Subarachnoid hemorrhage admissions retrospectively identified using a prediction model

OPEN

Shane W. English, MD
Lauralyn McIntyre, MD
Dean Fergusson, PhD
Alexis Turgeon, MD
Marlise P. dos Santos, MD
Cheemun Lum, MD
Michaël Chassé, MD, PhD
John Sinclair, MD
Alan Forster, MD
Carl van Walraven, MD

Correspondence to Dr. English: senglish@ohri.ca

ABSTRACT

Objective: To create an accurate prediction model using variables collected in widely available health administrative data records to identify hospitalizations for primary subarachnoid hemorrhage (SAH).

Methods: A previously established complete cohort of consecutive primary SAH patients was combined with a random sample of control hospitalizations. Chi-square recursive partitioning was used to derive and internally validate a model to predict the probability that a patient had primary SAH (due to aneurysm or arteriovenous malformation) using health administrative data.

Results: A total of 10,322 hospitalizations with 631 having primary SAH (6.1%) were included in the study (5,122 derivation, 5,200 validation). In the validation patients, our recursive partitioning algorithm had a sensitivity of 96.5% (95% confidence interval [CI] 93.9–98.0), a specificity of 99.8% (95% CI 99.6–99.9), and a positive likelihood ratio of 483 (95% CI 254–879). In this population, patients meeting criteria for the algorithm had a probability of 45% of truly having primary SAH.

Conclusions: Routinely collected health administrative data can be used to accurately identify hospitalized patients with a high probability of having a primary SAH. This algorithm may allow, upon validation, an easy and accurate method to create validated cohorts of primary SAH from either ruptured aneurysm or arteriovenous malformation. *Neurology*® 2016;87:1557–1564

GLOSSARY

AVM = arteriovenous malformation; **CI** = confidence interval; **DAD** = Discharge Abstract Database; **ICD** = *International Classification of Diseases*; **LR+** = positive likelihood ratio; **RP** = recursive partitioning; **SAH** = subarachnoid hemorrhage; **TOH** = The Ottawa Hospital; **TOHDW** = Ottawa Hospital Data Warehouse.

Primary subarachnoid hemorrhage (SAH) is an important but rare cause of cerebrovascular accidents that predominantly results from a ruptured saccular aneurysm or arteriovenous malformation (AVM)[1–4] and leads to devastating outcomes with less than a third of patients making a complete recovery.[5] Prospectively studying uncommon diseases like primary SAH can be very costly and time-consuming; this makes the retrospective study of such populations using health administrative data attractive.

The identification of patients in administrative databases is often accomplished with diagnostic coding (e.g., ICD codes). This is very problematic in rare diseases because these codes frequently lack the accuracy required to reliably identify rare cases.[6–8] Few studies have documented the accuracy of ICD codes in SAH.[9–17] The positive predictive value of the code ranged among the studies from 33% to 100% but these studies have questionable utility because they were small (1–247 patients with SAH). More importantly, the prevalence of SAH in these samples was much higher than that in the general hospital population; this will result in study positive predictive values being extensively overestimated, resulting in a high number of cases incorrectly

Supplemental data at Neurology.org

From the Department of Medicine (Critical Care) (S.W.E., L.M.), Clinical Epidemiology Program (S.W.E, L.M., D.F., A.F., C.v.W.), Ottawa Hospital Research Institute/The Ottawa Hospital; Department of Anesthesia (Critical Care) (A.T., M.C.), Hôpital de L'Enfant-Jésus, Quebec; and Departments of Medical Imaging (M.P.d.S., C.L.), Surgery (Neuro-Surgery) (J.S.), and Medicine (C.v.W, A.F.), The Ottawa Hospital, Canada.

Go to Neurology.org for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article. The Article Processing Charge was paid by the authors.

labeled as primary SAH. Only one study, involving 58 patients, reported the sensitivity and specificity of SAH codes (98% [95% confidence interval (CI) 90%–100%] and 92% [84%–96%], respectively).[14] This translates to a positive likelihood ratio (LR+) of 12.25. As a rare disease, primary SAH has a prevalence of only 0.06% among all hospitalizations (appendix e-1A at Neurology.org), resulting in very low pretest probability of disease. Thus, with such a low LR+, patients having these codes will have only a 0.74% probability of truly having the disease. These results demonstrate the dire need for better methods to accurately identify these patients retrospectively.

The objective of this study was to derive and validate a prediction model to accurately predict the probability that a patient truly had a primary SAH using variables widely available in health administrative data. Such a model would be widely generalizable and pertinent to the researcher. The improved accuracy of subsequent epidemiologic study has direct relevance to the bedside knowledge user including his or her understanding of natural history and prognostication.

METHODS

Study setting. The Ottawa Hospital (TOH) is a tertiary care university hospital with 1,150 beds that provides all neurosurgical services to the City of Ottawa and the Champlain Local Health Integrated Network (approximate catchment population of 1.2 million, or 10% of the population of Ontario[18]). The Ottawa Hospital Data Warehouse (TOHDW) is a collection of health datasets containing clinical and administrative data for all inpatient encounters at TOH. Included in the holdings of TOHDW is the Discharge Abstract Database (DAD), which is available from July 1, 2002.[19] The Discharge Abstract is an administrative dataset mandated by regulatory bodies that contains diagnostic, procedural, demographic, and administrative information for every hospitalization in Canada.[20] Most developed countries have similar databases for hospitalizations. The diagnostic codes utilized in the DAD are ICD-10CA. Procedures are captured according the Canadian Classification of Health Interventions codes.

Standard protocol approvals, registrations, and patient consents. Ethics approval for the study was obtained from the Ottawa Health Sciences Network Research Ethics Board using a waived consent model for this retrospective analysis.

Patient cohort used to derive and validate the prediction model. To derive and validate our prediction model, we used a complete cohort of previously identified primary SAH patients (reference population) admitted to TOH between July 1, 2002, and June 30, 2011.[21] We defined primary SAH as an SAH (supported by findings of CT head scan, lumbar puncture, or autopsy) that was the result of a ruptured aneurysm or AVM (from findings of angiography or autopsy). For the control group, we randomly selected from all TOH inpatient admissions over the same time period 2.5% of all patients greater than 17 years of age (figure 1). A random sample of 2.5% of inpatient admissions provided a cohort of approximately 10,000 admissions, which was the largest data sample achievable that maintained optimal function of the statistical software used for the analysis. Specific case-controls were not used to keep the sample as reflective of all hospital admissions against which the model is designed to identify primary SAH patients.

Variable selection for model derivation. The objective of our model was to predict primary SAH status using administrative data found in the DAD. Variables considered included demographic information (including patient age and sex) and hospitalization information (including admission service, diagnostic codes, hospital admission type [urgent or not], length of stay, surgical procedures coded, and transfers to intensive care unit). We also identified diagnostic codes and procedural codes that could be associated with SAH status for consideration of inclusion in the model by the following:

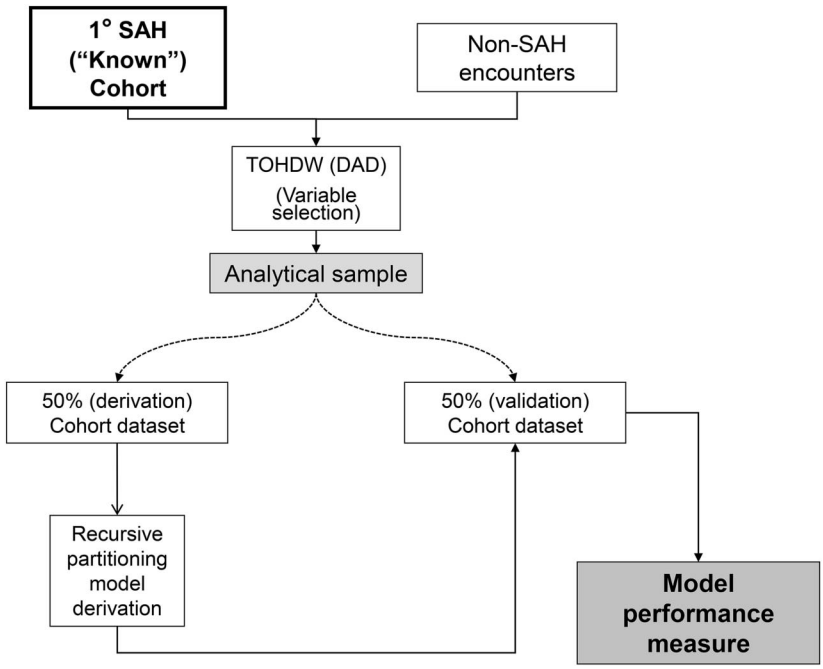
1. Collecting all diagnostic and procedural codes contained within the DAD for each encounter identified previously as primary SAH.
2. All diagnostic codes were grouped in order of frequency. This was repeated for procedural codes. Each diagnostic or procedural code with a frequency $\geq$10 (i.e., was observed in more than 10 encounters with primary SAH) was considered for inclusion in the model.
3. For each diagnostic and procedural code identified in (2), a binomial variable was created for each encounter identified in the derivation and validation datasets, where 1 represented the diagnosis or procedure as being present (or having occurred) and 0 as not. A missing variable would have been considered not present and assigned a 0.

Analysis. ***Model derivation and validation.*** We first constructed a multivariate logistic regression model but true cases were still not identified (results presented in appendix e-1B). We therefore conducted a recursive partitioning (RP) model. RP is a nonparametric regression method that clusters responses into homogenous groups.[22] It is useful in situations when the goal of correctly identifying a specific outcome group with high sensitivity is more important than overall accuracy.[23,24] Chi-square tests with potential predictor variables against the dependent variable (primary SAH) created successive partitioning. At each step, every potential variable was considered for entry into the model by performing a χ^2 test. The variable with the highest χ^2 statistic meeting statistical significance was entered into the model. The Fisher Exact test was used for variables in which at least one cell in the 2×2 contingency table had fewer than 5 observations.

A p value of 0.05 was used to determine statistical significance of the association between the potential predictor variable and primary SAH. In this fashion, a regression tree was constructed by creating branches at each splitting variable (node). Thus for each node, a branch was created for variable presence (variable 1) or absence (variable 0) of the splitting variables. This was repeated until no further variables met statistical significance for entry or each cell contained $\leq$1 of either outcome (primary SAH or no primary SAH). These were defined as terminal nodes. Variables that lacked clinical sensibility and did not contribute to a terminal node with at least 50% observed primary SAH events were removed.

Model performance. We measured model performance by generating 2×2 tables comparing predicted outcome with actual

Figure 1 Derivation and validation of a prediction model study design



This schema depicts the methods used to derive and validate a prediction model that identifies patient admissions with high probability of being the result of primary subarachnoid hemorrhage (SAH) using routinely collected health administrative data. DAD = Discharge Abstract Database; TOHDW = The Ottawa Hospital Data Warehouse.

outcome to determine sensitivity, specificity, and likelihood ratios. Three separate classifications were examined to determine the expected outcome predicted by the model: (1) the terminal node was classified as primary SAH if the observed event rate exceeded or equaled 50%; (2) the terminal node was classified as primary SAH if the observed event rate exceeded or equaled 75%; and (3) the terminal node was classified as primary SAH if the observed event rate exceeded or equaled 90%.

The performance of the algorithm was then tested against the validation set. We measured model performance (accuracy) by comparing expected and observed number of patients with primary SAH based on the classification with 2 × 2 tables to calculate sensitivity, specificity, and LR+ with 95% CIs. All analyses were completed using SAS 9.2.

RESULTS

Patient cohort. In total, 10,322 patient encounters were included, 631 of which had primary SAH (6.1%). These were randomly divided into 2 groups for model derivation (n = 5,122; 315 [6.1%] primary SAH) and validation (n = 5,200; 316 [6.1%] primary SAH).

Predictor variables. A total of 108 potential predictor variables were considered for inclusion in the model by considering all of the diagnostic and procedural codes that appeared more than 10 times in the discharge abstracts (DAD) of patients with primary SAH. Sixty-three variables were diagnostic codes, 38 were procedural codes, and 7 others were encounter characteristics (appendix e-1C). In addition to those variables found in the final models predicting primary SAH, we considered known risk factors (age, sex, hypertension), presenting characteristics (headache, visual disturbances, cardiac arrest), disease course variables (hydrocephalus, vasospasm, pulmonary edema, ventriculitis, seizures, stroke), and diagnostic/therapeutic variables (admitting service, brain imaging, externalized ventricular drain insertion).[25–29]

Logistic regression predictive model. Twelve variables were included in the multivariate logistic regression model (appendix e-1B, table e-B1). Using the validation dataset and predicted probability of primary SAH of ≥50%, the model had a sensitivity of 96.8% and specificity of 99.7%, leading to an LR+ of 312 (table 1).

Recursive partitioning model and predictive model performance. A total of 10 variables entered the final model with 12 splits and 13 terminal nodes (figure 2). Each terminal node consisted of patient groups with varying prevalence of primary SAH. The most highly discriminative predictor variable was the diagnostic code for SAH (I60). The presence of this code increased the probability of SAH from 6.1% to 97.6% (figure 2A). However, 7 patients (2.4%) with the code did not actually have a SAH diagnosis; none of these people had a code for therapeutic occlusion of an intracranial vessel (1JW51) and all had a hospital length of stay equal to or exceeding 48 hours (figure 2A). In the absence of introducing other variables, the sole use of the SAH diagnostic code (I60) would have missed 35 (11.1%) true cases of primary SAH (sensitivity of 88.9%). In the absence

Table 1 Model performance in the validation group

Prevalence of SAH required in terminal node to indicate SAH	Sensitivity, % (95% CI)	Specificity,% (95% CI)	LR+ (95% CI)	LR− (95% CI)	Probability[a] of truly having SAH if: Meets criteria for SAH (%)	Probability[a] of truly having SAH if: Does not meet criteria for SAH (%)
Logistic regression model						
≥50%	96.8 (94.3-98.3)	99.7 (99.5-99.8)	312 (190-523)	0.03 (0.02-0.06)	34.6	0.005
≥75%	94.0 (90.8-96.1)	99.7 (99.5-99.8)	353 (205-608)	0.06 (0.04-0.09)	37.1	0.010
≥90%	81.3 (76.7-85.2)	99.9 (99.7-99.9)	662 (297-1,475)	0.19 (0.15-0.24)	52.6	0.032
Recursive partitioning model						
≥50%	96.5 (93.9-98.0)	99.8 (99.6-99.9)	483 (254-876)	0.04 (0.02-0.06)	44.7	0.007
≥75%	96.5 (93.9-98.0)	99.8 (99.6-99.9)	483 (254-876)	0.04 (0.02-0.06)	44.7	0.007
≥90%	75.6 (70.6-80.0)	100 (99.9-100)	3,694 (520-26,244)	0.24 (0.20-0.30)	86.1	0.040

Abbreviations: CI = confidence interval; LR+ = positive likelihood ratio; LR− = negative likelihood ratio; SAH = subarachnoid hemorrhage.
This table compares the performance of the logistic regression and recursive partitioning (RP) model (figure 2) for the identification of patients with primary SAH using health administrative data. RP models cluster patients into homogenous groups based on the outcome. We used 3 different probability thresholds (column 1) for classifying patients in each prediction model as expecting to be with or without primary SAH. The probability of truly having or not having a primary SAH (final 2 columns) is based on the overall prevalence of SAH at the study hospital of 16.7 cases per 10,000 hospitalizations.
[a] Indicates posttest probability of having primary SAH assuming an SAH prevalence of 16.7/10,000 hospital admissions at The Ottawa Hospital. There was no difference in performance between SAH prevalence ≥50% and ≥75%.

of having the diagnostic code I60 (figure 2B), the probability of SAH in these patients was only 0.7%. However, the introduction of other diagnostic codes (n = 4), procedural codes (n = 3), and encounter characteristics (n = 2) created terminal nodes in which the probability of SAH varied between 75% and 100% (figure 2B, table 2).

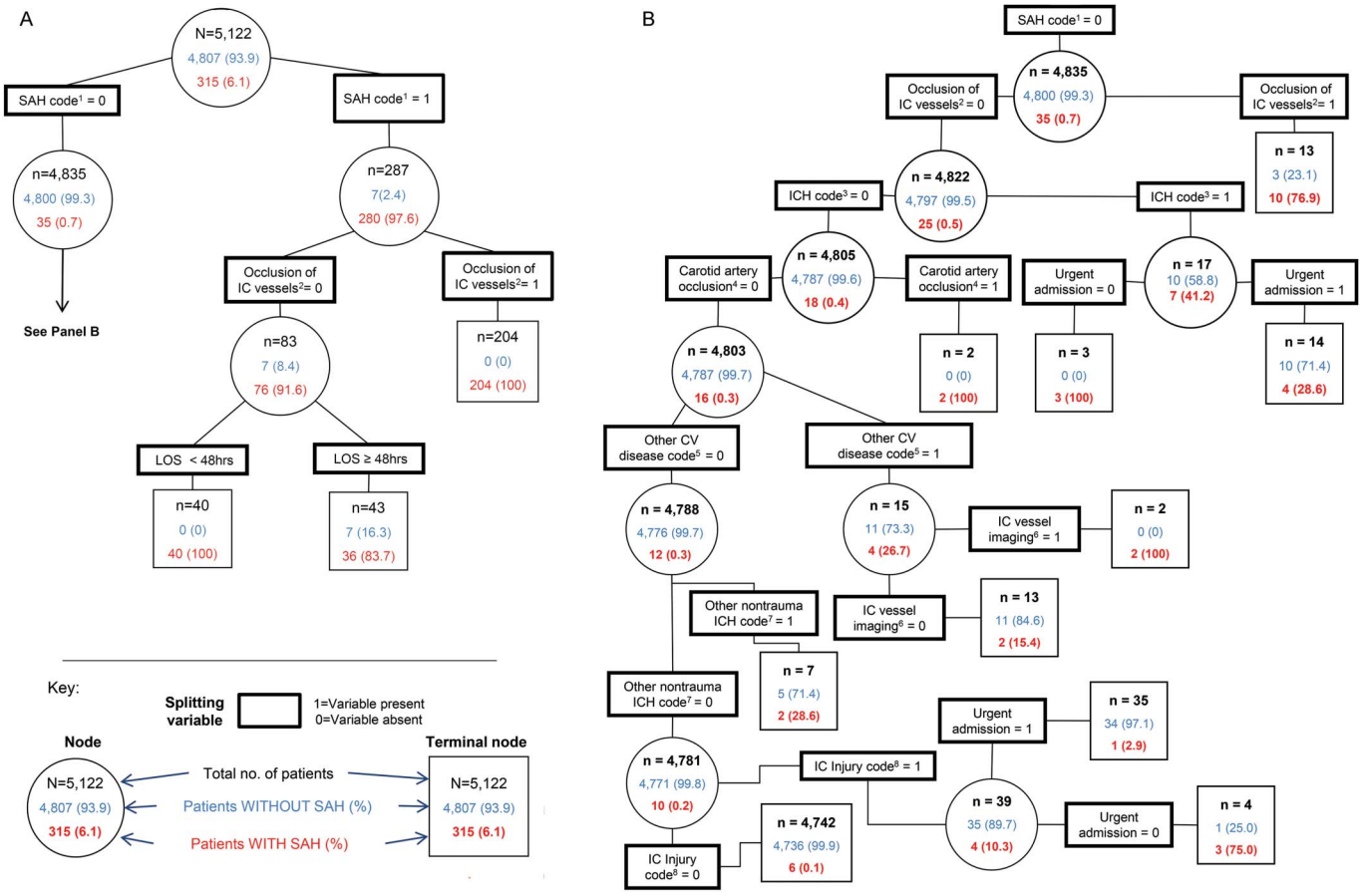
Performance characteristics of the model in the validation cohort varied based on the SAH prevalence required in terminal nodes to delineate SAH (table 1). Results were identical if the prevalence threshold of 50% or 75% was used in the recursive partitioning model with a sensitivity of 96.5% and a specificity of 99.8% (which translated to an LR+ of 483). Increasing the threshold to 90% yielded a higher specificity and increased the LR+ to 3,694; at a cost, however, of missing a quarter of all cases (sensitivity 75.6%). The performance of the logistic regression and recursive partitioning models were similar but the latter had a higher LR+, which led to greater posttest probability of identifying primary SAH patients (table 1). At TOH, the prevalence of primary SAH is 16.7/10,000 hospital admissions (unreported data); thus the probability of any admission being for primary SAH is 0.17%. Implementing the recursive partitioning model (with the 50% prevalence threshold), identified patients would have a 44.7% probability of truly having a primary SAH (table 1). Conversely, the odds of truly having primary SAH, if not identified by the model, are 7 in 100,000 admissions.

DISCUSSION We have created a model that uses health administrative data to accurately identify hospital encounters having a strong likelihood of primary SAH. This model can be used to screen large populations of patients with health administrative data to identify large cohorts of SAH using primary chart review. This study demonstrates that complete cohorts of rare diseases, such as primary SAH, can be identified using routinely collected health data facilitating timely, cost-effective, and large-scale epidemiologic studies.

Our study has several notable findings and implications. First, the methods implemented in this study were comprehensive, and are necessary to minimize misclassification that can arise, particularly in rare diseases, from the solitary use of diagnostic codes for case ascertainment. These methods effectively incorporate overlapping search strategies. Overlapping strategies have improved the accuracy of case ascertainment in other disease processes including all stroke, osteoporosis, and acute kidney injury.[20,30,31] In the current study, misclassification was minimized by including multiple variables in a recursive partitioning model. As such, not having a DAD diagnostic code for primary SAH was insufficient alone to exclude SAH. Such patients were still likely to have primary SAH if, for example, they had undergone occlusion of an intracranial vessel, thereby improving the sensitivity of the search strategy compared to diagnostic code alone. With this model, given a pretest probability of primary SAH of approximately 16.7/10,000 hospital admissions, the probability of a patient actually having primary SAH if deemed positive by the model is 45%, or nearly 1 in 2. This is a substantial improvement on 0.74% probability from a diagnostic code alone as demonstrated in the Introduction.

This leads to the second implication in that accurately and feasibly identifying large complete cohorts

Figure 2 Recursive partitioning (RP) model to identify primary subarachnoid hemorrhage (SAH)



(A, B) Each bolded box represents a splitting variable (which includes the presence or absence of a diagnostic code, a procedural code, or a hospitalization characteristic; the presence of a splitting variable is indicated by 1 and its absence by 0). Splitting variables successively partition the sample or node (presented in the ovals) until no further partitions are possible, creating a terminal node (rectangular boxes). Within each splitting node or terminal node, the number (and respective proportion) of patients truly with (SAH) and without (noSAH) primary SAH are presented. Corresponding codes for diagnoses or procedures denoted by superscripted numbers: 1 = I60, 2 = 1JW51, 3 = I61, 4 = 1JE51, 5 = I67, 6 = 3JW10, 7 = I62, 8 = S06. CV = cerebrovascular disease; IC = intracranial; ICH = intracranial hemorrhage; LOS = hospital length of stay (days); N = total number of patients.

of rare diseases like primary SAH, retrospectively, is possible. The ability to identify large, population-based cohorts of patients with a particular disease is of paramount importance to understand disease incidence, patient characteristics, risk factors, prognostic factors, and patient outcomes as well as health care resource utilization. For example, large-scale epidemiologic studies using cohorts derived from health administrative data with validated methods have been essential in our current understanding of severe sepsis[32–34] and cardiovascular disease.[35,36] Comprehensive case ascertainment methods are essential for accurate epidemiologic study[33]; we present an accurate method for primary SAH case ascertainment that could facilitate large-scale population-based study and importantly further our understanding of primary SAH disease epidemiology.

This study's strengths support its findings. First, we demonstrate that case identification that includes but is not solely reliant on diagnostic codes as part of a case ascertainment strategy may improve accuracy. However, diagnostic codes cannot be discounted altogether in retrospective case ascertainment. We previously demonstrated that using ICD codes as part of a case ascertainment strategy identified additional cases of SAH that would have otherwise been missed.[21] This is because the data necessary to positively identify such a patient were either unavailable or associated with a separate hospital encounter. In our validation dataset, use of the diagnostic code for SAH alone had a sensitivity of 88.9%, which improved to 96.5% when additional variables are considered. Others have used ICD codes as part of an algorithm, including the British SAH study by Pobereskin,[37] where multiple overlapping methods (including diagnostic codes, imaging results, and operative datasets) were used to identify their cohort. With the current model, we overcome the poor generalizability of these case ascertainment methods given that few institutions have access to and ability

Table 2 Pathway characteristics leading to high probability of primary subarachnoid hemorrhage (SAH) in the validation group

Pathway to terminal node			
SAH code[1] (present or absent)	Additional variables	Total patients in node	Total patients with primary SAH (proportion, %)
Present		287	97.6
Present	+ Occlusion of IC vessels[2] (1)	204	100
	+ LOS ≤48 h	40	100
Absent	+ Occlusion of IC vessels[2] (1)	13	76.9
	+ Intracranial hemorrhage code[3] (1), + urgent admission (0)	3	100
	+ Intracranial hemorrhage code[3] (0), + carotid artery occlusion[4] (1)	2	100
	+ Intracranial hemorrhage code[3] (0), + carotid artery occlusion[4] (0), + other cerebrovascular disease code[5] (1), + intracranial vessel imaging[6] (1)	2	100
	+ Intracranial hemorrhage code[3] (0), + carotid artery occlusion[4] (0), + other cerebrovascular disease code[5] (0), + other nontrauma intracranial hemorrhage code[7] (0), + intracranial injury code[8] (1), + urgent admission (0)	4	75

Abbreviations: IC = intracranial; LOS = length of stay.
This table describes the characteristics of patients from the validation set with high probability of having primary SAH from the recursive partitioning model presented in figure 2. 1 signifies the presence of the diagnostic or procedural code whereas 0 represents its absence. Corresponding codes for diagnoses or procedures denoted by superscripted numbers: 1 = I60, 2 = 1JW51, 3 = I61, 4 = 1JE51, 5 = I67, 6 = 3JW10, 7 = I62, 8 = S06.

to utilize such granular data. In our prediction model, only variables that are widely available in health administrative records were used, making the instituted methods potentially widely generalizable. Second, we utilized recursive partitioning methods to develop the model over logistic regression. This approach has been previously advocated by prediction rule experts when trying to completely separate 2 groups (e.g., those with and without a disease) using a highly sensitive rule.[24] Further, to avoid overfitting the data, separate derivation and validation sets were used, accomplishing the key first steps in developing a prediction rule.[38] Finally, we employed likelihood ratio formulation of the Bayes theorem to establish posttest probability of disease, which is less prone to future overestimations in disease probability caused by differences in disease prevalence.[39]

Potential study weaknesses should be kept in mind when interpreting its results. The retrospective nature of this study in identifying patients with a specific disease process will always have inherent limitations that can only be overcome with a rigorous prospective protocol. We have attempted to limit any misclassification bias by using multiple predictive variables in our algorithms. It is possible that center-related systematic errors in coding or center-specific tendencies in relation to diagnostic workup and therapeutic approaches biased the prediction algorithm, limiting its generalizability to other centers. Certainly, validation of the prediction model in other settings is necessary to ensure its accuracy using datasets generated from other health care centers.

This study demonstrates how routinely collected and widely available administrative data can be used to predict the probability that any given hospital encounter is the result of a primary SAH (from either a ruptured aneurysm or AVM), thereby facilitating the identification of a complete and accurate cohort of a rare disease. The validity of our method must be demonstrated using datasets from other institutions.

AUTHOR CONTRIBUTIONS

S. English: study concept and design, acquisition of data, statistical analysis and interpretation, drafting of manuscript. L. McIntyre: study design, analysis and interpretation of data, manuscript revision, project supervision. D. Fergusson: analysis and interpretation of data, critical review of manuscript. A. Turgeon: study design, interpretation of data, manuscript revision. M. dos Santos: acquisition of data, interpretation of data, manuscript revision. C. Lum: interpretation of data, critical revision of manuscript for intellectual content. M. Chassé: analysis and interpretation of data, manuscript revision. J. Sinclair: acquisition of data, critical revision of manuscript for intellectual content. A. Forster: Study design, manuscript revision. C. van Walraven: study concept and design, data analysis and interpretation, manuscript review, project supervision.

ACKNOWLEDGMENT

The authors thank Cathy Sun (uOttawa) for secondarily reviewing patient charts in creating the primary SAH and no SAH cohorts.

STUDY FUNDING

No targeted funding reported.

DISCLOSURE

S. English is the recipient of a Canadian Blood Services Fellowship Award during completion of this study. L. McIntyre, D. Fergusson, A. Turgeon, M. dos Santos, C. Lum, M. Chassé, J. Sinclair, A. Forster, and C. van Walraven report no disclosures relevant to the manuscript. Go to Neurology.org for full disclosures.

Received January 26, 2016. Accepted in final form June 29, 2016.

REFERENCES

1. Bonita R, Beaglehole R, North JD. Subarachnoid hemorrhage in New Zealand: an epidemiological study. Stroke 1983;14:342–347.
2. Ostbye T, Levy AR, Mayo NE. Hospitalization and case-fatality rates for subarachnoid hemorrhage in Canada from 1982 through 1991: The Canadian Collaborative Study Group of Stroke Hospitalizations. Stroke 1997;28:793–798.
3. Knekt P, Reunanen A, Aho K, et al. Risk factors for subarachnoid hemorrhage in a longitudinal population study. J Clin Epidemiol 1991;44:933–939.
4. van Gijn J, Rinkel GJE. Subarachnoid haemorrhage: diagnosis, causes and management. Brain 2001;124:249–278.
5. Le Roux PD. Anemia and transfusion after subarachnoid hemorrhage. Neurocrit Care 2011;15:342–353.
6. Manuel DG, Rosella LC, Stukel TA. Importance of accurately identifying disease in studies using electronic health records. BMJ 2010;341:c4226.
7. Williams GR, Jiang JG, Matchar DB, Samsa GP. Incidence and occurrence of total (first-ever and recurrent) stroke. Stroke 1999;30:2523–2528.
8. Stegmayr B, Asplund K. Measuring stroke in the population: quality of routine statistics in comparison with a population-based stroke registry. Neuroepidemiology 1992;11:204–213.
9. Leibson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. Stroke 1994;25:2348–2355.
10. Liu L, Reeder B, Shuaib A, Mazagri R. Validity of stroke diagnosis on hospital discharge records in Saskatchewan, Canada: implications for stroke surveillance. Cerebrovasc Dis 1999;9:224–230.
11. Phillips S, Cameron K, Chung C. Stroke surveillance revisited. Can J Cardiol 1993;9(suppl D):124–125.
12. Mayo N, Danys I, Carleton J. Validity of hospital discharge diagnosis of stroke. Can J Cardiol 1993;9 (suppl D):121–124.
13. Mayo NE, Chockalingam A, Reeder BA, Phillips S. Surveillance for stroke in Canada. Health Rep 1994;6:62–72.
14. Tirschwell DL. Validating administrative data in stroke research. Stroke 2002;33:2465–2470.
15. Broderick J, Brott T, Kothari R, et al. The Greater Cincinnati/Northern Kentucky Stroke Study: preliminary first-ever and total incidence rates of stroke among blacks. Stroke 1998;29:415–421.
16. Rosamond WD, Folsom AR, Chambless LE, et al. Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. Stroke 1999;30:736–743.
17. Roumie CL, Mitchel E, Gideon PS, Varas-Lorenzo C, Castellsague J, Griffin MR. Validation of ICD-9 codes with a high positive predictive value for incident strokes resulting in hospitalization using Medicaid health data. Pharmacoepidemiol Drug Saf 2008;17:20–26.
18. Bains N, Dall K, Hay C, Pacey M, Sarkella J, Ward M. Population Health Profile. Gloucester: Champlain LHIN; 2004.
19. Forster AJ, Taljaard M, Oake N, Wilson K, Roth V, van Walraven C. The effect of hospital-acquired infection with *Clostridium difficile* on length of stay in hospital. CMAJ 2012;184:37–42.
20. van Walraven C, Austin PC, Manuel D, Knoll G, Jennings A, Forster AJ. The usefulness of administrative databases for identifying disease cohorts is increased with a multivariate model. J Clin Epidemiol 2010;63: 1332–1341.
21. English SW, McIntyre L, Fergusson D, et al. Enriched administrative data can be used to retrospectively identify all known cases of primary subarachnoid hemorrhage. J Clin Epidemiol 2016;70:146–154.
22. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol Methods 2009;14:323–348.
23. Stiell IG, Wells GA, Vandemheen KL, et al. The Canadian C-spine rule for radiography in alert and stable trauma patients. JAMA 2001;286:1841–1848.
24. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: a review and suggested modifications of methodological standards. JAMA 1997;277:488–494.
25. Wartenberg KE, Schmidt JM, Claassen J, et al. Impact of medical complications on outcome after subarachnoid hemorrhage. Crit Care Med 2006;34:617–623; quiz 624.
26. Diringer MN, Bleck TP, Claude Hemphill J, et al. Critical care management of patients following aneurysmal subarachnoid hemorrhage: recommendations from the Neurocritical Care Society's Multidisciplinary Consensus Conference. Neurocrit Care 2011;15:211–240.
27. Connolly ES, Rabinstein AA, Carhuapoma JR, et al. Guidelines for the management of aneurysmal subarachnoid hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2012;43:1711–1737.
28. Bederson JB, Connolly ES, Batjer HH, et al. Guidelines for the management of aneurysmal subarachnoid hemorrhage: a statement for healthcare professionals from a special writing group of the Stroke Council, American Heart Association. Stroke 2009;40:994–1025.
29. Keyrouz SG, Diringer MN. Clinical review: prevention and therapy of vasospasm in subarachnoid hemorrhage. Crit Care 2007;11:220.
30. Lix LM, Yogendran MS, Leslie WD, et al. Using multiple data features improved the validity of osteoporosis case ascertainment from administrative databases. J Clin Epidemiol 2008;61:1250–1260.
31. Moore DF, Lix LM, Yogendran MS, Martens P, Tamayo A. Stroke surveillance in Manitoba, Canada: estimates from administrative databases. Chronic Dis Can 2008;29:22–30.
32. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med 2001;29:1303–1310.
33. Linde-Zwirble WT, Angus DC. Severe sepsis epidemiology: sampling, selection, and society. Crit Care 2004;8: 222–226.
34. Weycker D, Akhras KS, Edelsberg J, Angus DC, Oster G. Long-term mortality and medical care charges in patients with severe sepsis. Crit Care Med 2003;31:2316–2323.
35. Rinfret S, Behlouli H, Eisenberg MJ, Humphries K, Tu JV, Pilote L. Class effects of statins in elderly patients

with congestive heart failure: a population-based analysis. Am Heart J 2008;155:316–323.
36. Ivers NM, Schwalm JD, Jackevicius CA, Guo H, Tu JV, Natarajan M. Length of initial prescription at hospital discharge and long-term medication adherence for elderly patients with coronary artery disease: a population-level study. Can J Cardiol 2013;29:1408–1414.
37. Pobereskin LH. Incidence and outcome of subarachnoid haemorrhage: a retrospective population based study. J Neurol Neurosurg Psychiatry 2001;70:340–343.
38. Mcginn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. JAMA 2015;284:79–84.
39. Poses RM, Cebul RD, Collins M, Fager SS. The importance of disease prevalence in transporting clinical prediction rules. Ann Intern Med 1986;105:586–591.