Subarachnoid Hemorrhage: The OttawaHospital Experience

Dr. Shane English

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the MSc Degree in Epidemiology

Epidemiology and Community Medicine
Faculty of Medicine
University of Ottawa

Supervisors:
Dr. Carl van Walraven
Dr. Lauralyn McIntyre

Advisory Committee:
Dr. Dean Fergusson
Dr. Alexis Turgeon

Table of Contents:

Figure Index:

Table Index:

1 Abstract

Background: Primary subarachnoid hemorrhage (1°SAH) is an important disease that causes significant morbidity and mortality. The sparse Canadian epidemiologic literature on 1° SAH is outdated and relies ondiagnostic coding for case ascertainment whichmisses true cases and incorrectly labels non-cases.

Objectives: Primary objective was to identifyall patients with 1° SAH presenting to the Ottawa Hospital (TOH)between July 1, 2002 and June 30, 2011 by deriving and validating a search algorithm using an enriched administrative database. Secondary objectives included: 1) determine incidence and case-fatality rates (CFR) of 1° SAH at TOH; and 3) derive and validate a method to identify 1° SAH using routinely collected administrative data.

Methods:A cohort of 1° SAH patients were identified with a case-defining algorithm that was derived and validated using a combination ofcerebrospinal fluid analysis results and text-search algorithms of both cranial imaging and post-mortem reports. The incidence of 1° SAH was calculated using the total number of hospital encounters over the same time period. CFR was calculated by linking to vital statistic data of hospitalized patients at discharge.An optimal1° SAHprediction model was derived and validated using binomial recursive partitioning built with independent variables obtained from routinely collected administrative data.

Results: Using the case-defining algorithm, 831 patients were identified with a 1° SAH over the study period. Hospital incidence of 1° SAH was 17.2 events per 10,000 inpatient encounters (or 0.17% of encounters) with a case-fatality rate of 18.1%. A validated SAH prediction model based on administrative data using a recursive partitioning model had asensitivity of 96.5% (95% CI 93.9-98.0), a specificity of 99.8% (95%CI 99.6-99.9), and a +LR of 483 (95% CI 254-879). This results in a post-test probability of disease of 45%.

Conclusion: We identified almost all cases of 1° SAHat our hospital using an enriched administrative data. Accurately identifying such patients with routinely collected health administrative data is possible, providing important opportunities to examine and study this patient population. Further studies, involving multiple centres are needed to reproduce these results.

2 Terminology

Diagnosis:

Subarachnoid hemorrhage (SAH): blood in the subarachnoid space as demonstrated on radiographic imagining (namely CT), presence of red blood cells or xanthochromia in cerebrospinal fluid (CSF), or identified on post-mortem examination

Primary subarachnoid hemorrhage (1° SAH): SAH as the result of primary etiologies such as aneurysm (aSAH) or arteriovenous malformation (avmSAH) rupture; also includes perimesencephalic or idiopathic bleeds.

Non-primary subarachnoid hemorrhage: all other causes of subarachnoid hemorrhage not contained within primary subarachnoid hemorrhage, including but not limited to traumatic SAH, and secondary SAH post intraparenchymal hemorrhage.

aSAH: a primary subarachnoid hemorrhage as a result of a ruptured cerebral aneurysm

Aneurysm: defined by the A.D.A.M. Inc Encyclopedia as "an abnormal widening or ballooning of a portion of an artery due to weakness in the wall of the blood vessel" (ref: http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0002109/)

avmSAH: a primary subarachnoid hemorrhage as a result of a ruptured AVM

Arteriovenous malformation (AVM): defined by A.D.A.M. Inc Encyclopedia as "an abnormal connection between the arteries and veins in the brain that usually forms before birth." (ref: http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0001783/)

aSAH/avmSAH: a clinically important subgroup of 1° SAH (approximately 90%) with higher morbidity and mortality whose definitive treatment almost always requires therapeutic stabilization of the ruptured vessel by either neurosurgery or neuro-interventional radiology.

Perimesencephalic SAH: a form of 1° SAH with a distinct hemorrhagic pattern that is around the brainstem (pre-mid-brain and/or pons) with extension into the cisterns, with no other discernable cause on imaging

Incidental Aneurysm or AVM: an asymptomatic aneurysm or AVM found by chance on imaging commissioned for other reasons**.** For the purposes of this thesis, it also includes aneurysms or avm, found on imaging or at autopsy,that are non-contributory to a subarachnoid hemorrhage (outcome of interest), regardless of whether or not they are symptomatic or their finding was by chance alone.

Intracranial hemorrhage (ICH): represents any of a group of bleeds occurring within the cranial vault including epidural hemorrhage (EDH), subdural hemorrhage (SDH), subarachnoid hemorrhage (SAH), intraventricular hemorrhage (IVH) and intraparenchymal/intracerebral hemorrhage.

Administrative Data Research:

Discharge Abstract Database (DAD): an administrative dataset containing personal and clinical data generated for each unique inpatient encounters at every hospital nationally. The components which make up the discharge abstract can be found within The Ottawa Hospital Data Warehouse (TOHDW) in the DAD tables.

Gold Standard: widely accepted set of criteria, definitions or diagnostic modalities, felt to represent the true outcome against which new criteria, definitions or modalities may be compared or tested.

The Ottawa Hospital Data Warehouse (TOHDW): a series of administrative databases containing personal, clinical, administrative and facilities data of patient encounters at the different campuses of The Ottawa Hospital.

Patient Encounter: Any unique visit to hospital by a health care user. Any user may have numerous different encounters. For example a visit to the Emergency Department generates a patient encounter. This is represented in the OHDW by a unique identifier. A subsequent inpatient visit by the same user would generate a different encounter identified in the OHDW by a separate unique identifier.

stmWID: a unique, anonymized, identifier used in TOHDW to identify and classify different laboratory or radiology services

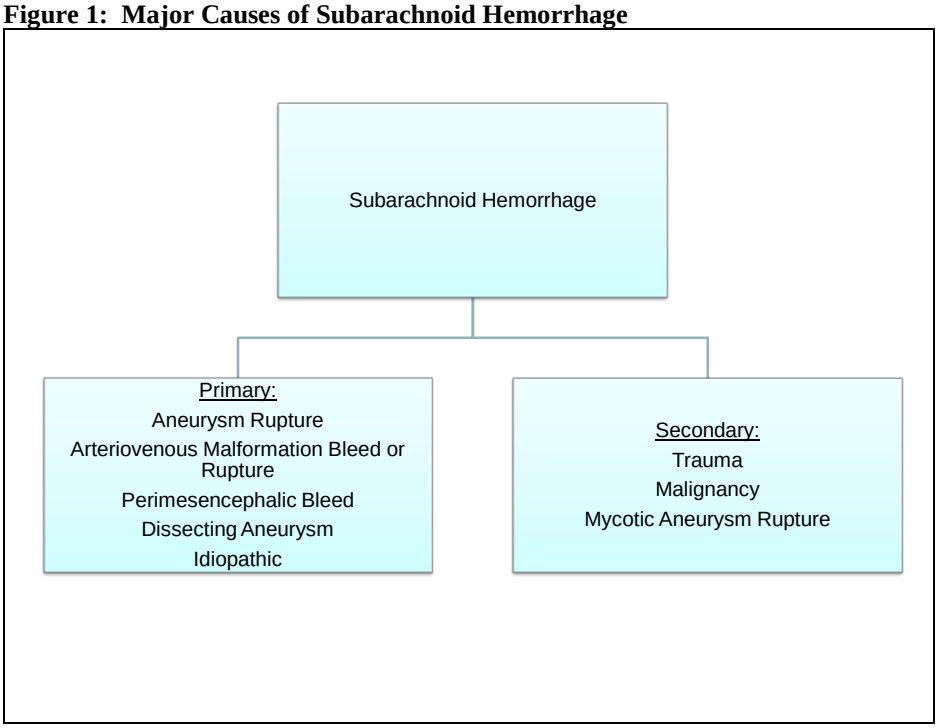
The Ottawa Hospital (TOH): an approximate 1100 bed tertiary care University associated hospital with regional services in trauma, neurosciences, oncology and a variety of other specialties. It is affiliated with the University of Ottawa.

Unique Patient: represents a single individual health care user. Any given unique patient may have any given number of hospital encounters.

3 Background

Primary subarachnoid hemorrhage (1° SAH) is a devastating illness. Although, it is predominantly the result of a ruptured arterial aneurysm[1–4], 1° SAH also occurs as the result of an arteriovenous malformation (AVM) rupture, a perimesencephalic bleed or other idiopathic causes in the absence of secondary causes such as trauma or malignancy (see Figure 1).Previous research has suggested that almost half of all patients presenting with SAHdie.[4–9]Those most affected are between the ages of 40 and 60 years,[7,8]resulting in a significant loss of meaningful life years.[10]Of those who survive the primary event, secondary brain injury due to edema and arterial vasospasmis seen in 50-75% of patients.[7,11]Secondary brain injury has further negative effects on the duration of recovery, health care costs, lost income,morbidity, and mortality.[7,12]SAH has a significant impact on thepatient, the community, and the utilization of healthcare resources.

Figure 1: Major Causes of Subarachnoid Hemorrhage



3.1 Subarachnoid Hemorrhage Incidence and Case-fatality Rate

Research in SAH has focusedon interventions to improve patient outcomes.The incidence of SAH has varied between studies. A recently published systematic review on *all* strokes showed a SAH incidence rangingbetween 2 and 16 cases per 100,000 person-years, depending on country of origin of study.[13]World Health Organization epidemiologic data from the MONICA Stroke study have suggested an SAH incidence ranging from 2.0 (95% CI 1.6 to 2.4) to 22.5 (95% CI 20.9 to 24.1) cases per 100,000 population, depending on country of study, but was limited to 11 European or Asian countries and to those aged 25-64 years.[14]A 2007 systematic review examining specifically SAH incidence demonstrated a range from 4.2 (95% CI 3.1 to 5.7) to 22.7 (95% CI 21.9 to 23.5) per 100,000 person-years.[15] Although this review included four North American studies, they all predated 1990. This may limit its interpretation given that the same review has demonstrated an overall decrease in incidence of 1° SAH over time. Nonetheless, when combined with the other western hemispheric studies, the SAH incidence rate was calculated at 9.1/100,000 person-years (95% CI 8.8 to 9.5).

Several systematic reviews and meta-analyses of observational studies have examined SAH case-fatality rates (CFRs). These studies have reported a significant decline in CFR from SAH over the past two decades.[9,17,21,22]Once again, little recent North American data have been published. In the extensive systematic review (including literature published to mid-2007) by Nieuwkamp and colleagues,[21] case-fatality rates for SAH varied significantly between studies ranging from 8.3% (95% CI 0 to 39) to 66.7% (95% CI 33 to 88). However, 65% of the patients included in this review came from 3 Japanese studies that had significantly lower case-fatality rates compared to patients from Europe, USA, Australia, or New Zealand. Only one 1988 study from North America (which reported a case fatality rate of 32% (95% CI 25 to 39)) was included in this review.[23] Although this review was an update of one published

previously in 1997 (containing studies from 1965 to 1995)[9], the authors changed inclusion criteria to include only prospective studies.

The more recent systematic review by Lovelock et al. (2010), which was published alongside their own cohort findings, included retrospective studies; in doing so, it only added a further two North American studies published in 1984[24] and 2006 (patient data from 1993 to 1997)[25].Case-fatality rates in these 2 studies were 67% and 27% respectively. A very recent study looking at ICU admissions in a number of US hospitals demonstrated that there has been a 77% decline in hospital CFR in SAH since 1988.[22] However, case ascertainment in this study is unclear, the majority of the observed decline appears to have occurred prior to 1995, and the cohort is described as patients with "surgery for subarachnoid hemorrhage", leaving it unclear if this represents only those patients who underwent surgical management or all aSAH patients.

Table 1: Summary of Key 1° SAH Incidence Literature

PUBLICATION AUTHOR (YEAR)	STUDY TYPE	SAMPLE SIZE (N)	CASE DEFINITION	TOTAL SAH CASES	INCIDENCE [95%CI]	POPULATION
Sandvei et al. (2011)[16]	PO	-126,729 patients at risk -2.1 million person years	Predominantly AD ICD codes, neuro-surgical registry prior to 1986, and death registry	214	10.3/100,000$^{\Delta}$	Norway
Lovelock et al. (2010)[17]	PO	91,106	Imaging, CSF and/orPM	38	0.07/1,000 [0.05-0.10]	UK
Feigin et al. (2009)[13]	SR/MA	82.2 million person years (18 studies)	Clinical presentation confirmed with imaging, CSF and/or PM	n/a	2-10 and 4-16/100,000 $^{\Delta}$ (high- and low/mid-income countries respectively)	Asia (1), Australasia (3), Europe (10), N. America (2), S. America (2)
Pobereskin et al. (2001)[18]	Retrospective Population based	Not stated	AD ICD codes, cause of death and neurosurgical registries	800	9.7/100,000 [7.5-12.6]	UK
Anderson C et al. (2000)[19]	Prospective Case-control population based	4.92 million person years	Clinical presentation confirmed with imaging, CSF and/or PM, secondary review of AD ICD codes	400	8.1/100,000 [7.4, 9.0]	Australasia
Ingall et al. (2000)[14]	Cross-sectional comparison of 11 population-base studies	35.9 million person-years of observation	All cases ID'd by AD ICD codes and confirmed with imaging, CSF and/or PM	3368	1.9-26/100,00	East Europe and Asia
Williams et al. (1999)[20]	RC	377544	AD ICD codes	5580	8.8/100,000*	Sample of AmericanHospital Admissions
Ostbye et al. (1997)[2]	RC	Not stated (10 provinces)	AD ICD codes	23140	8.0 and 11.2/100,000 (male, female)[n/a]	Canada

AD=administrative data, CI=confidence interval, ID=identified, MA=meta-analysis, PM=post-mortem, PO=prospective observational, RC=retrospective cohort, SR=systematic review, *=hospital incidence, $^{\Delta}$ crude, n/a=not available

Table 2: Summary of Key 1° SAH Case Fatality Rate Literature

PUBLICATION AUTHOR (YEAR)	STUDY TYPE	SAMPLE SIZE (N)	CASE DEFINITION	NUMBER SAH CASES	30 DAY CASE FATALITY RATE (CFR) (95% CI)	STUDY LOCATION (N)
Zimmerman et al. (2013)[22]	RC	482,601 ICU admissions	Not clear - "surgical management of SAH"	Not clear	~9%[#] (77% decline from 1988)	-numerous US centres
Sandvei et al. (2011)[16]	PO	-126,729 patients at risk -2.1 million person years	Predominantly AD ICD codes; neurosurgical registry prior to 1986 and death registry	214	36% (29-42%)	Norway
Lovelock et al. (2010)[17]	PO	91,106 patients at risk	Imaging, CSF and/or PM	38	43%	UK
	SR/MA	(31 studies)	Not clear	5187	8-67%*	Asia (3), Australasia (3), Europe (19), N. America (4), S. America (2)
Feigin et al. (2009)[13]	SR/MA	82.2 million person years (18 studies)	Clinical presentation confirmed with imaging, CSF and/or PM (WHO)	n/a	31.6%(high-income countries) 43.9% (low/mid income countries)	Asia (1), Australasia (3), Europe (10), N. America (2), S. America (2)
Nieuwkamp et al. (2009)[21]	SR/MA	8739 patients (33 studies)	Various (predominantly WHO)	8739	8.3-66.7% (N. America – 32%)	Asia (7), Australasia (8), Europe (21), N. America (1), S. America (2)

Anderson et al. (2000)[19]	Prospective Case-control population based	4.92 million person years	Clinical presentation confirmed with imaging, CSF and/or PM, secondary review of AD ICD codes	436	39.0%	Australasia
Ingall et al. (2000)[14]	Cross-sectional comparison of 11 population-base studies	35.9 million personyears of observation	AD ICD codes, confirmed with imaging, CSF and/or PM	3368	23-62%	East Europe and Asia
Pobereskin et al. (2001)[26]	Retrospec-tive population based	n/a	AD ICD codes, cause of death, and neurosurgical registries	800	44% (40-49%)	UK
Ostbye et al. (1997)[2]	RC	Not stated (10 provinces)	AD ICD codes	10626	37.9% [¥](n/a)	Canada
Hop et al. (1997)[9]	SR/MA	2155 patients (21 studies)	Various (predominantly WHO)	2155	32-67%	Asia (2), Australasia (2), Europe (12), N. America (5)

SR=systematic review, MA=meta-analysis, PO=prospective observational, RC=retrospective cohort,crude CFR ranging from 14-180days, n/a=not available,[#]=hospital discharge, [¥]=using national death statistics

Further European retrospective observational studies, not included in the above reviews, demonstrate the variance in incidence and CFR. In the UK, incidence of SAH is reported to be 9.7/100,000 person-years (95% CI 7.5 to 12.6) (7.4 males and 11.9 female) with a 30 day case-fatality rate of 44% (95% CI 40 to 49).[26] The analysis of 2 large Norwegian cohorts demonstrated a similar incidence rate of 10.3/100,000 person-years (95% CI 9.0 to 11.8), but demonstrated an increase of 2.5% (95% CI 0 to 4) per 5-year period, with a corresponding 30 day CFR of 36% (95% CI 29 to 42).[16]

In summary, estimates of 1° SAH incidence vary significantly and range from 2 to >20 cases per 100,000 population. Similarly, case fatality rates reported in the literature also demonstrate huge variation from 8% to >65% of cases.

3.2 Canadian SAH Epidemiologic Data

Very little data on the incidence or case-fatalityrate of SAH in Canadahave been published. Ostbye and colleagues presented data on Canadian hospital admissions prior to 1992 and suggested that SAH occurred at a rate of 8.0 per 100,000 men and 11.2 per 100,000 women with an in-hospital case-fatality rate varying with age between 14 and 38%.[2]The overall case-fatality rate was approximately 20%, which the authors themselves noted to be up to 50% lower than reported mortality statistics.[2]

The reliability of this study could be limited. These datapredated the introduction of more advanced SAH treatment modalities, including a shift to neuro-interventional aneurysm stabilization with coiling[27], and are therefore very dated. In addition, this study relied solely on administrative data to identify their cohort using diagnostic codes that had not been validated. Given how uncommon 1° SAH is, and the potential inaccuracy of diagnostic codes, it is very possible (in fact, it is quite likely) that many people in thiscohort

did not actually have aSAH, or that some patients with true SAH failed to be identified by the diagnostic code.The cited accuracy of the diagnostic codes used to identify the patient cohort was based primarily on positive predictive values derived from a population with different disease prevalence than the current study. Since accuracy of positive predictive value is affected by disease prevalence[28] the reliability of the case ascertainment may be limited. This is discussed further below in more detail.

More recently Canadian data on all stroke subtypes has been reportedby Moore et al. for the province of Manitoba. However, case ascertainment relied on search criteria that have not been validated.[29]Using only diagnostic codes for case ascertainment, SAH accounted for approximately 1.7% of all stroke in Manitoba from 1995/6 to 2003/4. Other Canadian work on all stroke in the province of Quebec demonstrated anannual rate of 151 per 100,000 population from 1981 to 1989 for all stroke subtypes.[30,31]Case ascertainment was again based entirely on diagnostic codes abstracted from administrative data. Of the 79,482 stroke discharges in this time period, 4,296 were reportedly due to SAH.

3.3 Case Definition from Diagnostic Codes

The creation of ICD (International Classification of Diseases) coding was intended to facilitate the logging, storage and utilization of disease diagnoses by converting words to an organized alphanumeric code.[32] Now in its 10th edition, ICD-10-CA was rolled out across most of Canada in 2001, replacing ICD-9-CM. This changeover occurred in April 2002 at the OttawaHospital. In ICD-10-CA, a practical classification of disease was implemented that organized diagnoses into groups based on one of: epidemiology; anatomical site; or causation (e.g.: injury). Cerebrovascular diseases are represented by the codes I60 to I69. Primary subarachnoid hemorrhage specifically is represented by I60. It

has 10 "sub-codes" (I60.0 to I60.9), each representing a different anatomical site of origin of aneurysm rupture, with the exception of I60.8 which represents “other”, specifically meningeal or ruptured arteriovenous malformation.[33] Secondary causes of SAH, including trauma have different codes (e.g.: S06.6). Previous to this 10th iteration, ICD-9-CM codes for stroke were 430-438. Specifically, 430 represented 1° SAH.

The sole use of diagnostic codes to define patient cohorts in administrative database research is fraught with difficulty.[20,34,35]Many studiesusingadministrative diagnostic codes to identify patients with stroke have been published.[2,20,26,29,30,35–37]The performance characteristics of these codes have varied significantly with a positive predictive values (PPV) of identifying incident stroke cases ranging from 68%[36] to >80%[20]with varying specificity and sensitivity. For example, Leibson et al.[37] found diagnostic codes in hospitalization datasets for both incident and recurrent stroke had a positive predictive value of only 60%. Further, 23% of incident stroke cases were missed by the coding and 36% of those coded were either non-incident cases or non-events. Using the Northern Sweden arm of the MONICA stroke study registry as a gold standard, a review of diagnostic codes amongst fatal and nonfatal cases of stroke demonstrated a false positive rate of 10% and 32% and false negative rate of 17% and 6% respectively.[35]Canadian work on stroke diagnostic code accuracy has demonstrated that accuracy is strongest for intracerebral hemorrhage (ICD 431) and ischemic stroke (ICD 434 and 436).[30]

As poor as these statistics appear, they likely *over*-estimate the accuracy of single administrative database codes to identify rare events like 1° SAH. Given that the pre-test probability of having a given disease is low in rare diseases, the chances of having the disease after being identified by a specific diagnostic code (post-test probability) remains surprisingly low. This literature is a good example of questionable epidemiology using

administrative data. Indeed, a meta-analysis of stroke case ascertainment using diagnostic codes where the actual diagnosis was confirmed with either stroke registry or chart review found that the identification of stroke using of all stroke-related diagnostic codes significantly over-estimated disease incidence (reported as a pooled OR 1.70 (1.53-1.88, 95%CI), where 1.0 suggests complete agreement, >1.0 an overestimate and <1.0 an underestimate).[29]Using a more specific algorithm (i.e. codes for specific stroke types) significantly decreased the overestimation but also dropped the sensitivity to as low as 54%[38]. Using more specific diagnostic code algorithms in combination with other database sources (including pharmacy records and physician billing codes)to identify specific diseases (a process itself which has not been validated) have been studied but the sensitivity and specificity of this algorithm was not published.[29]

Few studies have conducted a detailed examination of the validity of using diagnostic codes to identify SAH specifically (see Table 3). Ellekjaer et al.[36]concluded from their study using discharge data of stroke patients that diagnostic codes should not be used to identify the subtypes of stroke, including SAH, because of incidence overestimation. In a review by Williams et al.[20], the PPV of diagnostic codes for SAH was found to vary from 64-100%, with the higher values coming from the smaller studies (all under 30 patients). In Canadian work, Mayo et al.[30] summarized 4 studies[39–42] that have examined the accuracy of diagnostic coding. Although the probability that a patient with the diagnostic code for 1° SAH (ICD-9-CM 430) actually having the disease (based on clinical review) ranged from 33-100%, this patient population represented <1 to 7% of the patients examined.Further, the potential for missed cases was not accounted for and thus neither the specificity nor sensitivity can be calculated and is reported in only one study[43].

Table 3: Summary of Literature Describing the Accuracy of Diagnostic Codes for SAH

STUDY	TOTAL SAMPLE SIZE	NUMBER WITH ICD CODE(S) FOR SAH	PROPORTION OF THOSE WITH CODE TRULY HAVING SAH (PPV)	DIAGNOSTIC CODE ACCURACY, % (95% CI)
Liu L et al. (1993)[39]	683	14	92.9%	Not examined
Phillips SJ et al. (1993)[40]	301	3	33%	Not examined
Mayo N et al (1993)[41]	96	1	100%	Not examined
Mayo N et al (1993)[42]	3197	247	94.7%	Not examined
Leibson CL et al. (1994)[37]	364	11	100%	Not examined
Broderick J et al. (1998)[44]	Not stated	14	64%	Not examined
Rosamond WD et el. (1999)[45]	1185	22	86%	Not examined
Roumie CL et al. (1998)	231	2	100%	Not examined
Tirschwell et al. (2002)[43]	206	58	86%	Sn=98 (90-100) Sp=92 (84-96)

PPV=positive predictive value, CI=confidence interval, Sn=sensitivity, Sp=Specificity

3.4 Gold Standard Diagnosis of SAH

The diagnosis of SAH can be difficult. Therefore, numerous publications including the most recent edition of the American Heart Association/American Stroke Association Guidelines on the management of aneurysmalSAH,have suggested that its diagnostic work-up start with an unenhanced cranial computed-tomography (CT). Patients in whom this test does not indicate SAH should undergo lumbar puncture.[4,8,46–48]After establishing the presence of SAH, physicians are recommended to identify the source of bleedingwith CT-angiography (CTA) and/or digital subtraction angiography (DSA).[8]Thisdiagnostic approach

has also been utilized in the inclusion criteria of major randomized clinical trials involving 1° SAH patients.[10,49,50]

Implementing the above diagnostic criteria to identify a cohort of SAH patients retrospectively is not easily accomplished. Most centres do not have administrative datasets that include the primary data, or the capability to screen it, to make such a diagnosis. As a result, diagnostic codes are often relied upon to identify SAH. In the population-based retrospective observational study byPobereskin (2001)[26],the author applied these diagnostic criteria to patientsthat had beenscreen positive to a specific case-ascertainment algorithm.However, case-ascertainment still relied predominantly on diagnostic codessince patients were identified by one of: discharge diagnosis code (ICD-9-CM); national statistic death code; identification in an imaging database of angiograms; or procedure codes from an operating room database. Following this process, the authors found that 21 of 671 patients had beenmisidentified as having suffered a SAH (false positives). Further, they concluded that the sole use of primary diagnostic codes would have resulted in a false negative rate of 5%. It remains unclear from this study if other possible cases could have been missed given that diagnostic criteria were only applied to a specific screened (case-ascertainment algorithm) population that itself was not based on the diagnostic criteria.

In summary, few accurate data exist that describe the incidence and case-fatality rates of 1° SAH pertinent to our patient population. The scant existing Canadian data is not only outdated but relies on case-defining measures that are comprised entirely of diagnostic codes with highly questionable accuracy.A systematic, comprehensive, validated, and reproducible method to accurately identify patients with SAH in an administrative database is needed to retrospectively study this disease on a population-based level. Only once these

patients are accurately and completely identified can we rigorously study disease epidemiology, its natural history, management review and quality assurance and truly understand the burden of disease.

4 Objectives

This thesis will:

1. Define a "Gold Standard" SAH cohort by deriving and validating a search algorithm that uses diagnostic imaging reports, biochemical analyses of cerebrospinal fluid (CSF), and autopsy reports to attempt to identify every 1° SAH presenting to the Ottawa Hospital (TOH) between 2002 and 2011. This will require:
    a. The derivation and validation of a series of text-search algorithms to identify 1°SAH from radiology reports, autopsy reports, and biochemical analyses of cerebral spinal fluid (CSF) in the Ottawa Hospital Data Warehouse (TOHDW).
2. Calculate the incidence and hospital-based case-fatality rate of 1° SAH presenting to TOH between 2002 and 2011.
3. Derive and validate a multivariate logistic regression model to accurately predict the probability that a patient truly had a1° SAH based on variables available in the DAD.

Table 4 provides a general overview of thesis objective and relevant methods and results section.

Table 4: Thesis Objective and Corresponding Methods and Results Section

THESIS OBJECTIVE	METHODS SECTION	RESULTS SECTION
Derivation and validation of a1° SAH search algorithm	5.5	6.1-6.4
Identify cohort of all 1° SAH patients (9 year cohort)	5.6	6.5
Describe hospital incidence and case-fatality rates	5.8-5.9	6.6
Derive and validate a prediction model that predicts the probability that any given inpatient hospital encounter is the result of a 1° SAH	5.10	6.7

4.1 Hypotheses

I hypothesize that:

1. a text search algorithm can be developed that will, identify the vast majority of patients presenting to The Ottawa Hospital with 1° SAH. In order to ensure the fewest possible cases are missed, a higher sensitivity is expected, at the cost of specificity.
2. the hospital incidence of 1° SAH has remained stable over the last 10 years but case-fatality has decreased.
3. a multivariate logistic regression model can use variables in the Discharge Abstract Database to determine (with >90% accuracy) the probability that they had a 1° SAH.

5 Methods

5.1 Study Goals

This retrospective cohort study was completed with 2 over-riding study goals to meet the objectives of my thesis (see Section 4):

1. To accurately identify a patient cohort of 1° SAH patients presenting to the Ottawa Hospital from July 1, 2002 to June 30, 2011 (9 years) using an enriched administrative database (The Ottawa Hospital Data Warehouse), and
2. To develop a modelthat will predict the probability that someone truly has a 1° SAH using data from a health administrative database (the Discharge Abstract Database).

All programming and statistical analyses were completed on SAS 9.2 (SAS Institute Inc, North Carolina, USA; License: OHRI).

5.2 Study Setting

This is a retrospective cohort study using administrative data from a single centre - The Ottawa Hospital (TOH) - from July 1, 2002 to June 30, 2011. TOH is a tertiary care, academic hospital with approximately 1100 inpatient beds divided over two sites (a third site exists as an outpatient hospital).

5.3 The OttawaHospital Data Warehouse

The Ottawa Hospital Data Warehouse (TOHDW) is a collection of health datasets containing clinical and administrative data for all patient encounters (including in-patient and out-patient) at TOH. Encounter data is available from January 1996 for one of the inpatient sites and from August 1999 for the other. In addition to the Discharge Abstract Database (DAD) (which contains diagnostic, procedural, demographic and administrative information), TOHDW contains all laboratory, pharmacy and radiology information

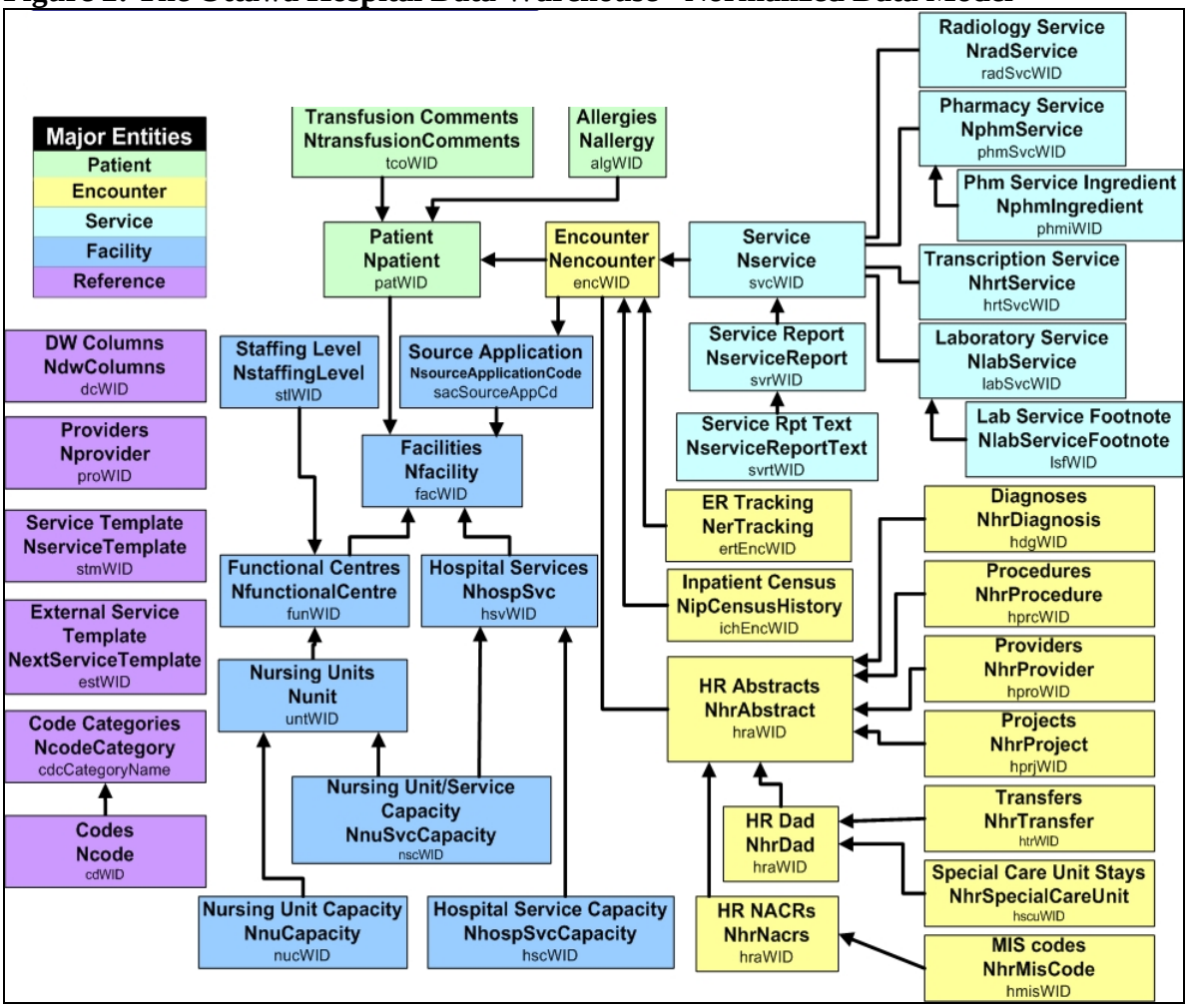
(available from both sites from July 1, 2002) as well as the patient registration system.[51] These data are all housed on a single, cross-linked data repository (see Figure 2).

Although the DAD is collected for all hospitalizations in Canada[52], the additional data in the TOHDW make analyses of primary data possible in a retrospective fashion. This has been done successfully a number of times in the literature.[51–53]Hence, not only are diagnostic codes available to identity potential patients with SAH, all radiological diagnostic and procedural imaging reports, autopsy reports and primary CSF laboratory results for each individual patient are also available for analysis.

Although TOHDW contains personal health information and identifiers (PHIs), these are encrypted and all patients have been assigned an anonymized unique identifier that is consistent between separate encounters. TOHDW is organized in a series of tables (see Figure 2), each linked using different unique identifiers. At the root of the database is the individual patient (information about which is housed within the "Patient" table). Each hospital encounter for a given patient has a unique encounter identifier (information about which is housed in the "Encounter" table) which is linkable to the "Patient" table using the unique patient identifier. A hospital admission at one campus of the hospital that involved a transfer to another campus during the same admission would result in 2 unique encounters that can be linked together to form a single hospitalization at TOH. Similarly, Emergency Room (ER) visits generate a unique encounter that is separate from the unique encounter that would ensue should the visit result in a hospital admission. For each encounter, data may be generated from the Emergency Room (ER), different hospital services (eg Radiology, Pharmacy, Laboratory etc), and the Discharge Abstract, each housed within their own tables with linkable identifiers. As such, there is much versatility in interrogating TOHDW since one may generate numerous datasets each with a different

unit of measure. For example, one may create a dataset where the unit of measure is a specific service (eg: CSF analysis), all encounters, or all patients. These may be analyzed separately or be subsequently linked.

Figure 2: The Ottawa Hospital Data Warehouse - Normalized Data Model



(ref: file:///T:/DW%20Documents/DWdocumentation/DWDocIndex.html)

5.4 Sample Size

A minimum sample of 300 patients with SAH will be required in order to ensure our search strategy has a sensitivity of 95% with a margin of error no greater than 2.5% (seeAppendix 1). Assuming that approximately 116 SAH cases are admittedeach year to TOH (approximation based on diagnostic codes and total yearly admissions, not published)

a minimum 3 years of hospital admissions should be included. We proposed using all9 years of admissions (the extent of data in TOHDW)which will ensure this requirement.

Assuming a Poisson distribution, a minimum of 1000 screen-negative cases will need to be reviewed to ensure that the true false negative proportion of our search strategy has an upper 95% confidence limit of ≤ 1% (see Appendix 1).

5.5 Derivation and Validation ofa 1° SAHSearch Algorithm

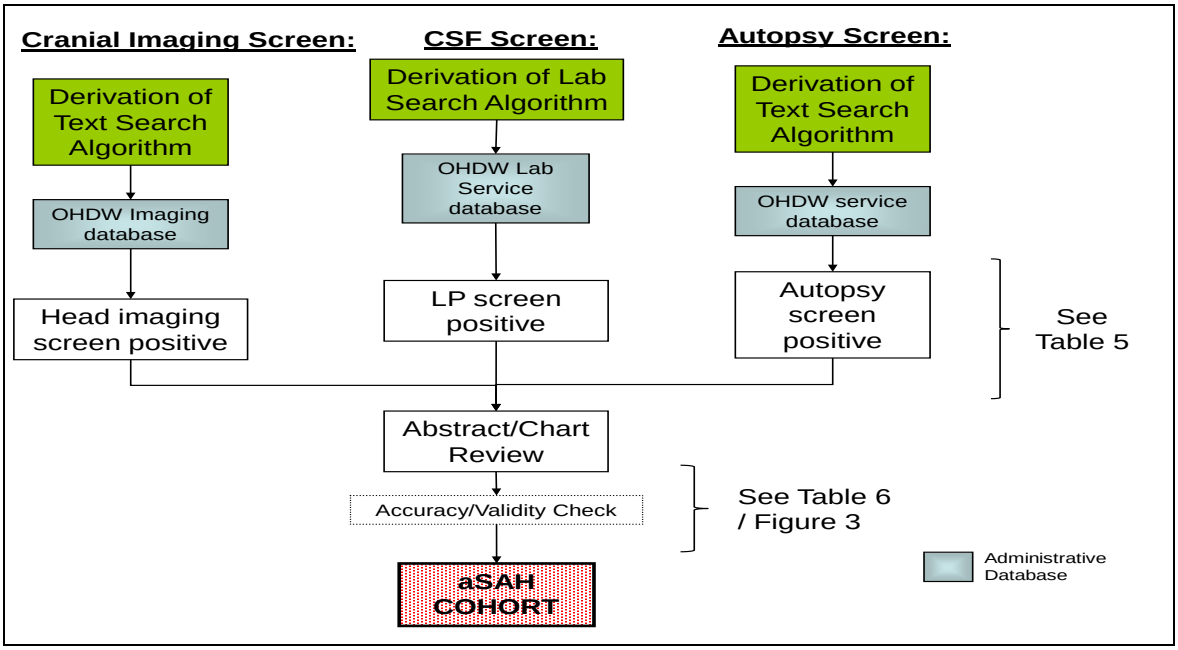
To identify all incident cases of 1° SAHamongst patients aged 18 years or older who present to TOH, three case-defining methods (seeFigure 3) were used:

a) Presence of SAH or ruptured aneurysm on head imaging (computed tomography (CT) head, CT angiogram, angiogram).

b) Presence of xanthochromia or red blood cells (>5x10^6/l on final sample) on lumbar puncture with evidence of primary source.

c) Autopsy report from postmortem examination showing 1° SAH.

Figure 3: Cohort Derivation



This approach was taken since it would be essentially impossible to definitely diagnose a case of 1° SAH without one of these three investigations being positive. This approach is supported by the most recent American Heart Association guidelines[7,8] and was used in two recent population-based British studies (Oxford Vascular Study (OXVASC) and Oxford Community Stroke Project (OCSP)[17]) as well as a Canadian Emergency Department study[54]. Primary subarachnoid hemorrhage was defined as blood in the subarachnoid space (as determined by one of the three methods in Figure 3) that results from: an aneurysm rupture; an arteriovenous malformation (AVM) rupture; a perimesencephalic bleed (a nonaneurysmal 1° SAH with a distinct peri-brainstem hemorrhage pattern with no discernable cause on imaging[55]); or an idiopathic cause. In particular, the diagnosis of 1° SAH required the exclusion of secondary causes including trauma and neoplasm. A clinically important subset of this population are patients with aneurysmal SAH (aSAH) or arteriovenous malformation SAH (avmSAH) as they import unique challenges that often include interventional or surgical management. These have been identified collectively in this study as aSAH/avmSAH patients. The majority of the literature that describes 1° SAH is in fact describing only this population subset, aSAH and avmSAH. It is thus the population subset of focus for this thesis although we also report on 1° SAH.

Table 5: Criteria Used to Determine Positive Tests in Screens

SCREEN	TEST	FINDING	DIAGNOSIS
Imaging Screen	CT* head	-blood in subarachnoid space	SAH
	CTA head	-blood in subarachnoid space	SAH
		-presence of aneurysm or AVM	Aneurysm or AVM
	Angiogram	-presence of aneurysm or AVM	Aneurysm or AVM
CSF** Screen	RBCs***	->$5x10^6$/l	SAH
	Xanthochromia	-present	SAH
Autopsy Screen		-blood in subarachnoid space	SAH
		-presence of aneurysm or AVM	Aneurysm or AVM

*Computerized Tomography **cerebrospinal fluid ***red blood cells

Table 5 demonstrates the criteria used to determine a positive test within each of the three screens illustrated in Figure 3. Effectively, the diagnosis of SAH is possible with the specified criteria in the imaging (specifically CT or CTA), cerebrospinal fluid (RBC or xanthochromia) or autopsy screens. The presence of an aneurysm or arteriovenous malformation was identified on either imaging (specifically CTA or angiogram) or autopsy.

These screening criteria were applied to data residing in TOHDW for all adults hospitalized at TOH between July 1, 2002 and June 30, 2011. The criteria listed in Table 5 were used to classify each patient with or without 1° SAH along with their subclassification (seeTable 6).

Figure4 presents these criteria graphically.

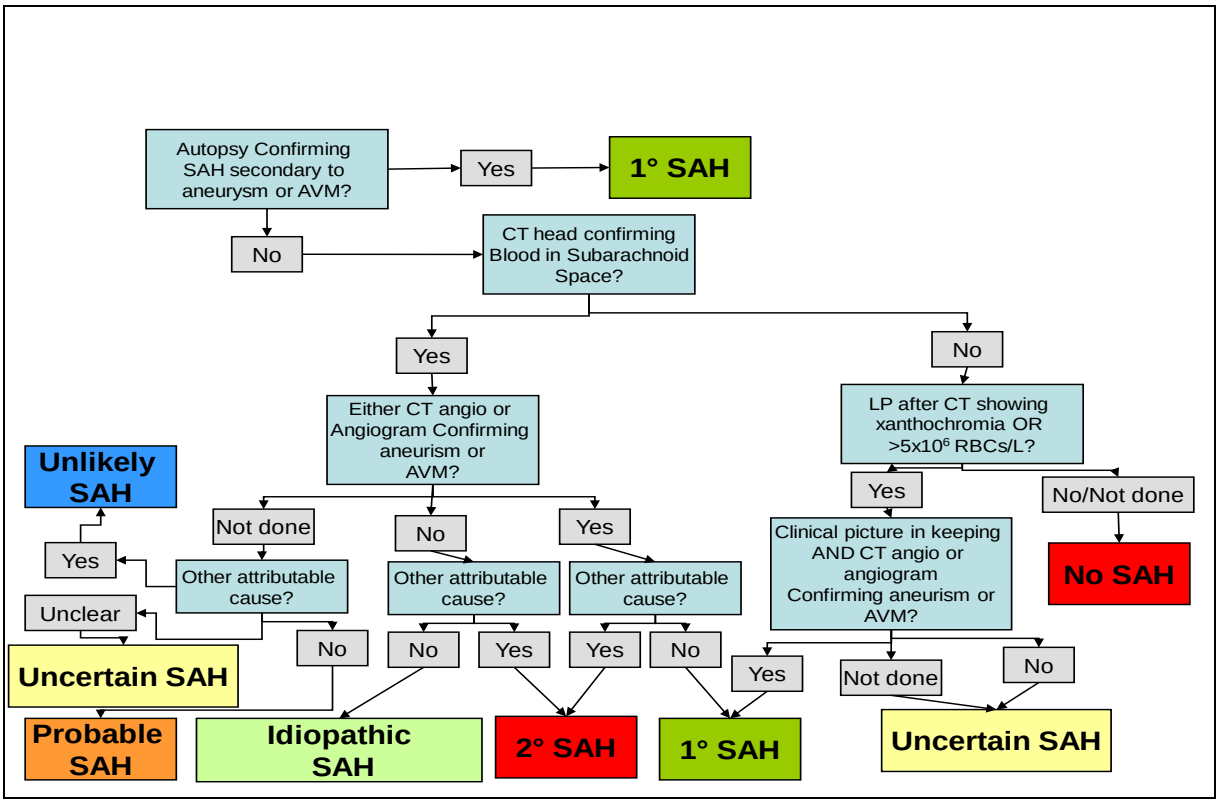
Table 6: Case Definition and Confirmation

DIAGNOSIS		SCREEN RESULT	CONFIRMATION
1° SAH		(CT+or LP+) and (CTA or angio +)	-chart review confirming 1° SAH
	OR	(CT+or LP+) and (postmortem showing aneurysm/AVM)	-chart and post-mortem report review confirming diagnosis
	OR	postmortem showing SAH+aneurysm/AVM	-post-mortem report and chart review
Idiopathic 1° SAH		(CT+) and (no aneurysm/AVM on CTA or angio) AND (no other attributable cause)	-chart review
Probable 1° SAH		(CT+or LP+) and (no aneurysm on CTA) and (no angio or PM done) and (no other attributable cause)	-chart review
Uncertain 1° SAH		(CT+or LP+) and (no CTA/angio/PM done) and (no other attributable cause)	-chart review

(CT=computed tomography, CTA=computed tomography angiography, LP=lumbar puncture, angio=angiography, SAH=subarachnoid hemorrhage, AVM= arteriovenous malformation, PM=post-mortem examination, +=positive)

To identify every 1° SAH at TOH during the study period, an overall search algorithm needed to be derived (see section 5.5.1) and then validated (see section 5.5.4). Given that the algorithm contains 3 sub-screens (one each for imaging, CSF analysis, and autopsy - see Figure 3), each individual sub-screen required separate derivation (see sections5.5.1.3to 5.5.1.5) and subsequent validation (see sections5.5.4.1 to 5.5.4.3) prior to being pooled together to form the overall search algorithmused to create the study cohort (see Figure 5and Figure 6). The description of the derivation and validation of this overall search algorithmfollows that of the individual sub-screens and is covered in sections 5.5.1.6 and 5.5.4.4 respectively.

Figure 4: Gold Standard 1° SAH Diagnostic Criteria



5.5.1 Derivation of the Overall Search Algorithm

5.5.1.1 Sources of Patient Data for the Search Algorithm Derivation

The search algorithm was derived using three sources of patient data: 1) a cohort of "known 1° SAH"patients treated at TOHbetween January 1 and December 31, 2007that were abstracted from the Interventional Neuroradiology Database; 2) a cohort of known 1° SAH patients treated between January 1 to December 31, 2007 that were abstracted from the Neurosurgical Database; and 3) a random sample of patients presenting to TOH over the same period in 2007 abstracted from TOHDW. The year 2007 was selected since it was believed to be the most complete for the first two databases above and was expected to yield the greatest number of patients with subarachnoid hemorrhage (approximately 100).The Interventional Neuroradiology Database is a retrospectively collected dataset of prospectively identified patients since 2003. It reportedly contains all patients that have undergone a neuro-interventional radiologic procedure for either cerebral aneurysm or arteriovenous malformation. The database is managed by the Neuro-Interventional Radiology group at TOH for the purposes of quality assurance. Similarly, the Neurosurgical Database is a prospectively managed by the Neurosurigical service at TOH and is a dataset reportedly containing all patients who have undergone surgical clipping of either an aneurysm or arteriovenous malformation. It serves as a log of clipping procedures for quality assurance.Permission to use these data was acquired from the custodians of both of these data collections.

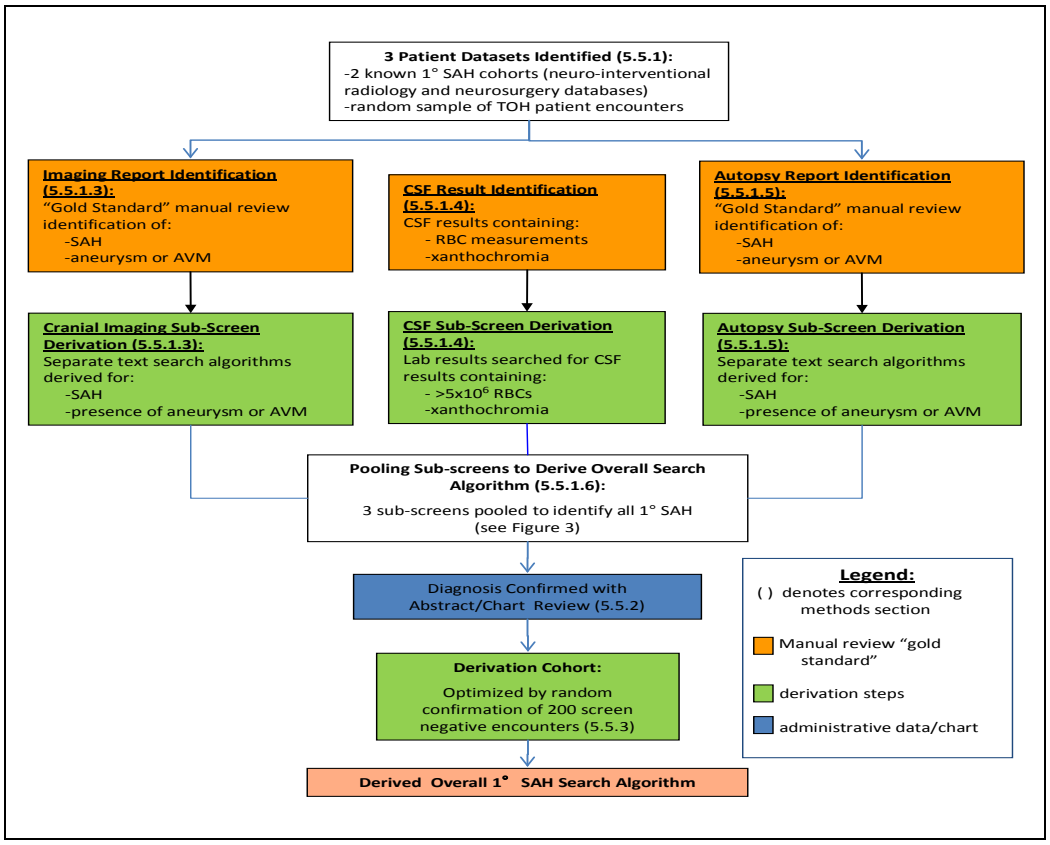
5.5.1.2 Rationale for Three Patient Data Sources

The first two datasources (Interventional Neuroradiology Database and Neurosurgical Database) - the "Known 1° SAH group" - were utilized to extract a large sample of patients with subarachnoid hemorrhage. Their test results were abstracted to

help derive the search algorithm. The third datasource (random sample of TOH admissions from TOHDW) - the "Random Sample Control Group" - was utilized for two purposes: 1) to have a random sample of patients without 1° SAH to contrast test findings and improve screen specificity; and 2) to potentially include patients with 1° SAH (not contained within the other 2 datasets) to explore potential differences in their test results and improve screen sensitivity. Given that there exists the potential for overlap of patients amongst the datasets (e.g. the same patient with 1° SAH appearing in both the Interventional Neuroradiology Dataset and the Neurosurgical Dataset), de-duplication was undertaken and analysis of the results were made at the unique patient level.

The derivation of the overall algorithm and its sub-screen components are described here (see Figure 5).

Figure 5: Derivation of Sub-Screens and Overall Search Algorithm



5.5.1.3 Cranial Imaging Screen Derivation

The presence of SAH or ruptured aneurysmon head imaging was determined by developing a text-search screening tool. This sub-screen forms part of the overall search algorithm for 1° SAH, and itself requires derivation and validation. The following steps weretaken to derive this screen:

a) Known 1° SAH Group: We retrieved the imaging text reports for all cranial CTs, CTAs and angiograms for the cohort of **known**1°SAH treated at TOH(formed by abstracting patients admitted in 2007 from the Interventional Neuroradiology Database and the Neurosurgical Database at TOH).

b) Random Sample Control Group: Using the OHDW database, it was first determined which identifiers (known as stmWIDs) are used to capture reports of all cranial CTs, CTAs and cerebral angiograms. This search was completed by manually compiling a list of all possible identifierscontained within TOHDW (see Appendix 2). A random sample of cranial imaging reports (including CT, CTA and angiogram) from inpatient or emergency room encounters from2007 that had one of these stmWIDsin the TOHDW were manually reviewed by the author.We randomly sampled 5000 reports, which is approximatelya 10% sample of all available reports.

c) "Gold Standard" SAH, Aneurysm and Arteriovenous MalformationIdentification Through Manual Review of Imaging Reports:Imaging reports from a) and b) were exported to a Microsoft Access (version 2003) database and manually reviewed to identify all cases of SAH by SE. Criteria that were used to identify SAH are described in Table 5. In addition, all processes that could identify a cause of SAH (eg: aneurysm,

arteriovenous malformation, traumatic history, malignancy)were also abstracted from the report. For the derivation of the Cranial Imaging Screen, this manual review served as the confirmation or "gold standard" against which the text search screening tool was tested.

d) Reports with SAH were reviewed in detail to identify specific termsusedby radiologists to specify or characterize the presence of 1° SAH (eg: ‘subarachnoid hemorrhage’, ‘aneurysm’, 'blood ' and 'Sylvian Fissure').

e) Imaging Report Text-search Screens Development:Using the terms identified in step (d), a text-search screening tool was created to identify 1°SAH patients from diagnostic imaging reports. To do so, assorted variations of term combinations (eg: “blood”, “hemorrhage”, “clot”) with modifiers (eg: ‘no’, ‘absent’, ‘without’) were tested, added, or removed, to optimize the sensitivity and specificity of the screening tool against the “Gold Standard” manual review performed in (c).Based on comparison with a manual review of the reports, a text-search screen was first derived to identify the presence of blood in the subarachnoid space, and then an additional screen was derived to identify the presence of aneurysm or arteriovenous malformation. Separate screens were necessary to determine SAH (presence of blood in the subarachnoid space) and potential causes (eg: aneurysm or AVM) since different information is acquired from different imaging modalities (eg: blood in the subarachnoid space is not generally discernable on angiogram or aneurysm is only exceptionally visible on a non-contrast CT scan). This was completed as follows:

i) using SAS 9.2 (SAS Institute Inc, North Carolina, USA; License: OHRI), an analytical dataset was created in which each radiology report represented a separate entry (row);

ii) Based on the manual review (Gold Standard) in step (c), columns were created for the different findings (eg: SAH, aneurysm, AVM);

iii) Each report was flagged as identifying or not ('1' or '0' respectively) each finding in (ii);

iv) The text-search screening tool was applied to the reports and populated new columns as identifying or not the same variables as in (ii) but this time by the screening tool;

v) the screen's operating characteristics were then calculated comparing the findings of the screening tool (iv) with those of the gold standard manual review (ii);

f) Testing the Text-search Screens:The text-search screens wereapplied against the "Known 1° SAH" and "Random Sample Control" cohorts(a and b above) and the operating characteristics maximized (first sensitivity; then specificity).The objective was toderive a text-search screen with high sensitivity (>95%) and then adjust it to optimize specificity without compromising sensitivity. Once the operating characteristics were sufficiently optimized, the derived screen was deemed complete.

g) A “Screen-positive”Report Resulted in a “Screen-positive”Encounter:Each of the image reports screened above were merged and grouped for each unique hospital encounter, such that any given hospital encounter may have any number of images associated with it. A positive screen imageon any imaging

study done for a particular encounter flagged that encounter as being image screen positive.

5.5.1.4 Derivation of CSF Screening Method

Red blood cells (RBCs) are normally absent from cerebrospinal fluid (CSF). Xanthochromia describes the discoloration of CSF caused by the pigments from breakdown products of RBCs in the CSF.[56]Its presence may take as long as 12 hours afterthe first presence of RBCs. Prior to this time, only red blood cells may be detected after an SAH.[56]The following steps were used to screenthe results of all CSF samples for indications of SAH as defined by the presence of xanthochromia or red blood cells:

a) CSF Results Identification:Using TOHDW, it was first determined which identifiers (stmWIDs) are used to capture CSF results, specifically presence of xanthochromia or RBC counts. This was completed by manually compiling a list of all possible identifiers contained within TOHDW (see Appendix 3). The results of all such identifiers from 2007 within TOHDW were exported and manually reviewed.

b) Using the identifiers identified in (a), all results from CSF samples taken fromJanuary1, 2007and December 31, 2007 were extracted using the lab services table within TOHDW. Only samples taken from patients captured by either an emergency room or inpatient encounter were included.

c) From the sample obtained in (b), it was determined how laboratory reports indicated thepresence of SAH withxanthochromia or red blood cells $>5x10^6$/l in the final CSF sample.

In this setting, lumbar puncture is a diagnostic and not therapeutic procedure. It is typically completed following a negative CT head scan (the mainstay of diagnosis of SAH, but its

sensitivity falls after 72 hours) as a sensitive measure to exclude SAH (seeFigure 4). Since CSF should indicate xanthochromia approximately 12 hours after the event,[54] only CSF samples drawn within 24 hours of the start of the encounter were considered. This was done to ensure that we considered all CSF samples drawn early in the diagnostic workup of a patient presenting with possible SAH, but not include the multiple samples drawn throughout the therapeutic course in hospital of this and other diagnoses.

Two separate screens were constructed: one to identify all CSF sampleswith $>5x10^6$/l RBCs[54] (to identify RBCs in CSF that have not had sufficient time to break down to form the byproducts that produce a positive result for xanthochromia) and another to identify all CSF samples positive for xanthochromia:

i. RBC Screen:Since CSF sample tube collection orders were not always recorded (nor were samples always processed in order), results were organized by encounter and date drawn. To improve the sensitivity of the search strategy, a downward trend of RBC count occurring was not considered (a downward trend from the first tube of CSF sample drawn to the last following lumbar puncture is suggestive of a traumatic tap). If the lowest RBC count performed on a given day exceeded $5x10^6$/l RBCs, the screen was considered positive (ie: even if there had been a drop of RBCs of a several hundred-fold from the first to the last tube collected).

ii. Xanthochromia Screen: all CSF samples tested for xanthochromia within 24 hours of the start of the encounter were examined. From (c) above, it was determined that xanthochromia could be reported in any of following test headers (ie: stmWIDs): 'Xanthochromia', 'CSF Appearance' and/or 'CSF color'. An algorithm was created to identify both possibilities of

positive tests using 'Xanthochromia', either as an outcome or as a descriptor (eg: "Xanthochromia: present", "xanthochromic"). Any such sample was identified as screen positive.

The 2 screens from section (i) and (ii) above were merged and grouped for each unique hospital encounter. Any positive screen flagged the encounter as being CSF screen positive. Unlike the sub-screen for radiology reports in which a manual review served as a "Gold Standard", laboratory result values of CSF analysis are directly exported to TOHDW and as such have no comparator to serve as a "Gold Standard".

5.5.1.5 Derivation ofAutopsy Screen

Post-mortem examination yielding a diagnosis of 1° SAH in this study was considered to be a definitive diagnosis. This was done to ensure that any patient with 1° SAH who died prior to undergoing or completing the usual diagnostic workup was included in our cohort. The following steps were taken to derive a screen for autopsy reports indicating SAH:

a) Autopsy Report Identification:Using TOHDW database, it was first determined which identifiers (stmWIDs) are used to capture the reports of all post-mortem examinations, both limited and complete. This was completed by manually compiling a list of all possible identifiers contained within the warehouse (see Appendix 4). Since the number of autopsies performed per year are relatively small (~100), and the diagnosis of SAH on post-mortem is definitive, all autopsies from 2007 within TOHDW were extracted and reviewed to increase the likelihood of capturing patients having suffered from 1° SAH.

b) "Gold Standard" SAH, Aneurysm and Arteriovenous MalformationIdentification Through Manual Review of Autopsy Reports:The 2007 reportswere manually screenedto identify any with thepresence of SAH or cerebral aneurysm or arteriovenous malformation. This served as the 'autopsy report Gold Standard'.

c) Autopsy Report Text-search screens Development:From (b), a text-search screen was created to identify1° SAH, optimizing first sensitivity and then specificity. High yield term - such as the use of the words "subarachnoid" and "hemorrhage" (or "haemorrhage" or "blood") or "cerebral" and "aneurysm" (or "aneurism"), etc. – were used to populate the screen.

d) The text-search screen was then applied to the cohort and any positive match was labeled 'Autopsy Screen Positive'.

5.5.1.6 Pooling Screens to Form an Overall 1°SAH Search Algorithm

Thethree TOHDW-based "sub-screens" from Sections 5.5.1.3to 5.5.1.5 were usedto identify all potential cases of 1° SAH at TOH with an overall search algorithm (Figure 5). This was completed as follows (seeFigure 4):

a) Any screen-positive encounter based on the autopsy was classified as'1° SAH'.

b) Datasets with the results of the radiology report screens for SAH and aneurysm/arteriovenous malformation were merged with the results of the screens for CSF RBCs and xanthochromia (see sections 5.5.1.3 and 5.5.1.4 respectively). Each observation in the merged dataset represented a patient encounter where each was labeled either positive or negative for the CSF RBC

screen, the CSF xanthochromia screen, or the radiology report screens.This overall screen algorithm then identified all encounters with:

i. SAH on radiology report with aneurysm/arteriovenous malformationalso reported in the radiology report text;

ii. No SAH on radiology report but CSF screen positive (with either RBC or xanthochromia) with aneurysm/arteriovenous malformationseen on cerebral imaging

Patients meeting criteria a) or b) (i) or (ii) were classified as '1° SAH'.

c) Encounters that were screen positive for SAH but had no aneurysm/arteriovenous malformation on imaging screen were labeled “SAH”. Patients who were screen negative for SAH but positive for aneurysm/arteriovenous malformation were classified with an “incidental aneurysm/arteriovenous malformation”.

5.5.2 Manual Chart Review Verification 1° SAH Status

From the criteria presented is section 5.5.1.6, patients could be classified with a '1° SAH', 'SAH', or 'Incidental Aneurysm/arteriovenous malformation'. The discharge summary and medical record of allpatients thus identified were reviewed to determine the true 1° SAH status (see Table 6). Gold Standard criteria for1° SAH was met if:

a) A diagnosis of 1° SAH was specified by the dictating physician as one of the patient's diagnoses (most responsible for admission or occurring as part of the admission) or cause of death,orthe report cited therapeutic interventions most in keeping with 1°SAH (eg: intravascular coiling or surgical clipping), or

b) Any criteria from5.5.2(a)were found in the progress notes of the patient's medical record (when a dictated report was not present in the patient's chart)

Patients meeting criteria in (a) or (b) were classified as truly having had a 1°SAH and comprised our "Gold Standard" cohort. People not meeting these criteria were classified as NOT having 1° SAH.

5.5.3 Optimizing the 1° SAH Search Algorithm

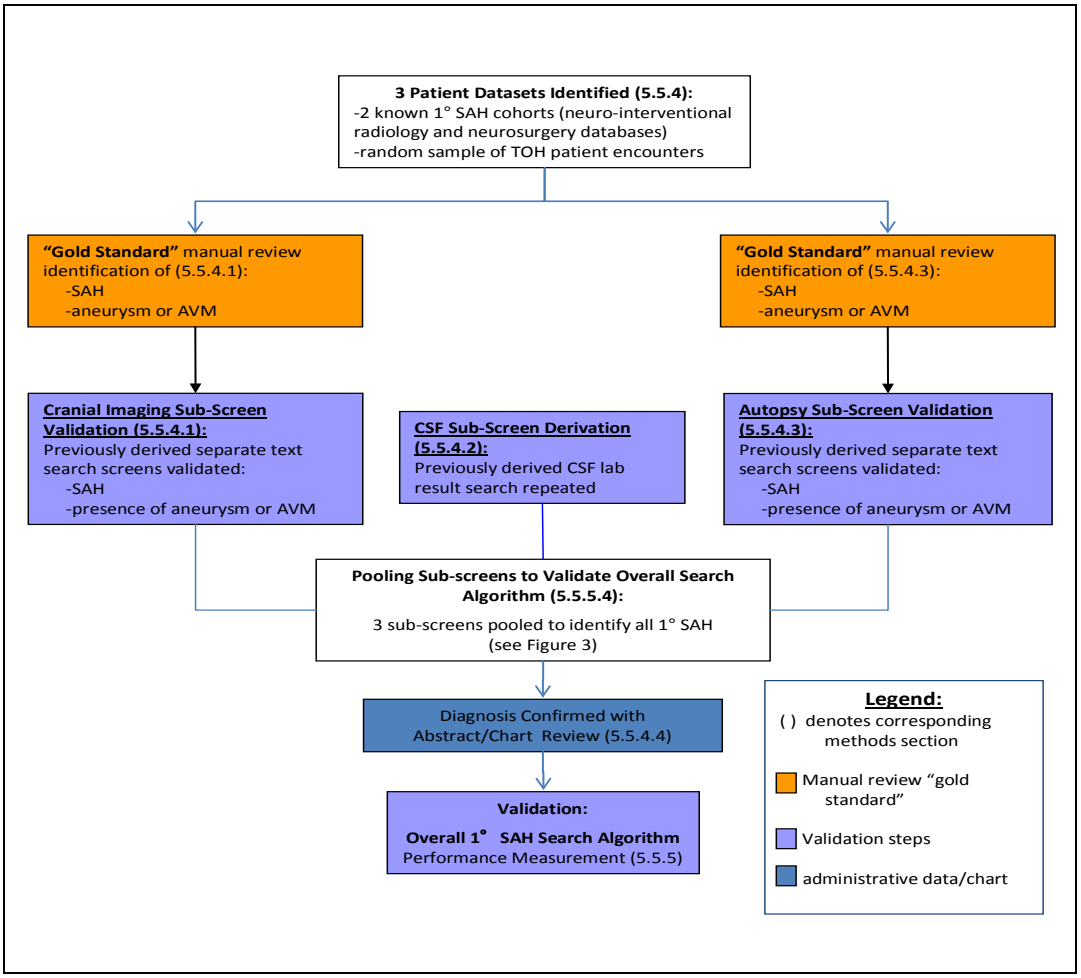
As discussed in Section 5.5.2, a manual review of each discharge summary or medical record of 'screen-positive' cases from the search above was undertaken to determine the 1° SAH status of all the 'screen-positive' cases. To attain a more complete assessment of the search algorithm (i.e. to measure false negative rate, sensitivity, specificity, and likelihood ratios), a random sample of approximately 200 patients who did***not*** meet screen-positive criteria from Section5.5.1.3(cranial imaging),5.5.1.4(CSF), or 5.5.1.5(autopsy) were selected. We selected 200 because our sample size calculation suggested needing 1000 screen-negatives for the 9 year cohort to ensure an upper 95% confidence limit for false negative rate of 1%. Given that this sample represented one of the 9 years of data to be collected, we calculated needing 111 (1000÷9) screen negatives, and rounded it up to the next even hundred. Any false-negative patient encounter was identified and examined to: determine the mechanism(s) by which the screening tool failed to identify them; and determine if the screening tool could be adapted to capture such patients.

From these data, 2x2 contingency tables were constructed and the operator performing characteristics (sensitivity; specificity; +likelihood ratios; and - likelihood ratios) of our search strategy on the derivation cohort were calculated.

5.5.4 Validation of Search Algorithm

For the validation of the overall 1° SAH Search Algorithm and its sub-screen components, we used datasets similar to the ones described in the derivation steps. However,for the validation, only encounters at TOH from January 1 to December 31, 2008 were used because it was believed to be the most complete years for the Interventional Neuroradiology Database and the Neurosurgical Database(to yield the most patients known to have 1° SAH patients). The validity of the algorithm was first checked by evaluating its sub-screen componentsin addition to the combined overall search algorithm (Figure 6). The datasets used and the validation of the sub-screens,as well as the overall 1° SAH Search Algorithm, are described below.

Figure 6: Validation of Sub-screens and Overall Search Algorithm



5.5.4.1 Cranial Imaging Screen Validation

As in the derivation steps, the validation of the cranial imaging screen was completed on imaging reports froma "known 1° SAH group" (identified from the Interventional Neuroradiology Database and the Neurosurgery Database) as well as a "Random Sample Control Group" comprised of a random sample of approximately 4000 text reports selected from all emergency room and inpatient encounters from 2008 (as previously described in section 5.5.1.3(b)). These reports were first manually reviewed and tagged for the presence or absence of SAH andaneurysm/arteriovenous malformation. This served as a gold standard reference against which the imaging text-search screen derived in 5.5.1.3 was tested and operational characteristics calculated.

5.5.4.2 CSF Screening Method for Validation of Overall Search Algorithm

The method of identifying CSF screen positives described in section 5.5.1.4 was repeated using results from all available samples drawn in 2008. The RBC screen and the xanthochromia screen were merged and grouped for each individual hospital encounter. Any positive screen flagged the encounter as being CSF screen positive. This was then used in the validation of the overall search algorithm in section 5.5.4.4.

5.5.4.3 Validation of Autopsy Report Case-defining Method

The autopsy report screening method was validated using all reports generated in 2008 at TOH. A manual review of all of the reports was first undertaken to identify any person with SAH, aneurysm, orarteriovenous malformation. This manual review and identification served as the gold standard against which the text-search screen was validated (as described in section 5.5.1.5).

5.5.4.4 Summary of the Validation of Overall 1°SAH Search Algorithm

Following the validation of the individualthreesub-screening methods, the sub-screens werepooled asdefinedpreviously (following optimization of the derived 1° SAH search algorithm in sections5.5.1.6and 5.5.3 respectively). The search algorithm classified patients as either having or not having 1° SAH, SAH or incidental aneurysm/arteriovenous malformation.Actual 1° SAH status was reviewed manually with the discharge summary or with manual chart review as in section 5.5.2. Primary SAH patients were identified using the same diagnostic criteria (seeFigure 4). Once again, in order to increase the likelihood of identifying a false negative encounter for 1° SAH (and to test the sensitivity of the algorithm), all SAH and incidental aneurism/arteriovenous malformation screen-positive encounters also underwent chart review verification (in addition to reviewing 1° SAH screen-positive encounters). As part of the accuracy assessment, an approximate additional random 200 encounters who did not meet screen positive criteria any of the 3 screens from 2008 were identified and manual chart verification undertaken.

5.5.5 1° SAH Search Algorithm Performance Measure

From the data obtained in section 5.5.4.4, 2x2 contingency tables were constructed and the operator performing characteristics (sensitivity, specificity, + and - likelihood ratios) of our search strategy on the validation cohort was calculated.

5.6 Identifying a Nine Year Cohort of 1° SAH

Using the search algorithm derived in 5.5.1 and validated in 5.5.4, all patients with 1° SAH presenting to TOH from July 1, 2002 to June 30, 2011 were identified as follows (see Figure 7):

a) All cranial imaging reports from emergency room visits and inpatient encounters during the defined period were abstracted from TOHDW using methods defined in

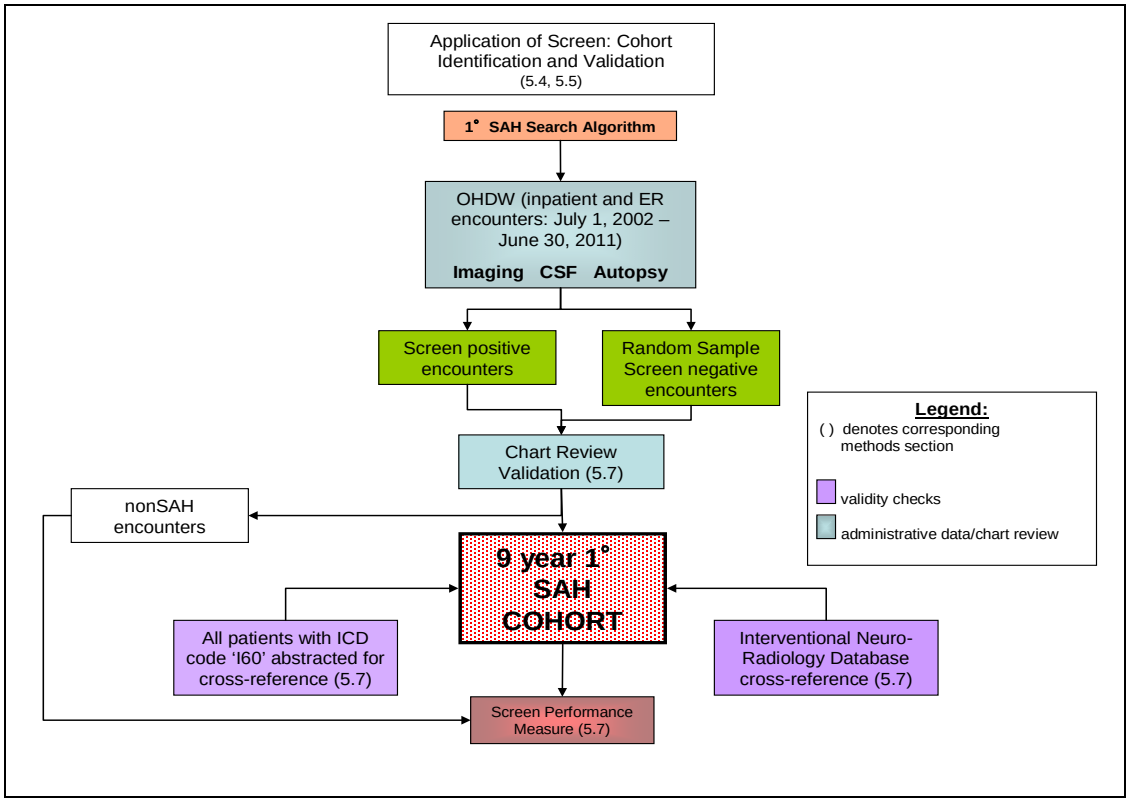
5.5.1.3. Similarly, the cranial imaging text screen was applied to identify all SAH, aneurysms, andarteriovenous malformations. If multiple reports were available for any given encounter, they were grouped with only one needing to be screen positive to classify the encounter as screen positive.

b) All CSF results from the study period were abstracted from TOHDW using the methods defined in 5.5.1.4. These results were first used to identify any with xanthochromia and then secondly for an RBC count $>5x10^6$ as described in 5.5.1.4. Any such sample was labeled as screen positive. Results from the 2 screens were merged and grouped at the encounter level such that any positive (xanthochromia and/or RBC screen) screen within a given encounter flagged the encounter as screen positive.

c) All autopsy reports over the same period were abstracted from TOHDW using the methods described in 5.5.1.5. The report text screen was then applied and any screen positive report was identified.

Similar to the methods described in 5.5.1.6and 5.5.4.4, the three screening methods were pooled and 1° SAH screen positive patients were identified using the same diagnostic criteria (seeFigure 4). Patients were excluded from the cohort if their age at time of encounter was less than 18years. All remaining 1° SAH screen positive encounters underwent chart review verification.

Figure 7: Primary SAH Cohort Identification (9 years)



5.7 Validity of Cohort

A manual review of each discharge summary or medical record of 'screen-positive' 1° SAH cases from the above searchstrategy wasundertaken to determine the true positive and false positive rates. To measure the accuracy of the search strategy for identifying all cases, a random sample of 1000 patients whodid not meet these screen criteria (as described in section 5.5) was selected for chart review verification.Based on sample calculations in section 5.4, finding no true cases of SAH among the 1000 screen negativeswould let us be 95% confident that the true proportion of false negative screens does not exceed 0.4%. Finding up to 5 false negative screened patients would provide an upper limit 95% confidence bound of the true false negative proportion of ≤1% (see also Appendix 1). The cranial CT head reports (if present), CSF analysis (if present), and autopsy reports (if

present), and the medical record of each of these patientswere reviewed to determine if they truly had a1°SAH. Selection of patients was stratified by year. From these data, 2x2 contingency tables were constructed and the operator performing characteristics(sensitivity, specificity, + and - likelihood ratios) of our search strategy were calculated.

The validity of the cohort was further tested in two ways:

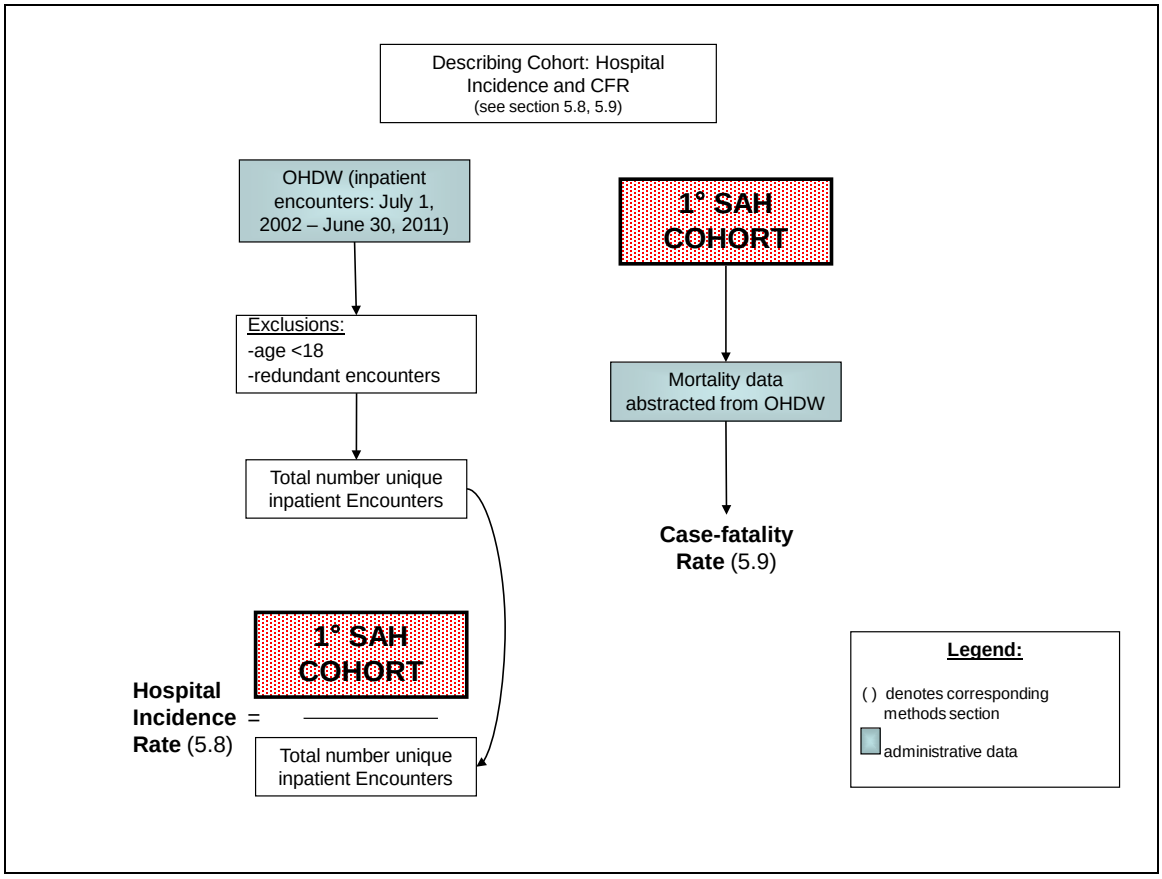
a) By cross-referencing the patients identified by the search algorithm with the existing Interventional Neuroradiology Database (2003 to 2011).
b) All patient encounters in TOHDW that had the diagnostic code for 1° SAH (ICD-10-CAI60) listed as a diagnosis for the encounter was abstracted from the database and their chart reviewed for verification.

Any patient having suffered a 1° SAH as identified by either of these methods and not identified by the search algorithm was labeled a false negative and entered into the performance characteristic measurement.

5.8 Hospital Incidence Rate

Hospital incidence rate was determined by calculating from TOHDW the total number of patient encounters that occurred over the entire study period to use as a denominator (see Figure 8). This was limited to those encounters for inpatientsexceeding 18 years of age.Psychiatric and rehabilitation admissions were excluded. To prevent double-counting,multiple encounters generated from the same hospital visit (e.g.transfer from one hospital site to another) and any two or more encounters whose discharge date and time was <12 hours from the admission date and time of a *subsequent* encounter for the same patient was considered a single encounter. Separate recurrent encounters by the same patient were not discounted.

Figure 8: Defining Hospital Incidence and Case-Fatality Rate



5.9 Hospital Case Fatality Rate at Hospital Discharge

Case fatality rates were determined using the vital status at hospital discharge for the cohort population identified in section 5.7 (see Figure 8). This was determined by linking the 1° SAH cohort from Section 5.7to TOH Discharge Abstract Database (DAD). The DAD accurately codes the vital status of all patients when discharged from the hospital.

5.10 1° SAH Prediction Model Development

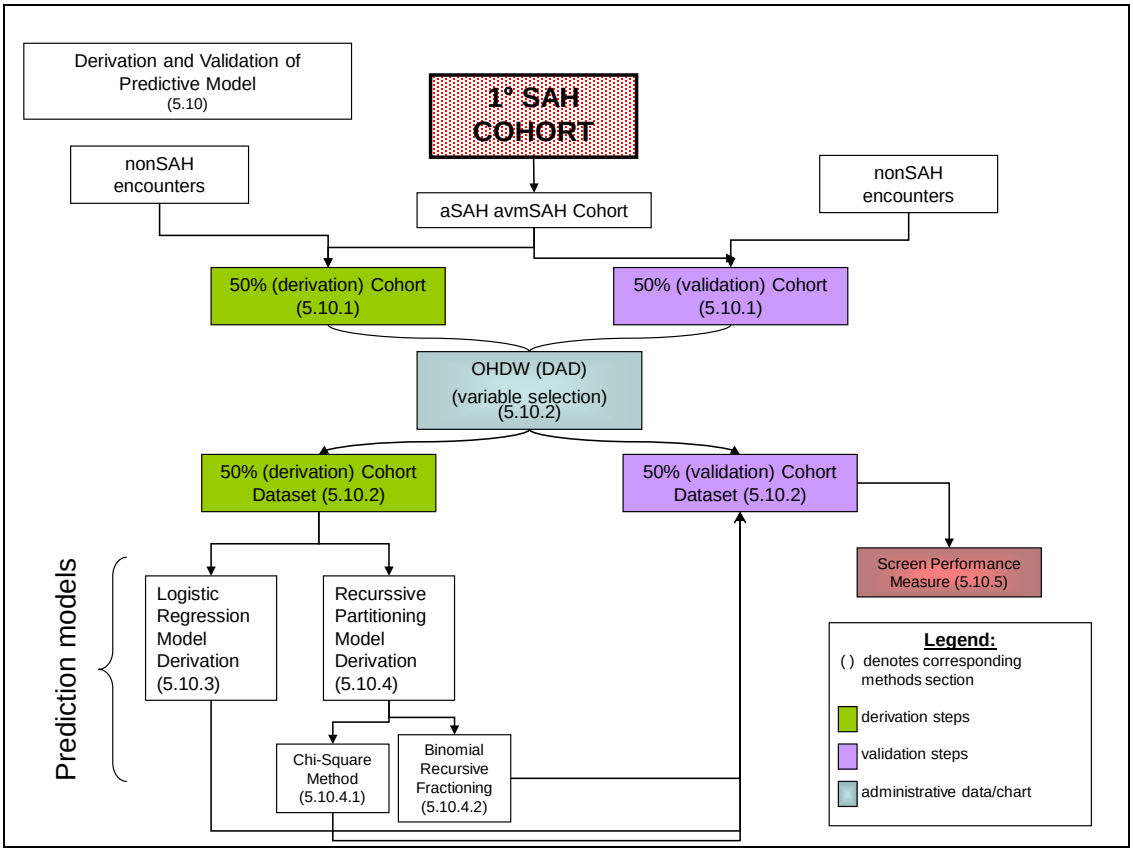
Patients with1°SAH as a result of aneurysm or arteriovenous malformation (aSAH/avmSAH) who were identified in the 1° SAH cohort (Section 5.6-5.7)served as the "Gold Standard"reference population of all aSAH/avmSAH at TOH between July 1, 2002and June 30, 2011. This subset of patients with SAH (aSAH/avmSAH) was selected

because they are of particular clinical importance, have a unique clinical course often involving neurosurgical and interventional neuroradiologic procedures and tend to have a higher morbidity and mortality then non-aneurysmal SAH patients. Patients who wereNOT listed in this cohort wereassumed to NOT have had anaSAH/avmSAH. This was given a binary variable (SAH / no SAH) whichserved as the outcome (dependent) variable for a logistic regression model that used covariates generated from the Discharge Abstract Database (DAD) as the predictor (independent) variables (see below).

5.10.1 Defining a Derivation and Validation Set

To derive and validate the prediction model, a derivation and validation set populated with cases and non-cases was needed. The steps followed are illustrated in Figure 9. Non-cases were identified from a random sample of 2.5% of all inpatient encounters during the study period from TOHDW;all encounters related to patients previously identified as having suffered an aSAH/avmSAH from our 1° SAH search algorithm, or age <18 years were removed. All remaining encounters were labeled “no SAH” (ie: outcome variable = 0). This cohort was merged with the known 9-yearaSAH/avmSAH cohort (ie: outcome variable = 1) and then randomly divided into 2 equal groups: one for derivation and the other for validation. All variables considered for the model were then abstracted from TOHDW (DAD) for both the derivation and validation set.

Figure 9: Derivation and Validation of a Prediction Model



5.10.2 Selection of Predictor Variables for Prediction Model

Predictor variables were generated from data available within the DAD. These included demographic and patient information (including patient age and sex) as well as hospitalization information (including admission service, diagnostic codes, hospital case mix group, length of stay, surgical procedures coded and transfers to intensive care unit). A list of diagnostic codes and procedural codes was generated by:

a) Collecting all diagnostic codes contained within the DAD for each encounter identified previously as aSAH/avmSAH

b) With the unit of measure being the encounter, all diagnostic codes were grouped in order of frequency. This was repeated for procedural codes.

Each diagnostic code or procedural code with a frequency ≥ 10 (i.e. was observed in more than 10 encounters with aSAH/avmSAH) was considered for inclusion in the model

c) For each diagnostic and procedural code identified in (b), a binomial variable was created for each encounter identified in the datasets in5.10.1, where "1" represented the diagnosis or procedure as being present (or having occurred) and "0" as not.

In a similar fashion as that described in (c) above, binomial variables for mortality, ICU admission, neurological admission (admission to Neurology or Neurosurgery Service) and urgent admission were also created.

5.10.3 Building the Logistic Regression Model

The following steps were used to create the model:

a) Logistic regression model was created using PROC LOGISTIC in SAS 9.2(SAS Industries Inc., North Carolina USA) with stepwiseselection of predictor variables and an alpha inclusion of 0.05. The more stringent alpha inclusion of 0.01 was also tested.

b) Clinically sensible interactions and those between the variables most strongly associated with the presence or absence of SAH were explored for collinearity.

Model discrimination was measured using the C-statistic with 95% confidence intervals. Model calibration was measured by comparing the expected and observed number of people with SAH after patients were categorized by deciles of predicted probability for SAH. The Hosmer-Lemeshow statistic was used to summarize model calibration. Once

the final model was achieved, its discrimination and calibration were tested on the validation set.

5.10.4 Building a Recursive Partitioning Model

Using the same independent variables identified in 5.10.2,a separate prediction model was built using recursive partitioning. Recursive partitioning is a simple nonparametric regression approach which allows successive splitting (recursive) of a sample by grouping (partitioning) like responses to different splitting variables (or predictor variables) together.[57]Recursive partitioning was completed using 2 methods for splitting:

i. Chi-squares method
ii. binomial recursive fractioning using Classification and Regression Tress (CART) (Salford Systems, California USA, version 6.6)

In both instances, the same derivation and validation sets as described in section 5.10.1were used.

5.10.4.1 Chi-squares Method

Using SAS 9.2, traditional 2x2 tables for each of the potential predictor variables (see Section 5.10.2) against the dependent variable (1° SAH) were created and the chi-square values calculated. Statistical significance of the association between the potential predictor variable and 1° SAH was established if its p-value was less than .05. The variable with the highest Chi-Square value that met statistical significance was entered into the model as a splitting variable. In this fashion, a tree was constructed by creating branches at each splitting variable by successively repeating these steps. Thus for each split, a branch was created for when the variable was present (variable=1) and for when the variable was

absent (variable=0). This was repeated until no further variables met statistical significance for entry or each cell contained ≤1 of either outcome (1° SAH or no 1° SAH).

5.10.4.2 Binomial Recursive Fractioning

Using CART®, a recursive partitioning tree was created with entry consideration for each of the potential predictor variables (see Section 5.10.2). Two separate statistical methods were considered for splitting (i.e. tree growing):

a) GINI: partitioningwas statistically determined by favoring the split that generated the most improvement in heterogeneity among daughter cells (buds) relative to the parent cell (node). (ref: Steinberg, Dan and Mikhail Golovnya. CART 6.0 User's Manual. San Diego, CA: Salford Systems, 2006)

b) Entropy: statistically similar to GINI, but tends to produce smaller terminal nodes (ie: with fewer observations), which may come at expense of accuracy. (ref: Steinberg, Dan and Mikhail Golovnya. CART 6.0 User's Manual. San Diego, CA: Salford Systems, 2006)

Trees were grown with a series of partitioning that were added and removed (pruning) to optimize the model with the optimal model determined by Receiver Operating Curves.

5.10.5 Prediction Model Performance

Model performance was measured by generating 2x2 tables to determine sensitivity, specificity and likelihood ratios in testing the expected outcome compared with observed outcome. Threeclassifications for determining expected outcomes were tested: one in which the terminal node was classified as representing patients with 1° SAH (expected event rate) if the observed event rate was ≥50%, another where the terminal node was considered to represent patients with 1° SAH if the observed event rate was ≥75% and

finally a more specific cut-off criteria of an observed event rate ≥90%. If the terminal node had an observed event rate of <50%,<75% or <90% respectively than the cell was classified as representing patients without 1° SAH. The model with the optimal performance characteristics was selected and manually programmed for validation using SAS 9.2. The performance of the algorithm was then tested against the validation set. Model performance (accuracy) was measured by comparing expected and observed number of patients with 1° SAH based on the classification with 2x2 tables to calculate sensitivity, specificity and positive likelihood ratios with 95% confidence intervals.

6 Results

6.1 Search Algorithm Derivation

Three separate screens were derived to make up the final search algorithm. The search algorithm was designed to identify all aSAH/avmSAH but all 1° SAH is alsoreported below.

6.1.1 Derivation of Imaging Case-defining Method

The text-search algorithm was derived using 263 cranial imaging reports abstracted from 69 patientencounters (representing 36 unique patients)from the Interventional Neuroradiology Database, 160 cranial imaging reports abstracted from 49 patient encounters(representing 30 unique patients) from the Neurosurgical Database, and 5814 cranial imaging reports abstracted from 5509 patient encounters (representing 5177 unique patients) randomly selected from all TOH inpatient and emergency room patient encounters in the year 2007.

Table 7displays the performance characteristics of the 2 image report text-search screens. For each of SAH and aneurysm/arteriovenous malformation, the performance of the text miner was assessed at the level an individual report, at all the reports contained within a given encounter and for all of the encounters for a given patient. Little change was observed in sensitivity, specificity, negative or positive predictive value at these levels. Although, sensitivity and specificity were each ≥95%, resulting in a very high negative predictive value, the positive predictive value for SAH and aneurysm/arteriovenous malformation was fair to low at 71% and 20% respectively.

Table 7: Image Report Text Miner Results

Text-search Objective	Unit of Analysis	Sample Size (n)	Sens-itivity (%)	95% CI	Spec-ificity (%)	95% CI	NPV	PPV
SAH	Image	5814	95.5	(91.8, 97.8)	98.5	(98.1-98.8)	99.8	70.9
	Encounter	5509	97.1	(93.3, 99,0)	98.7	(98.4-99.0)	99.9	70.9
	Patient	5177	96.6	(92.2, 98.9)	98.8	(98.5-99.1)	99.9	70.9
Aneurysm or AVM	Image	5814	100	(95.5, 100)	94.6	(94.0-95.2)	1	20.6
	Encounter	5509	100	(94.9, 100)	95.1	(94.5-95.7)	1	21.0
	Patient	5177	100	(94.1, 100)	95.2	(94.5-95.7)	1	19.7

SS=sample size, CI=confidence interval, NPV=negative predictive value, PPV=positive predictive value

6.1.2 Derivation of CSF Screening Method

Three CSF identifierswere identified in the TOHDW for xanthochromia in CSF samples. In 2007, 4702 CSF samples were analyzedfor xanthochromia, representing analyses from 516 unique encounters. Of these, 7% were positive for xanthochromia. Only 5% (26 cases) were positive within 24 hours of presentation.

Following manual review, only one stmWID identified RBC countsfrom CSF samples. From 2007, 3681 tests for CSF RBCs were reported. This represented 521 unique encounters, of which 51% (266 cases) had an RBC count $>5x10^6$/lwithin 24 hours of presentation.

6.1.3 Derivation of Autopsy Report Screening Method

Complete and/or partial autopsy reports are identified by 7 stmWIDs. Using these identifiers, a total of 88 reports were available from 2007. These were manually reviewed and zero reports identified either blood in the subarachnoid space and/or ruptured aneurysm or arteriovenous malformation. Reports of autopsies performed as a result of a coroner's inquest, although completed at the same facilities, are by law not available for review and furthermore are not housed within TOHDW. Consequently, the 88 reports above are all reports available that are unrelated to a coroner's inquest.

6.1.4 Pooling Screens to Form the Overall Search Strategy

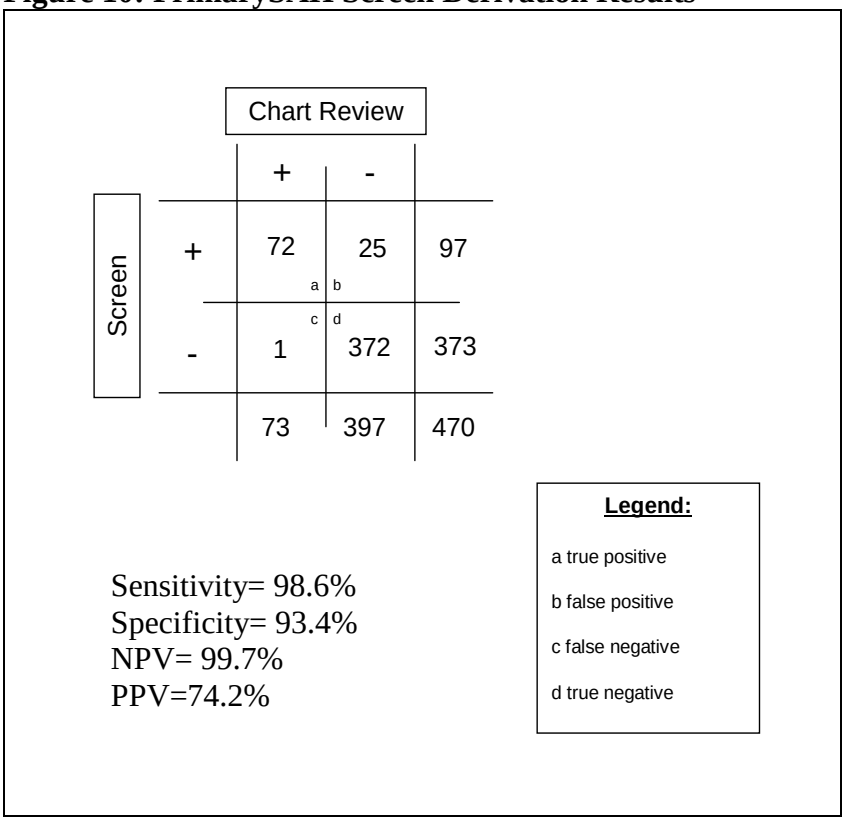
The 3 screens were pooled and patients with 1° SAH were identified following the pre-defined criteria presented inFigure 4. Applying this strategy to the three datasets (two known 1° SAH cohorts – Interventional Neuroradiology and Neurosurgery - and the random sample control group), a total of 198 patient encounters in 2007 were screen positive for 1° SAH. An additional 179 patient encounters were identified as screen negative for 1° SAH but were screen positive for either blood in the subarachnoid space or the presence of aneurysm or arteriovenous malformation on imaging.

6.2 Chart Review Verification of the Derived Screen

Chart review verification of the clinical diagnoses from the 377 patient encounters identified in 6.1.4 from the 3 datasets confirmed 177 true patient encounters with 1° SAH. This left 200 screen-positive patients without the diagnosis. From the patient encounters identified in the Imaging Case-Defining Method (section 6.1.1), an additional 211unique patients who were screen negative not only for 1° SAH but were also screen negative for blood in the subarachnoid space as well as the presence of aneurysm or arteriovenous malformationwere subsequently deemed to be true negatives on chart review (i.e. they did not have 1° SAH).

From chart review, all but one 1° SAH had been identified by the algorithm. The missed patient was identified in the Neurosurgical Database and had suffered a ruptured aneurysm in which blood was not identified in the subarachnoid space in the radiology reports but was, instead, identified in the patient's intraventricular space. Of the 73 patients with 1° SAH, 69 (95%) had suffered from ruptured aneurysm or arteriovenous malformation.

Figure 10: PrimarySAH Screen Derivation Results



6.3 Search Algorithm Validation

The results of the individual validation of each of the three components of the search algorithm as well as that of the final overall search algorithm are described here.

6.3.1 Validation of the Imaging Case-defining Method

The text-search screen was validated using 238 cranial imaging reports abstracted from 73 patient encounters (representing 36 unique patients) from the Interventional Neuroradiology Database, 195 cranial imaging reports abstracted from 37 patient encounters (representing 27 unique patients) from the Neurosurgical Database, and 4546 cranial imaging reports abstracted from 4413 patient encounters (representing 4194 unique patients) randomly selected from all TOH patient inpatient and emergency room encounters in 2008.

Table 8displays the performance characteristics of the 2 image report text search screens in the validation set. The performance of the text-search screen for SAH remained consistent in the validation set compared to the derivation with encounter sensitivity of 95.7% (95% CI 90.1-98.4%) and specificity of 98.4% (95%CI 98.0-98.8%). Negative predictive value remained high at almost 100% with a drop in positive predictive value to 65%. Similarly the aneurysm/arteriovenous malformation text-search screen at the encounter level had high sensitivity at 98.9% (95%CI 93.8-100%) with a specificity of 95.3% (95%CI 94.6-95.8%) while the positive predictive value was nearly 30%.

Table 8: Imaging Text-Search Screens Validation Results

TEXT-SEARCH OBJECTIVE	UNIT OF ANALYSIS	SAMPLE SIZE (N)	SENS-ITIVITY (%)	95% CI	SPEC-IFICITY (%)	95% CI	NPV	PPV
SAH	Image	4546	96.0	(91.4-98.5)	98.2	(97.8-98.6)	99.9	65.0
	Encounter	4413	95.7	(90.1-98.4)	98.4	(98.0-98.8)	99.9	66.0
	Patient	4194	95.3	(90.2-98.9)	98.6	(98.2-99.0)	99.9	68.7
Aneurysm/ AVM	Image	4546	98.9	(93.8-100)	95.0	(94.3-95.6)	1	28.2
	Encounter	4413	98.9	(93.8-100)	95.3	(94.6-95.8)	1	29.8
	Patient	4194	98.9	(93.8-100)	95.3	(94.6-95.9)	1	30.0

SS=sample size, CI=confidence interval, NPV=negative predictive value, PPV=positive predictive value

6.3.2 CSF Case-defining Method for Validation

Using the stmWIDs previously identified in section 6.1.2, all pertinent CSF data from 2008 was extracted. 3681 and 4863 test results (representing 521 and 522 patient encounters, respectively) were abstracted for CSF RBC count and xanthochromia testing respectively. Of these, 51% (266) and 6% (32) were positive for RBC count $>5x10^6$/l or xanthochromia within 24 hours of presentation respectively.

6.3.3 Validation of Autopsy Report Screening Method

A total of 101 partial or complete autopsy reports were available from 2008 and were screened using the text-search screens derived and validated from the radiology reports for blood in the subarachnoid space and presence of aneurysm or arteriovenous malformation (see sections 6.1.1 and 6.3.1). Two reports were identified as screen-positive and confirmed by manual chart review, as containing blood in the subarachnoid space,however, each were from secondary causes.

6.3.4 Pooling of the Search Strategies

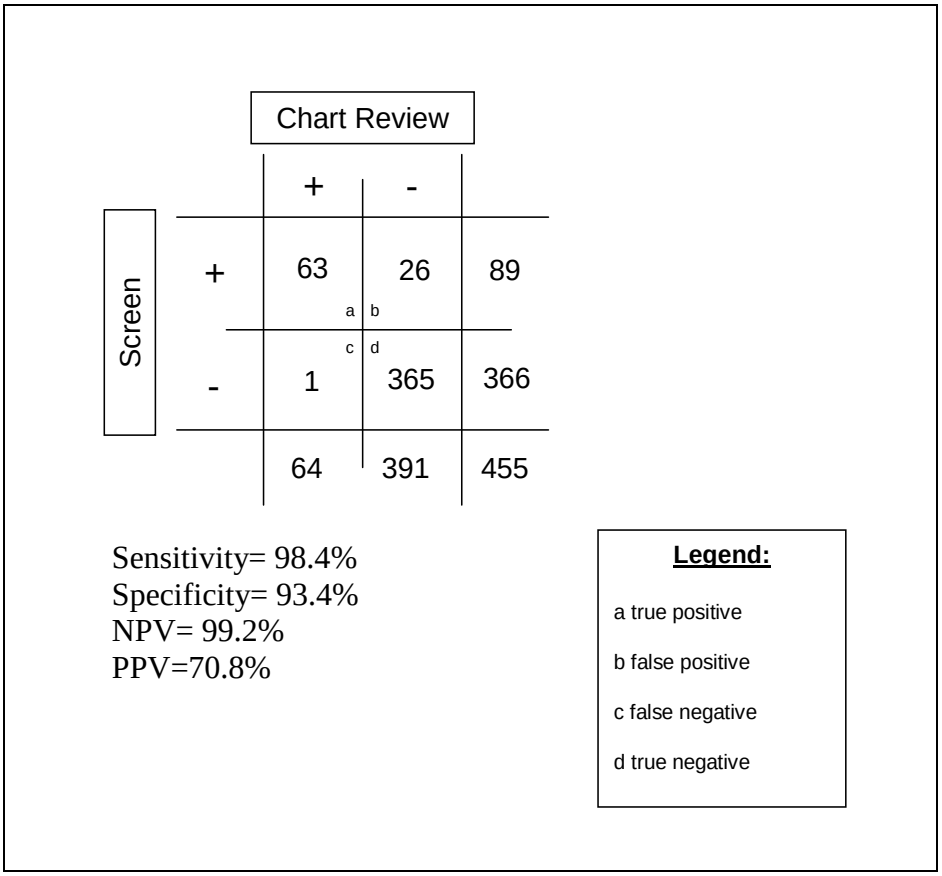
Pooling the 3 search strategies, and applying the predefined criteria displayed inFigure 4, a total of 171 patient encounters were screen positive for 1° SAH. An additional 161 patient encounters were identified as screen negativefor 1° SAH but screen positive for either blood in the subarachnoid space or presence of aneurysm or arteriovenous malformation on imaging (ie: non-1° SAH or incidental aneurysm or arteriovenous malformation).

6.4 Chart Review Verification to Validate Screen

Chart review verification of the clinical diagnoses from the 332 patient encounters from section 6.3.4 was conducted using the 3 datasets (2 known 1° SAH datasets – Interventional Neuroradiology dataset and Neurosurgery Dataset; and a random control group)and confirmed 141 true patient encounters with 1° SAH and191 without the diagnosis. An additional 222 unique patient encounters designated screen negative for blood in the subarachnoid space as well as for presence of aneurysm or arteriovenous malformation (as identified from the dataset utilized in the Imaging Case-Defining Method (section 6.3.1)) underwent chart review to ensure no falsely negative screen was overlooked.

The results of the analysis of the screen at the unique patient level following merging of the datasets and removal of duplication are presented in Figure 11.

Figure 11: Primary SAH Screen Validation Results



In this validation set, the screen for patients with 1° SAHidentified all but one case demonstrating a sensitivity of 98.4% (91.6-100,95% CI) and a specificity of 93.4% (90.4-95.6, 95% CI) when compared to the gold standard of chart review. The resultant negative predictive value of the screen was 99.7% suggesting a less than 1% chance of committing a type I error in a population with a high prevalence of 1° SAH.The one patient that the search algorithm failed to identify was again from the Neurosurgical Database. The patient had suffered a ruptured aneurysm but at time of admission had already undergone cranial imaging at a different institution and thus no reports were available for the screen to positively identify blood in the subarachnoid space. Of the 64 true cases of 1° SAH (as

confirmed in chart review), 56 (88%) were secondary to aneurysmal or arteriovenous malformation rupture, with 8 due to a perimesencephalic bleed or idiopathic causes.

6.5 The TOH 1° SAHCohort (a 9 Year Cohort)

The derived and validated search algorithm was then utilized to identify all patients with 1° SAH presenting to TOH from July 1, 2002 to June 30, 2011. This was completed by running each of the 3 separate componentscreens on the imaging reports, CSF results and autopsy reports as previously described (see Table 9). The three components were then combined as described in the derivation and validation steps above.

Table 9: Individual Screen Components for Cohort Identification

			N	ENCOUNTERS (N)	PATIENTS (N)
Imaging		**All**	136,353	117,818	81,386
		SAH +	9805	6774	5051
		Aneurysm/AVM +	9975	6759	5077
CSF	**RBC**	**All**	30,305 / 4658*	4653	4460
		>5x10^6/l	2341	2338	2272
	Xantho-chromia	**All**	38,984 / 7261*	4667	4471
		Present/ positive	236	191	188
Autopsy		**All**[£]	577	n/a	455
		SAH +	9	8	8

*restricted to first 24hours following start of encounter, [£] All autopsies include provisional reports and site specific post-mortem examinations

A total of 1699 unique patients at TOH were search algorithm positive for 1° SAH between 2002 and 2011. Over the 9 year period, a total of831 patients were true1° SAHbased on

chart review. Of these, 632 (76% of 1° SAH patients) had suffered from a ruptured aneurysm or arteriovenous malformation.Figure 12 summarizes the results of the search algorithm and the ensuing verification undertaken by chart review at an individual patient level suffering from aSAH/avmSAH. One patient in this cohort was less than 18 years of age at time of presentation (17 years).Forty-five patients had suffered a probable aneurysm rupture (seeFigure 4), in that their clinical history was in keeping with an aneurysm rupture, blood was seen in the subarachnoid space on neuro-imaging or CSF studies, but no further imaging to identify a potential aneurysm was completed.

An additional 758 unique patients who were search algorithm negative were used to check the validity of the cohort. In order to increase the chances of finding a false-negative, these negative-screen patients were populated with patient encounters that had positive sub-screens but were negative overall for the search algorithm for 1° SAH. In addition to the search algorithm negative samples of patients, this group was also populated with the review of all ICD-10 Code I60 positive patients and the Interventional Neuroradiology database (refer tomethods section 5.7) (see Table 10).

The sensitivity of the screen for 1° SAHwas 96.5% (94.8%-97.8%; 95% CI). There were 22screen-negative patients who truly had aneurysmal or arteriovenous malformation SAH basedon chart review. Of these 22 false negative patients, 21 were identified during the cross-reference validation with ICD-10 code I60 (subarachnoid hemorrhage) positive patients and 1 patient identified from the Interventional Neuroradiology Database (seeTable 11).

Figure 12: Chart Review Verification of 1° SAH Cohort

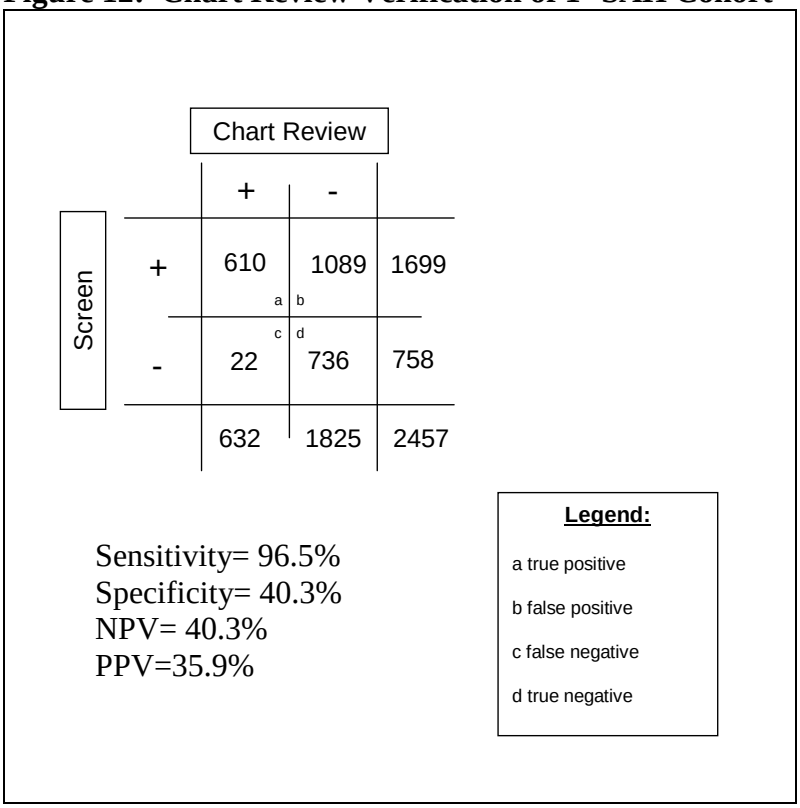


Table 10: Data Sources and Corresponding Contribution to Cohort

	Source	Encounters (N)	Patients (N)	True aSAH/ avmSAH on chart review	True False Negative (aSAH/ avmSAH) Patients
Base Cohort	1° SAH search algorithm	5383	1699	610	n/a
Additional Searches/ Validation steps*	IR database cross-reference	445	386	239	1
	ICD code for SAH cross-reference	1049	1009	484	22
	Search algorithm Negative	433	433	1	1

*Note: some patients identified by more than one method, ICD=International Classification of Disease Code, IR=interventional neuroradiology

Table 11: Examining False Negatives

		FALSE NEGATIVE PATIENTS (N)
Total:		22
Source:		
	I60 positive	21
	IR database	1
Reason missed:		
	No blood in subarachnoid space	9
	No imaging or LP at TOH	10
	Spelling error in dictated report not caught by search algorithm	2
	Other	1

IR=Interventional Neuroradiology

6.6 Hospital Incidence and Case Fatality Rate of aSAH/avmSAH

Over the 9 years included in the study, there were a total of 379,384 unique inpatient encounters, representing 226,669 unique patients (ie: accounting for repeat admissions for the same patient). Accounting for the fact that the first and last years of the study period (July 1, 2002 to June 30, 2011) only contain 6 months, the number of admissions per year were consistentwith less than 1% fluctuation, ranging from 41,613 to 42,566 (seeTable 12).

Table 12: Unique Patient Encounters by Year

ADMISSION YEAR	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
2002	21124	5.57	21124	5.57
2003	42018	11.08	63142	16.64
2004	42566	11.22	105708	27.86
2005	41942	11.06	147650	38.92
2006	42317	11.15	189967	50.07
2007	41760	11.01	231727	61.08
2008	41613	10.97	273340	72.05
2009	42340	11.16	315680	83.21
2010	42215	11.13	357895	94.34
2011	21489	5.66	379384	100.00

The number of incident aSAH/avmSAH cases ranged from 61 to 80 per year (see Table 13) with a corresponding hospital incidence of 14.5 to 19.2/10,000 admissions. The overall incidence of aSAH/avmSAH at TOH between 2002 and 2011 was 16.7 cases/10,000 admissions. Case fatality rates varied more significantly from 4% in 2006 to 24.6% in 2010, with an overall case fatality rate of 18.2% (see Table 13). Removing the outlier death rate of 4% from 2006, hospital mortality ranged from 15.6-24.6%.

Table 13: aSAH/avmSAHHospital Incidence and Mortality by Year of Admission

ADMISSION YEAR	FREQUENCY (N)	PERCENT (%)	HOSPITAL INCIDENCE (PER 10,000 ADMISSIONS)	DEATH (N)	CASE FATALITY RATE
2002	33	5.2	15.6	8	24.2
2003	63	10.0	15.0	14	22.2
2004	65	10.3	15.3	15	23.1
2005	68	10.8	16.2	14	20.6
2006	75	11.9	17.7	3	4.0
2007	80	12.7	19.2	14	17.5
2008	70	11.1	16.8	13	18.6
2009	77	12.2	18.2	12	15.6
2010	61	9.6	14.5	15	24.6
2011	40	6.3	18.6	7	17.5
Cumulative	632	100.00	16.7	115	18.20

6.7 Predictive Model

Two predictive models to estimate the probability that a given hospital admission is secondary to an aSAH/avmSAH were built. First, we report on the multivariate logistic regression model (section 6.7.1), followed by recursive partitioning models (section 6.7.4).

6.7.1 Derivation and Validation Set

As described in Methods section 5.10.1, two populated sets were created: one for derivation of the model and the other for its validation. A random 2.5% sample of all inpatient encounters covering the same time period of our study (July 1, 2002 to June 30, 2011), age (18 years and older), and not already identified in the cohort were merged with the existing 631 patients (one patient <18 years of age was removed). This group of 10,322 patient encounters was then randomly divided into 2 groups for derivation (n=5122) and validation (n=5200) of the model. The derivation set contained 315 unique patient encounters that were previously identified as having aSAH/avmSAH (gold standard). The validation set contained 316 unique patient encounters with aSAH/avmSAH.

6.7.2 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 which appeared more than 10 times the discharge abstracts (DAD) of those patients with aSAH/avmSAH. Variables were selected in order of frequency of appearance in the discharge abstracts of aSAH/avmSAH patients. Sixty-three variables were diagnostic codes, 38 procedural codes and 7 other encounter characteristics (see Table 14).

6.7.3 Logistic Regression Model

An univariate analysis of the 108 potential predictor variables was completed first in the derivation set (Table 14). Twenty-four variables failed to meet statistical significance and thus were dropped. The remaining 84 variables were then run in a univariate analysis using the validation set (see Table 14). Clinically relevant interactions were included in the derivation model, none of which reached statistical significance.

Table 14: Unadjusted (Univariate) OR of 1° SAH based on Diagnostic or Procedural Code and Hospital Encounter/Patient Characteristics

			ASSOCIATION OF CODE WITH 1° SAH			
			Derivation Set (1° SAH =315, n=5122)		Validation Set (1° SAH =316, n=5200)	
Category	**System**	**Variable/Diagnostic Code/Procedural Code**	**OR**	**95% CI**	**OR**	**95% CI**
Diagnosis	CNS	Subarachnoid Hemorrhage	>999.999	(<999.999,>999.999)	>999.999	(<999.999,>999.999)
		Hydrocephalus	162.048	(76.995, 341.057)	110.561	(60.407, 202.356)
		Cerebrovascular Disease	72.912	(39.211, 135.580)	110.307	(57.424, 211.891)
		Intracerebral hemorrhage	71.780	(35.577, 144.821)	22.399	(13.367, 37.534)
		Postprocedural disorder of CNS	113.067	(47.463, 269.347)	166.170	(65.507, 421.521)
		Cerebral edema	15.791	(9.533, 26.156)	10.235	(5.517, 18.986)
		Cerebral Infarction	7.801	(4.975, 12.233)	9.536	(6.126, 14.846)
		Vasospasm	25.814	(13.630, 48.886)	29.580	(15.823, 55.297)
		Disorientation/amnesia	2.631	(1.554, 4.456)	4.068	(2.646, 6.255)
		Hemiplegia	10.499	(6.498, 16.964)	10.159	(5.823, 17.723)
		Headache	34.257	(15.986, 73.409)	7.625	(3.683, 15.786)
		Speech disturbances	8.128	(4.807, 13.743)	4.060	(2.007, 8.211)
		Intracranial injury including concussion	2.577	(1.078, 6.160)	2.892	(1.106, 7.562)
		Visual disturbance	4.743	(1.538, 14.632)	6.893	(2.815, 16.880)
		Somnolence/stupor/coma	7.789	(3.122, 19.436)	11.368	(5.008, 25.803)
		Convulsions	3.745	(1.920, 7.304)	4.908	(2.303, 10.460)
		Other nontraumatic ICH	18.649	(5.660, 61.445)	11.041	(4.174, 29.205)
		Delirium	1.076	(0.431, 2.684)	--	--
		Stroke/TIA	2.367	(0.821, 6.822)	--	--
		Other aneurysm, dissection		()	4.587	(1.961, 10.730)
	Cardio-vascular	Essential Hypertension	2.971	(2.312, 3.816)	3.453	(2.706, 4.407)
		Atrial fibrillation/flutter	0.390	(0.192, 0.795)	0.729	(0.436, 1.221)
		Heart failure	0.601	(0.334, 1.084)	--	--
		Cardiac arrest	2.316	(0.974, 5.502)	--	--

		Pulmonary embolism	2.956	(1.376, 6.352)	0.657	(0.159, 2.713)
		Abnormal heart beats	3.031	(1.616, 5.685)	1.272	(0.507, 3.187)
		Acute MI	0.391	(0.183, 0.836)	0.805	(0.464, 1.395)
		Chronic ischemic heart disease	0.124	(0.051, 0.301)	0.147	(0.065, 0.330)
		Hypotension	1.178	(0.541, 2.563)	--	--
	Respir-atory	Pulmonary edema	4.541	(2.298, 8.973)	5.886	(3.222, 10.756)
		Respiratory failure	2.626	(1.527, 4.515)	2.841	(1.648, 4.897)
		Other mental disorder secondary to brain injury or physical disease	8.954	(4.491, 17.85)	21.011	(10.349, 42.658)
		Pneumonia	0.838	(0.452, 1.555)	--	--
		Pneumonitis due to solids/liquids	4.009	(2.047, 7.852)	3.948	(1.885, 8.270)
		Problems related to lifestyle (eg: EtOH, smoking)	0.193	(0.027, 1.393)	--	--
		Other COPD	0.540	(0.252, 1.159)	--	--
	Infection	Bacterial Agent as cause of Disease	4.427	(3.178, 6.166)	4.812	(3.510, 6.597)
		Strep/staph infection	3.937	(2.503, 6.192)	2.875	(1.781, 4.641)
		Bacterial pneumonia	8.897	(4.578, 17.292)	6.943	(3.585, 13.447)
		Fever unknown origin	1.108	(0.444, 2.765)	--	--
		Carrier of infectious disease	0.144	(0.036, 0.583)	0.308	(0.114, 0.835)
		Sepsis	0.978	(0.393, 2.433)	--	--
		Bacterial meningitis	>999.999	(<0.001, >999.999)	--	--
	GI/GU	Other disorder of Urinary System	4.892	(3.565, 6.712)	4.078	(2.973, 5.594)
		Dysphagia	4.381	(2.695, 7.121)	3.793	(2.271, 6.337)
		Postprocedural complication of GI system	1.132	(0.407, 3.146)	--	--
	Endocrine / Metabolic	Other disorders of Fluids/electrolytes/acid-base	3.093	(2.152, 4.446)	2.521	(1.740, 3.652)
		Other Anemias	1.657	(1.030, 2.666)	1.939	(1.239, 3.037)
		NIDDM	0.494	(0.308, 0.792	0.466	(0.288, 0.757)
		Postprocedure respiratory disorder	5.211	(3.142, 8.642)	4.793	(3.023, 7.599)
		Complications of other internal prosthetic devices	10.795	(6.057, 19.240)	9.200	(5.017, 16.870)
		Acute posthemorrhagic anemia	0.419	(0.132, 1.325)	--	--
		Other medical Care (incl transfusion, palliation)	0.565	(0.287, 1.110)	--	--
		Hypothyroid	0.692	(0.216, 2.211)	--	--
		Increased glucose	0.902	(0.438, 1.857)	--	--
		SIADH	20.649	(8.981, 47.477)	9.805	(3.189, 30.146)
	Misc	Problems related to health	5.607	(4.070, 7.725)	5.887	(4.288, 8.082)

		care facilities				
		Abnormal Reaction of patient to Surgical Operation/procedure	2.967	(2.259, 3.897)	3.179	(2.441, 4.140)
		Abnormal Reaction of patient to Other Medical Procedures	9.084	(6.634, 12.43)	9.080	(6.473, 12.736)
		Complications of procedures Not elsewhere classified	2.380	(1.636, 3.461)	2.112	(1.409, 3.167)
		Systemic sclerosis	<0.001	(<0.001, >999.999)	--	--
		Rash	<0.001	(<0.001, >999.999)	--	--
		Seborrheic dermatitis	15.306	(0.955, 245.274)	--	--
Proce-dures:	CNS Imaging:	CT Brain	48.430	(36.764, 63.797)	58.281	(43.981, 77.231)
		CT Head	7.968	(5.506, 11.531)	11.101	(7.875, 15.648)
		Diagnostic imaging - CNS	12.653	(5.871, 27.271)	4.505	(2.036, 9.967)
		MRI Brain	16.566	(11.868, 23.123)	14.995	(11.009, 20.423)
		CT other vessels of Neck	90.368	(49.048, 166.498)	104.575	(59.247, 184.583)
		MRI other vessels of head	26.925	(14.284, 50.755)	27.916	(15.993, 48.729)
		Diagnostic imaging – carotid artery	173.635	(22.373, >999.999)	56.556	(18.504, 172.860)
		Xray, carotid artery		()	192.161	(68.481, 539.210)
		Diagnostic imaging soft tissue neck	3.870	(1.570, 9.537)	6.620	(3.138, 13.968)
		Xray, intracranial vessels	313.430	(180.313, 544.821)	637.598	(294.187, >999.999)
		Diagnostic imaging – Carotid artery	54.546	(11.289, 263.562)	15.548	(2.183, 110.742)
	CNS Interven-tion:	Therapeutic occlusion intracranial vessels	>999.999	(>999.999, >999.999)	>999.999	(972.766, >999.999
		Includes EVD	254.563	(101.652, 637.496)	506.250	(158.522, >999.999)
		Imaging intervention – brain	17.929	(10.166, 31.620)	29.413	(14.749, 58.654)
		Repair meninges, dura	75.014	(30.726, 183.136)	54.850	(21.870, 137.559)
		Treatment spinal canal and meninges	291.204	(38.751, >999.999)	37.676	(14.378, 98.729)
		Pharmacotherapy IC vessels	173.635	(22.373, >999.999)	294.944	(39.241, >999.999)
		Cranium treatment	60.037	(19.804, 182.005)	62.347	(23.127, 168.079)
		Treatment – brain	154.214	(35.752, 665.194)	>999.999	(<0.001, >999.999)
		Dilation intracranial vessels	141.349	(17.851, >999.999)	>999.999	(<0.001, >999.999)
		Removal of device – ventricles of brain	62.606	(13.238, 296.080)	94.354	(11.338, 785.187)
		Repair intracranial vessels	>999.999	(<0.001,	--	--

				>999.999)		
		Drainage – meninges/dura	10.368	(3.668, 29.311)	10.045	(3.867, 26.090)
		Dilation ventricle of brain	93.316	(11.200, 777.494)	>999.999	(<0.001, >999.999)
		Management internal device – ventricles of brain	15.397	(3.095, 76.599)	110.611	(13.567, 901.816)
		Carotid artery treatment	>999.999	(<0.001, >999.999)	--	--
	Respir-atory Interven-tion:	Includes NIPPV, IPPV, Trach	9.013	(6.846, 11.865)	11.955	(9.207, 15.522)
		Bypass with externalization – trachea	13.578	(8.267, 22.300)	18.158	(10.405, 31.687)
		Includes bronch	1.806	(0.637, 5.121)	--	--
		Diagnostic intervention of the lung	<0.001	(<0.001, >999.999)	--	--
		Dilation trachea	61.813	(6.888, 554.695)	47.244	(9.496, 235.031)
	CVS interven-tion:	Implantation of internal device in Vena Cava	6.470	(4.879, 8.579)	7.412	(5.593, 9.822)
		Circulatory system treatment	3.770	(1.726, 8.232)	2.438	(1.091, 5.450)
		Implantation internal device – vein	46.152	(4.790, 444.642)	26.157	(6.223, 109.954)
	GI/GU Imaging/ interven-tion:	Stomach intervention (Including PEG)	9.830	(5.846, 16.530)	9.326	(5.390, 16.136)
	Misc	Implantation of internal device – tympanic membrane	<0.001	(<0.001, >999.999)	--	--
		Specimen collection for diagnostic imaging	>999.999	(<0.001, >999.999)	--	--
Encoun-ter / Patient Data:		UrgAdm	3.487	(2.557, 4.755)	3.523	(2.571, 4.829)
		ICUAdm	18.487	(14.242, 23.997)	22.522	(17.401, 29.149)
		age		()	1.004	(0.998, 1.010)
		Mortality	4.033	(2.886, 5.635)	6.088	(4.485, 8.263)
		Sex	1.508	(1.183, 1.922)	1.576	(1.237, 2.007)
		HospLOS	1.023	(1.018, 1.029)	1.029	(1.024, 1.034)
		ICULOS	1.184	(1.151, 1.218)	1.275	(1.231, 1.319)
		service	103.301	(74.177, 143.860)	100.168	(71.743, 139.855)

In a multivariate analysis using the 5122 encounters in the derivation set, independent predictors of a hospital encounter for aSAH/avmSAH included: having a

diagnostic code for SAH, intracerebral hemorrhage, intracranial injury including concussion, cerebrovascular disease, other nontraumatic intracranial hemorrhage, convulsions, bacterial pneumonia, or headache; or having the procedural code for therapeutic occlusion of intracranial vessels or diagnostic imaging of the carotid artery (see Table 15). If the discharge abstract contained the diagnostic code for atrial fibrillation/flutter or the code for urgent admission, an encounter was statistically significantly *less* likely to be related to aSAH/avmSAH.

Table 15: Adjusted OR for aSAH/avmSAH from Multivariate Logistic Regression Model

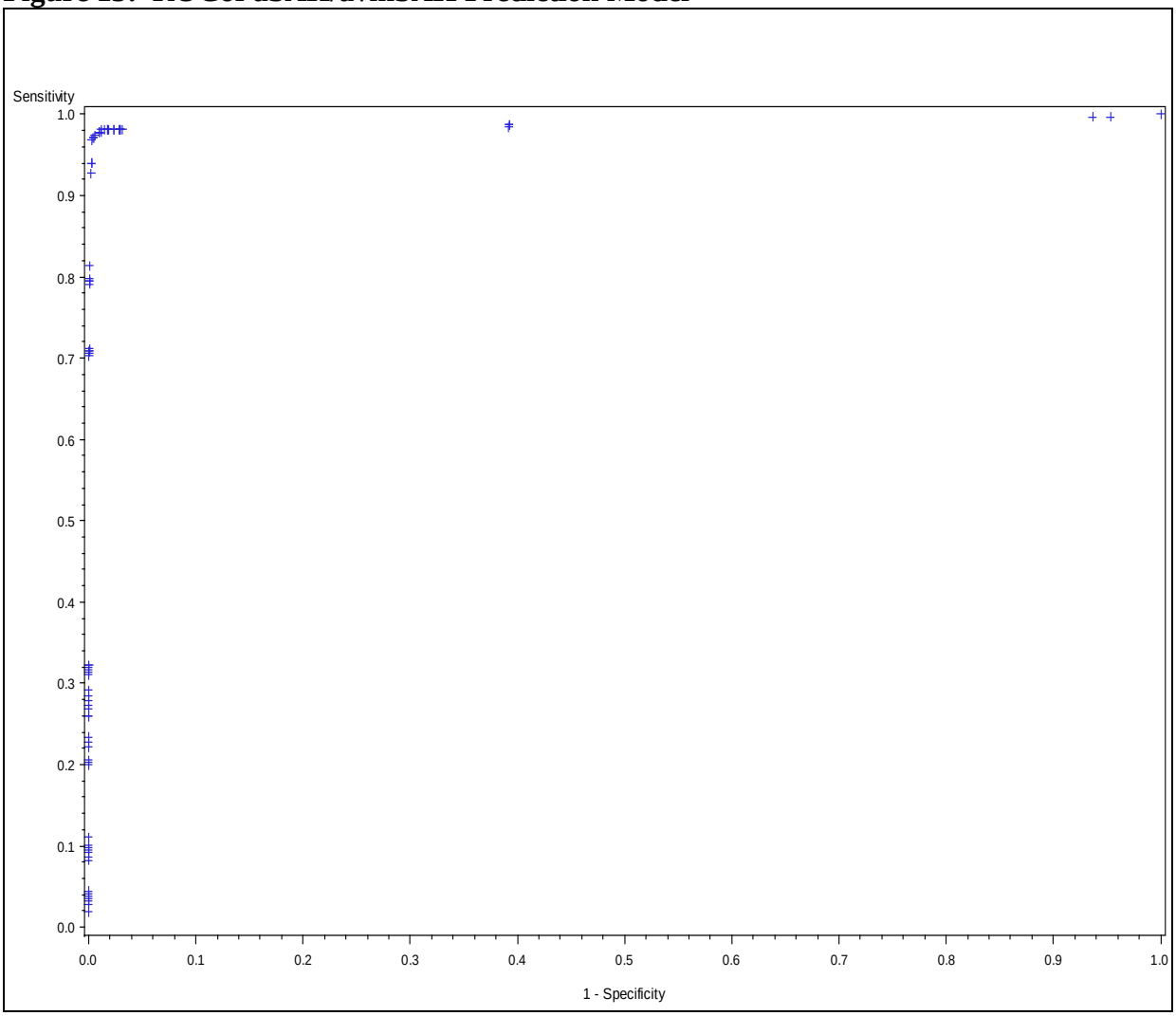
CATEGORY	VARIABLE/DIAGNOSTIC CODE/PROCEDURAL CODE	DERIVATION SET (1° SAH =315, N=5122)	
		OR	95% CI
Diagnosis:	Subarachnoid Hemorrhage	>999.9	(>999.999, >999.999)
	Intracerebral hemorrhage	>999.9	(567.900, >999.999)
	Intracranial injury including concussion	240.4	(47.873, >999.999)
	Cerebrovascular Disease	289.1	(50.883, >999.999)
	Other nontraumatic ICH	724.3	(47.178, >999.999)
	Convulsions	54.8	(9.112, 329.856)
	Bacterial pneumonia	395.5	(33.392, >999.999)
	Headache	470.0	(41.003, >999.999)
	Atrial fibrillation/flutter	0.04	(0.006, 0.306)
Procedures:	Therapeutic occlusion intracranial vessels	38.0	(7.376, 195.838)
	Diagnostic imaging – Carotid artery	>999.9	(55.082, >999.999)
Encounter / Patient Data:	Urgent Admission	0.07	(0.018, 0.253)

In the derivation set, this model had excellent calibration (Hosmer-Lemeshow test, p=0.5217) and an excellent discrimination with a c statistic of 0.998. The predictive model did not differ when the more stringent alpha inclusion of 0.01 was used.

In the validation set (comprised of 5200 encounters), the model was used to construct the receiver operating characteristic (ROC) curve presented in Figure 13. It

demonstrates excellent discriminatory performance for aSAH/avmSAH with an area under the curve of 0.988 (0.9783-0.9977, 95% CI).

Figure 13: ROCof aSAH/avmSAH Prediction Model



In the validation set, 93% of encounters without aSAH/avmSAH (true negatives) had a predicted probability of the outcome from the model of less than 10% (see Table 16). Further, 74% of the true positives (ie: patients with aSAH/avmSAH) had a predicted probability exceeding 90% based on the logistic model. The true positives were separated from the true negatives by the model with high and low probability respectively (see Figure

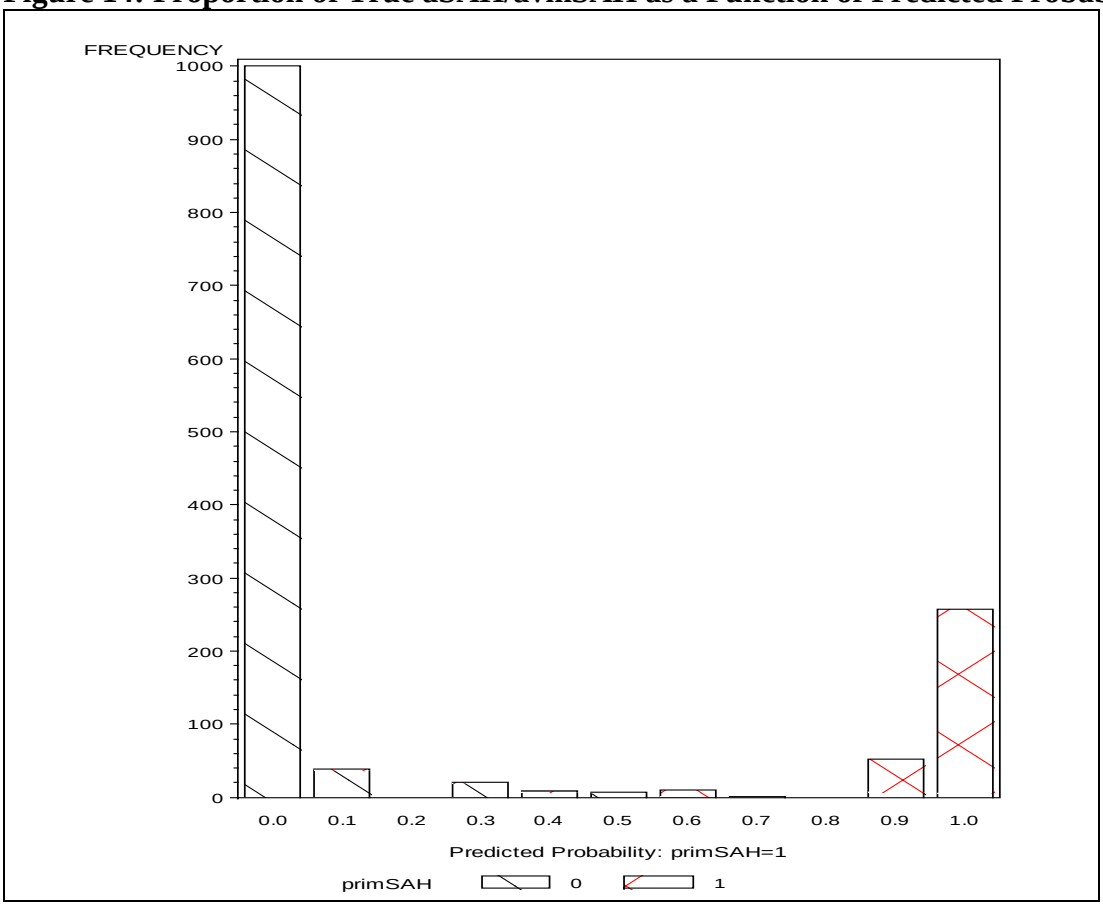
14). However, 7 true cases had a predicted probability of having aSAH/avmSAHthat was less than 10% with an additional 3 cases having a predicted probability that was less than 50%. In all 10 instances, the low predicted likelihood was likely due to cases not having a code for I60 (i.e. they were miscoded as *not* having subarachnoid hemorrhage). Of these 10 cases, 8 were aneurysmal SAH and 2 were related to arteriovenous malformations.

Table 17displays the performance characteristics of this model at arbitrary predicted probability cut points of 50%, 75%, and 90%. As probability cut-points increased from 50% to 90%, although specificity remained stable at >99%, sensitivity dropped substantially from 96.8% to 81.3% respectively. Despite the drop in sensitivity, increasing the probability cut-point to ≥90% was the only way to obtain a significant increase in positive likelihood ratio to 678. Given the limitation of this predictive model to accurately predict all cases of true aSAH/avmSAH, predictive models using recursive partitioning were explored.

Table 16: Predictive Model Performance at Various Predicted Probability Cut-offs

PREDICTED PROBABILITY OF ASAH/AVMSAH BASED ON LOGISTIC MODEL	PATIENTS WITHOUT DISEASE, N= 4884 (%)	PATIENTS WITH THE DISEASE, N=316(%)
<0.1	4835 (99.0)	7 (2.2)
0.1 – 0.2	1 (0)	0 (0)
0.2 – 0.3	0 (0)	0 (0)
0.3 – 0.4	24 (0.5)	2 (0.6)
0.4 – 0.5	9 (0.2)	1 (0.3)
0.5 – 0.6	1 (0)	9 (2.8)
0.6 – 0.7	1 (0)	0 (0)
0.7 – 0.8	0 (0)	0 (0)
0.8 – 0.9	7 (0.1)	40 (12.7)
0.9 – 0.95	0 (0)	5 (1.6)
>0.95	6 (0.1)	252 (79.7)

Figure 14: Proportion of True aSAH/avmSAH as a Function of Predicted Probability



6.7.4 Recursive Partitioning Model

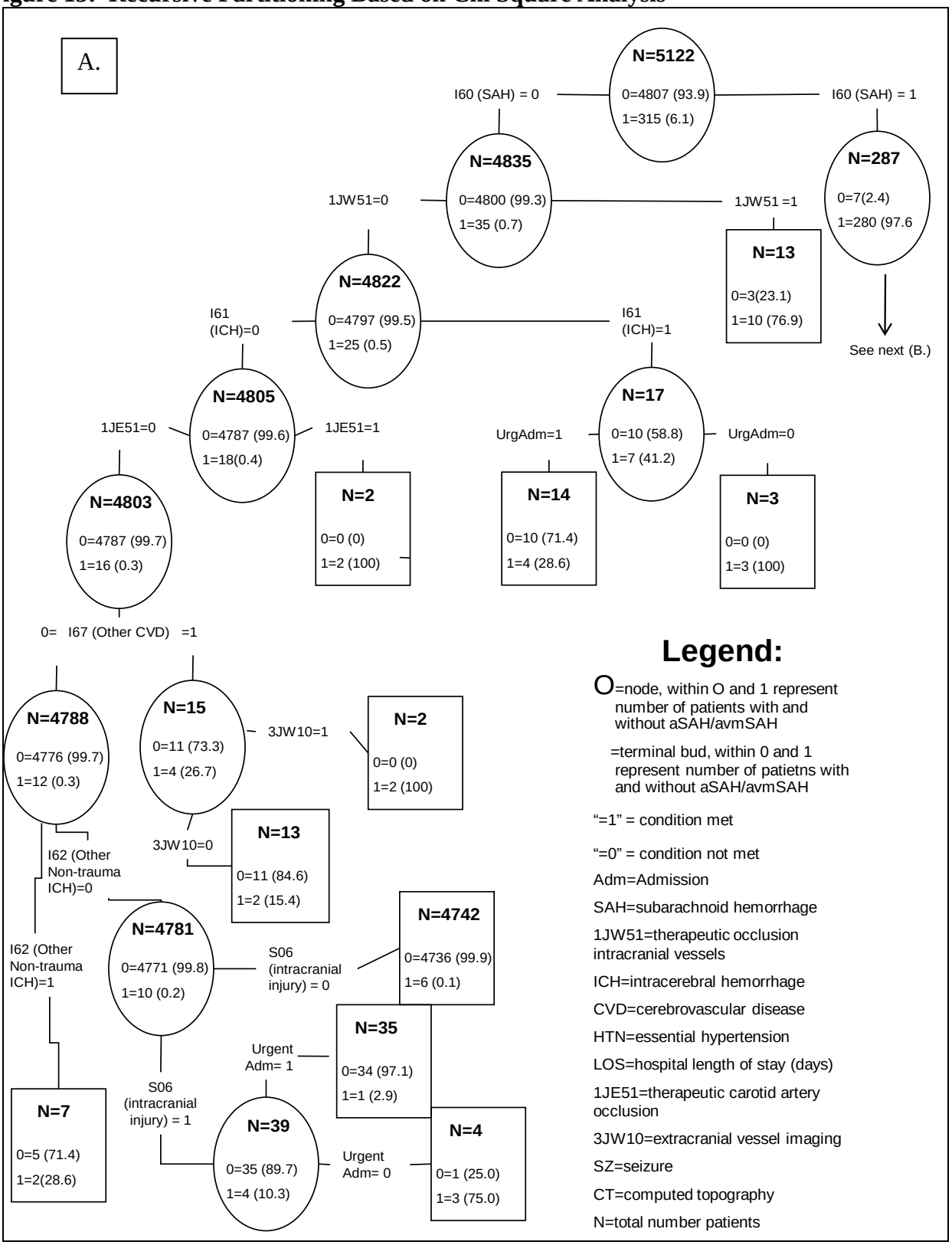
Two methods were instituted to derive a predictive model from recursive partitioning: a model where splits were determined using chi-square statistics and a binomial recursive fractioning model using CART statistical software.

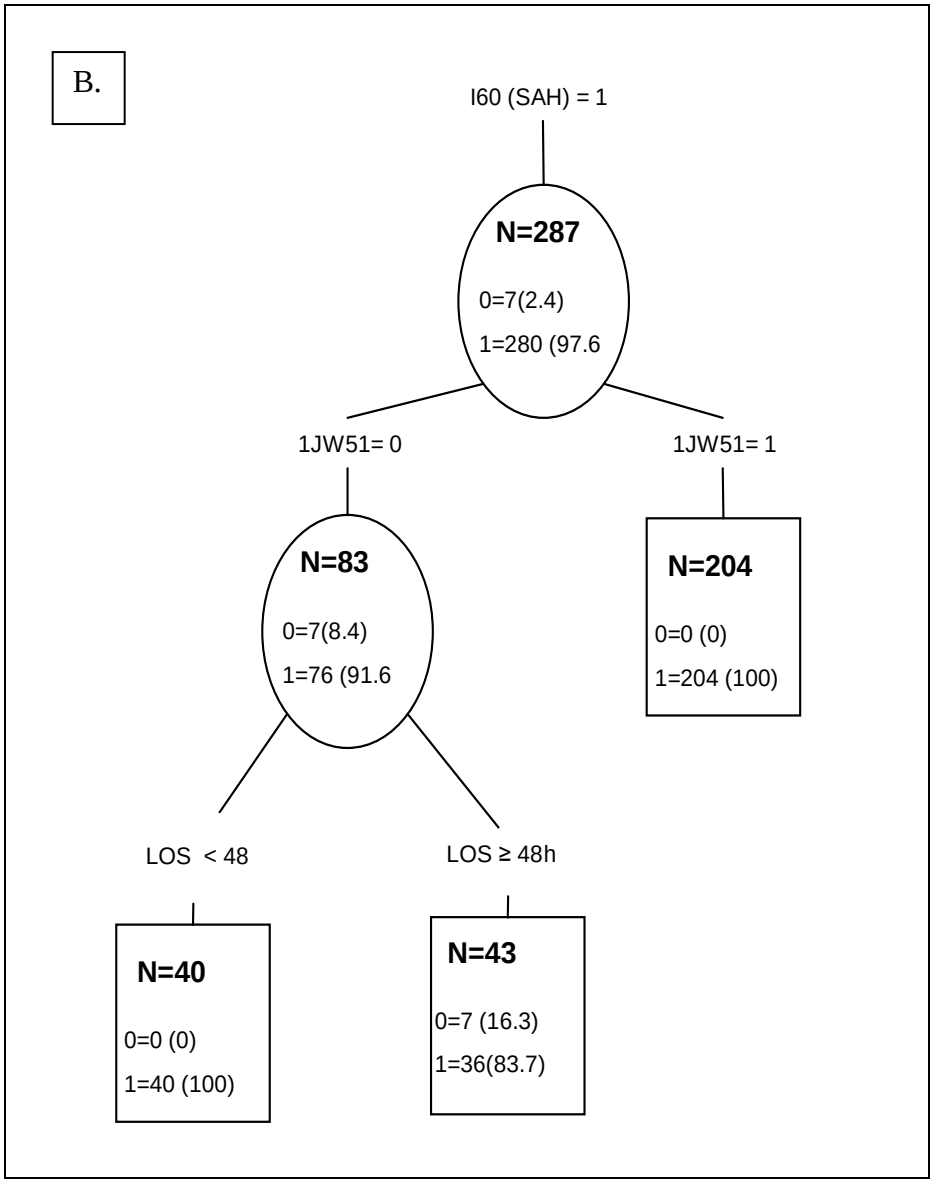
6.7.4.1 Recursive Partitioning based on Chi-square Statistics

The same predictor variables used to build the logistic regression model were used to create the recursive partitioning model. From the 2x2 chi-square tables, the variable with the highest chi-square value meeting the alpha significance of <.05 was entered into the model. For variables where less than 5 observations populated any cell in the 2x2 table, the Fisher's Exact Test was used. A total of 10 variables entered the model, creating 12

splits and 13 terminal nodes. Figure 15 A and B displays the tree created using this model of recursive partitioning. Each terminal node is comprised of patients with varying proportions truly with and without the disease. The performance characteristics of the Chi-square RP model using the validation set is displayed in Figure 15 and listed inTable 17. The arbitrary cut-offs of ≥50%, ≥75%, and ≥90% represent the proportion of patients within a terminal node who truly have the disease that is required for all observations within that node to be classified as aSAH/avmSAH.

Figure 15: Recursive Partitioning Based on Chi Square Analysis





6.7.4.2 Binomial Recursive Fractioning

The same derivation and validation sets, as well as predictor variables used in the logistic and chi-squared RP models were used in the binomial recursive fractioning. The GINI method of improving heterogeneity amongst resultant cells of a split was used for this model since the results had more clinical validity. With this method, 10 predictor variables were used to create splits, resulting in 11 terminal nodes again comprised of patients truly with and without the disease. Figure 16 displays the resulting tree.

The performance characteristics of this model using the validation set are presented in Table 17 using the same cut-points as previously described.

Figure 16: Recursive Partitioning Prediction Model for aSAH/avmSAH

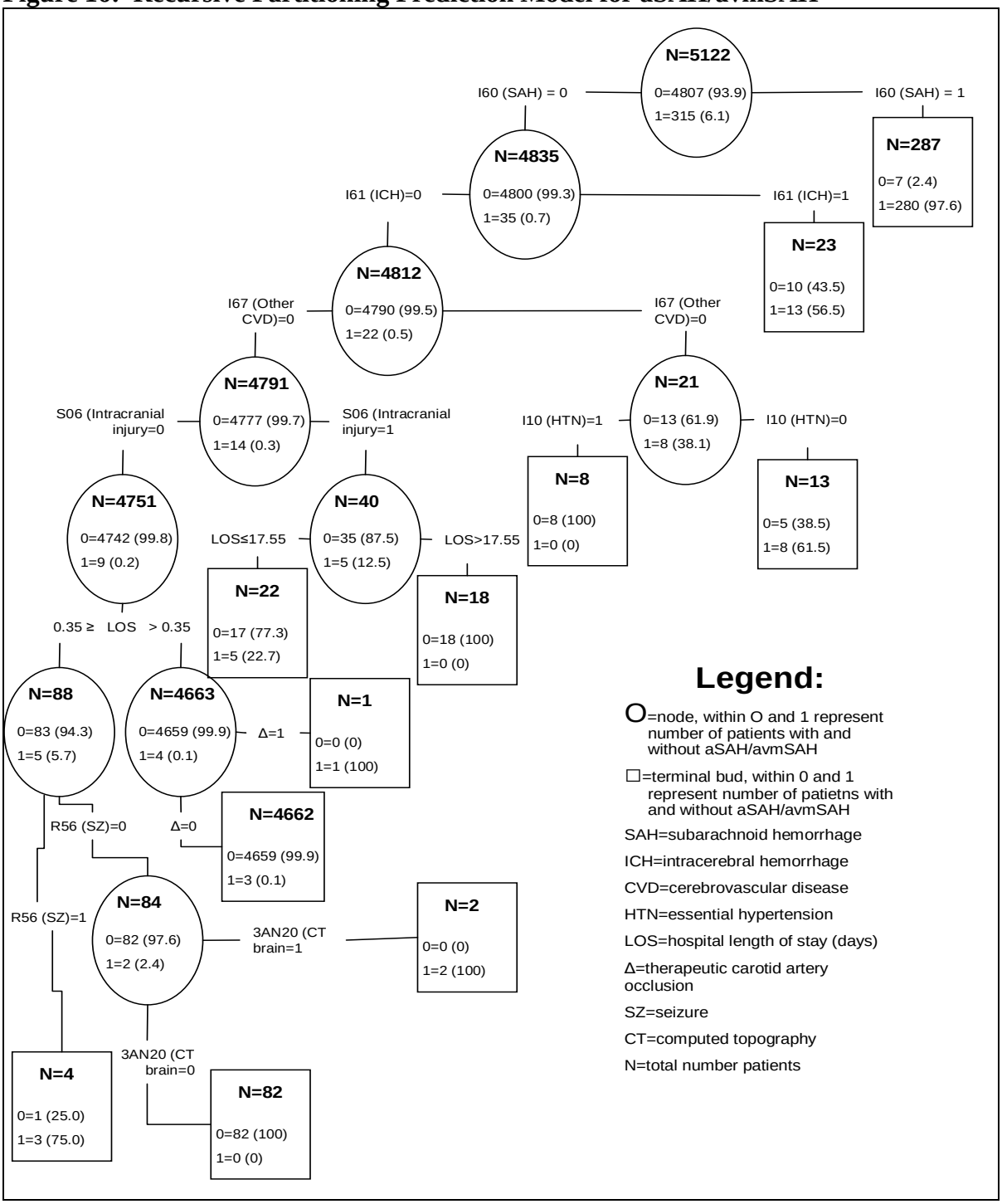


Table 17: Predictive Models Performance Characteristics Using Validation Set

MODEL:	PREDICTED PROBA-BILITY CUT-POINT	SENSITI-VITY (%)	95% CI	SPECIFI-CITY (%)	95% CI	PPV (%)	95% CI	NPV (%)	95% CI	LR+	95% CI	LR-	95% CI
Logistic Regression	≥50%	96.8	(94.3-98.3)	99.7	(99.5-99.8)	95.3	(92.4-97.1)	99.8	(99.6-99.9)	312	(190-523)	0.03	(0.02-0.06)
	≥75%	94.0	(90.8-96.1)	99.7	(99.5-99.8)	95.8	(93.0-97.5)	99.6	(99.4-99.8)	353	(205-608)	0.06	(0.04-0.09)
	≥90%	81.3	(76.7-85.2)	99.9	(99.7-99.9)	97.7	(95.1-99.0)	98.8	(98.5-99.1)	662	(297-1475)	0.19	(0.15-0.24)
Recursive Partitioning (ChiSq)	≥50%	96.5	(93.9-98.0)	99.8	(99.6-99.9)	96.8	(94.3-98.3)	99.8	(99.6-99.9)	483	(254-876)	0.04	(0.02-0.06)
	≥75%	96.5	(93.9-98.0)	99.8	(99.6-99.9)	96.8	(94.3-98.3)	99.8	(99.6-99.9)	483	(254-876)	0.04	(0.02-0.06)
	≥90%	75.6	(70.6-80.0)	100	(99.9-100)	99.6	(97.7-99.9)	98.5	(98.1-98.8)	3694	(520-26244)	0.24	(0.20-0.30)
Recursive Partitioning (CART)	≥50%	96.5	(93.9-98.0)	99.2	(98.9-99.4	88.7	(84.9-91.6)	99.8	(99.6-99.9)	121	(88-165)	0.04	(0.02-0.06)
	≥75%	91.5	(87.9-94.1)	99.9	(99.7-99.9)	98.0	(95.6-99.1)	99.4	(99.2-99.6)	744	(334-1657)	0.09	(0.06-0.12)
	≥90%	89.6	(85.7-92.5)	99.9	(99.7-99.9)	97.6	(95.1-98.8)	99.3	(99.1-99.5)	625	(298-1311)	0.11	(0.08-0.14)

6.7.5 Best Prediction Model

The objective of the prediction model was to accurately identify all patients with aSAH/avmSAH from all patient encounters at TOHDW. As such, the sensitivity of the tool is of primary importance. FromTable 17, it is clear that using a predicted probability cut-point of ≥50% yields similar sensitivities across the 3 models. Specificity is marginally better with both recursive partitioning models, but the model derived using recursive partitioning based on chi-square statistics with a predicted probability of ≥50% had the highest positive likelihood ratio and sensitivity. This is the strongest model to predict the probability of 1° SAH using administrative data.

With this model, given a pre-test probability of aSAH/avmSAH of approximate 16.7/10,000 TOH admissions (see section 6.6), the probability of a patient actually having aSAH/avmSAHif deemed positive by the model is 45%.

7 Discussion

7.1 Thesis Accomplishments

This work accomplishes several goals. First, a detailed search algorithm based on the accepted diagnostic criteria for 1°SAH[8] (including imaging, CSF results, and post-mortem examination) was derived and validated. This search algorithm was implemented on a dataset that included all hospital admissions and emergency room encounters over a 9 year period to identify a complete cohort of aSAH/avmSAH patients. The validity of this cohort was verified by chart review and cross-referenced using several checks and balances including previous patient registries and diagnostic coding (ICD10 code for SAH). The hospital incidence and case mortality rate of 1° SAH was calculated. Finally, using this cohort as a gold standard reference, a predictive model was derived and validated that accurately determines whether a particular patient encounter at TOH was due to an aSAH/avmSAH based on the characteristics available in the discharge abstract (routinely collected administrative data).

To my knowledge, this study is the first of its kind in the world to use such a rigorous case-defining method to examineincidence of aSAH/avmSAH. In fact, no other retrospective study identifying such patients exists in the literature. Previous retrospective work has all been based primarily on diagnostic coding as the case-defining method.[2]ABritish study by Poberskin et al.[18] used multiple overlapping methods(including diagnostic codes, imaging results, and operative datasets) to identify their cohort but this cohort was limited to a highly selected group of patients– namely those undergoing either angiography or aneurysm repair surgery, respectively. Patients who did not survive to the point of securing their aneurysm and those with inaccurate diagnostic coding would have

been missed - even though the diagnosis of aSAH was likely made based on antemortem imaging +/- CSF analysis. Only patients identified by cause of death would thus have been included, but disparities in documentation of death certificates is well known[58,59], limiting its utility. Further, any surviving miscoded patient who did not undergo intervention would have been missed. It is essential to include in epidemiologic studies as results would otherwise be biased in representing only those that survived long enough meet inclusion. Further any assessment of outcome would be largely understated.

7.2 Challenges with Studying "Rare" Disease

Cohorts derived retrospectively using diagnostic codes from administrative databases are subject to misclassification bias.[52] The degree of misclassification can only be determined if the probability of a patient with the code actually having the disease is known. As demonstrated in this work, this problem is especially prominent with rare diseaseswith low prevalence in the population. Even if the sensitivity and specificity of a diagnostic code (or algorithm) is high, the probabilitythat a particular patient with the disease code (or algorithm) truly has the disease may remain surprisingly low.[60]Sincethe positive predictive value of a test (or in this case the diagnostic code) is affected by the prevalence of the outcome in the population/sample in which it was derived, it should not be used to determine post-test probability in a population/sample whose prevalence of disease is distinct from the population used to calculate the positive predictive value.[28,60]

Instead, post-test probability should be measured using functions of sensitivity and specificity with the positive likelihood ratio. For example, let's assume the number of actual cases of aSAH in Canada in 1991 was 2592 (based on incidence of 9.6/100,000/yr,as previously described[2], with a population of 27 million(ref: http://www.statcan.gc.ca/c1996-

r1996/4129974-eng.htm accessed April 2013)). Given approximately 4 million Canadian hospital admissions in 1991[61], the proportion due to aSAH is 0.0006. This is the pre-test probability that a person admitted to hospital had a 1° SAH. Let's also assume that the sensitivity and specificity of the ICD code for aSAH is 98% and 92% respectively (the only published performance measures of this code - see Table 3)[43].Based on this, the positive likelihood ratio for the ICD code for aSAH is 12.25. With these data, the post-test probability that someone with an ICD code for aSAH truly having the disease is only .0074. Therefore, any admitted patient with an ICD code for aSAHhas a probability ofonly 0.74% of actually having the diagnosis. Given this, we estimate that only 16 of the 2186 people identified by Ostbye et al.[2]with aSAH in 1991-1992 actually had the diagnosis. This leads to question if then more than 2000 true cases were missed (if the incidence is in fact 9.6/100,000/year) of if perhaps incidence has just been over-estimated altogether.

7.3 Benefit of a Prediction Model Using Multiple Variables

Diagnostic codes cannot be discounted altogether in retrospective case-ascertainment. As demonstrated in this study, case-ascertainment that includes diagnostic coding as part of a strategy may improve accuracy. By examining ICD codes, we were able to identify an additional 21 cases for inclusion in our cohort that would have otherwise been missed. A large proportion of these cases were missed because the diagnosis was made on imaging performed at another institution and thus highlight a weakness in our original proposed algorithm (based solely on imaging, CSF analysis and post-mortems) rather than a flaw.

We demonstrated the use of diagnostic codes as part of a case-ascertainment strategy in our prediction model. Minimization of misclassification bias was achieved by

including multiple variables in both a logistic regression model and recursive partitioning model. In doing so, patients with aSAH/avmSAH who had been mislabeled by another diagnostic code were still considered for inclusion in the cohort. Specifically, the diagnostic codes for visual disturbances, coma, and hemiplegia were considered for entry into the models as they have previously been reported to be common misdiagnoses for stroke in the literature.[39] Further, not having a discharge abstract diagnostic code for aSAH/avmSAH was not sufficient alone to be excluded from the cohort, thereby improving the sensitivity of the search strategy. This effectively provides for overlapping search strategies, which has been demonstrated in the literature to improve accuracy of case ascertainment in other disease processes including all stroke, osteoporosis[62] and acute kidney injury.

7.4 Hospital Incidence and Case Fatality Rate

The finding of hospital incidence of aneurysmal/arteriovenous malformation SAH of 16.7/10,000 admissions or 61 to 80 cases per year was significantly lower than expected. Based on available data, we had anticipated almost 50% more cases. Given the rigour of the protocol that we utilized to identify patients in our cohort and the validation steps used to ensure the vast majority of cases were included, we think it unlikely that this difference is due to missed cases. Rather, we believe that this further strengthens the argument that previous incidence estimates are over-inflated by flawed case-ascertainment methodologies. Indeed, when diagnostic code case-ascertainment strategies for all stroke have been compared with prospective population stroke registries gathered from the same patient population, the former overestimate SAH incidence.[35,36] Although the methods employed

here were also retrospective, the more rigorous overlapping methods are likely more accurate than the previously published data based purely on diagnostic codes.

Interestingly, although the hospital case-fatality rate of 18% demonstrated in our cohort is again less than expected, it is similar to that reported in the previous Canadian SAH epidemiologic study[2]. Possible explanations for this low rate include: 1) misclassification error in the vital status in the datasets; 2) measured at time of discharge - mortality rate may have been higher if it were measured at a further time point (eg: 30 days or 3 months) and would include any death that occurred at another institution that the patient may have been transferred to; or 3) mortality has declined over the years with improved therapeutics and monitoring techniques. In the very recent study by Zimmerman and colleagues[22], they demonstrated a hospital CFR of only 9% with a 77% percent decline since 1988. Although there was no obvious decline in CFR over the nine years included in our cohort, the decreased mortality demonstrated in Zimmerman's study was less notable after 1995, an era not included in the current study.

7.5 Study Strengths

The strengths of this study lies in its rigourous methodology. Each of the components of the search algorithm were independently derived and validated. Further, the results of the subsequent search strategy that they served to compose were validated using chart review to confirm the diagnosis of aSAH/avmSAH. The validity was furthered by cross-referencing with other datasets likely to include potentially missed patients. It is unlikely that there exist further aSAH/avmSAH patients at TOH, during our study period, which were not identified by our search strategy. Given this, the cohort used to derive the

predictive model was complete. Three separate strategies were used to derive the best predictive model and were validated using different datasets.

7.6 Study Limitations

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. Systematic flaws and bias may not be entirely controlled for in a particular algorithm and in fact this algorithm may compensate for such possible systematic errors withinTOHDW or its data acquisition. Given that the search algorithm relied heavily on report text from imaging and autopsy and the predictive model on characteristics contained within the discharge abstract, the element of human error cannot be completely controlled for. For instance each image or autopsy report is generated by a clinician, often using voice recognition software. The discharge abstract is prepared by professional coders. An example of potential omissions from error is best displayed in one of the false negatively labeled patients who indeed had suffered from an aSAH. There were repeated grammatical errors in their radiology reports that ended in the report failing to be identified as having described SAH. The diagnostic code for SAH had been applied to the encounter and thus with the checks and balances of this study, this patient was identified and included in the cohort. It remains possible that other such errors not caught by the checks and balances exist. Further, systematic errors that may exist that have been compensated for by the algorithm may limit the generalizability of the prediction model to be used at other centres. Its utility and validity at other centres using different datasets needs to be tested.

The search algorithm utilized here included CSF analysis results performed within 24 hours of the start of the encounter. Although this approach is unlikely to miss a positive

result occurring in a patient who presents with SAH, it is possible that a patient already admitted to hospital for other reasons who subsequentlydevelops a SAH that is diagnosed only by CSF analysis (ie: no SAH on cranial imaging) would not have been included in our cohort. If however, the SAH was picked up on cranial imaging such a patient would still have been identified with our algorithm. The former example is thought to be too exceedingly rare to affect the results, but one must at least consider the possibility.

The 1˚ SAH Search Algorithm derived and validated in this work was used to identify a 9 year cohort of patients suffering from this diagnosis. Following chart review validation and the other additional checks and balances as described, a total of 22 cases of true 1˚ SAH were not identified by the search algorithm (false negatives). All were ICD coded as 1˚ SAH. The majority of cases missed (10 cases) were a result of patients having been transferred to TOH with imaging having either already been completed prior to transfer or the patient dying prior to further imaging being completed at TOH. In addition to 2 cases which were missed secondary to spelling/dictation errors in their imaging reports (described above), a further 9 cases were missed as no blood was described in the subarachnoid space specifically, even though they had indeed suffered a from a ruptured intra-cranial aneurysm. This highlighted the need for multiple overlapping case-defining methods including the use of ICD coding, such as the predictive model achieved from this work, which facilitated the identification of these cases.

7.7 Conclusion

In conclusion, this thesis has derived and validated a search strategy that accurately identifies all patients with aSAH/avmSAH presenting to TOH over a 9 year period resulting in a hospital incidence of 16.7/10,000 hospital admissions. The case-fatality rate at

hospital discharge for the cohort was 18.2%. The cohort identified using this algorithm was used to derive and validate a prediction tool that accurately identifies patients with aSAH/avmSAH based on characteristics contained within the discharge abstract. The positive likelihood ratio of the tool is 483, signifying in this patient population that a patient encounter at TOH predicted positive by the tool has an almost 1 in 2 chance of being related to an aSAH/avmSAH.

7.8 Future Work

Further work remains to be completed from this project. The validity of the search algorithm and predictive model needs to be tested in other settings with other administrative datasets. The real potential of being able to complete a reliable and rigorous population-based study on incidence of aSAH/avmSAH exists with the use of the predictive model should its validity withstand the test of generalizing to other datasets.

8 Supervisors and Their Roles:

8.1 Dr. Carl van Walraven:

As a scientist with the Ottawa Hospital Research Institute and leading expert in administrative database research, Dr. van Walraven served the dual role of methodologist and advisor in administrative database research.

8.2 Dr. Lauralyn McIntyre:

As a scientist with the Ottawa Hospital Research Institute and trained and practicing Critical Care Physician, Dr. McIntyre served the dual role of methodology guidance and content expert.

9 Category of Thesis:

This thesismeets criteria for two categories of thesis: Methodological Issues and Secondary Analysis of Existing Datasets. The former is fulfilled by the complex data collection involving the development of a detailed instrument to identify the SAH patient cohort. The latter was fulfilled with the analysis of both basic epidemiologic data as well as complex prediction modeling.

10 Ethics Review:

Expedited ethics review by the Ottawa Hospital Ethics Review Board was undertaken for the completion of these goals. Work was carried out under review number 2011370-01H.

11 Acknowledgements:

It is with the utmost gratitude that I would like to thank my supervisors, mentors, colleagues and friends Drs. Carl van Walraven and Lauralyn McIntyre. Your leadership, guidance and direction as well as unwavering support over not only my thesis but career is more appreciated than you will ever know.

I would like to thank all of the very talented people at the Ottawa Hospital Data Warehouse and ICES centre. Your willingness to help sort out the difficulties I continued to find myself in was truly appreciated. I remain incredibly grateful to my thesis advisory committee whose mentorship is invaluable to me; and to my clinical colleagues who have supported my research ambitions from the beginning.

Last, but most importantly, I would like to thank my ever supporting and caring family who has sacrificed most of all in this endeavour. This work is a testament to the support and strength they provide.

12 References:

1. Bonita R, Beaglehole R, North JD. Subarachnoid hemorrhage in New Zealand: an epidemiological study. Stroke: A Journal of Cerebral Circulation [Internet] 1983;14(3):342–7.

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; a journal of cerebral circulation 1997;28(4):793–8.

3. Knekt P, Reunanen a, Aho K, et al. Risk factors for subarachnoid hemorrhage in a longitudinal population study. Journal of Clinical Epidemiology 1991;44(9):933–9.

4. Gijn J Van, Rinkel GJE. Subarachnoid haemorrhage : diagnosis , causes and management. 2001;249–78.

5. Findlay JM. Current management of aneurysmal subarachnoid hemorrhage guidelines from the Canadian Neurosurgical Society. The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques 1997;24(2):161–70.

6. Smith M. Intensive care management of patients with subarachnoid haemorrhage. Current Opinion in Anaesthesiology 2007;20(5):400–7.

7. 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; a journal of cerebral circulation 2009;40(3):994–1025.

8. 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; a journal of cerebral circulation 2012;43(6):1711–37.

9. Hop JW, Rinkel GJ, Algra A, Van Gijn J. Case-fatality rates and functional outcome after subarachnoid hemorrhage: a systematic review. Stroke; a journal of cerebral circulation 1997;28(3):660–4.

10. Wong GKC, Poon WS, Chan MT V, et al. Intravenous magnesium sulphate for aneurysmal subarachnoid hemorrhage (IMASH): a randomized, double-blinded, placebo-controlled, multicenter phase III trial. Stroke; a journal of cerebral circulation 2010;41(5):921–6.

11. Keyrouz SG, Diringer MN. Clinical review: Prevention and therapy of vasospasm in subarachnoid hemorrhage. Critical Care 2007;11(4):220.

12. Dankbaar JW, Slooter AJ, Rinkel GJ, Schaaf IC Van Der. Effect of different components of triple-H therapy on cerebral perfusion in patients with aneurysmal subarachnoid haemorrhage: a systematic review. Critical care (London, England) 2010;14(1):R23.

13. Feigin VL, Lawes CMM, Bennett D a, Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. The Lancet2009;8(4):355–69.

14. Ingall T, Asplund K, Mähönen M, Bonita R. A multinational comparison of subarachnoid hemorrhage epidemiology in the WHO MONICA stroke study. Stroke: A Journal of Cerebral Circulation 2000;31(5):1054–61.

15. De Rooij NK, Linn FHH, Van der Plas JA, Algra A, Rinkel GJE. Incidence of subarachnoid haemorrhage: a systematic review with emphasis on region, age, gender and time trends. Journal of neurology, neurosurgery, and psychiatry 2007;78(12):1365–72.

16. Sandvei MS, Mathiesen EB, Vatten LJ, et al. Incidence and mortality of aneurysmal subarachnoid hemorrhage in two Norwegian cohorts, 1984-2007. Neurology 2011;77(20):1833–9.

17. Lovelock CE, Rinkel GJE, Rothwell PM. Time trends in outcome of subarachnoid hemorrhage: Population-based study and systematic review. Neurology 2010;74(19):1494–501.

18. Pobereskin LH. Incidence and outcome of subarachnoid haemorrhage: a retrospective population based study. Journal of Neurology, Neurosurgery & Psychiatry 2001;70(3):340–3.

19. Epidemiology of aneurysmal subarachnoid hemorrhage in Australia and New Zealand: incidence and case fatality from the Australasian Cooperative Research on Subarachnoid Hemorrhage Study (ACROSS). Stroke; a journal of cerebral circulation2000;31(8):1843–50.

20. Williams GR, Jiang JG, Matchar DB, Samsa GP. Incidence and Occurrence of Total (First-Ever and Recurrent) Stroke G. Rhys Williams, John G. Jiang, David B. Matchar and Gregory P. Samsa Stroke 1999;30;2523-2528.

21. Nieuwkamp DJ, Setz LE, Algra A, Linn FHH, De Rooij NK, Rinkel GJE. Changes in case fatality of aneurysmal subarachnoid haemorrhage over time, according to age, sex, and region: a meta-analysis. Lancet neurology 2009;8(7):635–42.

22. Zimmerman JE, Kramer A a, Knaus W a. Changes in hospital mortality for United States intensive care unit admissions from 1988 to 2012. Critical care 2013;17(2):R81.

23. Longstreth WT, Nelson LM, Koepsell TD, Van Belle G. Clinical course of spontaneous subarachnoid hemorrhage: A population-based study in King County, Washington. Neurology 1993;43(4):712–712.

24. Gross CR, Kase CS, Mohr JP, Cunningham SC, Baker WE. Stroke in south Alabama: incidence and diagnostic features--a population based study. Stroke; a journal of cerebral circulation 1984;15(2):249–55.

25. Labovitz DL, Halim AX, Brent B, Boden-Albala B, Hauser WA, Sacco RL. Subarachnoid hemorrhage incidence among Whites, Blacks and Caribbean Hispanics: the Northern Manhattan Study. Neuroepidemiology 2006;26(3):147–50.

26. Pobereskin LH. Incidence and outcome of subarachnoid haemorrhage: a retrospective population based study. Journal of Neurology, Neurosurgery & Psychiatry 2001;70(3):340–3.

27. Molyneux AJ, Kerr RSC, Yu L, et al. International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups, and . Lancet 2005;366:809–17.

28. Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. Critical care (London, England) 2004;8(6):508–12.

29. Moore DF, Lix LM, Yogendran MS, Martens P, Tamayo A. Stroke surveillance in Manitoba, Canada: estimates from administrative databases. Chronic diseases in Canada 2008;29(1):22–30.

30. Mayo NE, Chockalingam A, Reeder BA, Phillips S. Surveillance for stroke in Canada. Health reports / Statistics Canada, Canadian Centre for Health Information = Rapports sur la santé / Statistique Canada, Centre canadien d'information sur la santé 1994;6(1):62–72.

31. Mayo NE, Goldberg MS, Levy AR, Danys I, Korner-Bitensky N. Changing rates of stroke in the province of Quebec, Canada: 1981-1988. Stroke; a journal of cerebral circulation 1991;22(5):590–5.

32. ICD-10 10th Ed, Volume 2: Instruction Manual, WHO [Internet]. Available from: http://apps.who.int/classifications/icd10/browse/Content/statichtml/ICD10Volume2_en_2010.pdf

33. International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) Version for 2010 [Internet]. Available from: http://apps.who.int/classifications/icd10/browse/2010/en#/

34. Manuel DG, Rosella LC, Stukel T a. Importance of accurately identifying disease in studies using electronic health records. Bmj 2010;341(aug19 1):c4226–c4226.

35. Stegmayr B, Asplund K. Measuring stroke in the population: quality of routine statistics in comparison with a population-based stroke registry. Neuroepidemiology 1992;11(4-6):204–13.

36. Ellekjaer H, Holmen J, Krüger O, Terent A. Identification of incident stroke in Norway: hospital discharge data compared with a population-based stroke register. Stroke; a journal of cerebral circulation1999;30(1):56–60.

37. Leibson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. Stroke 1994;25(12):2348–55.

38. Reker DM, Hamilton BB, Duncan PW, Yeh SC, Rosen A. Stroke: who's counting what? Journal of rehabilitation research and development;38(2):281–9.

39. Liu L, Reeder B, Shuaib A, Mazagri R. Validity of stroke diagnosis on hospital discharge records in Saskatchewan, Canada: implications for stroke surveillance. Cerebrovascular diseases (Basel, Switzerland);9(4):224–30.

40. Phillips S, Cameron K, Chung C. Stroke surveillance revisited. Canadian Journal of Cardiology 1993;9(Suppl D):124–5.

41. Mayo N, Danys I, Carleton J. Validity of hospital discharge diagnosis of stroke. Canadian Journal of Cardiology 1993;9(Suppl D):121–4.

42. Mayo N, Danys I, Korner-Bitensky N, Carleton J, Wood-Dauphinee S. Determinants of Acute-care length of stay and discharge destination of stroke patients in the Montreal area.

43. Tirschwell DL. Validating Administrative Data in Stroke Research. Stroke 2002;33(10):2465–70.

44. 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; a journal of cerebral circulation 1998;29(2):415–21.

45. 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; a journal of cerebral circulation 1999;30(4):736–43.

46. Vermeulen M, Gijn J Van. The diagnosis of subarachnoid haemorrhage. 1990;365–72.

47. Cruickshank AM. Revision of national guidelines for cerebrospinal fluid analysis in suspected subarachnoid haemorrhage. Annals of clinical biochemistry 2008;45(Pt 3):236–7.

48. Tofteland ND, Salyers WJ. Subarachnoid hemorrhage. Hospital Physician 2007;(May):31–41.

49. Molyneux A, Kerr R, Stratton I, et al. International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised trial. Lancet 2002;360(9342):1267–74.

50. Van Den Bergh WM, Algra a, Van Kooten F, et al. Magnesium sulfate in aneurysmal subarachnoid hemorrhage: a randomized controlled trial. Stroke: A Journal of Cerebral Circulation 2005;36(5):1011–5.

51. 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 : Canadian Medical Association journal = journal de l'Association medicale canadienne 2012;184(1):37–42.

52. 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. Journal of clinical epidemiology 2010;63(12):1332–41.

53. Oake N, Taljaard M, Van Walraven C, Wilson K, Roth V, Forster AJ. The effect of hospital-acquired Clostridium difficile infection on in-hospital mortality. Archives of internal medicine 2010;170(20):1804–10.

54. Perry JJ, Stiell IG, Sivilotti ML a, et al. High risk clinical characteristics for subarachnoid haemorrhage in patients with acute headache: prospective cohort study. BMJ 2010;341:c5204–c5204.

55. Flaherty ML, Haverbusch M, Kissela B, et al. Perimesencephalic subarachnoid hemorrhage: incidence, risk factors, and outcome. Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association 2005;14(6):267–71.

56. Edlow JA, Bruner KS, Horowitz GL. Xanthochromia. Archives of pathology & laboratory medicine 2002;126(4):413–5.

57. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological methods 2009;14(4):323–48.

58. Iso H, Jacobs DR, Goldman L. Accuracy of death certificate diagnosis of intracranial hemorrhage and nonhemorrhagic stroke. The Minnesota Heart Survey. American journal of epidemiology 1990;132(5):993–8.

59. Corwin LE, Wolf PA, Kannel WB, McNamara PM. Accuracy of death certification of stroke: the Framingham Study. Stroke; a journal of cerebral circulation ;13(6):818–21.

60. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. American journal of epidemiology 1978;107(1):71–6.

61. Canadian Institute for Health Information. Hospital Trends in Canada — Results of a Project to Create a Historical Series of Statistical and Financial Data for Canadian Hospitals Over Twenty-Seven Years. 2005. Available from: https://secure.cihi.ca/free_products/Hospital_Trends_in_Canada_e.pdf

62. Lix LM, Yogendran MS, Leslie WD, et al. Using multiple data features improved the validity of osteoporosis case ascertainment from administrative databases. Journal of clinical epidemiology 2008;61(12):1250–60.

Signatures:

1. MSc Candidate:

_____ _____

Dr. Shane English **Date:**

2. Thesis Supervisors:

_____ _____

Dr. Carl van Walraven **Date:**

_____ _____

Dr. Lauralyn McIntyre **Date:**

Appendix 1: Sample Size

The following sample size calculations makes the following assumptions:

- Hospital incidence of SAH is estimated to be 116 per 50,000 admissions. This is based on 2007 data of 116 ICD-10CA related discharge diagnoses amongst 50,113 admissions over the year (not published).
- Goal sensitivity is 95% with margin of error of 2.5%

Sample Size Calculation based on Sensitivity:

Sensitivity = (true positive)/(true positive + false negative) = 95%

95% CI: sensitivity ± margin of error (ME) where: n = # patients without disease

ME = $1.96 \sqrt{[(\text{sensitivity})(1-\text{sensitivity})/n]}$

Then: $n = (\text{sensitivity})(1\text{-sensitivity})(Z_{1-(\alpha/2)}/ME)^2$

$= (0.95)(0.05)(1.96/0.025)^2$

$= 292$

Therefore, to ensure a minimum sensitivity of 95% ± 2.5%, our sample must include ≥ 300 true SAH cases. Given gross estimate of 116 cases per year, our sample must include a minimum of 3 years to achieve this sensitivity. This is ensured by the proposed 8.5 years to be included in the study.

The number of screen negative patients required to review to ensure true proportion of false negative rate lies ≤1% was calculated using a Poisson distribution. Table 1 shows the corresponding upper 95% confidence limit of this proportion given 'x' number of cases reviewed an 'y' number of true SAH cases found. These are calculated using an exact method for a confidence interval of a mean of a Poisson distribution by the following formula:[25] Given the event that four true SAH cases are found after reviewing 1000 screen negative cases, the upper confidence limit of the true proportion of false negatives is remains 1%.

Table 18: Upper Confidence Limits (95%) of the Differing Proportions of False Negative Screened Patients

# SCREEN NEGATIVES REVIEWED (x)	# OF PEOPLE IN A WITH A SAH (y)	PREVALENCE OF SAH IN REVIEWED SCREEN NEGATIVE POPULATION	UPPER LIMIT OF 95% CI*of SAH PREVALENCE IN SCREEN NEGATIVE POPULATION
100	0	0	0.036889
500	0	0	0.0073778
1000	0	0	0.0036889
5000	0	0	0.00073778
10000	0	0	0.00036889
1000	0	0	0.0036889
1000	1	0.001	0.0055716
1000	2	0.002	0.0072247
1000	3	0.003	0.0087673
1000	4	0.004	0.0102416
1000	5	0.005	0.0116683
1000	10	0.01	0.0183904
1000	20	0.02	0.0308884

*Poisson Distribution Calculator reference: http://statpages.org/confint.html (accessed March 16, 2011)

Appendix 2: List of potential cranial imaging report identifiers (stmWIDs) within the OHDW

9325	*CT HEAD
8049	*CTHEAD RES/STROKE CLASS
7765	*CT HEAD RES/STROKE ANCROD
7766	*CT HEAD RES/STROKE ATLANTIS
7767	*CT HEAD RES/STROKE BAY X37-02
387	*CT HEAD RESEARCH/CATIE
7764	*CT HEAD RES/STROKE FISSBIS
436	*CT HEAD RESEARCH/MHIP
437	*CT HEAD RESEARCH/RISPERIDONE
438	*CT HEAD CONTRAST
434	*CT HEAD W/O CONTRAST
7744	*CT HEAD VASC DEMENTIA
310	*CT NECK W&W/O CONTRAST
441	*CT NECK CONTRAST
440	*CT NECK W/O CONTRAST
435	*CT HEAD RESEARCH
7625	*CT HEAD COM RES MDC
439	*CT COMPLEX HEAD W/O CONTRAST
445	*CT COMPLEX HEAD W&W/O CONTRAS
7808	CT HEAD RES/STROKE FISSBISS
7730	CT HEAD VASC DEMENTIA
26499	CT PERFUSION
386	CT HEAD
35	CT HEAD
36	CT NECK
15014	CT HEAD RESEARCH
15575	CT HEAD W CONTRAST
24188	RES. CT Head /w Contrast G179
24187	RES. CT Neck /w Contrast G179
15964	CT HEAD
24834	CT-perfusion versus PET study
23215	CT BRAIN
23058	CT neck w/contrast
23305	CT HEAD
24392	RES CT Head /w Contrast 195
24417	RES CT Head /w Contrast 194
24896	RES CT Neck /w Contrast 215
14962	CT head contrast
15446	CT ANGIOGRAPHY
366	*SP ANGIO CEREB (NSEL-FC)
360	*SP ANGIO CEREB (NSEL-NFC)
363	*SP ANGIO CEREB (SEL-NFC-MAX 4
357	*SP ARTERIOGRAM BILATERAL
355	*SP ARTERIOGRAM UNILATERAL
10126	*ANGIOGRAM X1
10127	*ANGIOGRAM X2
10128	*ANGIOGRAM X3
7799	*SP ANEURYSM COILING/CEREBRAL

82	*SP ANGIO CEREBRAL
9351	ANGIO CEREBRAL X1
9356	ANGIO CEREBRAL X1 / ARCH
9352	ANGIO CEREBRAL X2
9357	ANGIO CEREBRAL X2 / ARCH
9353	ANGIO CEREBRAL X3
9358	ANGIO CEREBRAL X3 / ARCH
9354	ANGIO CEREBRAL X4
9359	ANGIO CEREBRAL X4 / ARCH
9355	ANGIO CEREBRAL X5
9360	ANGIO CEREBRAL X5 / ARCH
9476	IV DSA CEREBRAL
10434	ANGIO

Appendix 3: List of potential CSF result identifiers (stmWIDs) within the OHDW

4860	*CEREBROSPINAL FLUID ELECTROPH
8760	CSF Albumin
5214	CSF Glucose
7458	CSF Protein and Glucose
8757	CSF IgG
7342	CSF Lactate
7345	CSF LD
6050	CSF Total Protein
10301	CSF Albumin
6621	*VOLUME CSF
4774	Appearance - CSF
4813	Colour - CSF
4857	Spinal fluid count
4858	CSF differential
4901	CSF Differential
5651	Lymphocytes - CSF
5813	Neutrophils - CSF
5805	NRBC - CSF
5880	Others - CSF
5027	RBC - CSF
7967	CSF Tube # - Hematology
5622	WBC - CSF
6822	Xanthochromia - CSF
5675	CSF testing
4859	*CEREBRO-SPINAL FLUID CYTOLOGY
6821	*XANTHOCHROMIA BODY FLIUD

Appendix 4: List of potential Autopsy report identifiers (stmWIDs) within the OHDW

4069	*AUTOPSY SPECIMEN PROCEDURE
4264	*AUTOPSY BRAIN SPEC
4407	*NEURO AUTOPSY
4527	*AUTOPSY ATTENDING WORKLOAD
4532	*NUMBER OF AUTOPSY
4535	*AUTOPSY REVIEW PREPARATION
4538	*FINAL AUTOPSY ADDENDUM
4543	*FINAL AUTOPSY REPORT
5064	*FORENSIC AUTOPSY SPECIMEN
5787	*NEURO AUTOPSY STAINS
6972	Provisional Autopsy Consult
6973	Provisional Autopsy Comments
14565	FINAL DIAGNOSES (Text)
6974	Provisional Autopsy Diagnosis
14594	Final Autopsy Report
6975	Provisional Autopsy Report
14595	Cardiovascular Autopsy Report
14598	Neuropathology Autopsy Report
25680	Anatomical Pathology Report (Cyto/Patho/Autopsy)-Procedure