completely removes c_2. Consequently, the distance between these subsequences increases from from 0.2 to 1, as illustrated in figure 9.

Preliminary Results for the Learning Capability

For a preliminary evaluation of the learning capability, we tested APSS with learning rules that identify an early switch from intravenous to oral therapy, a key intervention in antimicrobial prescribing used at the end of the example presented in figure 1. A clinically valid recommendation requires the following indications: 72 consecutive hours of intravenous antimicrobial therapy, 48 hours of stabilized state of health (for example, normal levels of temperature and white blood cell), and 24 hours of concurrent oral therapy. The rules for recognizing patients who are eligible for an early switch involve nontrivial temporal constraints, making them a good test case for

the learning algorithm. In this experiment, this rule is not specified. The data set only contains inappropriate and appropriate labels specifying if a hospitalization contains or not a validated recommendation for early switch therapy. The objective is to demonstrate that our algorithm is capable of learning these clinical indications from alerts.

We created two data sets of different sizes and ratios of inappropriate sequences. The first was created with patients who received piperacillin-tazobactam (TAZO), the most prescribed intravenous antimicrobial at the Centre Hospitalier Universitaire de Sherbrooke. We created a smaller data set with patients who received metronidazole (METRO), an antimicrobial predominantly prescribed orally. They were partitioned into training and test data sets. Figure 10 describes their number of episodes, sequences, inappropriate sequences, and attributes. APSS preprocessed these data sets in 121.9 and 6.9 seconds, respectively.

TIM extracted an accurate and sensitive set of 35 rules. While precision was lower, it remained above APSS without TIM. Rules were presented to an infectious diseases specialist who evaluated their clinical relevance using a five-point Likert scale ranging from 1 — no relevance to 5 — excellent relevance. Excellent relevance required the presence of all three indications for early switch therapy. Figure 11 presents the scores; 63 percent of the rules were found to be clinically relevant (score ≥ 3). Interestingly, rules with high relevance scores also had the highest information content *(J-measure)*. On the other hand, rules with a relevance score of 1 were very specific and covered less than 1 percent of the test set. Removal of these rules from the rule set leads to little loss in alert coverage and accuracy.

Our expert noted the presence of patterns describing patient profiles associated with early switch therapy. Consider the conditions of the rule in figure 12 with a relevance score of 5. Noting *C* the conjunction of these conditions, we have the rule $C \rightarrow$ *inappropriate*. The three direct indications for early switch therapy are respected with prolonged intravenous (IV) treatment, normal levels of white blood cell (WCC), and concurrent oral treatment. However, it is complementary information observed in the rules that provided interesting insights into the prescribing practices of the Centre Hospitalier Universitaire de Sherbrooke. For example, in this rule, prolonged stay at the emergency room (ER), an older patient, salbutamol, and additional antimicrobial coverage with ciprofloxacin may indicate suspicion of pneumonia caused by resistant pathogens. Ten other rules targeted patients under postoperative antimicrobial prophylaxis, a practice not supported by medical evidence that will be addressed by the stewardship program. We also found that eight rules targeted patients with BMI ≥ 40. It could suggest that extended intravenous treatments are prescribed for very severely obese patients to ensure targeted concentrations are achieved. These patient profiles provide insight into the prescribing practices of the hospital and are of high interest for further investigation, as they identify subgroups of patients that could require closer monitoring or wards that could benefit from targeted in-service training.

We also compared TIM to three well-known learning methods for this type of problem. Figure 13 reports their respective number of rules, computation time, precision, recall, and accuracy. The first method used retrieval-only instance-based learning (IBL) where every known inappropriate sequence is used as a rule. The second method (CRL) used CN2's (Clark and Niblett 1989) general-to-specific search where individual rules are created by iteratively selecting the "best" condition. Conditions are added

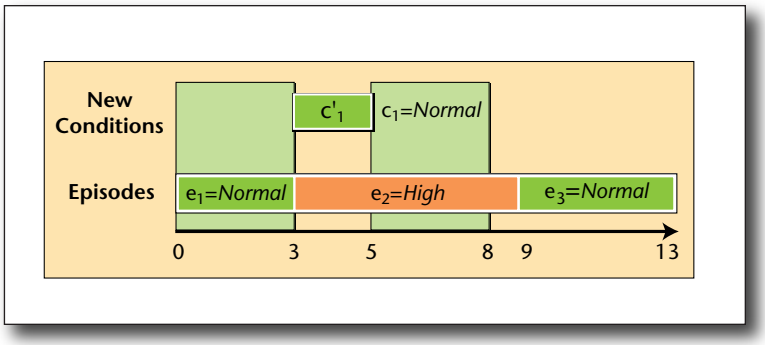

Figure 9. Example of Modified Conditions.

	Dataset	Episodes	Seq.	Inappr.	Attr.
METRO	Training	9,176	132	12	1206
	Test	19,182	278	46	
TAZO	Training	37,428	485	190	1581
	Test	68,188	947	413	

Figure 10. Description of Two Real World Data Sets.

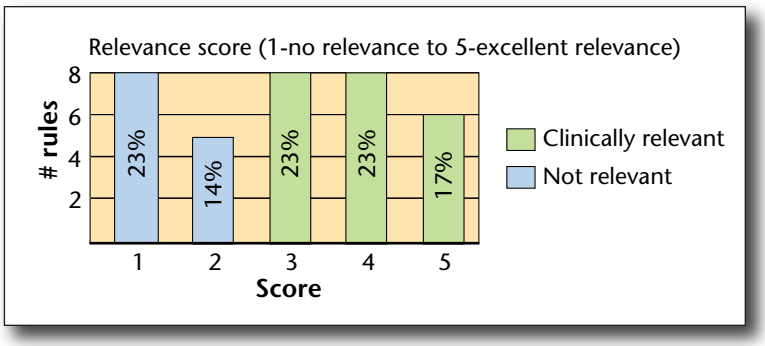

Figure 11. Relevance Score of the 35 Extracted Rules.

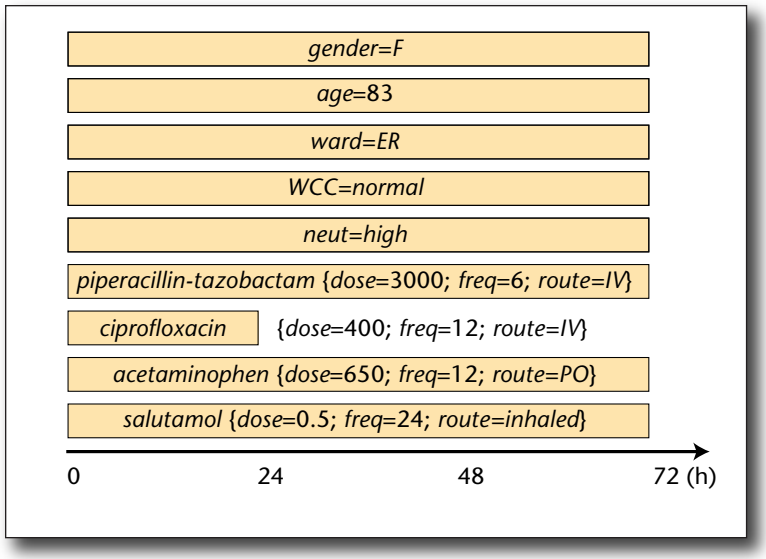

Figure 12. Conditions of a Clinically Relevant Rule.

	Algo.	#rules	Time (s)	Prec.	Rec.	Acc.
METRO	TIM	5	< 1	54	76	85
	IBL	12	< 1	30	96	63
	CRL	1	46	57	76	86
	ARM	8,074	15	44	33	82
TAZO	TIM	30	74	63	99	74
	IBL	190	10	59	99	70
	CRL	6	583	71	88	79
	ARM	614,652	17,865	66	81	74

Figure 13. Comparison of Four Algorithms on Two Data Sets.

to a rule until they no longer improve its *J-measure* or until every inappropriate instance is covered. The third method used an association rule mining (ARM) approach based on Apriori (Agrawal and Srikant 1994). Various strategies were used in CRL and ARM to focus on highly predictive rules for the inappropriate class. For example, ARM used candidate pruning on both support (METRO: $supp \geq 0.015$; TAZO: $supp \geq 0.02$) and confidence ($conf \geq 0.75$), and eliminated dominated patterns (Zaki, Lesh, and Ogihara 2000). We restricted ARM to rules of size 4 for the TAZO test.

Overall, TIM achieved relatively similar or better recall and accuracy than CRL and IBL, except for the recall metrics in METRO. IBL succeeded in classifying correctly most unseen inappropriate sequences in both tests. However, the wide coverage of its rules also incorrectly classified several appropriate sequences, penalizing greatly its precision and accuracy. In contrast, CRL achieved good accuracy on both data sets with fewer inappropriate prescription rules. However, they identified fewer inappropriate sequences in both tests. TIM combines the strengths of both previous methods. Performing a specific-to-general search and modifying every rule in parallel according to appropriate sequences speeds up the process, enabling TIM to outperform CRL by up to two orders of magnitude. TIM in addition better succeeds in identifying inappropriate sequences with equal or higher recall, without sacrificing much accuracy. Furthermore, TIM harnesses the ability of CRL to extract fewer rules than IBL. ARM performed poorly, being 30 to 200 times slower than TIM and producing many more rules, requiring heavy postprocessing to identify a subset of accurate rules.

Conclusion and Future Work

In this article, we presented APSS, a clinical decision support system that evaluates antimicrobial prescriptions and produces alerts for potentially inappropriate ones. Since its deployment in August 2010, APSS has been met by prescribers and decision makers with unanimous appreciation and recognition. We also presented an emerging machine-learning capability for APSS. The learning capability combines instance-based learning and rule induction to learn prescription classification rules from user feedback.

We discussed preliminary results demonstrating the rule-learning capability for appropriate early switch from intravenous to oral antimicrobial therapy. A majority of learned rules were found to be clinically relevant because they succeeded in identifying the clinical indications for early switch therapy. From these rules, a clinician identified emerging patient profiles associated with early switch recommendations providing further insight into the local prescribing practices and a potential for targeted interventions (for example, unsupported use of postoperative antimicrobial prophylaxis).

The next step is to pursue the experimentation of the learning capability before its release in the deployed version of APSS. Users will then be able to utilize the learning module to explore different rule sets and keep the rules they find clinically relevant and accurate. Learning from imbalanced data sets, where there are many more instances of some classes than others, is an important issue in domains such as ours, where inappropriate prescriptions are more the exception than the norm. Although the preliminary results of our algorithm seem encouraging, we have not yet characterized the algorithm with respect to the imbalanced data problem. This is on our agenda for future work. Other methods of temporal data mining could be integrated into the knowledge management tools in order to explore the vast quantity of data that we are accumulating and identify interesting patterns (that is, repetitive behaviors of interest) that could be further investigated by the antimicrobial stewardship team.

In the meantime, we are in the process of exporting APSS in other centers where we believe it will help reduce inappropriate antimicrobial prescribing and improve patient health. The revision process used by APSS could also be adapted to other drugs since it already manages the prescriptions of a patient, its vital signs, and laboratory and microbiology test results.

Acknowledgments

This project was partially funded by the Fonds de recherche du Québec – Santé, the Fonds de recherche du Québec – Nature et technologies, the Natural Sciences and Engineering Research Council of Canada, and the Centre Hospitalier Universitaire de Sherbrooke.

References

Aamodt, A., and Plaza, E. 1994. Case-Based Reasoning: Foundation Issues, Methodological Variation and System Approaches. *AI Communications* 7(1): 39–59.

Agrawal, R., and Srikant, R. 1994. Fast Algorithms for Mining Association Rules in Large Databases. In *Proceedings of the 20th International Conference on Very Large Data Bases,* 487–499. San Francisco: Morgan Kaufmann Publishers Inc.

Aha, D. W.; Kibler, D.; and Albert, M. K. 1991. Instance-Based Learning Algorithms. *Machine Learning* 6(1): 37–66. dx.doi.org/10.1007/BF00153759

Bellazzi, R.; Larizza, C.; Magni, P.; and Bellazzi, R. 2005. Temporal Data Mining for the Quality Assessment of a Hemodialysis Service. *Artificial Intelligence in Medicine* 34(1): 25–39. dx.doi.org/10.1016/j.artmed.2004.07.010

Chawla, N.; Japkowicz, N.; and Kotcz, A. 2004. Editorial: Special Issue on Learning from Imbalanced Data Sets. *ACM SIGKDD Explorations Newsletter* 6(1): 1–6. dx.doi.org/10.1145/1007730.1007733

Clark, P., and Niblett, T. 1989. The CN2 Induction Algorithm. *Machine Learning* 3(4): 261–283. dx.doi.org/10.1007/BF00116835

Concaro, S.; Sacchi, L.; Cerra, C.; Fratino, P.; and Bellazzi, R. 2009. Mining Healthcare Data with Temporal Association Rules: Improvements and Assessment for a Practical Use. In *Proceedings of the 12th Conference on Artificial Intelligence in Medicine,* AIME 2009, 16–25. Berlin: Springer-Verlag.

Dellit, T. H.; Owens, R. C.; McGowan, J. E.; Gerding, D. N.; Weinstein, R. A.; Burke, J. P.; Huskins, W. C.; Paterson, D. L.; Fishman, N. O.; Carpenter, C. F.; Brennan, P. J.; Billeter, M.; and Hooton, T. M. 2007. Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America Guidelines for Developing an Institutional Program to Enhance Antimicrobial Stewardship. Clinical Infectious Diseases 44(2): 159–177. dx.doi.org/10.1086/510393

Domingos, P. 1996. Unifying Instance-Based and Rule-Based Induction. *Machine Learning* 24(2): 141–168. dx.doi.org/10.1007/BF00058656

Gyssens, I. C.; van den Broek, P. J.; Kullberg, B. J.; Hekster, Y.; and van der Meer, J. W. 1992. Optimizing Antimicrobial Therapy. A Method for Antimicrobial Drug Use Evaluation. *Journal of Antimicrobial Chemotherapy* 30(5): 724–7. dx.doi.org/10.1093/jac/30.5.724

Hartge, F.; Wetter, T.; and Haefeli, W. E. 2006. A Similarity Measure for Case Based Reasoning Modeling with Temporal Abstraction Based on Cross-Correlation. *Computer Methods and Programs in Biomedicine* 81(1): 41–8. dx.doi.org/10.1016/j.cmpb.2005.10.005

Hsieh, T. C.; Kuperman, G. J.; Jaggi, T.; Hojnowski-Diaz, P.; Fiskio, J.; Williams, D. H.; Bates, D. W.; and Gandhi, T. K. 2004. Characteristics and Consequences of Drug Allergy Alert Overrides in a Computerized Physician Order Entry System. *Journal of the American Medical Informatics Association* 11(6): 482–91. dx.doi.org/10.1197/jamia.M1556

Montani, S.; Portinale, L.; and Leonardi, G. 2006. Case-Based Retrieval to Support the Treatment of End Stage Renal Failure Patients. *Artificial Intelligence in Medicine* 37(1): 31–42. dx.doi.org/10.1016/j.artmed.2005.06.003

Reichley, R. M.; Seaton, T. L.; Resetar, E.; Micek, S. T.; Scott, K. L.; Fraser, V. J.; Dunagan, W. C.; and Bailey, T. C. 2005. Implementing a Commercial Rule Base as a Medication Order Safety Net. *Journal of the American Medical Informatics Association* 12(4): 383–9. dx.doi.org/10.1197/jamia.M1783

Shahar, Y. 1997. A Framework for Knowledge-Based Temporal Abstraction. *Artificial Intelligence* 90(1–2): 79–133. dx.doi.org/10.1016/S0004-3702(96)00025-2

Smyth, P., and Goodman, R. M. 1991. Rule Induction Using Information Theory. In *Knowledge Discovery in Databases*. Menlo Park, CA: AAAI Press / The MIT Press. 159–176.

Valiquette, L.; Cossette, B.; Garant, M. P.; Diab, H.; and Pepin, J. 2007. Impact of a Reduction in the Use of High-Risk Antibiotics on the Course of an Epidemic of *Clostridium difficile*-Associated Disease Caused by the Hypervirulent NAP1/027 Strain. *Clinical Infectious Diseases* 45 Suppl 2: 112–121. dx.doi.org/10.1086/519258

Zaki, M.; Lesh, N.; and Ogihara, M. 2000. Planmine: Predicting Plan Failures Using Sequence Mining. *Artificial Intelligence Review* 14(6): 421–446. dx.doi.org/10.1023/A:1006612804250

Mathieu Beaudoin is a doctoral candidate in computer science and is also pursuing a master's degree in business intelligence, both at the Université de Sherbrooke. He has a mixed background including a B.Sc. in computer science and an M.Sc. in health sciences. His work focuses on artificial intelligence and machine learning and how they can be used to provide new solutions to medical problems and to improve clinical practice. He is the chief technology officer and a cofounder of Lumed Inc., a company that develops software solutions to optimize clinical practice and improve the quality of patient care.

Froduald Kabanza is a full professor in the Department of Computer Science at the Université de Sherbrooke. He joined the Université de Sherbrooke in 1993, after a postdoctoral fellowship at the University of Waterloo (Canada). He completed his Ph.D. in computer science at the Université de Liège (Belgium) in 1992. His research focuses on automated planning algorithms, autonomous systems, and intelligent decision support systems. He is the chief executive officer and the founder of Menya Solutions Inc., a company that develops AI software that can automate or support decision processes for businesses, governments, and educational institutions.

Vincent Nault is a doctoral candidate in health sciences, Faculty of Medicine and Health Sciences, and is also pursuing a master's degree in business intelligence, both at the Université de Sherbrooke. His work focuses on the development of clinical knowledge bases and on the integration and evaluation of clinical decision support systems. He is the chief executive officer and a cofounder of Lumed Inc.

Louis Valiquette is an associate professor in the Department of Microbiology and Infectious Diseases at the Faculty of Medicine and Health Sciences of the Université de Sherbrooke. He is also a consultant in Medical Microbiology and Infectious Diseases at the Centre Hospitalier Universitaire de Sherbrooke. He also completed an M.Sc. in clinical epidemiology at the University of Toronto. His research interests are medical informatics, antimicrobial stewardship, and the development of clinical tools to improve the management of nosocomial infections. He is the chief medical officer and a cofounder of Lumed Inc.