Using Data Mining to Predict Errors in Chronic Disease Care

Ryan M. McCabe; Gediminas Adomavicius, PhD; Paul E. Johnson, PhD; Gregory Ramsey; Emily Rund; William A. Rush, PhD; Patrick J. O'Connor, MD, MPH; JoAnn Sperl-Hillen, MD

Abstract

Development of data mining technologies to predict treatment errors in populations of patients represents a major advance in patient safety research. In the work presented here, we create a simulation test environment using characteristic models of physician decision strategies and simulated populations of patients with type 2 diabetes, employ a specific data mining technology that predicts encounter-specific errors of omission in representative databases of simulated physician-patient encounters, and test the predictive technology in an administrative database of real physician-patient encounter data. Two dominant decisionmaking strategies with different rates of treatment errors are identified: "feedback strategies" that use the results of past actions to guide treatment decision, and "feedforward strategies" that make treatment decisions based on anticipated future patient states. Evaluation of data mining results shows that the predictive tools developed from simulated treatment data can predict errors of omission in clinical patient data. The methods developed in this work have the potential for wide use in identifying decision strategies that lead to encounter-specific treatment errors in chronic disease care.

Introduction

Reducing the incidence of medical errors is an ongoing area of interest to the health care community. Much of the focus in recent work has been on reducing errors of commission (inappropriate actions), such as amputating the wrong limb.[1] Less attention has been given to errors of omission (i.e., failure to act when patients are not at evidence-based goals), a particular area of interest in the treatment of chronic diseases, such as type 2 diabetes.[2] Two major challenges in preventing errors of omission are (1) identifying patterns of behavior that predict future errors and (2) determining appropriate actions to take to prevent errors when error-prone behavior patterns are identified. Such actions might include changing the way individual physicians manage patients and matching physicians' abilities to patients with characteristics they are most successful in treating.

The health care community is placing greater emphasis on the use of data sources to improve the quality of care delivered to patients.[3, 4] Available data sources include administrative and clinical records, particularly with the advent of the electronic medical record (EMR). However, extracting meaningful information from large databases is a challenging task. Data mining, the extraction of useful and potentially actionable information from data, has been identified as a tool for culling health care databases for information to improve the quality of care. The majority

of recent data mining research directed toward improving quality of care has focused on detecting outcomes associated with patient/illness characteristics.[5, 6] Less work has directly examined outcomes and physician actions.

Methods

The present study applies data mining to identify patterns of physician decisionmaking used to treat patients with the goal of predicting errors of omission. In this approach, we conducted a simulation study of the clinical environment of type 2 diabetes to model alternative physician treatment strategies and develop a representative database of treatment records reflecting the use of these strategies to treat populations of simulated patients. The resulting database was used to employ a specific form of data mining technology—decision trees—that enabled accurate prediction of errors of omission across a range of patients and physician treatment characteristics. The resulting decision trees were then evaluated by using them to predict errors in an administrative database of actual patient records.

Simulation Environment

The simulation environment used in the present work comprised a computational model for patients with type 2 diabetes, plus a set of physician decision strategies for treating patients with this disease. Clinical responses of individual patients and populations of patients' responses to treatment were modeled. Each part of the simulation environment is described here.

Patient model. A computational model of a patient with type 2 diabetes was developed in prior research.[7] This model is composed of an inference structure and a rule network that responds to treatment actions in a manner characteristic of diabetes patients in real clinical encounters. Changes in levels of blood glucose, blood pressure, and lipids are simulated in response to actions taken over time. The patient model accounts for specific effects on outcomes from various oral medications, insulin, and other treatments, including medical nutrition therapy. The model captures the physiologic effects of drugs in the form of dose-response curves taken from the clinical literature.[8, 9] The patient model includes three sources of individual patient variation: (1) seasonal fluctuations in adherence to the treatment regimen following a 1-year cycle,[10] (2) daily fluctuations in adherence,[11] and (3) an assigned rate of disease progression.[12]

Individual patients. Changes in patient blood glucose values (i.e., glycated hemoglobin and fasting plasma glucose, hereafter referred to as A1c and FPG, respectively) are computed in two parts: (1) the effect of treatment actions (hereafter referred to as "moves") is computed as determined by a dose-response table, where each medication dose in the table has a maximum expected effect on patient blood glucose values (i.e., A1c and FPG); and (2) the percent of maximum effect is computed at any point in time based on the time-course of respective medications. Time-dependent values for percent maximum effect of medications are computed for both A1c values and FPG values.

Adherence to a treatment regimen indirectly affects A1c level, where the level of adherence assigned to each patient determines the percentage of the regimen followed. For example, if a patient has a prescription for 2,000 mg of metformin and is modeled to be 75 percent adherent,

then the patient model receives 2,000 mg x 0.75, or 1,500 mg, of metformin throughout the period in question. Blood pressure and lipid values are similarly represented.

Patient populations. Characteristic properties of populations of patients with type 2 diabetes were modeled using an administrative database of treatment records. Working in conjunction with physicians from a major health care organization and their clinical database, 11 key health indicators for patients with diabetes were identified and represented: A1c, low-density lipoprotein (LDL), high-density lipoprotein (HDL), triglycerides, systolic blood pressure, weight, creatinine, height, depression, stress, and level of adherence to treatment regimen (estimated from the timeliness of prescription fills). Generated populations of synthetic patients in the study reported here are based on distributions of these attributes.

Physician Treatment Strategies

Physician process model. The physician process model (PPM) referred to in the present research is a dynamic decision model designed to bring patient blood glucose, blood pressure, and lipid values to evidence-based goals (e.g., A1c values <7 percent). The PPM manages patient state by processing information and making corresponding moves. Moves take three forms: (1) treatment moves, which include medications and referrals to specialists; (2) scheduling moves, which stipulate when future visits will occur; and (3) information seeking moves, which are made up of orders for medical tests (e.g., A1c) to support decisionmaking.

Because patient states change as a function of patient variables and decisions made by the physician model, each treatment strategy requires a series of real-time decisions. As described by Brehmer, making such decisions is analogous to a process of achieving control of a system to obtain desired outcomes (goals).[13] PPMs for treating patients with type 2 diabetes are based on the internal model control paradigm,[14] shown in Figure 1. This model is composed of processes and computations (represented as squares and circles, respectively). The system being controlled is a patient (model) with type 2 diabetes.

The "inverse model" shown in Figure 1 is a process that generates treatment moves, schedules visits, and orders tests. The inverse model determines what moves to make based on the patient's "distance to goal" and whether prior moves have had their expected effect on patient state, as

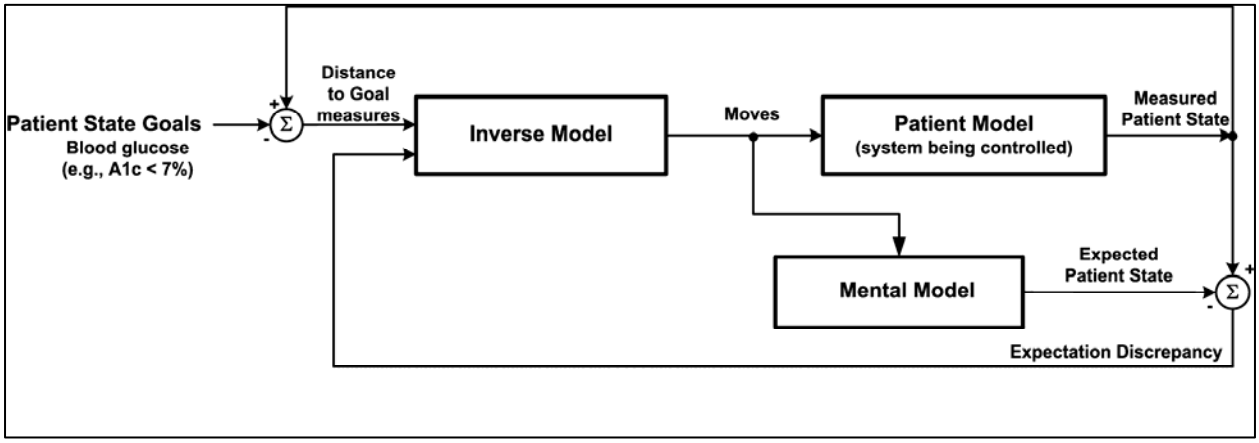

Figure 1. Physician process model for managing patients with type 2 diabetes.

measured by the "expectation discrepancy." Distance to goal is the amount of change in A1c (or other measures) required to bring the patient to the desired A1c level. Expectation discrepancy is defined as the difference between the patient's current A1c and expected A1c at a given point in time, as estimated by the mental model. The "mental model" represents parameters that affect estimations of patient state, including patient adherence to treatment (low or high) and dose-response curves (i.e., representations of patient responses to doses of various medications over time). The mental model generates an expectation of patient A1c as a consequence of moves that are made. The "patient model" receives moves from the inverse model and generates the next state of the patient being treated.

Brehmer has suggested two decision strategies to control dynamic processes: feedback and feedforward.[13] A feedback strategy uses information based on current patient state to make decisions. A feedforward strategy uses current patient state information *and* anticipated future patient states to make decisions. Two versions of the feedback strategy and two versions of the feedforward strategy were created for the purposes of the present study. These versions are labeled "weak" and "strong." Each version is described in detail below.

Feedback strategies. A common heuristic for feedback control from the literature on decision making is "anchoring and adjustment."[15, 16] "Anchoring" is accomplished by selecting a control variable to use as a basis for making decisions. The PPM uses previous blood glucose value as an anchor. "Adjustment" is the process of making decisions relative to past decisions. For the PPM, this equates to titrating (adjusting) the dose of a single medication.

The weak version of the feedback strategy represents a physician who infrequently titrates (adjusts) diabetes-related medications. This version of the feedback strategy involves encounters with the patient every 90 days, but treatment moves are made only when A1c is either above goal (i.e., A1c = 7 percent) and has stopped decreasing, or it is increasing compared to the previous encounter. The strong version of the feedback strategy involves encounters with the patient every 90 days; treatment moves are made at every encounter when the patient is above the A1c goal (7 percent). The 90-day scheduling parameter was chosen because most oral medications require 90 days for complete absorption.

Feedforward strategies. The feedforward strategy estimates the amount of medication needed to treat the patient, based on the distance of the patient's A1c from the goal. The feedforward strategy has a time constraint of 1 year to reach the A1c goal. Given the time constraint, the feedforward strategy determines whether enough visits can be scheduled to administer the amount of medication needed to reach the goal. If the strategy estimates that time is insufficient, given standard scheduling of visits and titration of single medications, the model would schedule visits more often, and when appropriate, would administer and titrate more than one medication (including insulin).

Unlike feedback strategies that use only A1c values in their decisionmaking, feedforward strategies use both FPG and A1c values when determining progress toward goal. Because feedforward strategies can make medication moves before prior moves have finished taking their full effect, these strategies must manage this "delay of feedback" by combining dose-response

curves in order to prevent overmedicating the patient. The two versions of the feedforward strategy manage this in different ways.

The weak version of the feedforward strategy calls for a single medication move at every encounter when the patient is above goal. When the patient's A1c is ≤7.5 percent, the weak version calls for treatment moves only when the A1c has stopped decreasing or is increasing, compared to the previous encounter. By using this rule, the weak version does not need to combine dose-response curves in order to avoid overmedicating the patient, possibly resulting in hypoglycemia. In the weak feedforward version, visits are scheduled every 60 or 90 days. This allows enough time between moves for medications to have the bulk of their effect, thus avoiding the need to combine dose-response curves.

The weak feedforward strategy uses a relatively simple (linear) representation of a dose-response curve to anticipate the amount of change in A1c, which is to be expected for a given treatment move. On each encounter, the weak feedforward PPM compares the anticipated A1c with the measured A1c to infer patient adherence (Figure 1).

The strong version of the feedforward strategy calls for moves at every encounter when the patient is above goal. This version of the feedforward strategy combines dose-response curves in order to estimate delayed effects of medications. Because this version of the feedforward strategy can combine dose-response curves, visits can be scheduled as frequently as every 30 days. If results are less than expected from a prior move, and if patient adherence is not an issue, an additional medication can be introduced. More than one medication can then be titrated during a given visit. The strong feedforward strategy has a more complex (piecewise linear) representation of the patient's responses to treatment moves, which is needed for combining dose responses from more than one treatment move.

Experiment

An experiment was conducted to determine the effect of the two versions of each model when treating simulated populations of diabetes patients. In the research reported here, the focus was on the way each PPM managed patient blood sugar values. (Blood pressure and lipid values were treated but are not reported.) All models were tested on their ability to bring patients' A1c values to <7 percent within 1 year. Four metrics were used to evaluate each of the models at the end of 1 year of treatment: (1) proportion of patients reaching goal, (2) average final A1c values, (3) average costs to treat patients, and (4) proportion of encounters with errors of omission.

Simulated patients. Using statistical characteristics in an existing database of real patients as a guide, a population of 10,000 synthetic patients was created. Patients were blocked into six cells (or groups) combining their initial A1c value (L, M, and H) and level of adherence (L or H). Initial A1c categories were defined as follows: (1) low A1c: A1c < 8 percent, (2) medium A1c: A1c between 8 and 10 percent, and (3) high A1c: A1c >10 percent. Adherence to treatment level was divided into two categories: (1) low adherence: adherence < 75 percent), and (2) high adherence: adherence >75 percent). Each of the six cells was labeled using a two-letter designator, where the first letter specified A1c range, and the second letter specified level of adherence (e.g., cell MH represents patients with medium initial A1c and high adherence to treatment).

The number (percentages) of simulated patients in each cell was as follows: LL, 900 (9 percent); LH, 6,000 (60 percent); ML, 400 (4 percent); MH, 1,900 (19 percent); HL, 200 (2 percent); and HH, 600 (6 percent). These numbers reflect the actual distribution of patients in the administrative (clinical) database. Average initial A1c values within each A1c range were low, 6.7 percent; medium, 8.8 percent; high, 11 percent. Average initial adherence was 60 percent in low adherence cells and 90 percent in high adherence cells.

Procedure. Each physician model treated 10,000 simulated patients for 1 year of elapsed time. Both weak and strong feedback models scheduled visits every 90 days. Weak feedforward models scheduled visits every 90 days, except when results from the past move did not decrease A1c by the anticipated amount. In these cases, visits were scheduled 60 days from the previous visit.

The strong feedforward model scheduled visits every 60 days, when A1c changed by the anticipated amount and 30 days between visits when A1c changed by less than the anticipated amount. All models used the same drug formulary for treating blood glucose. The formulary was composed of metformin, sulfonylurea, thiazolidinedione (TZD), and insulin. The models prescribed medications in the order in which they were listed. Simulation treatment records for each model were evaluated for errors of omission using the following error definition:

- **High and medium A1c cells:** If a patient's current A1c is >8 percent and has not dropped by at least 0.5 percent A1c over the past 120 days, then flag an error on this encounter.
- **Low A1c cells:** If a patient's A1c is <8 percent but above goal (7 percent), and a treatment move has not been made in the past 120 days, plus a treatment move is not made on the current encounter, then flag an error on this encounter.

Results

Table 1 summarizes results for all models at the end of 1 year of treating 10,000 individual patients. The results are summarized in four categories.

Percent to Goal Within 1 Year

Feedforward strategies brought more patients to goal within 1 year than feedback strategies. Within each strategy type, strong versions brought more patients to goal than weak versions. These findings hold for all six patient cells. Feedforward-weak and feedback-strong strategies brought similar numbers of patients to goal in all but the HH cell, where feedforward-weak brought 5 percent of patients to goal and feedback-strong brought 0 percent to goal. In both high A1c cells, only the feedforward-strong strategy was able to bring patients to goal.

Table 1. Performance of physician process models treating 10,000 synthetic patients for 1 year

Cell designation[a]	Feedback		Feedforward	
	Weak	Strong	Weak	Strong
Percent to goal by cell (%)				
LL	84.6	94.8	98.4	100.0
LH	91.9	99.1	99.8	100.0
ML	9.8	26.5	26.5	90.5
MH	20.6	42.9	49.5	98.4
HL	0.0	0.0	0.0	44.5
HH	0.0	0.0	5.0	87.3
Average cost[b] to treat by cell ($)				
LL	811	866	1,210	1,390
LH	791	836	1,165	1,325
ML	1,646	1,697	2,044	2,648
MH	1,639	1,691	2,008	2,303
HL	1,973	2,045	2,526	4,200
HH	1,961	2,033	2,524	3,839
Average final A1c by cell (%)				
LL	6.7	6.7	6.5	6.2
LH	6.7	6.6	6.4	6.1
ML	7.9	7.6	7.5	6.8
MH	7.7	7.4	7.2	6.6
HL	10.1	9.7	9.2	7.7
HH	9.8	9.1	8.5	6.8
Percent encounters with errors by cell (%)				
LL	1.0	0.9	0.0	0.0
LH	1.2	1.2	0.0	0.0
ML	23.6	17.0	9.0	3.0
MH	16.0	7.9	2.3	0.8
HL	53.1	52.2	41.6	19.3
HH	45.8	34.1	27.0	2.4

a Cells are labeled using a 2-letter designator, where the first letter specifies initial A1c range (L, M, or H), and the second letter specifies level of patient adherence to treatment (L or H).

b Average costs included prescription costs, charges for office visits, lab orders, and referrals to specialists; inpatient and emergency treatment costs were not included. Prescription costs were based on Average Wholesale Prices.[17]

Cost to Treat for 1 Year

Feedforward strategies were generally more successful in treating patients and required more resources than feedback strategies. Strong versions of strategies required more resources than weak versions. This trend applied to all six patient cells. Even though the feedback-strong strategy and feedforward-weak strategy brought approximately the same number of patients to goal, they did it at very different costs. The feedback-strong strategy incurred 28 percent, 16 percent, and 20 percent lower costs, compared to the feedforward-weak strategy in the low, medium, and high A1c ranges, respectively.

Average Final A1c After 1 Year

Feedforward models outperformed feedback models within patient cells. High-adherence patients also achieved the same or lower A1c than low-adherence patients. These differences became greater as the initial A1c of the patients increased.

Errors of Omission

Feedback strategies committed more errors of omission than feedforward strategies. For a given strategy, weak versions committed more errors than strong versions. All models committed fewer errors in the low A1c cells (feedforward strategies committed no errors). However, in medium and high A1c cells, patients' adherence to treatment influenced the occurrence of errors. Each model committed more errors in low-adherence cells than in high-adherence cells.

In the following sections, treatment data and error patterns were used to develop predictive tools for both synthetic and real clinical patient data.

Predicting Errors in Synthetic Patient Data

Data mining is an area of computer science that uses algorithms to find patterns in datasets. One use of data mining involves the training and testing of classification functions. Common types of functions include decision trees, artificial neural networks, and support vector machines.[18] Each type of function, referred to as a classifier, uses a different method for categorizing input data. A classifier reads a single vector of input data and outputs a classification of these data as one type or another. (For example, in this research, a classifier would predict the presence or absence of an error of omission in the next physician/patient encounter.)

Building classifiers requires a training set of data that forms the classifier and a testing set of data that tests how well the classifier works. Each vector of training data is labeled (i.e., the class label is known). Thus, the classifier is trained by developing a mathematical model that minimizes the number of inaccurate classifications. The present research uses decision trees[18] as classifiers, based on their use of logical rules to discover insights into the relationships between input variables and class labels.

Classifiers are data mining models of relationships between attributes and classes. Such models are considered either explanatory or predictive in nature.[19] Explanatory models are formed to account for patterns in the available data (e.g., a linear regression model that explains some of the variance in the independent variable). Predictive models are formed to predict classes from

input data that have not yet been seen. The work described here develops data mining classifiers using information gathered at a current physician/patient encounter to predict an error of omission on the next encounter.

Predicting errors in real physician-patient encounter databases is often problematic because of missing data, which makes it difficult to identify consistent patterns of physician decisionmaking that lead to error. For example, previous attempts in this research to train classifiers using available real encounter data yielded single-class decision trees that predicted that every next encounter would result in an omission error. An alternative approach for learning predictive patterns of physician decisionmaking is to train classifiers on synthetic data, where issues such as noisy, insufficient, or missing data can be avoided.

To predict errors of omission in a database of physician-patient treatment encounters, the set of classifier attributes was constrained to reflect those data available in both the synthetic and real patient data sets. These attributes were as follows:

Input vector	Classification label
• Time elapsed since last encounter • Time until next scheduled encounter • Monthly rate of change in A1c between last encounter and current encounter • Current reported A1c • Previous medication move • Current medication move	• Error of omission on next encounter

The goal of this work was to predict whether an error of omission would occur on the next encounter. Encounters were defined in terms of clinical visits. The time between encounters was measured in days. A1c was reported when ordered. If a current A1c was not available, the last available A1c was used. Medication moves covered the range of oral medications, including metformin, sulfonylurea, and TZD. A medication move was measured from one encounter to the next by changes in existing prescriptions for each oral medication using values of {-1, 0, 1, 2, 3}. A negative number represented a decrease in the prescribed dose of a medication. A zero indicated the current prescription remained unchanged. Positive numbers indicated the number of medication doses that were increased.

Data on each of the six attributes and the class label (i.e., presence of error in the next encounter) were collected from simulated encounter data and assembled to form vectors of input data to train decision trees. The attributes formed a collection of commonly available information representing a momentary patient state across three clinical encounters (i.e., previous, current, and next) and a set of physician moves connected to that state transition. The resulting set of

input data was combined across all patients with an initial A1c ≥8.0 and for all four physician models.

Each of the four previously defined PPMs treated this patient data set for 1 year of simulated time. The resulting encounter data were combined into a database, scored for errors on a per-encounter basis, and conditioned for classification (i.e., the records containing the classifier input attributes mentioned earlier and the class label were created). One decision tree was trained for all physician/patient combinations, where the patient's initial A1c was ≥8.0 percent.

A 10-fold cross-validation training method was used to generate a decision tree. Cross-validation is a method that divides the data set into the number of folds (i.e., sub-datasets) specified, using one fold for testing and the other folds for training. The classifier then rotates through the folds, holding each one out as the test set in order to create the best general classifier for the data.[20] The decision tree was generated by the J4.8 tree learning algorithm (which is part of Weka data mining tool)[18] that analyzed the input data described above and determined which attributes best predicted errors. This algorithm arranged these attributes into a decision tree composed of a root decision node, subsequent decision nodes, and leaf nodes as shown in Figure 2.

The decision tree shown in Figure 2 is used for prediction by starting at the root of the tree and traversing a series of branches until a leaf node is reached. The value of that leaf node determines

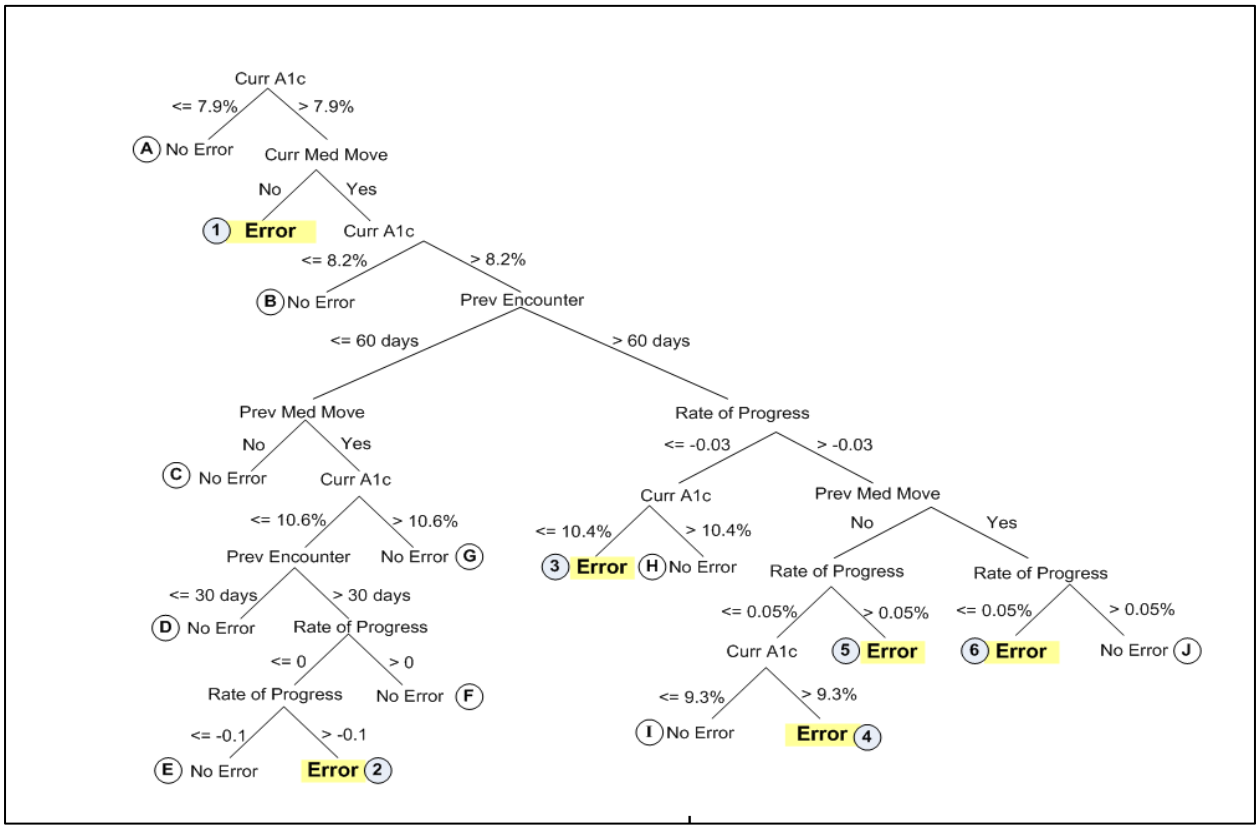

Figure 2. Decision tree generated using encounter data from combined outcomes of all models. Circled numbers identify leaf nodes where errors are predicted to occur on the next visit. Similarly, circled letters identify leaf nodes where next visits are predicted to result in no errors being committed.

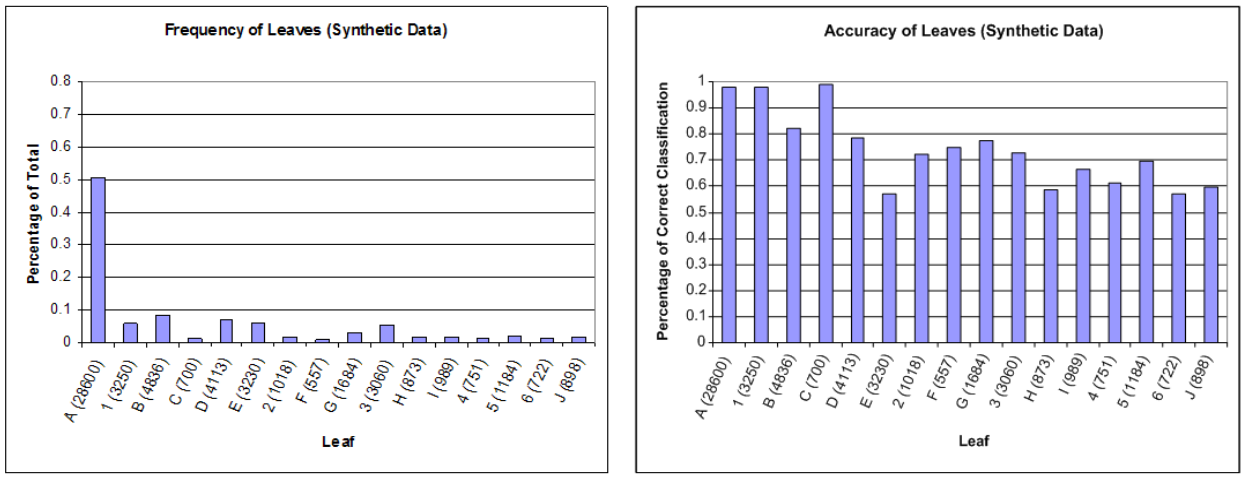

Figure 3. Histograms show the way synthetic data coursed through the decision tree during the cross-validation process and the accuracy with which each leaf node predicted "error" or "no error." The 16 leaf nodes from the tree in Figure 2 are laid flat across the x-axis in a left-to-right organization, with letters denoting a prediction of "no error" and numbers denoting "error." The total number of data points held in each leaf node is in parentheses.

the prediction the decision tree makes for that encounter. Circled numbers identify leaf nodes where errors of omission are predicted to occur on the next visit. Circled letters identify leaf nodes where next visits are predicted to result in no errors being committed.

The branch paths from each decision tree can be stated as logical rules to classify encounters. For example, tracing a path in Figure 2 from the root node right and then left at the next node creates the following rule: "If a patient's current A1c is >7.9 and no medication move is made at a current encounter, then an error is likely to occur in the next encounter with this patient." Each path can be similarly restated.

A given leaf node does not indicate that an actual error did occur at the next encounter for each corresponding input vector. Rather, it predicts that an error is likely to occur at the next encounter. Sometimes this prediction is accurate; sometimes it is not. Each leaf node is not necessarily used the same number of times throughout the dataset.

The left histogram in Figure 3 indicates the number of data records that end up in each of the 16 leaf nodes of the decision tree in Figure 2. The decision tree, when tested on the synthetic data, predicts the occurrence of "error" and "no error" as the positive and negative prediction, respectively. A total of 56,465 physician-patient encounters were used to predict errors on next visits. The tree in Figure 2 correctly predicted error 8,042 times (true positives) and incorrectly 5,093 times (false positives). No error was correctly predicted 40,814 times (true negatives) and incorrectly predicted 2,516 times (false negatives).

Classification results are summarized in Table 2. Data mining literature offers a number of different metrics to evaluate the predictive performance of decision trees.[18, 20] The traditional way to measure predictive accuracy is to calculate the percentage of correctly predicted class labels (both error and no error), i.e., accuracy = (TP + TN) / (TP + TN + FP + FN).

As shown in Table 2, the recall measure of the predictive performance is the percentage of errors correctly predicted out of all the errors that actually occurred; i.e., recall = TP / (TP + FN). Precision is the percentage of the actual errors among all the encounters that were classified as errors by the decision tree; i.e., precision = TP / (TP + FP). The F-measure metric is the harmonic mean of the recall and precision metrics and is computed as:

$$F\text{-}measure = 2 \times (recall \times precision) / (recall + precision)$$

The F-measure, which quantifies the overall reliability of the decision tree, represents a tradeoff between correctly predicting errors and overpredicting errors. It is in the range of 0 to 1 and is commonly used in data mining literature to evaluate classifiers.

Table 2. Summarized results of training and testing decision trees on synthetic data

Number of attributes in tree	Attributes used to predict errors	Counts [N (%)]	Performance metrics
6	Current A1c Days since last encounter Days to next encounter Med adjustment – current visit Med adjustment – previous visit Monthly rate of change in A1c	TP = 8042 (14) FP = 2516 (5) FN = 5093 (9) TN = 40,814 (72)	Accuracy = 87% Recall = 0.61 Precision = 0.76 F-measure = 0.68 Kappa statistic = 0.59
8	Current A1c Days since last encounter Days to next encounter Oral adjustment – current visit Oral adjustment – previous visit Insulin adjustment – current visit Insulin adjustment – previous visit Monthly rate of change in A1c	TP = 9382 (17) FP = 2665 (5) FN = 3753 (7) TN = 40,665 (72)	Accuracy = 89% Recall = 0.71 Precision = 0.78 F-measure = 0.75 Kappa statistic = 0.67
11	Current A1c Days since last encounter Days to next encounter Metformin adjustment – current visit Metformin adjustment – previous visit Other oral adjustment – current visit Other oral adjustment – previous visit Current creatinine level Insulin adjustment – current visit Insulin adjustment – previous visit Monthly rate of change in A1c	TP = 10,248 (18) FP = 1967 (3) FN = 2887 (5) TN = 41,363 (73)	Accuracy = 91% Recall = 0.78 Precision = 0.84 F-measure = 0.81 Kappa statistic = 0.75

TP = true positives; FP = false positives; FN = false negatives; TN = true negatives

The Kappa statistic in Table 2 measures the similarity between the trained classifier and a random classifier. A value of 0 indicates that a random classifier predicts as well as the trained classifier, whereas a value of 1 indicates no agreement between a random classifier (i.e., chance prediction) and the trained classifier (i.e., a Kappa value of 1 is a strong indication of an accurate classifier).

The predictive accuracy (in terms of F-measure) of the decision tree in Figure 2 is 0.68. Attempts were next made to improve this rate by providing additional information to the data mining model. The model shown in Figure 2 was constructed using information present in the database of real patient encounter records available for this study. However, the synthetic models (described in the previous section) used to simulate physician/patient clinical encounters maintain a larger set of information, including the actual amounts of different types of insulin administered to each patient and the estimates of the patients' adherence to treatment. To determine whether training predictive tools using such information could increase the accuracy of error predictions, additional information was incrementally provided.

Two alternate decision trees were evaluated. In the first case, attributes for medication adjustments were further divided into oral medication adjustments and insulin adjustments. Adding these attributes increased the decision tree predictive ability (in terms of F-measure) from 0.68 to 0.75. In the second case, a decision tree was created to reflect the fact that PPMs are able to stop the use of contraindicated medications. For the formulary used in the simulation study, the PPMs use creatinine levels to determine whether metformin is contraindicated. In the second alternate tree, attributes for oral medication adjustments were subdivided into metformin adjustments vs. non-metformin adjustments, and an attribute was added for creatinine levels. These additional attributes improved the error prediction ability to 0.81.

Table 2 shows that errors can be predicted more accurately by tracking additional patient data. However, since the present research tests whether predictive tools developed in a synthetic database of patient records can be transferred to an available database of real patient records, the decision trees tested on the real (administrative) data were restricted to currently available attributes.

Predicting Errors in Clinical Patient Data

The decision tree developed with synthetic data (Figure 2) was tested on an administrative database of clinical treatment records. The test was designed to predict errors in real encounter data using a decision tree constructed from synthetic models. A database of deidentified data for 129 real primary care physicians treating 12,708 real diabetes patients over the course of 1 year served as the test bed. Data records contained date of encounter, last reported A1c value, and types and doses of prescribed oral medications (metformin, sulfonylurea, TZD). Although records indicated the use of insulin, titrations of insulin were not available. In cases where A1c values were not reported on a given encounter, the last measured A1c was used to fill in the missing value.

The clinical data were prepared using a process similar to that used in preparing synthetic treatment data. All encounters not coded as diabetes-related encounters using available ICD-9

codes and any duplicate entries were removed. Patients with initial A1c ≥8 percent were extracted and scored with the error definition.

In the real clinical data set, the first encounter for a patient was used to establish a baseline for that patient. The first usable medication move was in the second encounter, which is analogous in the synthetic database to the first encounter. Due to this difference, the first, second, and final encounters listed for each physician/patient treatment path were used to collect information, but could not be used as inputs for error prediction.

The initial clinical database contained 77,514 encounter records, spanning 129 physicians and 12,708 patients. Usable encounter data included 3,389 records spanning 101 physicians and 799 patients. For patients with initial A1c ≥10, 83 physicians saw an average of 2.5 patients for 4.5 usable encounters per year. For patients with initial A1c ≥8 percent but <10 percent, 97 physicians each saw 6.1 patients 4.2 times. Patients treated by physicians were excluded from these averages if they had fewer than four diabetes-related encounters within the year of clinical data examined.

Figure 4 shows two histograms resulting from testing the decision tree with real clinical data. A total of 3,389 physician-patient instances were used to predict errors. The tree in Figure 2 correctly predicted error 1,844 times (true positives) and incorrectly predicted error 923 times (false positives). No error was correctly predicted 122 times (true negatives) and incorrectly predicted 500 times (false negatives). The recall for errors was 67 percent (compared to 61 percent in synthetic data); the F-measure was 0.72 (compared to 0.68 for the synthetic data). As shown by the histogram on the right, most of the data are classified by the first two leaf nodes. Comparison with the histogram on the right in Figure 3 shows that, in the real database, fewer patients fell below 7.9 percent over the course of the year, and physicians made fewer medical moves, compared with simulated data.

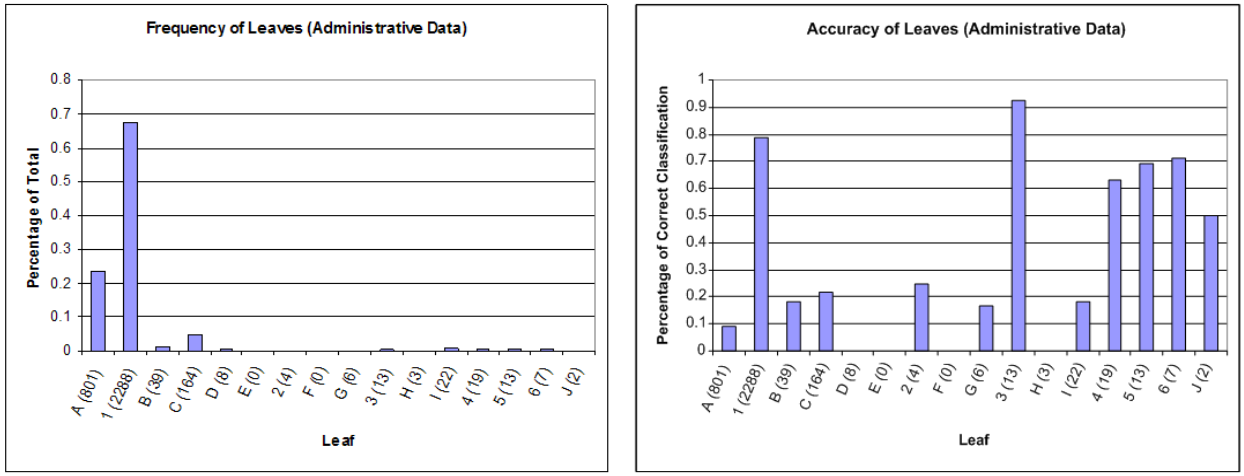

Figure 4. Histograms show the way real clinical data coursed through the decision tree (learned using synthetic data) and the accuracy with which each leaf node predicted "error" or "no error." The 16 leaf nodes from the tree in Figure 2 are laid flat across the x-axis in a left-to-right organization with letters denoting a prediction of "no error" and numbers denoting "error." Total number of data points held in each leaf is in parentheses.

The classification accuracy of individual leaf nodes in the decision tree varies, due in part to the small number of examples. Focusing on the first two leaf nodes (the most common nodes) it is clear that in the real (clinical) administrative data, having a patient's A1c ≤7.9 percent does not prevent an error from occurring on the next encounter. Not making a medical move when a patient's A1c is >7.9 percent is highly predictive of errors of omission for next encounter. It is noteworthy that the decision tree developed in a synthetic environment was able to accurately predict errors that were detected in some of the more complex branches of the tree.

Conclusion

Simulation of four physician process models (PPMs) treating populations of individual synthetic patients with type 2 diabetes over the course of 1 year showed heterogeneity in outcome effects for different patient subgroups in terms of average resulting A1c, percent of patients reaching goal, costs of treatment, and proportion of encounters with errors of omission. From this subgroup analysis, tradeoffs in reaching evidence-based goals at the lowest possible costs were made in each subgroup. Low A1c patients were driven to goal easily by all the PPMs but at lowest cost by the feedback-weak model. Medium A1c patients were treated most effectively by the feedforward-strong model. Significant per patient cost savings were achieved by using a feedback-strong model rather than a feedforward-weak model, without a large dropoff in performance. In the high A1c cell, only the feedforward-strong model was effective at bringing a significant percentage of patients to evidence-based goals.

A specific data mining model (a decision tree) was developed to predict errors of omission that occurred throughout the patient population by virtue of different treatment strategies. Analysis of the prediction results and the specific branch paths used by the decision tree indicated where particular patterns of physician decisionmaking led to errors. The decision tree used to predict errors was formed by restricting the leaf node size to no less than 500 physician/patient encounters. This restriction resulted in a relatively small tree that could be used to interpret predictions generated by the model. The ability to interpret how an error is predicted has implications for the formation of policies based on these predictions.

An important aspect of the work presented here is the fact that physician and patient models were used to generate simulated encounter data from which data mining tools could be developed. The ability to use these tools to predict errors of omission in real (and synthetic) patient data suggests that future developments in this kind of work have the potential to enable identification and correction of physician decisionmaking strategies that lead to encounter-specific treatment errors.[21] Application of this work to the improvement of care for patients with type 2 diabetes, as well as chronic diseases more generally, is promising.

Acknowledgments

This work was supported in part by grants from the Agency for Healthcare Research and Quality, the National Institutes of Health (NIDDK), the University of Minnesota Digital Technology Center, and the Carlson School of Management Medical Institute Leadership Initiative.

Author Affiliations

Computer Science, University of Minnesota (Mr. McCabe); Information and Decision Sciences, University of Minnesota (Dr. Adomavicius, Dr. Johnson, Mr. Ramsey); Carlson School of Management, University of Minnesota (Ms. Rund); HealthPartners Medical Group and HealthPartners Research Foundation (Dr. Rush, Dr. O'Connor, Dr. Sperl-Hillen).

Address correspondence to: Ryan McCabe, Carlson School of Management, University of Minnesota, 321 19th Avenue, S, Mail Code 7041, Minneapolis, MN 55455; telephone: 612-626-0320; fax: 612-626-1316; e-mail: Mcca0502@umn.edu.

References

1. Institute of Medicine. To err is human: Building a safer health system. Washington, DC: National Academies Press; 2000.
2. O'Connor P, Sperl-Hillen J, Johnson P, et al. Identification, classification, and frequency of medical errors in outpatient diabetes.. Advances in patient safety: From research to implementation. Vol 1: Research findings. AHRQ Publication No. 05-0021-1 Rockville, MD: Agency for Healthcare Research and Quality; 2005. Available at: www.ahrq.gov/downloads/pub/advances/vol1/OConnor.pdf. Accessed March 28, 2008.
3. Johnson PE, Veazie PJ, Kochevar L, et al. Understanding variation in chronic disease outcomes. Health Care Manag Sci 2002: 5: 175-189.
4. Veazie PJ, Johnson PE, O'Connor PJ, et al. Making improvements in the management of patients with type 2 diabetes: A possible role for the control of variation in glycated hemoglobin. Med Hypotheses 2005: 64: 792-801.
5. Siau K. Health care informatics. IEEE Trans Inf Technol Biomed 2003; 7: 1-7.
6. Breault J, Goodall C, Fos P. Data mining a diabetic data warehouse. Artif Intell Med 2002; 26: 37-54.
7. Dutta P, Biltz GR, Johnson PE, et al. SimCare: A model for studying physician decision making activity. Advances in patient safety: From research to implementation. Vol 4: Programs, tools, and products. AHRQ Publication No. 05-0021-4. Rockville, MD: Agency for Healthcare Research and Quality; 2005. Available at: www.ahrq.gov/downloads/pub/advances/vol4/Dutta.pdf. Accessed March 28, 2008.
8. Mazze R, Strock E, Simonson G, et al. Staged diabetes management: Detection and treatment. Minneapolis: International Diabetes Center; 2001.
9. Defronzo R. Pharmacologic therapy for type 2 diabetes mellitus. Ann Intern Med 1999; 131: 281-303.
10. Tseng C, Brimacombe M, Xie M, et al. Seasonal patterns in monthly hemoglobin A1c values. Am J Epidemiol 2005; 161: 565-574.
11. Derr R, Garrett E, Stacy G, et al. Is HbA1c affected by glycemic instability? Diabetes Care 2003; 26: 2728-2733.
12. Levy J, Atkinson AB, Bell PM, et al. Beta-cell deterioration determines the onset and rate of progression of secondary dietary failure in type 2 diabetes mellitus: The 10-year follow-up of the Belfast Diet Study. Diabet Med 1998; 15: 290-296.
13. Brehmer B. Strategies in real-time dynamic decision making. In: Hogarth R, ed. Insights in decision making: A tribute to Hillel J. Einhorn. Chicago: University of Chicago Press; 1990.
14. Garcia C, Morari M. Internal model control. 1. A unifying review and some new results. Ind Eng Chem Proc Des Dev 1982; 21: 308-323.
15. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. Science 1974; 185: 1124-1131.

16. Sterman J. Misperceptions of feedback in dynamic decision making. Organ Behav Hum Decis Process 1989; 43: 301-335.

17. Sitagliptin (Januvia™) for type 2 diabetes. Medical Lett Drugs Ther 2007; 49: 1-3.

18. Witten I, Frank E. Data mining: Practical machine learning tools and techniques with java implementations. San Francisco: Morgan Kaufmann Publishers; 2000.

19. Shmueli G, Koppius O. Predictive vs. exploratory modeling in IS research. Conference on Information Systems and Technology; 2007 Nov 3-4; Seattle, WA.

20. Mitchell, T. Machine learning. Boston: McGraw Hill; 1997.

21. Green L, Seifert C. Translation of research into practice: Why we can't "just do it." J Am Board Fam Med 2005; 18: 541-545.