Validity of Simple Algorithms to Identify Recurrence of Intracerebral Hemorrhage in Two Danish Nationwide Registries

Mie Micheelsen Jensen [ID][1]
Stine Munk Hald [ID][1,2]
Line Marie Buch Kristensen[1]
Nils Jensen Boe [ID][1,2]
Frederik Severin Gråe Harbo [ID][3]
David Gaist [ID][1,2,4]

[1]Department of Neurology, Odense University Hospital, Odense, Denmark; [2]Neurology Research Unit, Department of Clinical Research, University of Southern Denmark, Odense, Denmark; [3]Department of Radiology, Odense University Hospital, Odense, Denmark; [4]Open Patient Data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark

Purpose: Danish registries could be an attractive resource for studies of recurrent intracerebral hemorrhage (re-ICH). We developed and validated algorithms to identify re-ICH in the Danish Stroke Registry (DSR) and the Danish National Patient Registry (DNPR).

Patients and Methods: Using multiple sources, we followed-up an inception cohort with verified first-ever spontaneous ICH (n = 2528) for their first re-ICH in 2009–2018 (study period). We used verified cases of re-ICH (n = 124) as the gold standard to assess the performance of register-based algorithms for identifying re-ICH. For each cohort member, we traced events of re-ICH (ICD-10-code I61) in the study period according to DSR and DNPR, respectively. For each registry, we tested algorithms with a blanking period (BP) – ie, a period immediately following the index ICH during which outcome events were ignored – of varying length (7 days–360 days). The algorithm with the shortest BP that returned a positive predictive value (PPV) of ≥80% was considered optimal. We also calculated negative predictive value (NPV), sensitivity, and specificity of each algorithm and [95% confidence intervals] for all proportions.

Results: The optimal algorithm for DSR (BP 30 days) had a PPV of 89.5% [82.2–94.0], NPV 98.8% [98.2–99.1], sensitivity 75.8% [67.6–82.5], and specificity 99.5% [99.2–99.7]. The optimal algorithm for DNPR (BP 120 days) had a PPV of 80.6% [71.7–87.2], NPV 98.1% [97.5–98.6], sensitivity 63.7% [55.0–71.6], and specificity 99.2% [98.8–99.5].

Conclusion: Simple algorithms accurately identified re-ICH in DSR and DNPR. Compared with DNPR, DSR achieved higher PPV and sensitivity with a shorter BP. The proposed algorithms could facilitate valid use of DSR and DNPR for studies of re-ICH.

Keywords: stroke, recurrent stroke, intracerebral hemorrhage, epidemiology, validity, register-based research

Introduction

Intracerebral hemorrhage (ICH) represents 10–15% of incident strokes in high-income countries, the remaining being primarily due to ischemic stroke.[1] In most cases, ICH is spontaneous, ie, not related to trauma, underlying vascular malformation or tumor. Spontaneous ICH (s-ICH) is largely due to cerebral small vessel disease, predominantly deep perforator arteriopathy (also termed hypertensive arteriopathy) and cerebral amyloid angiopathy.[2] ICH is the deadliest form of stroke.[3] Almost 40% of patients admitted with ICH die within 30 days of admission.[4,5] Additionally, patients surviving ICH face an increased risk of suffering recurrent ICH (re-ICH).[6–8] The risk of re-ICH is highest in the first year after

Correspondence: David Gaist
Department of Neurology, Odense University Hospital, J.B. Winsløwsvej 4, Odense C, 5000, Denmark
Tel +45 65412485
Fax +45 65413389
Email dgaist@health.sdu.dk

the index ICH,[9,10] but remains increased for years thereafter, particularly in lobar ICH.[7] To ensure better long-term prognosis and outcome for survivors of ICH, research on the associated risk factors and causes of re-ICH is greatly warranted. However, due to the relative rarity of the disease, large-scale prospective studies of ICH can be difficult and costly to conduct. Large administrative registries and clinical databases may provide attractive alternative data sources for the study of rare diseases such as ICH. In Denmark, two nationwide registries can be used for research on ICH: The Danish Stroke Registry (DSR)[11] and the Danish National Patient Registry (DNPR).[12]

We recently reported on the high validity of first-ever ICH-diagnoses in DSR and DNPR.[13,14] However, accurate identification of re-ICH in these registries may pose distinct methodological challenges, compared with first-ever ICH. When tracking a patient with an index ICH event (eg, first-ever ICH) in a registry it can be difficult to determine which, if any, of the patient's subsequent admissions coded as ICH represent re-bleeds (re-ICH), as opposed to admissions for causes related to the index ICH (eg, transfers between wards, re-admissions, or admissions for complications to the index ICH). The extent of such misclassification likely varies by source, reflecting the purpose for data collection. DNPR is a registry originally established to serve administrative purposes, and is therefore more likely to be prone to the above-mentioned sources of misclassification of re-ICH, compared with DSR, a regularly audited clinical database of patients admitted with acute stroke.[11] Conversely, it is conceivable that DNPR offers a higher degree of completeness than DSR, as recently demonstrated for first-ever ICH-diagnoses in these registries.[13,14]

The difficulties outlined above could lead to misclassification of the number of ICH events and could ultimately result in overestimation of the burden of re-ICH in the individual patient as well as misclassification of the date of symptom onset (onset date) of re-ICH, thereby affecting risk estimates.[15] Yet to our knowledge, research on the use of registries to identify cases of recurrent stroke is very limited. Only one study[16] from the US has previously reported on the accuracy of registry data for identifying re-ICH ([Supplementary Material and Supplementary Figure S1](#)). Therefore, we conducted this study with the purpose to develop and validate algorithms to identify cases of re-ICH using data from DSR and DNPR; and, further, to assess the accuracy of the onset date of re-ICH identified by the algorithms.

Patients and Methods
Setting and Data Sources

Data for this study were based on hospital admissions for residents of the Region of Southern Denmark (RSD) (1.2 million inhabitants); a geographically well-defined area in Denmark. Within RSD there are four dedicated stroke units as well as a neurosurgical department to which all patients suspected of stroke are principally admitted or transferred.

Data were obtained from two nationwide registries: DSR and DNPR. DSR is a clinical database that was established in 2003 to monitor the quality of stroke care.[11] It is mandatory for all departments involved in stroke care to report standardized data on all admissions for acute stroke to DSR.[11,17] DNPR is an administrative registry, established in 1977, to which all hospitals in Denmark report standardized data, including diagnosis codes according to the International Classification of Diseases, version 10 (ICD-10), on all hospital contacts (inpatient, outpatient and emergency department (ED)).[12]

In accordance with Danish law regarding register-based research, the study was approved by the Region of Southern Denmark (j.nr. 18/51966) and informed consent was waved.[18]

The Gold Standard

In a previous study, we validated cases of first-ever s-ICH among residents of RSD during the period January 1, 2009 through December 31, 2017.[14] We defined s-ICH as ICH not attributable to prior trauma, hemorrhagic transformation of an ischemic stroke, or an alternative explanation (ie, tumor, vascular malformation, or cerebral sinus thrombosis – but not use of antithrombotic drugs).[13,14,19] A total of 2528 patients with verified first-ever s-ICH were included in the inception cohort. Each patient from the inception cohort was followed-up for events of re-ICH by identifying all admissions registered with an ICH-diagnosis (ICD-10 code I61) within DSR and DNPR (inpatient diagnosis, any coding position) from the date of the index ICH through December 31, 2018, which secured at least one year of follow-up. To ensure completeness, we also identified all admissions within DNPR registered with diagnostic codes for a range of other cerebrovascular diseases (ICD-10 codes in parentheses) including subarachnoid hemorrhage (I60.9), subdural hematoma (I62.0 and S06.5), unspecified intracranial hemorrhage (I62.9), acute ischemic

stroke (I63 and I66), unspecified stroke (I64), other specified cerebrovascular disease (I67.8), transient ischemic attack (G45.9), and vascular syndrome (G46). Within DSR, we also identified admissions for ischemic stroke and unspecified stroke. For each of the identified events the diagnosis of s-ICH as well as the onset date of re-ICH were validated by manual review of the full medical record and brain scan reports. Validation was performed by study personnel (two medical students and two physicians) supervised by a neurologist and a radiologist, both with special interest in stroke. The recurrence status of ICH and the onset date of re-ICH ascertained by revision of the medical record and brain scan reports were considered the gold standard.

For patients who suffered more than one verified re-ICH during follow-up, we only included the first of these events in the analysis.

To calculate follow-up from index ICH to verified re-ICH (or to censoring due to emigration, death, or end of study period), we linked our data to information retrieved from the Danish Civil Registration System,[20] an administrative registry with all Danish citizens' unique personal identification number (CPR-number), which is continuously updated with information on vital status and place of residence. The CPR-numbers, also recorded in our data, enabled unambiguous linkage between the sources.

Study data were collected and managed using REDCap electronic data capture tools hosted by Open Patient data Explorative Network (OPEN).[21,22] The data can only be accessed via a secured virtual private network (VPN) and only by approved collaborative partners. The data is not publicly available due to the Danish data protection legislation as the data contains information that could compromise the privacy of the research participants.

The Algorithms

We developed two algorithms with the purpose of identifying cases of re-ICH based exclusively on registry data from DSR and DNPR, respectively.

When using the algorithms, all chronologically consecutive hospital contacts were considered as belonging to a single admission episode if the contacts were not separated by a gap (ie, discharge date of a hospital record corresponded to the admission date of the next hospital record). We excluded episodes that included the admission date of the index ICH, as such admissions are directly related to the index ICH (eg, representing transfers between wards) according to our experience from previous

studies[13,14] and the validation efforts involved in this study.

In DSR, an algorithm-based re-ICH event (Algo-re-ICH) established exclusively through registry information (ie, not verified through medical records) was defined as the first inpatient contact registered with an ICH-diagnosis following the index ICH. Using data from DSR, the algorithm identified events of interest occurring in the inception cohort during the follow-up period from January 1, 2009 through December 31, 2018. The admission dates corresponding to events identified by the algorithm were defined as the onset date of Algo-re-ICH (Figure 1). We iteratively repeated the above process using a blanking period (BP) of 7, 30, 60, 90, 120, 150, 180 and 360 days, respectively. A BP is a time period immediately after an inception event (here index ICH), during which outcome events (here possible admissions for re-ICH) are ignored (Figure 2).[23,24] The underlying assumption when using a BP is that misclassification of recurrent events is to a certain degree time-dependent, ie, more distant admissions are more likely to represent true recurrent events compared with admissions immediately after the index event.

We applied the same approach as outlined above with the algorithm using data from DNPR. In DNPR, however, Algo-re-ICH was defined as the first inpatient contact registered with an ICH-diagnosis in primary coding position following the index ICH.

For each registry, we defined the algorithm with the shortest BP that returned a positive predictive value (PPV) of ≥80% to be optimal.

Statistical Analysis

We assessed the validity of the algorithms by calculating the PPV, negative predictive value (NPV), sensitivity, and specificity with the gold standard as reference. We used Wilson score method to estimate 95% confidence intervals (CI) for all proportions.

We also calculated the proportion of onset dates identified by the algorithms that coincided (same date) or occurred within 1, 7, 30, or 90 days of the gold standard onset date, respectively.[25]

It could be argued that patients who die or emigrate during a given BP should not be included in the analyses as such patients by definition cannot suffer a re-ICH in the follow-up ensuing after the BP in question. Therefore, in a subanalysis, we calculated PPV, NPV, sensitivity and specificity of re-ICH in each registry for the optimal BP
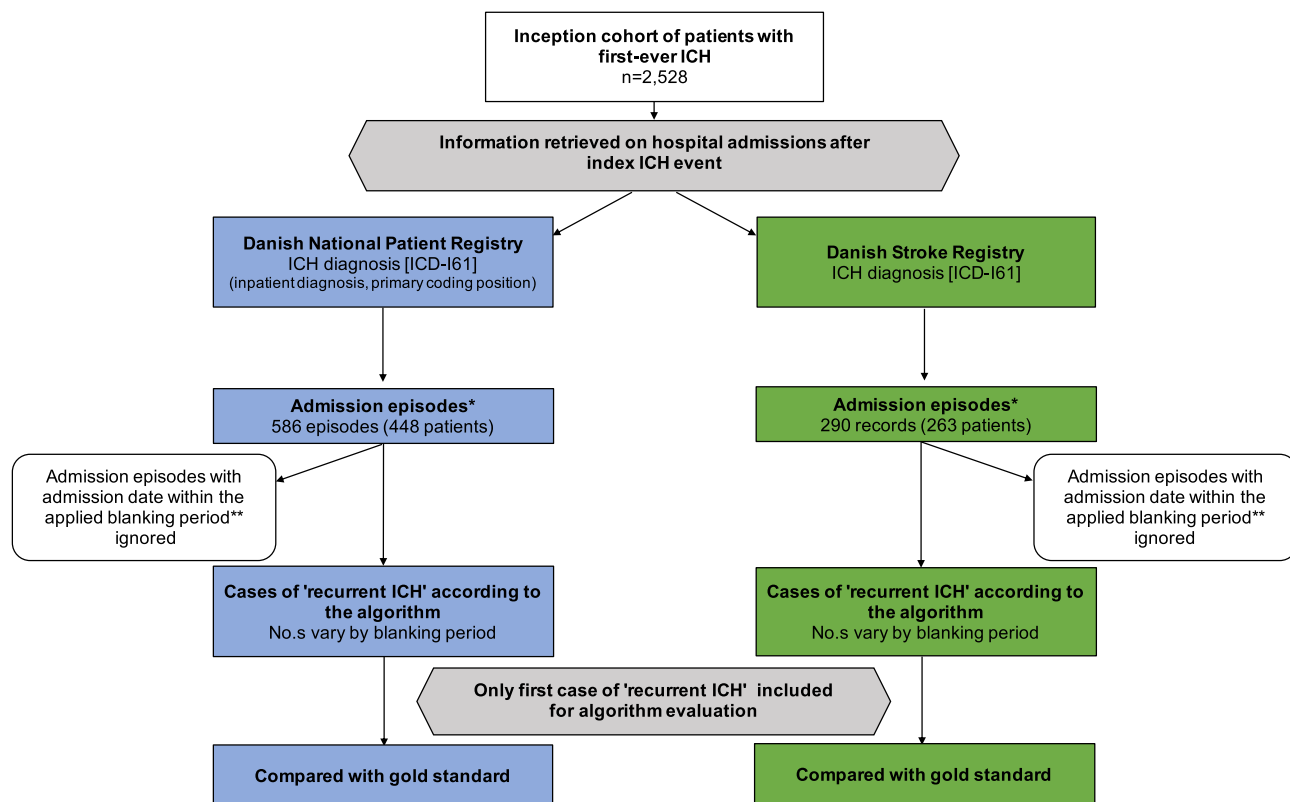
Figure 1 Flowchart illustrating the development and validation of algorithms for identifying cases of recurrent intracerebral hemorrhage using data from the Danish National Patient Registry and the Danish Stroke Registry. *Multiple consecutive records considered as single admission episode, if no gap between discharge date of one record and admission date of the following record. **Period immediately following the index ICH, during which outcome events are ignored; varying lengths (7–360 days) tested. **Abbreviation**: ICH, intracerebral hemorrhage.
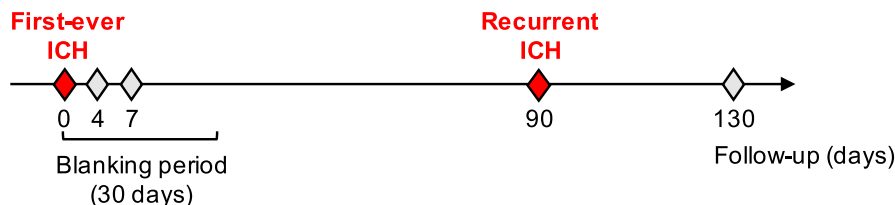

Figure 2 An example of use of an algorithm with a blanking period to identify recurrent intracerebral hemorrhage based exclusively on registry data. In the example, a blanking period of 30 days is used. The patient had her first-ever ICH on day 0. According to registry data (eg, Danish Stroke Registry), she was subsequently readmitted (ie, had inpatient contacts registered with ICD-10 code I61) 4 times for ICH in the study period. When applying the algorithm, the first two re-admissions (day 4 and day 7) were ignored as they occurred during the 30-day blanking period. The first admission occurring after the blanking period (day 90) was classified as her first recurrent ICH event. As follow-up stopped after the first episode of recurrent ICH, the last admission on day 130 was ignored. **Abbreviation**: ICH, intracerebral hemorrhage.

and after excluding patients that had died or emigrated during the BP.

All statistical analyses were done using Stata 16.1 (Stata Corp, TX).

Results

The inception cohort comprised 2528 patients, after exclusion of a small number of patients (n = 18) with insufficient information on admissions during follow-up (mostly untraceable medical records). Re-ICH could be verified in 124

patients of the inception cohort leaving 2404 patients not having a re-ICH during follow-up. The gender distribution and median age were similar between patients when stratified by gold standard re-ICH (Table 1). The mean follow-up after index ICH was 2.4 years (median 1.37 years). However, patients that had a re-ICH were followed up for longer than those with no re-ICH (median follow-up, 4.18 vs 1.21 years). This difference was primarily due to the high short-term all-cause mortality of patients with ICH, as demonstrated by the substantial reduction in the difference in follow-up between

Table 1 Characteristics of the Study Population (n = 2528) Stratified on the Recurrence Status of Intracerebral Hemorrhage According to the Gold Standard[a]

Population Characteristics	ICH Recurrence		No ICH Recurrence	
	n	(%)	n	(%)
Number	124	(100)	2404	(100)
Sex				
Men	61	(49.2)	1247	(51.9)
Women	63	(50.8)	1157	(48.1)
Age, median (IQR)	72.3	(72.2–80.0)	74.5	(64.5–82.9)
Follow-up time in years, median (IQR)				
All follow-up[b]	4.18	(1.75–6.55)	1.21	(0.00–3.89)
Follow-up from day 31 onwards[c]	4.45	(1.87–6.57)	3.75	(1.64–5.51)

Notes: [a]Numbers are n (%) unless otherwise stated. [b]Time from first-ever ICH to the first of the following events: recurrence, emigration, death, or end of study period (December 31, 2018). [c]Follow-up began on day 31 post-ICH and ended as described in footnote above for a total of 1587 patients (122 with ICH recurrence and 1465 with no ICH recurrence).
Abbreviations: ICH, intracerebral hemorrhage; IQR, interquartile range.

patients with re-ICH vs patients with no re-ICH in an analysis restricted to those that survived the first 30 days post-ICH (median follow-up: 4.45 years vs 3.75 years).

We assessed the performance of the algorithms by comparison with the gold standard and evaluated the algorithms' concordance with the gold standard when using varying lengths of BPs (Tables 2 and 3). When testing the algorithms without a BP, they yielded very low PPVs of 37.5% (95% CI = 31.9–43.6) and 23.0% (95% CI = 19.3–27.1) for DSR and DNPR, respectively. The length of the BP was directly associated with PPV (ie, longer BP resulted in higher PPV). However, application of a BP of more than 60 days had little impact on the PPV in DSR (Table 2), but more so in DNPR (Table 3). Even so, in DNPR, application of the maximum BP

of 360 days only improved the PPV by approximately 10% compared with the 60-day BP, and less than 5% compared with the optimal BP (see below). As expected, the length of the BP was inversely associated with sensitivity in both registries with longer BPs incrementally returning lower sensitivities, eg, with the maximum BP of 360 days resulting in sensitivities below 50% for both registries. In both DSR and DNPR, BP-length had little impact on NPV and specificity, which were both high in both registries.

Overall, the algorithm using data from DSR achieved a higher PPV with a shorter BP when compared with the algorithm using data from DNPR. To illustrate, applying a BP of only 7 days returned a PPV of 75.6% (95% CI = 67.4–82.2). Similar results with the algorithm for DNPR (ie,

Table 2 Accuracy of the Algorithm Using Data from the Danish Stroke Registry to Identify Cases of Recurrent Intracerebral Hemorrhage When Applying a Blanking Period of Varying Length[a]

Algorithm Performance in the Danish Stroke Registry				
Length of Blanking Period	Positive Predictive Value % (95% CI)	Negative Predictive Value % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
0 days	37.5 (31.9–43.6)	98.9 (98.3–99.2)	79.0 (71.0–85.3)	93.2 (92.1–94.2)
7 days	75.6 (67.4–82.2)	98.8 (98.3–99.2)	77.4 (69.3–83.9)	98.7 (98.2–99.1)
30 days[b]	**89.5 (82.2–94.0)**	**98.8 (98.2–99.1)**	**75.8 (67.6–82.5)**	**99.5 (99.2–99.7)**
60 days	91.5 (84.1–95.6)	98.4 (97.9–98.9)	69.4 (60.8–76.8)	99.7 (99.3–99.8)
90 days	91.4 (83.9–95.6)	98.4 (97.8–98.8)	68.5 (59.9–76.1)	99.7 (99.3–99.8)
120 days	91.7 (83.8–95.9)	98.1 (97.5–98.6)	62.1 (53.3–70.2)	99.7 (99.4–99.9)
150 days	91.0 (82.6–95.6)	97.8 (97.2–98.3)	57.3 (48.5–65.6)	99.7 (99.4–99.9)
180 days	90.8 (82.2–95.5)	97.8 (97.1–98.3)	55.6 (46.9–64.1)	99.7 (99.4–99.9)
360 days	90.6 (81.0–95.6)	97.3 (96.6–97.9)	46.8 (38.2–55.5)	99.8 (99.5–99.9)

Notes: [a]Period immediately following the index ICH, during which outcome events are ignored. [b]Optimal algorithm marked in bold.

Table 3 Accuracy of the Algorithm Using Data from the Danish National Patient Registry to Identify Cases of Recurrent Intracerebral Hemorrhage When Applying a Blanking Period of Varying Length[a]

Algorithm Performance in the Danish National Patient Registry[b]				
Length of Blanking Period	Positive Predictive Value % (95% CI)	Negative Predictive Value % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
0 days	23.0 (19.3–27.1)	99.0 (98.5–99.3)	83.1 (75.5–88.6)	85.6 (84.2–87.0)
7 days	46.1 (39.6–52.7)	99.0 (98.5–99.3)	81.5 (73.7–87.3)	95.1 (94.2–95.9)
30 days	67.4 (59.3–74.5)	98.9 (98.4–99.2)	78.2 (70.2–84.6)	98.0 (97.4–98.5)
60 days	75.4 (66.9–82.3)	98.5 (98.0–99.0)	71.8 (63.3–78.9)	98.8 (98.3–99.2)
90 days	78.7 (70.1–85.4)	98.4 (97.8–98.8)	68.5 (59.9–76.1)	99.0 (98.6–99.4)
120 days[c]	**80.6 (71.7–87.2)**	**98.1 (97.5–98.6)**	**63.7 (55.0–71.6)**	**99.2 (98.8–99.5)**
150 days	82.4 (73.3–88.9)	98.0 (97.4–98.5)	60.5 (51.7–68.6)	99.3 (98.9–99.6)
180 days	83.5 (74.2–89.9)	97.8 (97.2–98.3)	57.3 (48.5–65.6)	99.4 (99.0–99.7)
360 days	85.5 (75.3–91.9)	97.4 (96.6–97.9)	47.6 (39.0–56.3)	99.6 (99.2–99.8)

Notes: [a]Period immediately following the index ICH, during which outcome events are ignored. [b]Inpatient diagnosis, primary coding position. [c]Optimal algorithm marked in bold.

PPV of 75.4% (95% CI = 66.9–82.3)) were achieved with a BP of 60 days.

We considered the algorithms to be optimal when using the shortest BP possible and still returning a PPV $\geq$ 80%. In DSR, the optimal algorithm (BP 30 days) had a PPV of 89.5% (95% CI = 82.2–94.0), NPV 98.8% (95% CI = 98.2–99.1), sensitivity 75.8% (95% CI = 67.6–82.5), and specificity 99.5% (95% CI = 99.2–99.7). In DNPR, the optimal algorithm (BP 120 days) returned a PPV of 80.6% (95% CI = 71.7–87.2), NPV 98.1% (95% CI = 97.5–98.6), sensitivity 63.7% (95% CI = 55.0–71.6), and specificity 99.2% (95% CI = 98.8–99.5) (Tables 2–4).

For the optimal algorithms, we evaluated the concordance of the onset date of re-ICH identified by the algorithm and the gold standard. For DSR 79.8% of the onset dates identified by the algorithm matched the gold standard, 88.3% of the onset dates matched within one day of the gold standard, and 91.5% matched within 7 days of the gold standard. For DNPR, 74.7% of the onset dates identified by the algorithm matched the gold standard, 83.5% of the onset dates matched within one day of the gold standard, and 88.6% of the onset dates matched within 7 days of the gold standard. After 30 days, the proportion of onset dates identified by the algorithms matching the gold standard did not change (Table 5).

Subanalysis

In a subanalysis of validity estimates for the optimal algorithms (DSR: 30-day BP; DNPR: 120-day BP) we excluded patients who died during the BP (none had emigrated). The subanalysis returned higher sensitivity estimates (DSR 91.3% (95% CI = 84.2–95.3); DNPR 81.4% (95% CI = 72.6–87.9)) and slightly

Table 4 Concordance of the Recurrence Status of Intracerebral Hemorrhage Identified by the Optimal Algorithms and the Gold Standard

		Recurrence by Gold Standard		
		Yes	No	Total
Danish Stroke Registry				
Recurrence by optimal algorithm (30 day blanking period)	Yes	94	11	105
	No	30	2393	2423
	Total	124	2404	2528
Danish National Patient Registry[a]				
Recurrence by optimal algorithm (120 day blanking period)	Yes	79	19	98
	No	45	2385	2430
	Total	124	2404	2528

Note: [a]Inpatient diagnosis, primary coding position.

Table 5 Accuracy of the Optimal Algorithms for Estimating the Onset Date of Intracerebral Hemorrhage Recurrence

Onset Date Estimated by Algorithm	Proportion Matching the Onset Date According to Gold Standard (%)	
	Danish National Patient Registry[a]	Danish Stroke Registry
Same date	74.7	79.8
Within 1 day	83.5	88.3
Within 7 days	88.6	91.5
Within 30 days	92.4	94.7
Within 90 days	92.4	94.7

Note: [a]Inpatient diagnosis, primary coding position.

lower specificity estimates (DSR 98.9 (95% CI = 98.0–99.4); DNPR 97.9 (95% CI = 96.8–98.7)) than the main analysis (Tables 2 and 3). Predictably, PPV and NPV were unchanged, as calculation of these measures does not depend on data from the BP.

Discussion

For this study, we developed and validated algorithms to identify cases of re-ICH using data from two Danish nation-wide registries. The algorithms accurately identified cases of re-ICH as well as the onset date of re-ICH using data from DSR and DNPR, respectively. However, the PPV of the algorithms increased with the length of the applied BP, at the expense of the sensitivity, which decreased. The algorithm using data from DSR yielded a higher PPV with a considerably shorter BP when compared with the algorithm using data from DNPR (ie, BP of 30 days vs 120 days). This difference was to some degree anticipated due to the different purposes of the two registries; whereas DNPR primarily serves as an administrative registry,[12] DSR serves as a clinical database for monitoring stroke care.[11] Therefore, events of re-ICH are expected to be coded with greater accuracy in DSR compared with DNPR.

In previous studies,[13,14] we found a high validity of ICH-diagnoses within DSR and DNPR for identifying cases of first-ever ICH. In this study, we found an equally high validity of ICH-diagnoses within DSR and DNPR for identifying cases of re-ICH when using algorithms to account for the methodological difficulties associated with identifying recurrent events. Our study also documents that use of a BP is essential for accurate classification of re-ICH events, as the algorithms where we did not apply a BP returned very low PPVs (<40%); conversely, even a very short BP of 7 days provided marked improvement of PPVs in both registries, particularly DSR. Striking the right balance between PPV and sensitivity will depend on the aim of the individual project. For example, a high PPV is important in studies of prognosis after re-ICH, while in studies of the incidence of re-ICH, high sensitivity is

also vital. We note that compared with the main analysis, the subanalysis, returned markedly higher sensitivity estimates for both DSR (75.8% vs 91.3%) and DNPR (63.7% vs 81.4%). However, the higher sensitivity in the subanalysis is not only conditional on ignoring re-ICH events during the BP (as in all main analyses) but also on excluding all patients that died or emigrated during the BP. Based on these results and in order to avoid immortal person-time, we suggest that cohort studies of the risk of re-ICH based on DSR or DNPR that employ a BP should begin follow-up on the date the BP ends and only include cohort members that are eligible at start of follow-up (ie, exclude patients that died or emigrated during the BP). Another consideration is the time frame of interest for a project considering use of DSR or DNPR, as the risk of re-ICH is reported to be highest in the first year after an index ICH.[9] For projects focusing on, eg, the first year post-ICH, use of DSR would seem more optimal as even a short blanking period provides valid results. Conversely, the markedly longer blank-ing period necessary to achieve a high PPV for re-ICH in DNPR may be less of a concern in studies interested in the period after the first year post-ICH. We arbitrarily chose to focus on PPV ≥80% for selection of an optimal length of BP in the two algorithms for DSR and DNPR, respectively, but provide values for other BPs that could guide researchers' choices, depending on their aims.

To our knowledge, this is the first study on the validity of DSR and DNPR for identifying re-ICH. Similarly, no research on identifying re-ICH has been conducted using data from other Scandinavian registries. In general, very little research on the subject is available (Supplementary Material and Supplementary Figure S1). One US-based study[16] using hos-pital discharge codes to identify cases of re-ICH found a PPV of 53%. However, the study was not specifically designed for identifying re-ICH. As a result, the study did not consider the abovementioned methodological difficulties. In particular, no BP was used, which could, at least in part, explain why the

study found a considerably lower PPV compared with our study.

Our study has some strengths. First, we developed algorithms specifically designed for identifying re-ICH thereby being able to account for the abovementioned methodological difficulties associated with using registry data for identifying recurrent events. Second, the algorithms were validated by comparison to a gold standard which was established by reviewing the full medical record and brain scan reports for each case thereby ensuring a high degree of accuracy. Additionally, when establishing the gold standard, we included events registered with an ICH-diagnosis as well as events registered with a wide range of diagnoses for other cerebrovascular diseases thus ensuring completeness of the gold standard. Third, the inception cohort comprised residents of RSD; a geographically well-defined area in Denmark, which previously has been shown to be representative for the Danish population in terms of sociodemographic and health-care characteristics,[26] thus ensuring generalizability of the study to the entire Danish population.

Our study has some limitations. First, due to the registry approach used for follow-up, patients with re-ICH that were not admitted to the hospital might have been missed. However, due to the structure of the Danish health-care system where all patients suspected of stroke are referred for acute evaluation at a hospital, we believe that this source of bias only had a minor impact on our results. Additionally, patients dying from re-ICH before hospital admission are not included in our data sources. We considered supplementing our data with information from the Cause-of-Death Registry in an effort to include patients dying from re-ICH before hospitalization. However, the only way to reliably establish ICH as cause of death in such patients would be an autopsy. Since the autopsy rate in Denmark is low,[27] it is unlikely that inclusion of autopsy-verified cases of re-ICH would have affected our estimates notably. Supporting this notion is a finding in a previous large Danish study of more than 2800 cases of ICH,[14] where only a single patient's ICH was established through autopsy (Stine M. Hald, personal communication). Second, some patients might have been missed due to incorrect coding with a non ICH-diagnosis; however, previous studies have found the degree of misclassification of ICH within the Danish registries to be low;[13,28,29] nonetheless, we screened within other cerebrovascular diagnosis codes for ICH, which probably reduced the impact of this potential source of bias. Third, as we did

not include outpatient contacts and ED contacts registered with an ICH-diagnosis during the follow-up period for case ascertainment, events recorded exclusively as outpatient/ED contacts might have been missed. However, we know from previous research that this rarely is the case in the Danish setting.[14] Fourth, we excluded events corresponding to 18 patients during case ascertainment; however, as this exclusion was primarily due to medical records being untraceable because of individual hospitals' storage policies, this source of bias is not systematic and should not affect our estimates of, eg, PPV, although it may have had a minor impact on our estimates of sensitivity. Finally, when collecting data for the gold standard, we did not gather information on inter-rater agreement. However, the study personnel were closely monitored by a senior researcher and were urged to discuss any cases of doubt with the study neurologists and radiologists. Also, in case of doubt, original brain scan images were re-evaluated. We therefore believe that our gold standard is of sufficiently high quality for the purposes of this study.

Conclusion

We developed and validated simple, register-specific algorithms which accurately identified cases of re-ICH and the onset date of re-ICH in DSR and DNPR, respectively. The algorithm using data from DSR more accurately identified events of re-ICH while using a shorter BP compared with the algorithm using data from DNPR.

The results from our study can be used to guide the use of data from DSR and DNPR in further research on the associated risk factors and causes of re-ICH as well as the prognosis of re-ICH.

Acknowledgments

Author Contributions

DG and SMH conceived the study. MMJ, SMH, LMBK and NJB acquired the data. MMJ performed the data analysis and wrote the first draft. All authors have contributed to data analysis, drafting or revising the article, have agreed on the journal to which the article will be submitted, given final approval of the version to be published, and agreed to be held accountable for all aspects of the work.

Funding

Disclosure

References

1. Hostettler IC, Seiffge DJ, Werring DJ. Intracerebral hemorrhage: an update on diagnosis and treatment. *Expert Rev Neurother.* 2019;19 (7):679–694. doi:10.1080/14737175.2019.1623671

2. McGurgan IJ, Ziai WC, Werring DJ, Al-Shahi Salman R, Parry-Jones AR. Acute intracerebral haemorrhage: diagnosis and management. *Pract Neurol.* 2020:practneurol-2020-002763. doi:10.1136/practneurol-2020-002763

3. GBD 2016 Stroke Collaborators. Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 2019;18(5):439–458. doi:10.1016/S1474-4422(19)30034-1

4. van Asch CJ, Luitse MJ, Rinkel GJ, van der Tweel I, Algra A, Klijn CJ. Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. *Lancet Neurol.* 2010;9(2):167–176. doi:10.1016/S1474-4422(09)70340-0

5. González-Pérez A, Gaist D, Wallander M-A, McFeat G, García-Rodríguez LA. Mortality after hemorrhagic stroke: data from general practice (the health improvement network). *Neurology.* 2013;81(6):559–565. doi:10.1212/WNL.0b013e31829e6eff

6. Hanger HC, Wilkinson TJ, Fayez-Iskander N, Sainsbury R. The risk of recurrent stroke after intracerebral haemorrhage. *J Neurol Neurosurg Psychiatry.* 2007;78(8):836–840. doi:10.1136/jnnp.2006.106500

7. Banerjee G, Wilson D, Ambler G, et al. Longer term stroke risk in intracerebral haemorrhage survivors. *J Neurol Neurosurg Psychiatry.* 2020;91(8):840–845. doi:10.1136/jnnp-2020-323079

8. Li L, Luengo-Fernandez R, Zuurbier SM, et al. Ten-year risks of recurrent stroke, disability, dementia and cost in relation to site of primary intracerebral haemorrhage: population-based study. *J Neurol Neurosurg Psychiatry.* 2020;91(6):580–585. doi:10.1136/jnnp-2019-322663

9. Poon MTC, Fonville AF, Al-Shahi Salman R. Long-term prognosis after intracerebral haemorrhage: systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry.* 2014;85(6):660–667. doi:10.1136/jnnp-2013-306476

10. Schmidt LB, Goertz S, Wohlfahrt J, Melbye M, Munch TN. Recurrent intracerebral hemorrhage: associations with comorbidities and medicine with antithrombotic effects. *PLoS One.* 2016;11(11): e0166223. doi:10.1371/journal.pone.0166223

11. Johnsen SP, Ingeman A, Hundborg HH, Schaarup SZ, Gyllenborg J. The Danish stroke registry. *Clin Epidemiol.* 2016;8:697–702. doi:10.2147/ CLEP.S103662

12. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish national patient registry: a review of content, data quality, and research potential. *Clin Epidemiol.* 2015;7:449–490. doi:10.2147/CLEP.S91125

13. Hald SM, Kring Sloth C, 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

14. Hald SM, Kring Sloth C, Agger M, et al. The validity of intracerebral hemorrhage diagnoses in the Danish patient registry and the Danish stroke registry. *Clin Epidemiol.* 2020;12:1313–1325. doi:10.2147/ CLEP.S267583

15. Cea Soriano L, Gaist D, Soriano-Gabarró M, García Rodríguez LA. The importance of validating intracranial bleeding diagnoses in the health improvement network, United Kingdom: misclassification of onset and its impact on the risk associated with low-dose aspirin therapy. *Pharmacoepidemiol Drug Saf.* 2019;28(2):134–139. doi:10.1002/ pds.4561

16. Jones SA, Gottesman RF, Shahar E, Wruck L, Rosamond WD. Validity of hospital discharge diagnosis codes for stroke: the Atherosclerosis Risk in Communities Study. *Stroke.* 2014;45 (11):3219–3225. doi:10.1161/STROKEAHA.114.006316

17. Mainz J, Krog BR, Bjørnshave B, Bartels P. Nationwide continuous quality improvement using clinical indicators: the Danish national indicator project. *Int J Qual Health Care.* 2004;16(Suppl 1):i45–i50. doi:10.1093/intqhc/mzh031

18. Thygesen LC, Daasnes C, Thaulow I, Brønnum-Hansen H. Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. *Scand J Public Health.* 2011;39(7 Suppl):12–16. doi:10.1177/1403494811399956

19. Samarasekera N, Fonville A, Lerpiniere C, et al. Influence of intracerebral hemorrhage location on incidence, characteristics, and outcome: population-based study. *Stroke.* 2015;46(2):361–368. doi:10.1161/STROKEAHA.114.007953

20. Pedersen CB. The Danish civil registration system. *Scand J Public Health.* 2011;39(7 Suppl):22–25. doi:10.1177/1403494810387965

21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377–381. doi:10.1016/j.jbi.2008.08.010

22. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform.* 2019;95:103208. doi:10.1016/j. jbi.2019.103208

23. Quinn GR, Severdija ON, Chang Y, Dallalzadeh LO, Singer DE. Methodologic differences across studies of patients with atrial fibrillation lead to varying estimates of stroke risk. *J Am Heart Assoc.* 2018;7:12. doi:10.1161/JAHA.117.007537

24. Sahlén A, Varenhorst C, Lagerqvist B, et al. Outcomes in patients treated with ticagrelor or clopidogrel after acute myocardial infarction: experiences from SWEDEHEART registry. *Eur Heart J.* 2016;37(44):3335–3342. doi:10.1093/eurheartj/ehw284

25. Rasmussen LA, Jensen H, Virgilsen LF, Hölmich LR, Vedsted P. A validated register-based algorithm to identify patients diagnosed with recurrence of malignant melanoma in Denmark. *Clin Epidemiol.* 2021;13:207–214. doi:10.2147/CLEP.S295844

26. Henriksen DP, Rasmussen L, Hansen MR, Hallas J, Pottegård A. Comparison of the five Danish regions regarding demographic characteristics, healthcare utilization, and medication use–a Descriptive Cross-Sectional Study. *PLoS One.* 2015;10(10):e0140197. doi:10.1371/journal.pone.0140197

27. Petri CN. Decrease in the frequency of autopsies in Denmark after the introduction of a new autopsy act. *Int J Qual Health Care*. 1993;5 (4):315–318. doi:10.1093/intqhc/5.4.315

28. Gaist D, Vaeth M, Tsiropoulos I, et al. Risk of subarachnoid haemorrhage in first degree relatives of patients with subarachnoid haemorrhage: follow up study based on national registries in Denmark. *BMJ*. 2000;320(7228):141–145. doi:10.1136/bmj.320.7228.141

29. Poulsen FR, Halle B, Pottegård A, García Rodríguez LA, Hallas J, Gaist D. Subdural hematoma cases identified through a Danish patient register: diagnosis validity, clinical characteristics, and pre-admission antithrombotic drug use. *Pharmacoepidemiol Drug Saf*. 2016;25(11):1253–1262. doi:10.1002/pds.4058