

Validation of ICD-10-CM Diagnosis Codes for Identification of Patients with Acute Hemorrhagic Stroke in a National Health Insurance Claims Database

This article was published in the following Dove Press journal:
Clinical Epidemiology

Meng-Tsang Hsieh^{1-3,*}
Kuo-Chang Huang^{4,*}
Cheng-Yang Hsieh^{5,6}
Tzu-Tung Tsai¹
Li-Ching Chen¹
Sheng-Feng Sung⁷⁻⁹

¹Stroke Center and Department of Neurology, E-Da Hospital, Kaohsiung, Taiwan; ²School of Medicine for International Students, College of Medicine, I-Shou University, Kaohsiung, Taiwan; ³Institute of Clinical Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwan; ⁴Division of Neurosurgery, Department of Surgery, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan; ⁵Department of Neurology, Tainan Sin Lau Hospital, Tainan, Taiwan; ⁶School of Pharmacy, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan; ⁷Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan; ⁸Department of Information Management and Institute of Healthcare Information Management, National Chung Cheng University, Chiayi County, Taiwan; ⁹Department of Nursing, Min-Hwei Junior College of Health Care Management, Tainan, Taiwan

*These authors contributed equally to this work

Correspondence: Sheng-Feng Sung
Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan
Tel +886 5 276 5041 Ext 7283
Fax +886 5 278 4257
Email richard.sfsung@gmail.com

Purpose: The performance of the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes for identifying acute hemorrhagic stroke in Taiwan's National Health Insurance claims database has not been assessed. This study aimed to construct and validate the case definitions for acute hemorrhagic stroke based on ICD-10-CM diagnostic codes.

Patients and Methods: From January 2018 to December 2019, all inpatient records with ICD-10-CM code of I60 or I61 in any field of the discharge diagnoses were retrieved from the hospitalization claims data and all hospitalizations with a final diagnosis of subarachnoid hemorrhage (SAH) or intracerebral hemorrhage (ICH) were identified from the stroke registry databases. The clinical diagnosis in the stroke registry was treated as the reference standard. For hospitalizations not recorded in the stroke registry, manual review of the medical records and images was done to ascertain the diagnosis. The positive predictive value (PPV) and sensitivity of various case definitions for acute hemorrhagic stroke were estimated.

Results: Among the 983 hospitalizations, 860, 111, and 12 were determined to be true-positive, false-positive, and false-negative episodes of acute hemorrhagic stroke, respectively. The PPV and sensitivity of the ICD-10-CM codes of I60 or I61 for identifying acute hemorrhagic stroke were 88.6% and 98.6%, respectively. The PPV increased to 98.2%, whereas the sensitivity decreased to 93.1% when acute hemorrhagic stroke was defined as hospitalizations in which the primary diagnosis field contained I60 or I61. Hemorrhagic transformation of ischemic stroke and concomitant cerebrovascular diseases other than SAH or ICH were the main reasons for a false-positive and false-negative diagnosis of acute hemorrhagic stroke, respectively.

Conclusion: This study demonstrated the performance of ICD-10-CM codes for identifying acute hemorrhagic stroke and may offer a reference for future claims-based stroke studies.

Keywords: administrative claims data, diagnosis, ICD-10-CM, hemorrhagic stroke

Introduction

Hemorrhagic stroke, including nontraumatic subarachnoid hemorrhage (SAH) and intracerebral hemorrhage (ICH), is the second most common type of stroke, with a global lifetime risk of 8.6% among adults aged over 25 years.¹ Despite a worldwide decline in death rates from all kinds of stroke since 1990, hemorrhagic stroke accounts for the greatest burden of stroke among adults aged between 20 and

64 years,² particularly in the low- and middle-income countries.³ Meanwhile, even in the high-income countries, either early or long-term case-fatality rates due to ICH have not always improved over the past decades.^{4,5} Therefore, more clinical research is needed to optimize the treatment and thereby improve the outcomes of patients with hemorrhagic stroke.

Administrative claims data, which are derived from electronic data collected by health-care providers and insurers for the purpose of health-care management, have nowadays provided an affordable and efficient way in conducting research on stroke⁶ or other cardiovascular diseases.⁷ For example, Taiwan's National Health Insurance (NHI) claims database, one of the largest comprehensive administrative claims databases in the world,⁸ has been widely used in stroke research. A bibliometric analysis found that stroke was the second most common research topic in studies using this database.⁹ However, most of the stroke studies have focused on ischemic stroke rather than hemorrhagic stroke.⁹ A potential explanation for this may be that researchers are less confident in performing research on hemorrhagic stroke because unlike an ischemic stroke, the validity of the diagnosis codes for ascertaining hemorrhagic stroke has never been assessed in Taiwan's NHI claims database.⁸

Therefore, this study aimed to construct and validate case definitions for hemorrhagic stroke based on the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes using claims data from two major stroke centers in Taiwan.

Patients and Methods

Study Settings

This study was conducted in the two NHI-contracted hospitals in southern Taiwan: the Ditmanson Medical Foundation Chia-Yi Christian Hospital, a regional teaching hospital with approximately 650 stroke admissions annually, and the E-Da Hospital, a would-be medical center with approximately 1000 stroke admissions annually. The stroke centers of both hospitals have participated in the Taiwan Stroke Registry (TSR) Program¹⁰ for more than a decade, and their quality of stroke care is certified by the Joint Commission of Taiwan. The study protocol was independently approved by the Institutional Review Board of the Ditmanson Medical Foundation Chia-Yi Christian Hospital (IRB2020118) and the Institutional Review Board of the E-Da Hospital (EMRP-109-013).

The requirement for informed consent was waived due to the retrospective design. The study data were kept with confidentiality to ensure the privacy of the study participants. This study was conducted in accordance with the Declaration of Helsinki.

Data Sources and Record Linkage

This study used stroke registry data and administrative claims data from both hospitals. The stroke registry of each hospital prospectively registered consecutive patients hospitalized for stroke within 10 days of symptom onset according to the design of the TSR.¹⁰ Patient demographics, comorbidities, personal and medical history, stroke severity as assessed by the National Institutes of Health Stroke Scale, treatments received, hospital course, final diagnoses, and functional status as assessed using the modified Rankin Scale were collected. The hospitalization claims data reported to the Administration of NHI were retrieved from the hospital information system. Both data sources were linked through patients' unique identifier and the date of admission.

Study Sample

Patients discharged between January 2018 and December 2019 with a final diagnosis of acute hemorrhagic stroke, including SAH and ICH, were identified from the stroke registry. Even though Taiwan switched to ICD-10-CM coding in 2016, this study period was chosen since the E-Da Hospital did not enroll all patients hospitalized for stroke into the stroke registry before January 2018. Those with an in-hospital stroke were retained. Patients with traumatic intracranial hemorrhage were excluded from the study, except those cases with concomitant acute hemorrhagic stroke and traumatic intracranial hemorrhage. Patients with missing information regarding date of birth, sex, or personal identifier were also excluded from the study. Multiple hospitalizations for the same stroke episode were combined into a single hospitalization. Conversely, multiple hospitalizations for the same patient due to different stroke episodes were counted as individual hospitalizations. Similarly, all patients discharged during the same period with an ICD-10-CM diagnosis code of I60 or I61 in any field of the discharge diagnoses were identified from the hospitalization claims data. Hospitalization records from either data sources were merged to form the study sample (Figure 1A).

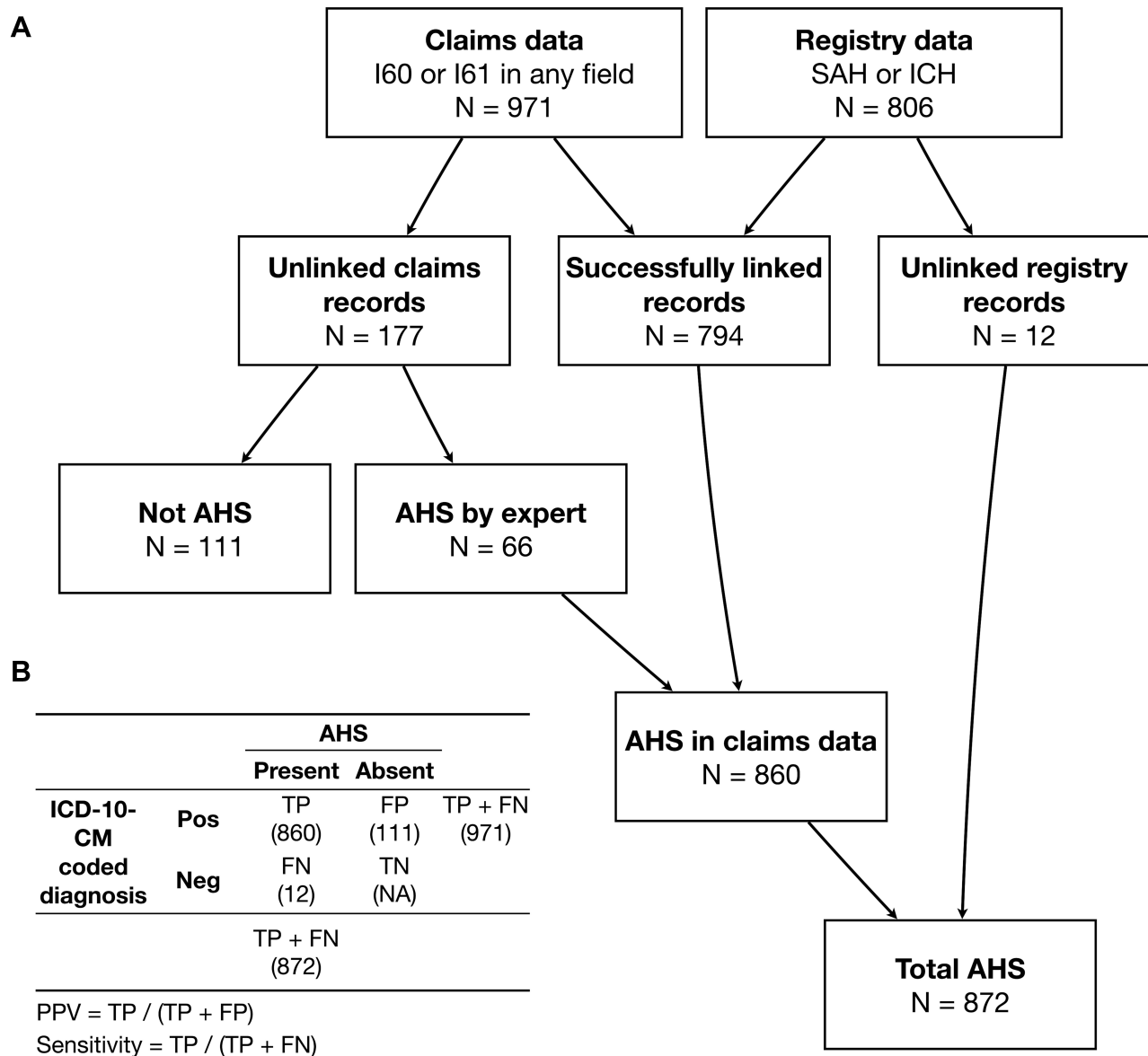


Figure 1 Process of case ascertainment (A) and calculation of the PPV and sensitivity (B).

Abbreviations: AHS, acute hemorrhagic stroke; FN, false negative; FP, false positive; ICD-10-CM, International Classification of Diseases, Tenth Revision, Clinical Modification; ICH, intracerebral hemorrhage; Neg, negative; Pos, positive; PPV, positive predictive value; SAH, subarachnoid hemorrhage; TN, true negative; TP, true positive.

Ascertainment of Acute Hemorrhagic Stroke

Figure 1A shows the process of case ascertainment. According to the TSR program,¹⁰ SAH was defined as

characteristic clinical manifestations of neurologic deficit, usually with sudden onset of severe headache and either subarachnoid blood on brain computed tomography or/and non-traumatic bloody (or xanthochromic) cerebrospinal fluid with or without confirmation of the existence of aneurysm by magnetic resonance angiography or conventional cerebral angiography.

ICH was defined as

non-traumatic abrupt onset of symptoms with relevant focal neurological deficit with or without headache or altered level of consciousness with a focal collection of blood within the brain parenchyma on computed tomography or magnetic resonance imaging that was not a hemorrhagic conversion of a cerebral infarction.

Cases of SAH or ICH must meet these criteria before being entered the stroke registry.

The clinical diagnosis in the stroke registry was treated as the reference standard because the TSR used several

quality assurance processes to ensure data quality, including web-based examination and on-site auditing.¹⁰ Any hospitalization in the claims data that could be successfully linked to one of the hospitalizations in the registry data was considered a true episode of acute hemorrhagic stroke. For hospitalizations in the claims data that could not be linked to any of the hospitalizations in the registry data, their electronic medical records and brain images were reviewed separately in each hospital by an experienced stroke neurologist (SFS in the Chia-Yi Christian Hospital and MTH in the E-Da Hospital) to determine whether these were true episodes of acute hemorrhagic stroke.

For hospitalizations determined not to be true episodes of acute hemorrhagic stroke, the reasons were further categorized into the following groups: (1) hemorrhagic transformation of acute ischemic stroke (AIS); (2) subacute, chronic, or remote SAH or ICH (presenting more than 10 days after symptom onset); (3) subdural hemorrhage; (4) a tentative diagnosis of acute hemorrhagic stroke, which was later excluded after clinical evaluation and imaging studies; and (5) other diagnoses (eg, brain tumor with hemorrhage and neonatal intracranial hemorrhage).

Data Analyses

This study examined various case definitions for identifying cases of acute hemorrhagic stroke including (1) ICD-10-CM code of I60 or I61 as the primary diagnosis; (2) I60 or I61 as the primary or first secondary diagnosis; (3) I60 or I61 as the primary, first secondary, or second secondary diagnosis; and (4) I60 or I61 in any field of the diagnosis. Similarly, case definitions for identifying subtypes of acute hemorrhagic stroke were investigated, including the ICD-10-CM codes of I60 and I61 for identifying cases of SAH and ICH, respectively.

The positive predictive value (PPV) was defined as the proportion of hospitalization records with an ICD-10-CM coded diagnosis for which the hospitalization was confirmed to be a true episode of acute hemorrhagic stroke (Figure 1B). Sensitivity was defined as the proportion of hospitalizations due to acute hemorrhagic stroke matching the ICD-10-CM coded diagnosis in the administrative claims data (Figure 1B). PPVs and sensitivities were calculated and their 95% confidence intervals (CIs) for binomial proportions were estimated using the Clopper–Pearson exact method. For subgroup analysis, whether patient age, sex, and medical comorbidities, including

hypertension and diabetes, would affect the validity of ICD-10-CM coding in the claims data was additionally examined. PPVs and sensitivities between stroke subtypes and subgroups were compared using the chi-square test. A two-tailed *p* value of <0.05 was considered to be statistically significant. Data analyses were performed using Stata 15.1 (StataCorp, College Station, Texas).

Results

From January 2018 to December 2019, a total of 806 distinct hospitalizations due to acute hemorrhagic stroke were identified from the registry data. All hospitalization records contained valid information regarding date of birth, sex, and personal identifier and were thus eligible for record linkage. During the same period, 971 distinct hospitalizations with an ICD-10-CM code of I60 or I61 in any field of the discharge diagnoses were identified from the hospitalization claims data. Among them, 161 and 865 hospitalizations were coded with I60 and I61, respectively; that is, 55 hospitalizations were coded with both I60 and I61. The linking of hospitalization records between both data sources (Figure 1A) resulted in 794 successfully linked hospitalizations, which were regarded as true-positive episodes of acute hemorrhagic stroke. Among the remaining 177 hospitalizations in the claims data that could not be linked to any record in the registry data, 66 were considered true-positive episodes whereas 111 were determined to be false-positive episodes of acute hemorrhagic stroke after manual review. In addition, 12 hospitalizations from the registry data could not be identified in the claims data and were thus regarded as false-negative episodes of acute hemorrhagic stroke. Consequently, the PPV and sensitivity of ICD-10-CM diagnosis of acute hemorrhagic stroke were 88.6% (95% CI, 86.4–90.5%) and 98.6% (95% CI, 97.6–99.3%).

Table 1 shows the validity of various case definitions to identify cases of acute hemorrhagic stroke as a whole or cases of SAH or ICH separately. The case definition of using the ICD-10-CM code of I60 or I61 as the primary diagnosis to identify acute hemorrhagic stroke yielded a PPV and sensitivity of 98.2% and 93.1%, respectively. This case definition missed a proportion of acute hemorrhagic stroke hospitalizations because conditions other than hemorrhagic stroke were coded as the primary diagnosis. Table 2 summarizes the reasons for these missed cases. Among them, the most common reason was concomitant cerebrovascular diseases other than SAH or ICH (*n* = 22), followed by infection or inflammation (*n* = 12).

Table 1 Validity of Various Case Definitions for Identifying Cases of Acute Hemorrhagic Stroke, SAH, and ICH

| Case Definitions | TP | FP | FN | PPV (95% CI), % | Sen (95% CI), % |
|---|-----|-----|----|------------------|------------------|
| Acute hemorrhagic stroke (SAH or ICH) | | | | | |
| I60 or I61 as primary diagnosis | 812 | 15 | 60 | 98.2 (97.0–99.0) | 93.1 (91.2–94.7) |
| I60 or I61 as primary or first secondary diagnosis | 843 | 68 | 29 | 92.5 (90.6–94.2) | 96.7 (95.3–97.8) |
| I60 or I61 as primary, first secondary, or second secondary diagnosis | 856 | 79 | 16 | 91.6 (89.6–93.3) | 98.2 (97.0–98.9) |
| I60 or I61 in any field of diagnosis | 860 | 111 | 12 | 88.6 (86.4–90.5) | 98.6 (97.6–99.3) |
| SAH | | | | | |
| I60 as primary diagnosis | 123 | 7 | 27 | 94.6 (89.2–97.8) | 82.0 (74.9–87.8) |
| I60 as primary or first secondary diagnosis | 135 | 14 | 15 | 90.6 (84.7–94.8) | 90.0 (84.0–94.3) |
| I60 as primary, first secondary, or second secondary diagnosis | 141 | 16 | 9 | 89.8 (84.0–94.1) | 94.0 (88.9–97.2) |
| I60 in any field of diagnosis | 142 | 19 | 8 | 88.2 (82.2–92.7) | 94.7 (89.8–97.7) |
| ICH | | | | | |
| I61 as primary diagnosis | 708 | 17 | 83 | 97.7 (96.3–98.6) | 89.5 (87.2–91.6) |
| I61 as primary or first secondary diagnosis | 761 | 48 | 30 | 94.1 (92.2–95.6) | 96.2 (94.6–97.4) |
| I61 as primary, first secondary, or second secondary diagnosis | 772 | 56 | 19 | 93.2 (91.3–94.9) | 97.6 (96.3–98.5) |
| I61 in any field of diagnosis | 785 | 80 | 6 | 90.8 (88.6–92.6) | 99.2 (98.4–99.7) |

Abbreviations: CI, confidence interval; FN, false negative; FP, false positive; ICH, intracerebral hemorrhage; PPV, positive predictive value; SAH, subarachnoid hemorrhage; Sen, sensitivity; TP, true positive.

Table 2 Reasons for False-Negative Episodes of Acute Hemorrhagic Stroke When I60 or I61 as the Primary Diagnosis Was Used to Identify Acute Hemorrhagic Stroke (N = 60)

| Reason | n |
|---|----|
| Concomitant cerebrovascular diseases other than SAH or ICH ^a | 22 |
| Infection or inflammation | 12 |
| Malignancy | 7 |
| Concomitant traumatic accident | 6 |
| Consciousness change | 2 |
| Extracranial aneurysm, dissection, or injury | 2 |
| Heart diseases | 2 |
| Miscellaneous ^b | 7 |

Notes: ^aCerebrovascular diseases other than SAH or ICH, including cerebral aneurysm with or without embolization (n=8), acute ischemic stroke without hemorrhagic transformation (n=5), unspecified nontraumatic intracranial hemorrhage (n=3), arteriovenous malformation (n=3), subdural hemorrhage (n=2), and personal history of stroke (n=1); ^bMiscellaneous, including seizure or epilepsy (n=3), hydrocephalus (n=2), peripheral artery occlusive disease (n=1), and head deformity (n=1).

Abbreviations: ICH, intracerebral hemorrhage; SAH, subarachnoid hemorrhage.

In contrast, when acute hemorrhagic stroke was defined as I60 or I61 in the first three fields of discharge diagnoses, the sensitivity increased to 98.2% with a drop of PPV from 98.2% to 91.6%. When ICD-10-CM codes were used to identify cases of SAH or ICH alone, the values of PPV were lower for SAH than for ICH without reaching statistical significance (*p* values ranging from 0.053 to 0.314). The sensitivity values for SAH were significantly lower than those for ICH (*p* values ranging from <0.001 to 0.017). The sensitivity was as low as 82.0% when I60 as

the primary diagnoses was used to ascertain the diagnosis of SAH.

Table 3 shows the reasons for the false-positive episodes of acute hemorrhagic stroke. More than half of the episodes were due to hemorrhagic transformation of AIS. Around one-fifth of them were due to subacute, chronic, or remote SAH or ICH. Another one-fifth of the false-positive episodes received other diagnoses (Table 4) but were miscoded as an acute hemorrhagic stroke. Of the 12 false-negative episodes of acute hemorrhagic stroke, 4 were coded as I67.1 (cerebral aneurysm, non-ruptured), 1 was coded as I62.9 (nontraumatic intracranial hemorrhage, unspecified), and 7 had diagnoses codes unrelated to cerebrovascular diseases (from I60 to I61).

Table 5 shows the results for the subgroup analyses that compared the PPVs and sensitivities between subgroups according to patient age, sex, and the presence of hypertension

Table 3 Reasons for False-Positive Episodes of Acute Hemorrhagic Stroke (N = 111)

| Reason | n (%) |
|---|-----------|
| Hemorrhagic transformation of AIS | 59 (53.2) |
| Subacute, chronic, or remote SAH or ICH | 24 (21.6) |
| Subdural hemorrhage | 1 (0.9) |
| Acute hemorrhagic stroke as a tentative diagnosis but excluded by studies | 3 (2.7) |
| Other diagnoses miscoded as acute hemorrhagic stroke | 24 (21.6) |

Abbreviations: AIS, acute ischemic stroke; ICH, intracerebral hemorrhage; SAH, subarachnoid hemorrhage.

Table 4 Final Diagnoses of Miscoded Cases (N = 24)

| Diagnosis | n |
|--|----|
| Brain tumor with hemorrhage | 12 |
| Neonatal intracranial hemorrhage | 3 |
| Intracranial aneurysm without hemorrhage | 2 |
| Traumatic intracranial hemorrhage | 1 |
| Cerebral venous thrombosis with hemorrhagic infarction | 1 |
| AIS | 1 |
| Brain abscess | 1 |
| Subdural empyema | 1 |
| Ventriculomegaly | 1 |
| Ventriculitis | 1 |

Abbreviation: AIS, acute ischemic stroke.

or diabetes. The PPV in the young subgroup (91.7%) was significantly higher than that in the elderly group (84.9%), while the sensitivities were comparable between subgroups. The PPV and sensitivity for patients without hypertension were significantly lower than for those with hypertension.

Discussion

This study assessed the validity of various case definitions of ICD-10-CM diagnoses of acute hemorrhagic stroke as a whole, SAH alone, or ICH alone. The PPVs for SAH or ICH were generally comparable to those reported by similar studies that examined the validity of ICD codes for diagnosing SAH or ICH in other administrative databases.^{11–15} According to a systematic review of validation studies of ICD codes for acute stroke, the PPVs for SAH and ICH were $\geq 93\%$ and $\geq 89\%$, respectively, in at least half of the studies.¹¹ In more recent studies from the Czech

Republic,¹² South Korea,¹³ Denmark,¹⁴ and Australia,¹⁵ using the primary diagnosis code only, the PPVs ranged from 85% to 91% and sensitivities from 73% to 90% for SAH. For ICH, the PPVs ranged from 75% to 92% and sensitivities from 60% to 85%. However, due to the variation in study settings and reference standards used across studies, a direct comparison of results may be difficult.

This study suggested that, when only the primary diagnosis code was used, the PPV was maximized at the cost of low sensitivity. As the number of diagnosis code fields used for case ascertainment increased, the sensitivity increased as the PPV decreased. The study findings were similar to our previous study that examined the performance of ICD-10-CM codes for identifying AIS in Taiwan's NHI claims database.¹⁶ Nevertheless, discussing the potential implications for researchers interested in hemorrhagic stroke using this claims database is worthwhile. In contrast to AIS, acute hemorrhagic stroke mainly encompasses SAH (I60) and ICH (I61). Among the study sample, more than 5% (55/971) of the hospitalizations due to acute hemorrhagic stroke had both SAH and ICH listed in the discharge diagnoses. The co-occurrence of both conditions forced each condition to compete for the primary diagnosis field, thereby explaining why the PPV and sensitivity using either I60 or I61 as the primary diagnosis to identify acute hemorrhagic stroke were higher than those using I60 as the primary diagnosis to identify SAH or using I61 as the primary diagnosis to identify ICH. Besides, since SAH is usually associated with cerebral aneurysms, I60 was less likely to be coded as the primary diagnosis when the amount of SAH was small and cerebral

Table 5 Subgroup Analyses According to Patient Age, Sex, and the Presence of Hypertension or Diabetes

| Subgroup | TP | FP | FN | PPV (95% CI), % | P | Sen (95% CI), % | P |
|----------------------------|-----|----|----|------------------|--------|-------------------|--------|
| Age | | | | | 0.001 | | 0.887 |
| Elderly (≥ 65 years) | 376 | 67 | 5 | 84.9 (81.2–88.1) | | 98.7 (97.0–99.6) | |
| Young (<65 years) | 484 | 44 | 7 | 91.7 (89.0–93.9) | | 98.6 (97.1–99.4) | |
| Sex | | | | | 0.951 | | 0.352 |
| Male | 532 | 69 | 9 | 88.5 (85.7–91.0) | | 98.3 (96.9–99.2) | |
| Female | 328 | 42 | 3 | 88.6 (85.0–91.7) | | 99.1 (97.4–99.8) | |
| Hypertension | | | | | <0.001 | | <0.001 |
| Yes | 672 | 67 | 3 | 90.9 (88.6–92.9) | | 99.6 (98.7–99.9) | |
| No | 188 | 44 | 9 | 81.0 (75.4–85.9) | | 95.4 (91.5–97.9) | |
| Diabetes | | | | | 0.441 | | 0.130 |
| Yes | 241 | 35 | 1 | 87.3 (82.8–91.0) | | 99.6 (97.7–100.0) | |
| No | 619 | 76 | 11 | 89.1 (86.5–91.3) | | 98.3 (96.9–99.1) | |

Abbreviations: CI, confidence interval; FN, false negative; FP, false positive; PPV, positive predictive value; Sen, sensitivity; TP, true positive.

aneurysms were considered the principal clinical problem. Such a coding convention might lead to a relatively low sensitivity (82.0%) when using I60 as the primary diagnosis to identify SAH.

More than half of the false-positive episodes (53.2%) for the case definition using I60 or I61 in any diagnosis field to identify acute hemorrhagic stroke were attributed to the hemorrhagic transformation of AIS. As the rate of thrombolytic therapy for AIS increases over the years,^{17,18} hemorrhagic transformation of AIS will become more prevalent and may warrant more claims database studies to provide real-world evidence of disease burden and treatment patterns.¹⁷ Nevertheless, no existing ICD-10-CM codes that can be used to code this condition were found. Whether hemorrhagic transformation of AIS can be identified from Taiwan's NHI claims database by the coexistence of the ICD-10-CM codes of acute hemorrhagic stroke and AIS in the discharge diagnoses awaits further investigation. For now, the accuracy of using ICD codes to identify hemorrhagic transformation of AIS remains questionable.^{19,20}

The case definitions tested in this study showed different tradeoffs between the PPV and sensitivity, as shown in a study that investigated various case definitions for stroke in Korean NHI claims data.¹³ In this regard, the study results presented in Table 1 may help researchers choose an ICD-10-CM coding definition that suits their research purposes to identify an acute hemorrhagic stroke. For studies where a validated method for identifying all patients with acute hemorrhagic stroke, SAH, or ICH is crucial, the case definitions that use ICD-10-CM codes in the first three diagnosis fields to identify the corresponding condition are highly recommended. Despite the expense of a drop in PPV values as compared to those case definitions that use only the primary diagnosis, these case definitions had a much higher sensitivity, which is beneficial for enhancing study inclusiveness and ascertaining exposures.²¹ Conversely, case definitions with high PPV that use only the primary diagnosis may be more appropriate when it is important for researchers to identify a "pure" cohort of patients with acute hemorrhagic stroke, SAH, or ICH that need not be representative of every patient with that condition.²¹

Of note, the PPV of ICD coding for acute hemorrhagic stroke in the young subgroup was significantly higher (91.7% vs 84.9%) than that in the elderly group. A previous study on ICD coding for AIS reported a similar finding, which may be explained by the higher

prevalence of comorbidities among the elderly.¹⁶ Compared to young adults, elderly subjects carry a higher risk of hemorrhagic transformation of AIS,^{22,23} and a higher risk of brain tumor.²⁴ Furthermore, elderly people are naturally more likely to have chronic or remote hemorrhagic stroke. All of the above conditions might be erroneously coded as acute hemorrhagic stroke (Tables 3 and 4), leading to a higher false-positive rate for acute hemorrhagic stroke in the elderly subgroup. Besides, the PPV and sensitivity for patients without hypertension were significantly lower than for those with hypertension. Up to 80% of cases of ICH have hypertension,⁴ which is probably the most prevalent and relevant risk factor for acute hemorrhagic stroke. Therefore, for patients without hypertension, bleeding in the brain may be more likely due to etiologies other than acute hemorrhagic stroke (eg, brain tumor with hemorrhage).

This study has several limitations worth mentioning. First, the study design did not allow us to discover patients with acute hemorrhagic stroke who had been missed by both the stroke registry and ICD codes (false negatives). For example, if only ICD codes are used to identify patients for inclusion in the stroke registry, the estimate of false negatives will be low. In fact, patients with stroke were screened using a computer algorithm using both ICD-10 codes and free-text diagnosis fields in the electronic medical records, followed by a manual review by stroke case managers. Therefore, the underestimation of false negatives is likely to be minor in this study. Second, data on the proportion of patients who did not have an acute hemorrhagic stroke and were not coded as such (true negatives) were not available. Therefore, providing the specificity and the negative predictive value is impossible. Third, data from the stroke registry databases were used as the reference standard for the diagnosis of hemorrhagic stroke. Although the registry data were collected by trained study nurses following the standard protocol of the TSR and the data quality was audited,¹⁰ verifying the accuracy of the stroke registry may still be necessary by performing a cross-sectional study. Fourth, hospitalizations that could not be identified in the registry data were manually reviewed by only one stroke neurologist at each site. Finally, the study sample comes from only two stroke centers. They may not be representative of all hospitals contracted with Taiwan's NHI since the validity of diagnostic codes may vary across hospitals.¹³ Further multi-center validation studies may be needed to confirm or compare these findings.

Conclusion

This validation study assessed the validity of various ICD-10-CM definitions for acute hemorrhagic stroke including SAH and ICH in Taiwan's NHI claims database using combined data from two centers. The sensitivity and PPV values varied across different case definitions; therefore, researchers should consider the tradeoffs between the accuracy measures. Thus, the results of this study may help future studies apply appropriate case definitions for acute hemorrhagic stroke that are most suited to their study purposes.

Acknowledgments

This research was funded by the E-Da Hospital (grant number EDAHP107021) and the Ditmanson Medical Foundation Chia-Yi Christian Hospital Research Program (grant number R 109-37-1). The funders of the research had no role in the design and conduct of the study, interpretation of the data, or decision to submit for publication. The authors would like to thank Ms Li-Ying Sung for English language editing and Enago for the English language review.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Feigin VL, Nguyen G, Cercy K; GBD 2016 Lifetime Risk of Stroke Collaborators. Global, regional, and country-specific lifetime risks of stroke, 1990 and 2016. *New Engl J Med*. 2018;379(25):2429–2437. doi:10.1056/nejmoa1804492
2. Krishnamurthi RV, Moran AE, Feigin VL, et al. Stroke prevalence, mortality and disability-adjusted life years in adults aged 20–64 years in 1990–2013: data from the Global Burden of Disease 2013 Study. *Neuroepidemiology*. 2015;45(3):190–202. doi:10.1159/000441098
3. Krishnamurthi RV, Moran AE, Forouzanfar MH, et al. The global burden of hemorrhagic stroke: a summary of findings from the GBD 2010 Study. *Glob Heart*. 2014;9(1):101–106. doi:10.1016/j.gheart.2014.01.003
4. Zahuranec DB, Lisabeth LD, Sánchez BN, et al. Intracerebral hemorrhage mortality is not changing despite declining incidence. *Neurology*. 2014;82(24):2180–2186. doi:10.1212/wnl.0000000000000519
5. Béjot Y, Grelat M, Delpont B, et al. Temporal trends in early case-fatality rates in patients with intracerebral hemorrhage. *Neurology*. 2017;88(10):985–990. doi:10.1212/wnl.0000000000000681
6. Ung D, Kim J, Thrift AG, et al. Promising use of big data to increase the efficiency and comprehensiveness of stroke outcomes research. *Stroke*. 2019;50(5):1302–1309. doi:10.1161/strokeaha.118.020372
7. AlHajji M, Alqahtani F, Alkhouli M. Contemporary trends in the utilization of administrative databases in cardiovascular research. *Mayo Clin Proc*. 2019;94(6):1120–1121. doi:10.1016/j.mayocp.2019.03.022
8. Hsieh C-Y, Su -C-C, Shao S-C, et al. Taiwan's National Health Insurance Research Database: past and future. *Clin Epidemiol*. 2019;11:349–358. doi:10.2147/clep.s196293
9. Sung S-F, Hsieh C-Y, Hu Y-H. Two decades of research using Taiwan's National Health Insurance claims data: bibliometric and text mining analysis on PubMed. *J Med Internet Res*. 2020;22(6):e18457. doi:10.2196/18457
10. Hsieh F-I, Lien L-M, Chen S-T, et al. Get with the guidelines-stroke performance indicators: surveillance of stroke care in the Taiwan stroke registry: get with the guidelines-stroke in Taiwan. *Circulation*. 2010;122(11):1116–1123. doi:10.1161/circulationaha.110.936526
11. McCormick N, Bhole V, Lacaille D, Avina-Zubieta JA. Validity of diagnostic codes for acute stroke in administrative databases: a systematic review. *PLoS One*. 2015;10(8):e0135834. doi:10.1371/journal.pone.0135834
12. Sedova P, Brown RD, Zvolsky M, et al. Validation of stroke diagnosis in the national registry of hospitalized patients in the Czech Republic. *J Stroke Cerebrovasc Dis*. 2015;24(9):2032–2038. doi:10.1016/j.jstrokecerebrovasdis.2015.04.019
13. Park TH, Choi JC. Validation of stroke and thrombolytic therapy in Korean National Health Insurance claim data. *J Clin Neurol*. 2016;12(1):42–48. doi:10.3988/jcn.2016.12.1.42
14. Hald SM, Sloth CK, Hey SM, et al. Intracerebral hemorrhage: positive predictive value of diagnosis codes in two nationwide Danish registries. *Clin Epidemiol*. 2018;10:941–948. doi:10.2147/clep.s167576
15. Ryan OF, Riley M, Cadilhac DA, et al. Factors associated with stroke coding quality: A comparison of registry and administrative data. *J Stroke Cerebrovasc Dis*. 2021;30(2):105469. doi:10.1016/j.jstrokecerebrovasdis.2020.105469
16. Hsieh M-T, Hsieh C-Y, Tsai -T-T, Wang Y-C, Sung S-F. Performance of ICD-10-CM diagnosis codes for identifying acute ischemic stroke in a National Health Insurance claims database. *Clin Epidemiol*. 2020;12:1007–1013. doi:10.2147/clep.s273853
17. Hsieh C-Y, Chen C-H, Chen Y-C, Yang Y-HK. National survey of thrombolytic therapy for acute ischemic stroke in Taiwan 2003–2010. *J Stroke Cerebrovasc Dis*. 2013;22(8):e620–e627. doi:10.1016/j.jstrokecerebrovasdis.2013.07.043
18. Hsieh C-Y, Wu DP, Sung S-F. Trends in vascular risk factors, stroke performance measures, and outcomes in patients with first-ever ischemic stroke in Taiwan between 2000 and 2012. *J Neurol Sci*. 2017;378:80–84. doi:10.1016/j.jns.2017.05.002
19. Mullen MT, Moomaw CJ, Alwell K, et al. ICD9 codes cannot reliably identify hemorrhagic transformation of ischemic stroke. *Circ Cardiovasc Qual Outcomes*. 2013;6(4):505–506. doi:10.1161/circoutcomes.113.000215
20. Uchino K. Can stroke complications be distinguished from comorbid stroke in administrative data? *JAMA Cardiol*. 2018;3(11):1126. doi:10.1001/jamacardio.2018.2903
21. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *J Clin Epidemiol*. 2012;65(3):343–349.e2. doi:10.1016/j.jclinepi.2011.09.002
22. Okada Y, Yamaguchi T, Minematsu K, et al. Hemorrhagic transformation in cerebral embolism. *Stroke*. 1989;20(5):598–603. doi:10.1161/01.str.20.5.598
23. Larrue V, von Kummer R, Muller A, Bluhmki E. Risk factors for severe hemorrhagic transformation in ischemic stroke patients treated with recombinant tissue plasminogen activator. *Stroke*. 2001;32(2):438–441. doi:10.1161/01.str.32.2.438
24. Barnholtz-Sloan JS, Ostrom QT, Cote D. Epidemiology of brain tumors. *Neurol Clin*. 2018;36(3):395–419. doi:10.1016/j.ncl.2018.04.001

Clinical Epidemiology

Dovepress

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>