Developing Indicators of Inpatient Adverse Drug Events Through Nonlinear Analysis Using Administrative Data

Jonathan R. Nebeker, MS, MD,*†‡§ Paul R. Yarnold, PhD,¶ Robert C. Soltysik, MS,‖
Brian C. Sauer, PhD,†‡ Shannon A. Sims, MD, MS,** Matthew H. Samore, MD,†‡§**
Randall W. Rupper, MD, MPH,*†‡ Kathleen M. Swanson, RPh, MS,†† Lucy A. Savitz, PhD, MBA,†‡§‡‡
Judith Shinogle, PhD,¶¶‖‖ and Wu Xu, PhD†‡***

Background: Because of uniform availability, hospital administrative data are appealing for surveillance of adverse drug events (ADEs). Expert-generated surveillance rules that rely on the presence of International Classification of Diseases, 9th Revision Clinical Modification (ICD-9-CM) codes have limited accuracy. Rules based on nonlinear associations among all types of available administrative data may be more accurate.

Objectives: By applying hierarchically optimal classification tree analysis (HOCTA) to administrative data, derive and validate surveillance rules for bleeding/anticoagulation problems and delirium/psychosis.

Research Design: Retrospective cohort design.

Subjects: A random sample of 3987 admissions drawn from all 41 Utah acute-care hospitals in 2001 and 2003.

Measures: Professional nurse reviewers identified ADEs using implicit chart review. Pharmacists assigned Medical Dictionary for Regulatory Activities codes to ADE descriptions for identification of clinical groups of events. Hospitals provided patient demographic, admission, and ICD9-CM data.

Results: Incidence proportions were 0.8% for drug-induced bleeding/anticoagulation problems and 1.0% for drug-induced delirium/ psychosis. The model for bleeding had very good discrimination and sensitivity at 0.87 and 86% and fair positive predictive value (PPV) at 12%. The model for delirium had excellent sensitivity at 94%, good discrimination at 0.83, but low PPV at 3%. Poisoning and adverse event codes designed for the targeted ADEs had low sensitivities and, when forced in, degraded model accuracy.

Conclusions: Hierarchically optimal classification tree analysis is a promising method for rapidly developing clinically meaningful surveillance rules for administrative data. The resultant model for drug-induced bleeding and anticoagulation problems may be useful for retrospective ADE screening and rate estimation.

Key Words: hierarchically optimal classification tree analysis, patient safety indicators, surveillance rules, administrative data

(Med Care 2007;45: S81–S88)

Although adverse drug events (ADEs) are one of the largest and most costly patient safety problems in the United States,[1,2] they have not been productively targeted for identification using administrative data. The Agency for Healthcare Research and Quality (AHRQ) has developed 20 provider-level patient safety indicators (PSIs) to be used with administrative data.[3] The provider-level PSIs are intended to serve as a screening tool for surveillance of potentially preventable complications of inpatient hospital care. However, nearly three-quarters of the PSIs are related to surgical and obstetrical complications and only one—complications of anesthesia—targets ADEs.[4]

The focus on indicators for surgical complications and lack of indicators for ADEs reflects the more complete coverage of surgeries by codes in the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). A relatively comprehensive ICD-9-CM code set facilitates the identification of both the population at risk for surgical complications and the complications themselves. In contrast, few ICD-9-CM codes specifically describe drug exposure. Drug complication codes are rarely used and have been found to have poor validity.[5,6] Therefore, to identify ADEs, it is necessary to look beyond drug complication codes to other types of administrative data.

In modeling ADE indicators using administrative data, the challenge is to identify predictive covariates when both the independent and dependent ICD-9-CM variables are sparse. Incidence proportions of specific types of ADEs are often low—in the <1% range. Given approximately 20,000 ICD-9-CM codes in use, individual codes possibly predictive of a type of ADE are also sparsely distributed across all admissions. Linear statistical techniques such as logistic regression have been unable to produce accurate models with these data.

There are 2 approaches that may facilitate modeling of ADE surveillance rules using administrative data. One is to expand the number of candidate covariates beyond ICD-9-CM codes and Diagnosis-Related Groups (DRGs) to include all available non-ICD-9-code data, such as age, admission source, and length of stay. These data have a relatively small number of possible response categories and are present in most records. Current PSIs use all these data to identify the risk pool but only use ICD-9-CM codes to identify adverse events. Another approach is to use a partitioning analysis rather than a linear statistical method. Hierarchically optimal classification tree analysis (HOCTA) is an exact, nonparametric, nonlinear statistical model.[7] In applied research, HOCTA has yielded predictive models possessing superior accuracy and providing richer solutions than prior results using linear models.[8–11]

In the present study, we applied HOCTA to all available administrative data to develop surveillance rules for identifying inpatient ADEs manifesting as delirium or bleeding problems.

METHODS

Description of Data

The data for this project were derived from a random sample of 3987 patients in all 41 acute-care hospitals in Utah in 2001 and 2003. The purpose of the study was rate estimation for iatrogenic inpatient adverse events and validation of ICD-9-CM codes for these events. The focus of this study was ADEs. The methods are described in more detail elsewhere.[12] Briefly, professional nurse reviewers reviewed all charts for adverse events. Up to 5 adverse events were evaluated for each admission. Nurses coded causative drugs into major classes. One of the physician authors verified drug-class assignments manually. Specially trained pharmacists coded all reactions into Medical Dictionary for Regulatory Activities version 9.0 (MedDRA) terms for classification of the outcome.[13]

Variable Selection

The investigators selected 2 of the most common ADEs, delirium and psychosis, for the symptom-based ADEs and bleeding and anticoagulation problems for the sign-based ADEs. The decision to target symptom- and sign-based ADEs related to the constraints on coding administrative data. Medical coders may only code events documented in physician notes. Physicians are less likely to recognize and document symptom-based or subjective drug complications.[14,15] Delirium is especially likely to be unrecognized.[16] In contrast, physicians do recognize laboratory-based or objective ADEs, such as coagulation problems or bleeding.[14,15] We therefore hypothesized that ICD-9-CM codes would not be predictive of drug-induced delirium, but would be predictive of bleeding and anticoagulation problems.

Another distinction between the selected ADEs is that a population at high risk for exposure can be established for agents causing bleeding, but not for agents causing delirium. A wide variety of commonly used drug classes cause delirium. Not only would it be arduous to enumerate these classes and determine ICD9 codes corresponding to indications for them, but also the resulting at-risk cohort would contain a large majority of the admissions in the main cohort. Fewer classes of drugs commonly cause bleeding. Major mechanisms for bleeding include inhibition of clotting and breakdown of gastric or intestinal mucosa. The commonly used drug classes corresponding to these mechanisms are warfarins, heparins, antiplatelet agents, nonsteroidal anti-inflammatory drugs (NSAIDs), and systemic glucocorticoids. However, administrative data lack pharmacy exposure data. The absence of definitive exposure data is also a problem in infection control, where device exposures are rarely available for rate estimation of catheter-related infections.

Because we lacked pharmacy data to determine the at-risk cohort, we estimated drug exposure using ICD-9-CM codes. We derived a list of codes reflecting indications where guidelines[17,18] dictate the use of drugs that increase the bleeding risk. We grouped these codes into general categories (eg, coronary/cerebral vascular disease, selected vascular/orthopedic procedures, and inflammatory conditions) and created a dummy variable equal to 1 if at least 1 code was present in that category. Relevant to our results is the category of acquired or induced coagulopathies that comprised seven ICD-9-CM diagnostic codes 286.5, 286.7, 286.9, 287.4, 790.92, E934.2, V58.61, and 1 procedure code 99.10. We created the at-risk variable, which indicated the presence of any of the codes of any exposure category. The complete list of codes is available from the authors.

The hierarchical structure of MedDRA allowed events to be sorted into related groups for the dependent variables. The outcome for bleeding and anticoagulation problems comprised the following high-level terms: red blood cell analyses, anemias not elsewhere classified (NEC), hemorrhages NEC, coagulation and bleeding analyses, platelet analyses, thrombocytopenias, and anemia deficiencies. The outcome for delirium and psychosis comprised the following high-level terms: confusion and disorientation, perception disturbances, deliria, anxiety symptoms, mental disorders NEC, psychotic disorder NEC, behavior and socialization disturbance, decreased physical activity levels, delusional symptoms, depressive disorders, dissociative states, and emotional and mood disturbances NEC.

The independent variables were derived from the raw administrative data submitted to the Utah Department of Health. For ICD-9-CM codes, there were 16 fields: 1 primary diagnosis field, 8 secondary diagnosis fields, 6 procedure fields, and 1 external cause of injury code (E-code). These fields included adverse event codes. To reduce the sparseness

of the variable, we grouped ICD-9 codes into single-level codes of the Clinical Classification System (CCS) using SAS code available from the AHRQ web site.[19] Non-ICD-9-code data described demographics and admission information: age, length of stay, sex, race, admission source, and discharge disposition. Race was omitted due to an excessive proportion of missing values.

For the models of the 2 events, we planned to test explicitly the performance of ICD-9-CM codes for drug complications and poisonings. We created dummy variables that aggregated the codes. Variables were set equal to 1 if any of these codes were present and 0 otherwise. The complications and poisoning codes were selected using text searches of the ICD-9-CM dictionary and from a list compiled by iterative expert review.[12] For deliria, we used codes for drug-induced mental status changes: 292.0, 292.11, 292.12, 292.2, 292.81, 292.82, 292.83, 292.84, 292.89, and 292.9. For bleeding and anticoagulation problems, we used codes for poisoning and adverse effects: 964.2, 964.3, 964.4, 964.5, 964.9, E858.2, E934.2, E934.3, E934.4, E934.5, E934.8, and E934.9. Additionally, we created a dummy variable indicating the presence of any code for anemia, thrombocytopenia, or hemorrhage: 287.4, 287.5, 285, 285.1, 285.8, 285.9, and 459.0. Test characteristics of these dummy variables are reported for aggregate variables if component codes were present in either the primary or secondary position, as submitted for billing purposes. Because analyses on related data demonstrated that primary diagnoses often represented outpatient conditions,[5,12] we also evaluated aggregate variables with component codes present only in the secondary position.

For each model, we conducted sensitivity analysis by exploring the variation in the model's performance among hospitals having an a priori reasonable number of admissions in the analysis cohort—50 for bleeding and 100 for delirium—based on a desire to have a number of either ≥50 or the inverse of the ADE's prevalence in the analysis cohort. We reported the median and range of values for the selected hospitals.

Approach to Analysis

We selected HOCTA because it offers several advantages over linear models. Data sets that are not of full rank (ie, that have R^2 values near 100%), and thus induce computational instability in linear models, are not problematic for HOCTA due to its hierarchical nature: if no attributes continue to improve the HOCTA model at a given step in the procedure, then the procedure is terminated at that point (and no "bouncing betas" are possible). Similarly, sporadic missing data are not problematic for HOCTA: an observation with missing data is omitted from the model only if the observation is missing data on an attribute used to classify the observation on that node and downstream. Finally, the extreme skew in the present data (in which ADEs are rare) is not problematic for HOCTA, as it is for maximum-likelihood or general-linear-model-based methods, because HOCTA explicitly maximizes accuracy. Thus, HOCTA models seek to maximize prediction of both classes, rather than the class with the largest sample size or the greatest variance. These features of

the analytic framework make HOCTA an excellent fit for generating surveillance rules from administrative data.

HOCTA was conducted via Optimal Data Analysis software.[7,20] Like all methods derived within the optimal discriminant analysis paradigm,[7] HOCTA *explicitly maximizes accuracy* in the analysis sample. HOCTA models begin with the variable providing best accuracy for the entire sample, followed by the variable that adds incrementally most beyond the first, and so forth until no variables add incrementally to the accuracy achieved. To reduce overfitting, attributes were entered into the model only if the minimum denominator at each branch of the attribute exceeded 20 observations; a sequentially rejective Bonferroni-type procedure ensured an experimentwise Type I error of $P < 0.05$ for all of the attributes in the model[7]; and its performance in accurately classifying observations did not decline in jackknife analysis. Jackknife analysis was performed at each node by holding out each observation, developing a significant model using the same attribute, and then classifying the held-out observation with the resulting model. If the combined accuracy of jackknife classifications was at least as good as the whole-sample model, the node was considered jackknife stable. The purpose of this jackknife procedure is to reduce bias introduced by idiosyncratic values and reduce the shrinkage in accuracy when the model is applied to other data.[21]

In addition to computing the positive predictive value (PPV) of the model, results within the optimal discriminant analysis paradigm are typically reported in terms of the Effect Strength for Sensitivity (ESS), the targeted measure of accuracy. In a system with a binary class variable (eg, ADE present or absent), ESS is computed as: ESS = [(sensitivity + specificity)/2 − 50]/50. ESS is a standardized index of model classification accuracy on which 0 represents the level of classification accuracy that is expected by chance, and 1.0 represents perfect, errorless classification; negative values indicate classification performance worse than expected by chance.[7] We used Stata version 9.2 (College Station, TX) to determine other performance measures such as the area under the receiver operator curve (ROC AUC), a more commonly used measure of discrimination, in this case equal to (sensitivity + specificity)/2,[22] and relative risk (RR). Nonparametric confidence intervals for ESS, sensitivity, and PPV were computed via bootstrap analysis involving a resample with replacement for 50% of the number of observations in the analysis cohort for 10,000 iterations.[21]

RESULTS

We identified 1676 admissions with diagnoses indicative of drugs potentially causing bleeding and anticoagulation problems. The at-risk cohort contained 29 of the 30 admissions with ADEs in the full cohort of 3987 admissions. Table 1 lists the test characteristics and measures of association for selected independent variables. Inspection of the 30 most accurate variables for the first node (not all listed in Table 1) revealed that the attributes with the highest ESS were dummy variables for discharge disposition and thresholds for length of stay. Only 5 variables were individual CCS codes, whereas

TABLE 1. Bivariate Analysis for the First Attribute of the Model for Bleeding and Anticoagulation Problems

Selected Variables	ESS	Prevalence (%)	Sensitivity (%)	Specificity (%)	ROC AUC	PPV (%)	RR	*P**
Aggregate event variable of ICD-9-CM code in secondary position for bleeding or anticoagulation problem[†]	0.36	13	48	88	0.68	6.5	6.28	0.0000
Length of stay ≥5 d	0.33	47	79	54	0.67	2.9	4.35	0.0004
Acute posthemorrhagic anemia (CCS D-60)	0.30	5	34	95	0.65	11.6	9.73	0.0000
Length of stay ≥8 d	0.30	19	48	81	0.65	4.3	3.91	0.0001
[†]Complications of surgical procedures or medical care (CCS D-238)	0.28	10	38	90	0.64	6.4	5.31	0.0000
Discharged to another short-term hospital	−0.26	60	34	39	0.37	1.0	0.35	0.0043
Age ≥75 yr	0.24	35	59	66	0.62	2.9	2.66	0.0066
Exposure variable of acquired/induced coagulopathy[†]	0.21	7	28	94	0.61	7.1	5.32	0.0000
Blood transfusion (CCS P-222)	0.21	7	28	94	0.61	7.1	5.32	0.0000
Length of stay ≥4 d[†]	0.21	69	90	31	0.60	2.2	3.83	0.0167
Cardiac dysrhythmias (CCS D-106)[†]	0.21	21	41	79	0.60	3.4	2.64	0.0070
Discharged to home under home health services	0.20	15	34	85	0.60	4.0	2.96	0.0033
Aggregate exposure variable of ICD-9-CM codes for coronary or cerebral vascular disease	0.15	41	55	60	0.57	2.3	1.79	0.1093
Aggregate exposure variable of ICD-9-CM codes for non-vascular surgery for which anticoagulation is indicated	0.14	10	24	90	0.57	4.1	2.78	0.0130
Aggregate exposure variable of ICD-9-CM codes for vascular surgery for which anticoagulation is indicated	0.13	8	21	92	0.56	4.5	3.03	0.0104
Coagulation and hemorrhagic disorders (CCS D-62)	0.13	5	17	96	0.56	6.6	4.39	0.0009
Gastrointestinal hemorrhage (CCS D-153)	0.12	5	17	95	0.56	5.6	3.71	0.0038
E codes: Place of occurrence (CCS D-2621)	0.11	7	17	94	0.55	4.5	2.94	0.0204
Male[†]	−0.11	45	34	55	0.45	1.3	0.65	0.2593
Aggregate exposure variable of ICD-9-CM codes for inflammatory conditions for which anticoagulation is indicated	0.10	14	24	86	0.55	2.9	1.91	0.1249
Discharge to home or self-care	0.10	4	14	96	0.55	6.0	3.84	0.0066
Pleurisy; pneumothorax; pulmonary collapse (CCS D-130)	0.10	8	17	93	0.55	4.0	2.56	0.0452
Aggregate ADE variable of ICD-9-CM codes in any position for drug-induced bleeding or anticoagulation problem	0.10	1	10	99	0.55	23.1	14.76	0.0000

*P value is for the relative risk.
[†]Variable included in the final model.
ICD-9-CM indicates International Classification of Diseases, 9th Revision Clinical Modification; ROC AUC, receiver operator characteristic area under the curve; PPV, positive predictive value; ESS, effect strength for sensitivity; RR, relative risk; CCS, Clinical Classification System single level code; D-x, diagnostic code; P-x, procedure code.

8 variables were aggregated CCS codes for bleeding and for exposure to drugs potentially causing bleeding. The performance of an aggregate variable comprising all ICD9-CM codes for drug-induced bleeding or anticoagulation problems had good PPV at 23% and was strongly associated with the target problems (RR = 14.7), but had poor sensitivity and ESS—both at 10%.

Figure 1 depicts the model for ADEs manifested as bleeding or anticoagulation problems. We found a classification tree using 8 nodes. The tree segregated into 2 parts depending on whether there was a secondary ICD-9-CM code present for a bleeding or anticoagulation problem. If there was such a code, an ADE was predicted in 2 cases: where there was no nonhemorrhagic anemia and either an induced coagulopathy or a complication of surgical procedures or medical care. If there was no secondary code present for a bleeding or anticoagulation problem, an ADE was predicted if either length of stay (LOS) was ≤4 days and there was a complication; or if LOS was >4 days and the patient was a woman with cardiac dysrhythmias or anyone with a disorder

of lipid metabolism. The model had excellent discrimination: ESS = 0.747 [95% confidence interval (CI): 0.746–0.748], sensitivity = 86.4% (95% CI: 85.9–86.8), specificity = 89%, and ROC AUC = 0.87. The prevalence of the ADE in the at-risk cohort was 1.7%. Although the positive likelihood ratio was good at 7.5, PPV was 11.5% (95% CI: 11.4–11.7). In the sensitivity analysis, there were 8 hospitals with more than 50 at-risk admissions (range, 52–113 admissions). The median point estimates of various performance characteristics were as follows: ESS = 0.84 (range, 0.61–0.94%); sensitivity = 100% (range, 75–100%); specificity = 85% (range, 77–93%); ROC AUC = 92% (range, 80–97%); PPV 13% (range, 5–33%).

For the delirium model, we used the entire cohort of 3987 admissions. Table 2 lists the test characteristics and measures of association for selected independent variables at the first node. As with the bivariate analysis for bleeding, non-ICD-9-CM variables ranked highest in accuracy. CCS codes ranked no higher than the 47th most accurate variable. Aggregate variables of ICD-9-CM codes that indicated drug-
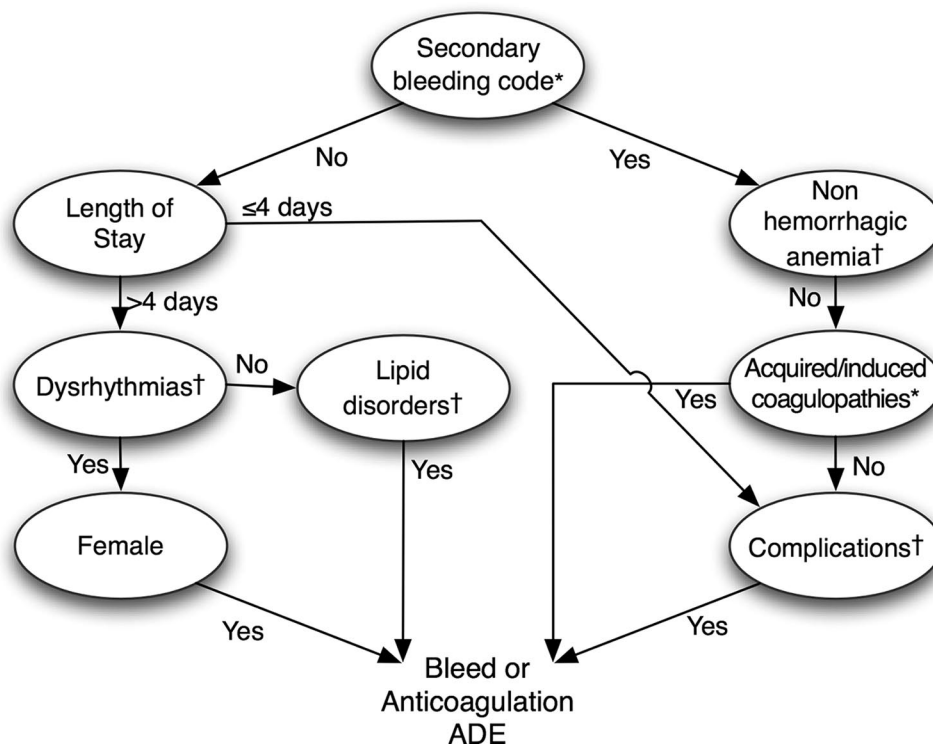
FIGURE 1. Classification tree model for bleeding and coagulation problems. The entire sample is included in the first node, and the unclassified portion of the population is carried down the tree. Secondary bleeding code indicates a code for anemia, thrombocytopenia, or hemorrhage in the secondary position. * Signifies aggregate variables constructed by the authors. † Indicates Clinical Classification System codes.

TABLE 2. Bivariate Analysis for the First Attribute of the Model of Drug-Induced Delirium

Selected Variables	ESS	Prevalence (%)	Sensitivity (%)	Specificity (%)	ROC AUC	PPV (%)	RR	P*
Length of stay >3 d†	0.44	51	94	50	0.72	2	16.52	0.0000
Length of stay >5 d†	0.42	22	64	78	0.71	3	6.20	0.0000
Age ≥55 yr	0.42	39	81	61	0.71	2	6.49	0.0000
Length of stay >4 d	0.41	32	72	69	0.70	2	5.60	0.0000
Age ≥75 yr†	0.37	19	56	81	0.68	3	5.35	0.0000
Length of stay >6 d	0.34	17	50	84	0.67	3	5.06	0.0000
Discharged to another short-term hospital	−0.24	77	53	23	0.38	1	0.34	0.0007
Age <20 yr	−0.24	26	3	74	0.38	0	0.08	0.0013
Congestive heart failure; nonhypertensive (CCS D-108)	0.18	10	28	90	0.59	3	3.48	0.0003
Discharged to home under home health services	0.17	8	25	92	0.59	3	3.83	0.0002
Liveborn (CCS D-218)	−0.15	15	<1	85	0.42	0	0.00	0.0112
Normal pregnancy and/or delivery (CCS D-196)	−0.14	14	<1	86	0.43	0	0.00	0.0148
Fluid and electrolyte disorders (CCS D-55)	0.14	14	28	86	0.57	2	2.32	0.0196
Any ICD-9-CM code indicative of drug-induced delirium or psychosis	0.14	<1	14	100	0.57	28	35.56	0.0000
Admitted through the emergency room	−0.12	94	81	6	0.44	1	1.61	0.1654
Substance-related mental disorders (CCS D-67)	0.12	8	19	92	0.56	2	2.83	0.0094
Any ICD-9-CM code indicative of drug-induced delirium or psychosis in the secondary position	0.11	<1	11	100	0.55	31	38.21	0.0000
Diabetes Mellitus (CCS D-50)†	0.08	4	11	97	0.54	3	3.41	0.0134

*P value is for the relative risk.
†Variable included in the final model.
ICD-9-CM indicates International Classification of Diseases, 9th Revision Clinical Modification; ROC AUC, receiver operator characteristic area under the curve; PPV, positive predictive value; ESS, effect strength for sensitivity; RR, relative risk; CCS, Clinical Classification System single level code; D-x, diagnostic code.
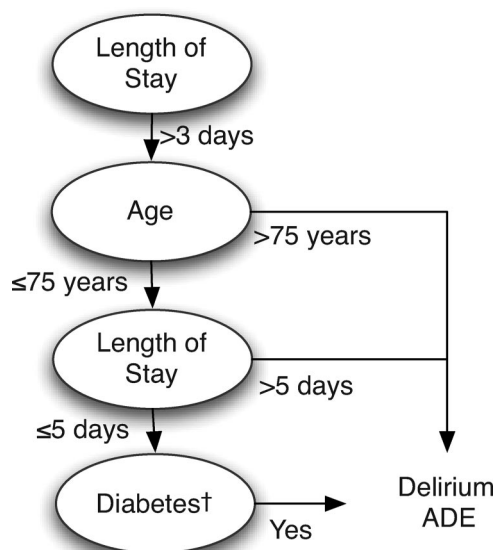
FIGURE 2. Classification tree model for delirium and psychosis. The entire sample is included in the first node, and the unclassified portion of the population is carried down the tree. † Indicates Clinical Classification System codes.

induced delirium or psychosis were not accurate or sensitive in classifying which admission had an ADE (ESS and sensitivity = 0.14 for codes in any position), yet they had among the highest measures of association with an ADE (RR = 49–55) and good predictive value (PPV = 28–31%).

Figure 2 depicts the model for ADEs manifested as delirium and psychosis. Using 4 nodes, the HOCTA model predicted drug-induced delirium if LOS was ≥3 days and the patient was older than 75 years or had diabetes mellitus; or if LOS was >5 days. The model had good accuracy: ESS = 0.652 (95% CI: 0.651–0.654); sensitivity = 94.4% (95% CI: 94.2–94.6), specificity = 71%; and ROC AUC = 0.83. The prevalence of the ADE was 0.95%, or 9.5 events per 1000 admissions. Despite a positive likelihood ratio of 3.2, PPV was low at 2.93% (95% CI: 2.90–2.96). Forcing inclusion of the aggregate variable indicative of a drug-induced delirium degraded the HOCTA model's performance. Of the 41 hospitals in the main cohort, the sensitivity analysis included 11 sites having over 100 admissions (range, 112-296). In this analysis, the median point estimates of various performance characteristics were as follows: ESS = 0.67 (range, −0.26 to 0.70); sensitivity = 1% (range, 0–1%); specificity = 69% (range, 63–73%); ROC AUC = 84% (range, 36–85%); PPV 4% (range, 0–8%). These medians and ranges did not change substantially when the required minimum number of admissions was raised to 150.

DISCUSSION

We explored the ability of HOCTA to produce accurate models for ADEs when it was applied to all available administrative data. We sought models that could be used to screen charts for finding targeted ADEs and to be used as PSIs for estimation of ADE rates. HOCTA was able to produce

clinically meaningful, accurate models for both drug-induced bleeding problems and delirium. The model for bleeding problems used an at-risk cohort and was superior to the model for delirium that addressed all admissions. In contrast to current AHRQ PSIs, none of the models included aggregate variables limited to adverse event codes.

The bleeding model exemplified the ability of HOCTA to produce clinically meaningful rules.[23] Two nodes combined as a marker of the targeted ADE: a secondary code for bleeding and a following exception for nonhemorrhagic anemia. Female sex is an independent risk factor for bleeding on anticoagulation.[24] Codes for complications and longer LOS are nonspecific indicators of complications. An aggregate variable for induced coagulopathies doubled as an indicator of more intensive exposure to causative drugs and as a marker for the targeted ADE. Cardiac dysrhythmias and lipid disorders—the latter being a surrogate for coronary artery disease—are additional indicators of more intensive anticoagulant exposure. These variables for exposure help compensate for our imperfect estimation of drug exposure. As explained in the Methods section, we established the at-risk cohort using ICD-9-CM codes likely to be associated with the administration of drugs potentially causing bleeding. Because the at-risk indicator was positive if any of the selected codes were present, all codes had equal weight. The indicators of more intensive exposure define partitions at increased risk.

In addition to being clinically meaningful, the bleeding model may also be useful. The model's accuracy was excellent and sensitivity was very good. When applied to small samples of admissions from individual hospitals, none of these performance values decreased by more than 25%. We expect increased stability with adequate sample sizes. The model's main limitation was moderately low PPV, which reflected the very low incidence proportion of events. Nonetheless, high sensitivity and accuracy with acceptable PPV make the model suitable for retrospective screening of patient charts for quality improvement efforts. The number needed to screen to find 1 admission with drug-induced bleeding or anticoagulation problems is just over 8 compared with over 100 without the model. The model's favorable performance characteristics and its reliance on markers for the targeted ADE may allow it to be used as a PSI to estimate rates of ADEs. The model would overestimate the true ADE rate by a factor of 8, but the crude rate of predicted ADEs could be adjusted to account for imperfect PPV and sensitivity. Further study is necessary to establish the reliability and appropriate risk adjustment of the model before it is used as a PSI.[25]

The delirium model was clinically meaningful but did not perform well. It was largely based on age and LOS, 2 known risk factors for the development of delirium.[16] Longer LOS is also a nonspecific marker for complications during the hospitalization. Diabetes mellitus may function in the model as an indicator of higher morbidity from chronic disease. Aside from its clinical interpretation, the quantitative performance of this model was disappointing. Sensitivity was excellent, but PPV was very low. Moreover, the sensitivity analysis of a small sample of admissions at each hospital demonstrated wide performance variation. Despite this model's

low PPV and concerns about its generalizability for rate estimation, its high sensitivity may allow it to be a useful tool for screening patient charts. The number needed to screen to find 1 admission with drug-induced delirium is about 30, less than the 100 that would otherwise be necessary. This benefit is modest, but the PPV is within the range of other models in current use for ADE surveillance.[26] This sensitive model may best be used in conjunction with other surveillance rules designed for retrospective chart review.[27] Finally, the delirium model is not suitable for a PSI. A very low PPV would require a high discount factor of nearly 97% to estimate the rate of drug-induced delirium. This high discount factor and the lack of clinically specific markers for delirium decrease the likelihood that the model will be responsive to changes in the true rates of ADEs.

This report illustrates several lessons for modeling sparse ADEs in administrative data. First, the estimation of an at-risk cohort facilitated an improved model for bleeding and anticoagulation problems. Without an at-risk cohort, the best model we created for drug-induced bleeding had performance characteristics approximating those of the delirium model. Second, grouping the codes into clinically meaningful predictors also improved the resulting model's performance by reducing the sparseness of relevant, positive variables. Third, code-related markers for ADEs performed better if the codes were submitted for billing in a secondary position. This reflects the finding of others that ADE codes in the primary diagnostic position often represent outpatient ADEs.[12] Finally, non-ICD9-code administrative data such as age, length of stay, and discharge disposition may be the most important variables of models, as was the case for delirium.

Although questions remain about the validity of models generated in this report, the models have been empirically derived and in contrast, all of the AHRQ PSIs were developed and validated by expert consensus with empirical support.[4,5] Many have undergone empiric validation of PPVs. At the time of this writing, however, there was only 1 report of a validated PSI using a cohort sample.[28] The sensitivity and specificity of the other PSIs remain unknown. As with the HOCTA models, questions remain about responsiveness of PSIs to changes in rates of the targeted event.[25]

There are several limitations to this study. First, the models were developed and validated for inpatients in Utah hospitals and may not replicate in other states with different coding practices and data structures. Second, model validation was limited to jackknife methods for model development. Whole-sample validation, through exploration of all possible models using bootstrap methods at the stage of model development, is not computationally practical. Using a split-sample validation set was not practical due to low prevalence of the dependent variables. Third, the models may be over fit. We attempted to minimize this outcome by using procedures to ensure jackknife stability and experimentwise statistical significance. The requirement of at least 20 observations at each branch further reduces the likelihood of over fitting. HOCTA is not subject to the empirically based rule of thumb of 10 cases per variable developed for linear models.[29]

There are 16 cells in the design matrix of the HOCTA model for bleeding. There would be 2^8 or 256 cells in the design matrix if logistic regression were used; most of these cells would have been empty. Fewer cells in the HOCTA design matrix highlights the suitability of HOCTA for modeling data with sparse independent and dependent variables.[30] Finally, the ICD-9-CM codes for exposure to drugs potentially causing bleeding were derived from literature review and aggregated into 1 exposure indicator. This exposure variable would be better specified as an empirically derived and validated model itself.

In conclusion, these results demonstrate the ability of HOCTA to produce accurate and clinically meaningful models for selected ADEs from large data sets with sparsely positive independent and dependent variables. Non-ICD-9-code administrative data describing patient demographics and hospitalization information may be central to these models. In contrast, ICD-9-CM codes designed for ADEs and poisoning lacked discrimination and were excluded from these models. Development of models was greatly facilitated by focusing on the group estimated to be exposed to drugs capable of causing the targeted ADE. The application of HOCTA to all available administrative data is a promising approach for the development and validation of PSIs for ADEs.

REFERENCES

1. Institute of Medicine. *To Err Is Human: Building a Safer Health System.* Washington, DC: National Academy Press; 1999.
2. Institute of Medicine. *Preventing Medication Errors: Quality Chasm Series.* Washington, DC: National Academy Press; 2006.
3. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. *JAMA.* 2003;290: 1868–1874.
4. Agency for Healthcare Research and Quality. *AHRQ Quality Indicators—Guide to Patient Safety Indicators.* AHRQ Pub. 03-R203, Version 3.0a (May 1, 2006). Rockville, MD: Agency for Healthcare Research and Quality, US Department of Health & Human Services; 2006.
5. Lawthers AG, McCarthy EP, Davis RB, et al. Identification of in-hospital complications from claims data. Is it valid. *Med Care.* 2000; 38:785–795.
6. Weingart SN, Iezzoni LI, Davis RB, et al. Use of administrative data to find substandard care: validation of the complications screening program. *Med Care.* 2000;38:796–806.
7. Yarnold PR, Soltysik RC. *Optimal Data Analysis: A Guidebook with Software for Windows.* 1st ed. Washington, DC: American Psychological Association; 2005.
8. Arozullah AM, Parada J, Bennett CL, et al. A rapid staging system for predicting mortality from HIV-associated community-acquired pneumonia. *Chest.* 2003;123:1151–1160.
9. Arozullah AM, Yarnold PR, Weinstein RA, et al. A new preadmission staging system for predicting inpatient mortality from HIV-associated *Pneumocystis carinii* pneumonia in the early highly active antiretroviral therapy (HAART) era. *Am J Respir Crit Care Med.* 2000;161:1081–1086.
10. Kucera CM, Greenberger PA, Yarnold PR, et al. An attempted prospective testing of an asthma severity index and a quality of life survey for 1 year in ambulatory patients with asthma. *Allergy Asthma Proc.* 1999;20:29–38.
11. Yarnold PR, Michelson EA, Thompson DA, et al. Predicting patient satisfaction: a study of two emergency departments. *J Behav Med.* 1998;21:545–563.
12. Hougland P, Xu W, Pickard S, et al. Performance of international classification of diseases, 9th revision, clinical modification codes as an adverse drug event surveillance system. *Med Care.* 2006;44:629–636.
13. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf.* 1999;20:109–117.
14. Hurdle JF, Weir CR, Roth B, et al. Critical gaps in the World's largest

electronic medical record: ad hoc nursing narratives and invisible adverse drug events. *AMIA Annu Symp Proc*. 2003;2003:309–312.

15. Nebeker JR, Hoffman JM, Weir CR, et al. High rates of adverse drug events in a highly computerized hospital. *Arch Intern Med*. 2005;165:1111–1116.

16. Potter J, George J. The prevention, diagnosis and management of delirium in older people: concise guidelines. *Clin Med*. 2006;6:303–308.

17. UpToDate: [Web site] Published by UpToDate. 2007. Available at: http://www.uptodate.com/. Accessed March 1, 2007.

18. Proceedings of the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy: evidence-based guidelines. *Chest*. 2004;126:172S–696S.

19. Agency for Healthcare Research and Quality. HCUP CCS. Healthcare Cost and Utilization Project (HCUP). 2007. Available at: http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Accessed March 1, 2007.

20. Yarnold PR. Discriminating geriatric and nongeriatric patients using functional status information: an example of classification tree analysis via UniODA. *Educ Psychol Meas*. 1996;56:656–667.

21. Lachenbruch PA. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*. 1967;23:639–645.

22. Cantor SB, Kattan MW. Determining the area under the ROC curve for a binary diagnostic test. *Med Decis Making*. 2000;20:468–470.

23. Kyriacou DN, Yarnold PR, Stein AC, et al. Discriminating inhalational anthrax from community-acquired pneumonia using chest radiograph findings and a clinical algorithm. *Chest*. 2007;131:489–496.

24. Shireman TI, Howard PA, Kresowik TF, et al. Combined anticoagulant-antiplatelet use and major bleeding events in elderly atrial fibrillation patients. *Stroke*. 2004;35:2362–2367.

25. Rosen AK, Zhao S, Rivard P, et al. Tracking rates of patient safety indicators over time: lessons from the Veterans Administration. *Med Care*. 2006;44:850–861.

26. Jha AK, Kuperman GJ, Teich JM, et al. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *J Am Med Inform Assoc*. 1998;5:305–314.

27. Rozich JD, Haraden CR, Resar RK. Adverse drug event trigger tool: a practical methodology for measuring medication related harm. *Qual Saf Health Care*. 2003;12:194–200.

28. Romano PS, Yasmeen S, Schembri ME, et al. Coding of perineal lacerations and other complications of obstetric care in hospital discharge data. *Obstet Gynecol*. 2005;106:717–725.

29. Harrell FE, Lee KL, Matchar DB, et al. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep*. 1985;69:1071–1077.

30. Greenland S, Schwartzbaum JA, Finkle WD. Problems due to small samples and sparse data in conditional logistic regression analysis. *Am J Epidemiol*. 2000;151:531–539.