Reliability of intracerebral hemorrhage classification systems: a systematic review

Reliability of intracerebral hemorrhage classification systems: a systematic review

Journal:	*International Journal of Stroke*
Manuscript ID	IJS-10-15-4490.R1
Manuscript Type:	Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Rannikmae, Kristiina; University of Edinburgh, Centre for Clinical Brain Sciences Woodfield, Rebecca; University of Edinburgh, Department of Clinical Brain Sciences Anderson, Craig; The George Institute for global Health, Charidimou, Andreas; UCL Institute of Neurology and The National Hospital for Neurology and Neurosurgery, Queen Square , Stroke Research Group, Department of Brain Repair and Rehabilitation Chiewvit, Pipat; Siriraj Hospital, Mahidol University, Department of Radiology Greenberg, Steven Jeng, Jiann-Shing; National Taiwan University Hospital, Department of Neurology Meretoja, Atte; FLOREY NEUROSCIENCE INSTITUTES, Palm, Frederick; Klinikum Ludwigshafen, Neurology Putaala, Jukka; Helsinki University Central Hospital, Neurology Rinkel, Gabriel; University Medical Center Utrecht, Neurology Rosand, Jonathan; Massachusett General Hospital, Neurology Rost, Natalia; Massachusetts General Hospital, Neurology Strbian, Daniel; Helsinki University Central Hospital, Neurology Tatlisumak, Turgut; Helsinki University Central Hospital, Department of Neurology; Tsai, Chung-Fen; Cardinal Tien Hospital, Neurology; University of Edinburgh, Centre for Clinical Brain Sciences Wermer, Marieke; Leiden University Medical Centre, Neurology Werring, David; University College London, National Hospital for Neurology and Neurosurgery Yeh, Shin-Joe; National Taiwan University Hospital, Stroke Center and Department of Neurology Al-Shahi Salman, Rustam Sudlow, Cathie; University of Edinburgh, Division of Clinical Neurosciences
Keywords:	Brain bleed, Hemorrhage, Intracerebral hemorrhage, Neurology, Reliability, Systematic review, Classification

Reliability of intracerebral hemorrhage classification systems: a systematic review

Kristiina Rannikmäe[1] MD, Rebecca Woodfield[1] MD, Craig S Anderson[2] MD PhD, Andreas Charidimou[3] MD PhD, Pipat Chiewvit[4] MD, Steven M Greenberg[5] MD PhD, Jiann-Shing Jeng[6] MD PhD, Atte Meretoja[7,8] MD PhD MSc (Stroke Med), Frederic Palm[9] MD PhD, Jukka Putaala[7] MD PhD, Gabriel JE Rinkel[10] MD FRCPE, Jonathan Rosand[5,11,12] MD MSc, Natalia S Rost[11] MD, Daniel Strbian[7] MD PhD MSc FESO, Turgut Tatlisumak[7,13,14] MD PhD, Chung-Fen Tsai[15] MD PhD, Marieke JH Wermer[16] MD, David Werring[3] MD PhD, Shin-Joe Yeh[6] MD, Rustam Al-Shahi Salman[1] PhD FRCPE, Cathie LM Sudlow[1,17,18] DPhil FRCPE

[1]Centre for Clinical Brain Sciences, University of Edinburgh, UK
[2]The George Institute for Global Health, Royal Prince Alfred Hospital and the University of Sydney, Australia
[3]Stroke Research Group, Department of Brain Repair and Rehabilitation, UCL Institute of Neurology and National Hospital for Neurology and Neurosurgery, UK
[4]Department of Radiology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Thailand
[5]Department of Neurology, Massachusetts General Hospital, USA
[6]Stroke Center and Department of Neurology, National Taiwan University Hospital, Taiwan
[7]Department of Neurology, Helsinki University Central Hospital, Finland
[8]Departments of Medicine and the Florey, Royal Melbourne Hospital, University of Melbourne, Australia
[9]Department of Neurology, Städtisches Klinikum Ludwigshafen, Germany
[10]Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, The Netherlands
[11]Center for Human Genetic Research, Massachusetts General Hospital, USA
[12]Program in Medical and Population Genetics, Broad Institute, USA
[13]Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Sweden
[14]Department of Neurology, Sahlgrenska University Hospital, Sweden
[15]Department of Neurology, Cardinal Tien Hospital, School of Medicine, Fu-Jen Catholic University, Taiwan
[16]Department of Neurology, Leiden University Medical Center, The Netherlands
[17]Institute for Genetics and Molecular Medicine, University of Edinburgh, UK
[18]UK Biobank

Corresponding author: Professor Cathie LM Sudlow
Email: cathie.sudlow@ed.ac.uk
Address:
Centre for Clinical Brain Sciences
The University of Edinburgh
Chancellor's Building
49 Little France Crescent
Edinburgh EH16 4SB
Telephone: +44(0)131 537 2622
Fax: +44(0)1313325150

Word count: 4741

Tables: Table 1

Figures: Figure1; Figure 2; Figure 3

Online supplement: Supplemental appendix; Tables I-IV

Key words: Cerebral Hemorrhage, Cerebrovascular Disorders, Reproducibility of Results, Reliability, Systematic review, Classification system;

Cover title: Intracerebral hemorrhage classification systems' reliability

ABSTRACT

Background: Accurately distinguishing non-traumatic intracerebral hemorrhage (ICH) subtypes is important since they may have different risk factors, causal pathways, management and prognosis. We systematically assessed the inter- and intra-rater reliability of ICH classification systems.

Methods: We sought all available reliability assessments of anatomical and mechanistic ICH classification systems from electronic databases and personal contacts until October 2014. We assessed included studies' characteristics, reporting quality and potential for bias; summarized reliability with kappa value forest plots; and performed meta-analyses of the proportion of cases classified into each subtype.

Summary of review: We included 8 of 2152 studies identified. Inter- and intra-rater reliabilities were substantial to perfect for anatomical and mechanistic systems (inter-rater kappa values: anatomical 0.78-0.97 [6 studies, 518 cases], mechanistic 0.89-0.93 [3 studies, 510 cases]; intra-rater kappas: anatomical 0.80-1 [3 studies, 137 cases], mechanistic 0.92-0.93 [2 studies, 368 cases]). Reporting quality varied but no study fulfilled all criteria and none was free from potential bias. All reliability studies were performed with experienced raters in specialist centers. Proportions of ICH subtypes were largely consistent with previous reports suggesting that included studies are appropriately representative.

Conclusions: Reliability of existing classification systems appears excellent, but is unknown outside specialist centers with experienced raters. Future reliability comparisons should be facilitated by studies following recently published reporting guidelines.

INTRODUCTION

Non-traumatic intracerebral hemorrhage (referred to in this paper as ICH) accounts for 10-20% of strokes worldwide.[1] Although age-standardized mortality rates for hemorrhagic stroke have decreased in the past two decades, the absolute number of those affected is increasing.[2] Risk factors, causal pathways, investigations, management and prognosis may differ between ICH subtypes, and stratified approaches to treatment may be appropriate.[3] Hence studies of ICH need to use classification systems that distinguish subtypes accurately. Such studies also need to be very large for adequate statistical power.[4]

Existing ICH classification systems are 'anatomical' or 'mechanistic'. Anatomical systems classify hemorrhages according to their anatomical origin or location as lobar, deep, infratentorial, intraventricular, and various combinations or modifications of these. Mechanistic systems integrate this anatomical information with clinical symptoms, signs and investigations to assign a subtype based on presumed mechanism. Commonly used categories include hypertension, cerebral amyloid angiopathy (CAA), anticoagulation and structural causes.

This study was part of an initiative to develop scalable methods for sub-classifying ICH in large, population-based, prospective research studies, such as the UK Biobank. An ideal classification system for large-scale research use would assign the maximum number of cases to determined, valid subtypes without sacrificing reliability or accuracy, and would be applicable in a range of different clinical settings. Here we systematically sought and assessed the performance of existing ICH classification systems, focusing on their inter- and intra-rater reliability. We also assessed the

proportion of cases assigned to each subtype to ensure that the included studies were appropriately representative.

METHODS

Search strategy

We searched Ovid Medline and Embase (until October 2014) for studies that assessed the inter- or intra-rater reliability of ICH classification systems in human adults, combining search terms for hemorrhagic stroke, classification systems, and reliability (appendix e-1). We included conference abstracts and foreign language articles, checked the bibliographies of all relevant studies and reviews identified, searched Google Scholar for relevant citations and contacted researchers in the field for information about unpublished studies. One author (stroke research fellow KR) assessed eligibility by reviewing all titles, abstracts and, where necessary, the full texts of potentially relevant articles, resolving uncertainties through discussion and mutual consensus with a second author (professor of neurology and clinical epidemiology CLMS).

Study inclusion/exclusion criteria and contact with authors

We included all studies that reported inter- or intra-rater reliability of any anatomical or mechanistic ICH classification system. To minimize publication and reporting bias (whereby positive results are more likely to be published or reported than negative ones), we contacted the authors of all studies which reported the proportion of cases classified into mutually exclusive categories, to obtain unpublished reliability data. We also contacted the authors of all included studies to obtain additional information about study characteristics that may affect reliability.

We excluded studies that used a classification system based on features other than anatomical or mechanistic (e.g., based on severity or prognosis), and studies conducted in highly selected patient populations (e.g., ICH in one anatomical territory) or among cases with selected clinical features (e.g., including only ICH cases with an epileptic seizure at onset).

Data extraction

We divided included studies into those that had used either an anatomical or a mechanistic classification system. From each study, we extracted data on: study characteristics; study population characteristics; classification system; raters classifying the cases; number of cases classified into each category; methods and results of reliability assessment.

Inter- and intra-rater reliability

We displayed available inter- and intra-rater reliability measures (kappa [κ] statistics and 95% confidence intervals) on a forest plot. Kappa was considered to indicate slight- (κ= 0.01–0.20), fair- (κ =0.21-0.40), moderate- (κ=0.41–0.60), substantial- (κ=0.61–0.80) or almost perfect reliability (κ=0.81–0.99).[5] We assessed the quality of reliability reporting of included studies, using criteria based on the Guidelines for Reporting Reliability and Agreement Studies (GRRAS)[6], and potential risk of bias, using criteria developed specifically for this study, relating to study design features that we considered might influence the reliability results (see first column of Supplemental Table I).

Meta-analyses of proportions of ICH subtypes

We extracted data and performed random effects proportion meta-analyses for the more commonly used categories for anatomical classifications (supratentorial versus

(vs) infratentorial; lobar vs any other location) and mechanistic classifications (attributed vs not to: hypertension; CAA; anticoagulant use; vascular structural cause; undetermined). We assessed heterogeneity and the effect of study mean age (≥70 vs <70 years), country (Europe/USA/Australia vs Asia) and hospital- vs population-based study design. We performed analyses with StatsDirect (http://www.statsdirect.com/).

RESULTS

From 2152 publications screened we identified 20 potentially eligible studies.[7-26] We contacted the authors of all 20 studies, eventually including 8 studies with reliability data available in the publication or direct from the authors (Figure 1).[7,8, 21-26] Six of the eight included studies provided additional unpublished information.[7, 8, 22, 24-26]

Reliability of classification systems

Six studies provided data about the reliability of an anatomical classification system[8,21, 23-26] and three about the reliability of a mechanistic classification system[7, 22, 25] (Table 1, Figure 2).

Study characteristics

Most studies were hospital- rather than population-based (5/6 anatomical and 2/3 mechanistic system studies). There were a median of 76 cases for anatomical and 142 for mechanistic studies, mean age was 57-75 years for anatomical and 61-71 years for mechanistic studies, and 48-66% were male. Studies were conducted in Europe (4/6 anatomical, 2/3 mechanistic), Asia (1/6 anatomical, 1/3 mechanistic) and the USA (1/6 anatomical). Classifications were performed retrospectively in most studies (4/6 anatomical and 3/3 mechanistic). The time interval between the two intra-rater reliability ratings ranged from 2-6 months (anatomical) and 15 months to two years (mechanistic). There were 2-6 raters for anatomical and 2-3 for mechanistic studies, including neuroradiologists, neurologists (some with special stroke expertise) and neurosurgeons in anatomical and all stroke neurologists in mechanistic studies. In 5/6 anatomical studies, CT scans were clearly available to each rater. Raters had access to medical records and imaging reports in all mechanistic studies. Completeness of investigation varied across mechanistic studies, but most cases had a CT brain scan,

15-20% an MRI brain scan and 25-30% intracranial blood vessel imaging (CTA, MRA or DSA) (Table 1).

Quality of reporting and measures used to reduce bias

No study satisfied all the GRRAS criteria.[6] Many lacked details about the subject population, rater experience with the classification system and factors relating to the rating process (see Supplemental Table II). No study had used all possible measures to reduce potential bias. The commonest potential sources of bias were that the raters came from the same institution and were aware of being compared to other raters (see Supplemental Table I).

Inter-rater and intra-rater reliability

Anatomical classification systems. Six studies provided data about inter-rater reliability,[8,21,23-26] and three about intra-rater reliability[23,24,26] (Figure 2).

Inter-rater reliability was substantial to almost perfect for classifying ICH as lobar vs any other location in four studies (κ range 0.78-0.97),[21, 23-25] and for classifying ICH into 4-5 categories (lobar, deep, cerebellar, brainstem ± multiple location categories) in two studies (κ 0.81-0.87).[8,26] Intra-rater reliability for lobar versus any other ICH was almost perfect in two studies (κ 0.85-1),[23, 24] and substantial in one study of ICH classified into 4 categories (κ=0.8).[26]

Mechanistic classification systems. Three studies provided data about inter-rater reliability[7,22,25] and two[7, 22] about intra-rater reliability of a mechanistic classification system. All assessed SMASH-U[7] (structural vascular lesions [S], medication [M], amyloid angiopathy [A], systemic disease [S], hypertension [H], or undetermined

[U]) or modifications of this. Classification rules for SMASH-U can be found in a recent publication[7] and in Supplemental Table IV. Both inter-rater and intra-rater reliability were almost perfect (κ ranges 0.89-0.93 and 0.92-0.93, respectively) (Figure 2). There were insufficient data for meta-analyses of reliability estimates or to draw reliable conclusions about factors potentially affecting reliability. (Table 1, Figure 2)

Proportions of ICH subtypes

Data were available for meta-analyses of ICH subtype proportions from three studies that used an anatomical classification system[8, 25,26] (Figure 3A) and three that used a mechanistic system[7, 22, 25] (Figure 3B). The number of studies included in different meta-analyses varied, depending on availability of required data. Detailed study characteristics are shown in Supplemental Table III. The pooled proportion of ICH cases classified as lobar was 0.32 (95% CI 0.24-0.41), with moderate heterogeneity between studies (I^2=60%). The proportion was smaller in one study, including younger cases recruited from Asia.[8] The pooled proportion classified as supratentorial was 0.90 (95% CI 0.84-0.94) (Figure 3A). The pooled proportion classified as hypertensive was 0.47 (95% CI 0.32-0.62), CAA-related 0.20 (95% CI 0.12-0.29), undetermined cause 0.11 (95% CI 0.04-0.20), due to anticoagulant use 0.09 (95% CI 0.02-0.20) and due to a vascular structural cause 0.06 (95% CI 0.03-0.08). There was substantial heterogeneity between studies (I^2 values 87-99%) (Figure 3B). The proportion classified as due to CAA was higher and the proportion undetermined lower in one study including older cases.[25]

DISCUSSION

Inter- and intra-rater reliabilities of existing anatomical and mechanistic classification systems appeared to be substantial to almost perfect. Reporting quality was variable with no study completely following the GRRAS guidelines,[6] probably because measuring reliability was not the primary aim for all studies and the guidelines were published only recently. Furthermore, since no study had used all possible measures to reduce potential bias, reliability may have been over-estimated. All raters in these studies were experts in their field, limiting the generalizability of the results to less expert raters who might usefully contribute to large-scale research studies. Finally, the majority of the studies were conducted in Caucasian participants, which limits the generalizability of the results to other ethnicities. The proportions of ICH subtypes were largely consistent with previous reports,[27] suggesting that the included studies are representative.

The included classification systems have some limitations. For anatomical systems, these include: classification based on presumed site of origin of ICH in some studies and on ICH location in others; unclear and/or variable category definitions; and few systems with a separate category for bleeds in multiple or uncertain locations. For mechanistic systems, limitations include: assumptions about causal pathways (e.g., hypertension is commonly considered to be causally associated with deep ICH location and CAA with lobar ICH, despite doubts about the nature and/or strength of these associations);[28,29] the dependence on investigations undertaken to identify the potential cause (which vary considerably among specialties and countries, and with age, ICH location and blood pressure);[30] varying definitions of primary and spontaneous ICH; inability to assign a proportion of cases to a determined subtype; and that most cases do not have a single cause, but several interacting contributory factors.[27]

To our knowledge, there are no prior published systematic reviews of the reliability of ICH classification systems. Other strengths include a thorough search strategy, and rigorous assessment of study characteristics, quality and bias indicators. In addition, although we found relatively few relevant published studies, through contacting authors we were able to include additional unpublished reliability results. Finally, our study has highlighted the limitations of existing classification systems, which should help ensure their further refinement where needed and their appropriate use in diverse clinical and research settings.

We may have missed some publications where reliability was assessed but buried in a few words within the body of the text. We attempted to address this by manually searching through relevant review papers and reading full texts of all potentially relevant publications. Limited available data mean that conclusions about potential factors affecting the reliability and proportions of ICH subtypes are also inherently limited. Finally, although reliability is an important feature of a classification system, it does not necessarily correlate with diagnostic accuracy or validity, which would require reference to a 'gold' standard.

While both anatomical and mechanistic systems appear to have excellent reliability, for large population-based, prospective epidemiological studies, anatomical classification systems are likely to be more: feasible (less information from investigations is required); scalable (automated or semi-automated classification may be possible); and appropriate for many prospective studies of potential causes of ICH (free of assumptions about causal pathways). Developing such methods for use at scale will require clear definitions, classification protocols, and categories for multiple and uncertain locations. Mechanistic systems such as the SMASH-U have the advantage of already having a very clear set of rules which probably contributes to their excellent reliability.[31] However, the validity of mechanistic systems could be further improved by integrating categories for cases with an uncertain and multiple overlapping mechanisms. Such systems are likely to be appropriate for stratifying patients for

clinical trials; some case-control studies; and in clinical practice to encourage a more systematic mechanistic work-up. The feasibility of collecting the additional information required for mechanistic classification in large, prospective, population-based studies needs further assessment, since it would complement the simpler information required for anatomical sub-classification and – potentially – allow nested case-control studies based on not only anatomical but also mechanistic information.

To conclude, existing classification systems appear to have excellent reliability in the settings in which they have been tested, but their reliability is unknown outside highly specialized centers with experienced readers. Future comparisons will be facilitated by studies following published GRRAS reporting guidelines.[6]

Authors acknowledge M Moragas (Consultant Neurologist, NHS Lothian) for help with translating a Spanish article.

Funding. K Rannikmäe: Edinburgh Stroke Research and Amenities Endowments Fund, UK Biobank; **C Sudlow:** Scottish Funding Council, UK Biobank; **R Woodfield:** UK Biobank, Chest Heart and Stroke Scotland; **R Al-Shahi Salman:** Medical Research Council; **A Charidimou:** Greek State Scholarship Foundation, Stroke Association, British Heart Foundation; **D Werring:** Higher Education Funding Council for England, Stroke Association, British Heart Foundation; **CS Anderson:** National Health and Medical Research Council of Australia. Details of funding for **P Chiewvit, S Greenberg, J-S Jeng, A Meretoja, F Palm, J Putaala, G Rinkel, J Rosand, NS Rost, D Strbian, T Tatlisumak, MJH Wermer** and **S-J Yeh** found in references 7,8,16,22,24,25.

Conflicts of interest: None.

Authors' individual contributions: K Rannikmäe was involved in the original design and conceptualization of the study, and with supervision from **C Sudlow** and input from **R Al-Shahi Salman** and **R Woodfield** performed the systematic review, collected the data, performed the meta-analyses, interpreted the results, drafted the manuscript and revised it prior to submission. **R Al-Shahi Salman** and **R Woodfield** contributed to designing the study, assisted with interpreting the data and revised the manuscript for intellectual content. **C Sudlow** supervised the design and conceptualization of the study, the data collection process, the analysis and interpretation of the data and revised the manuscript for intellectual content. **C Anderson, A Charidimou, P Chiewvit, S Greenberg, J-S Jeng, A Meretoja, F**

Palm, J Putaala, G Rinkel, J Rosand, NS Rost, D Strbian, T Tatlisumak, MJH Wermer, D Werring and **S-J Yeh** contributed to the original data collection and revised the manuscript for intellectual content. **C-F Tsai** advised about statistical methods for the proportion meta-analyses and subgroup analyses and revised the manuscript for intellectual content. **A Charidimou, MJH Wermer, F Palm, P Chiewvit, J-S Jeng, A Meretoja, J Putaala, G Rinkel, D Strbian, T Tatlisumak, D Werring** and **S-J Yeh** also performed the initial reliability analysis.

References

1. Feigin VL, Lawes CM, Bennett DA, Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: A systematic review. *Lancet Neurol* 2009; 8:355-369.

2. Krishnamurthi RV, Feigin VL, Forouzanfar MH, et al. Global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990-2010: Findings from the global burden of disease study 2010. *Lancet Glob Health* 2013; 1:e259-281.

3. Al-Shahi Salman R, Labovitz DL, Stapf C. Spontaneous intracerebral haemorrhage. *BMJ* 2009; 339:b2586.

4. Burton PR, Hansell AL, Fortier I, et al. Size matters: Just how big is big?: Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol* 2009; 38:263-273.

5. Viera AJ, Garrett JM. Understanding Interobserver Agreement: The Kappa Statistic. *Fam Med* 2005; 37:360-363.

6. Kottner J, Audigé L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; 64:96-106.

7. Meretoja A, Strbian D, Putaala J, et al. SMASH-U: A proposal for etiologic classification of intracerebral hemorrhage. *Stroke* 2012; 43:2592-2597.

8. Chiewvit P, Danchaivijitr N, Nilanont Y, Poungvarin N. Computed tomographic findings in non-traumatic hemorrhagic stroke. *Journal Med Assoc Thai* 2009; 92:73-86.

9. Zahuranec DB, Sánchez BN, Brown DL, et al. Computed tomography findings for intracerebral hemorrhage have little incremental impact on post-stroke mortality prediction model performance. *Cerebrovasc Dis* 2012; 34:86-92.

10. Wijman CA, Venkatasubramanian C, Bruins S, Fischbein N, Schwartz N. Utility of early MRI in the diagnosis and management of acute spontaneous intracerebral hemorrhage. *Cerebrovasc Dis* 2010; 30:456-463.

11. Barton CW, Hemphill JC III. Cumulative dose of hypertension predicts outcome in intracranial hemorrhage better than American Heart Association guidelines. *Acad Emerg Med* 2007; 14:695-701.

12. Nilsson OG, Lindgren A, Brandt L, Säveland H. Prediction of death in patients with primary intracerebral hemorrhage: a prospective study of a defined population. *J Neurosurg* 2002; 97:531-536.

13. Lovelock CE, Anslow P, Molyneux AJ, et al. Substantial observer variability in the differentiation between primary intracerebral hemorrhage and hemorrhagic transformation of infarction on CT brain imaging. *Stroke* 2009; 40:3763-3767.

14. Labovitz DL, Hauser WA, Sacco RL. Prevalence and predictors of early seizure and status epilepticus after first stroke. *Neurology* 2001; 57:200-206.

15. Takahashi O, Cook EF, Nakamura T, Saito J, Ikawa F, Fukui T. Risk stratification for in-hospital mortality in spontaneous intracerebral haemorrhage: a classification and regression tree analysis. *QJM* 2006; 99:743-750.

16. Rost NS, Smith EE, Chang Y, et al. Prediction of functional outcome in patients with primary intracerebral hemorrhage: the FUNC score. *Stroke* 2008; 39:2304-2309.

17. Mayda-Domaç F, Misirli H, Yilmaz M. Prognostic role of mean platelet volume and platelet count in ischemic and hemorrhagic stroke. *J Stroke Cerebrovasc Dis* 2010; 19:66-72.

18. Lee SH, Bae HJ, Kwon SJ, et al. Cerebral microbleeds are regionally associated with intracerebral hemorrhage. *Neurology* 2004; 62:72-76.

19. Anderson CS, Chakera TMH, Stewart-Wynne EG, Jamrozik KD. Spectrum of primary intracerebral haemorrhage in Perth, Western Australia, 1989-90: incidence and outcome. *J Neurol Neurosurg Psychiatry* 1994; 57:936-940.

20. Díaz-Guzmán J, Egido-Herrero JA, Fuentes B, et al. Incidencia de ictus en España: estudio Ibericus. Datos del estudio piloto. *Rev Neurol* 2009; 48:61-65.

21. Ziai W, Gill R, Ullman N, Gandhi D, Hanley D. Reliability of spontaneous intracerebral hemorrhage localization on admission and follow-up computed tomography. *Abstract at the Neurocritical Care Conference: 9th Annual Meeting of the Neurocritical care Society. Montreal, QC, Canada* 2011; 1 Suppl 1:pp S102.

22. Yeh S-Y, Tang S-C, Tsai L-K, Jeng J-S. Pathogenetical subtypes of recurrent intracerebral hemorrhage: designations by SMASH-U classification system. *Stroke* 2014; 45:2636-2642.

23. Bhattathiri SP, Gregson B, Prasad KS, et al. Reliability assessment of computerized tomography scanning measurements in intracerebral hematoma. *Neurosurg Focus* 2003; 15:E6.

24. Wermer MJ, Rinkel GJ, van Rooij WJ, Witkamp TD, Ziedses des Plantes BG, Algra A. Interobserver agreement in the assessment of lobar versus deep location of intracerebral haematomas on CT. *J Neuroradiol* 2002; 29:271-274.

25. Palm F, Henschke N, Wolf J, et al. Intracerebral haemorrhage in a population-based stroke registry (LuSSt): incidence, aetiology, functional outcome and mortality. *J Neurol* 2013; 260:2541-2550.

26. Charidimou A, Wilson D, Shakeshaft C, et al. The Clinical Relevance of Microbleeds in Stroke study (CROMIS-2): rationale, design, and methods. Int J Stroke 2015; doi: 10.1111/ijs.12569 [Epub ahead of print].

27. Warlow C, van Gijn J, Dennis M, et al. What caused this intracerebral haemorrhage? In *Stroke: practical management*. 3rd Ed. Blackwell Publishing; 2008:411-456.

28. Jackson CA, Sudlow CL. Is hypertension a more frequent risk factor for deep than lobar supratentorial intracerebral haemorrhage? *J Neurol Neurosurg Psychiatry* 2006; 77:1244-1252.

29. Samarasekera N, Smith C, Al-Shahi Salman R. The association between cerebral amyloid angiopathy and intracerebral haemorrhage: systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry* 2012; 83:275-281.

30. Cordonnier C, Klijn CJ, van Beijnum J, Al-Shahi Salman R. Radiological investigation of spontaneous intracerebral hemorrhage. systematic review and trinational survey. *Stroke* 2010; 41:685-690.

31. Woodfield R, Grant I, UK Biobank Stroke Outcomes Group, UK Biobank Follow-Up and Outcomes Working Group, Sudlow CLM. Accuracy of Electronic Health Record

Data for Identifying Stroke Cases in Large-Scale Epidemiological Studies: A Systematic Review from the UK Biobank Stroke Outcomes Group. *PLoS ONE* 2015; 10(10): e0140533.

Table 1. Characteristics of studies assessing reliability of an ICH classification system.

First author Country Year	Classification system			Population						Raters						Data			
								Timing¶		Number			Blinding#			Blinding††			
	Categories*	Clear rules	Developed locally	Cases† Ethnicity % male	Description	Age (mean ±SD)	Method‡	Inter	Intra	Inter	Intra	Expertise§	1	2	Institutions **	Hx	Out-come	Same‡‡	Information available¶¶
ANATOMICAL SYSTEMS																			
Chiewvit[8] Thailand 2009	·Lobar ·Thalamic-ganglionic ·Cerebellum ·Brainstem ·Multiple location	n/a	n/a	84 Thai 63	Hospital case review	57±17	R	?	n/a	n/a	2	n/a	Experts	Y	N	Single	Y	Y	Y	Access to CT scans
Wermer[24] Netherlands 2002	·Lobar ·Deep	n/a	n/a	50/25 Caucasian 54	Selected (based on ICH volume) hospital case review	67	R	1-10 years	2 months	n/a	3	1	Experts	Y	N	Multiple	Y	Y	Y	Access to CT scans
Bhattathiri[23] UK 2003	·Lobar ·BG/ thalamus ·Internal capsule	n/a	n/a	43 ? ?	Selected (opportunity sample) trial case review	?	R	?	Min. 2 months	n/a	6	6	Experts	Y	?	?	Y	Y	Y	Assume access to CT scans
Ziai[21] USA 2011 *Abstract*	·Lobar ·Caudate ·Globus pallidus ·Putamen ·Thalamus ·Primary IVH	n/a	n/a	145 ? ?	Presumed hospital-based	?	?	?	n/a	n/a	3	n/a	Experts	?	?	?	?	?	?	?
Palm[24] Germany 2013	·Lobar ·Deep	n/a	n/a	127 Caucasian 50.7	Population based stroke registry	71.3± 13.5	P	Most within 24 hours	n/a	n/a	2	n/a	Experts	Y	N	Single	Y	Y	Y	Access to CT scans
Charidimou[26] UK 2015	·Lobar ·Deep ·Cerebellar ·Brainstem	n/a	n/a	69 92% Caucasian 48	Hospital-based, selected subset of cases	74.9± 12.3	R	>1 month	~6 months	n/a	3	1	Experts	Y	N	Multiple	Y	Y	Y	Access to CT scans

First author Country Year	Classification system			Population						Raters						Data			
	Categories*	Clear rules	Developed locally	Cases† Ethnicity % male	Description	Age (mean ±SD)	Method‡	Timing¶ Inter	Timing¶ Intra	Investigations performed	Number Inter	Number Intra	Expertise§	Blinding# 1	Blinding# 2	Institutions **	Blinding†† Hx	Blinding†† Out-come	Same‡‡	Information available¶¶
MECHANISTIC SYSTEMS																				
Meretoja[7] Finland 2012	·Structural lesion ·Systemic/ other disease ·Medication ·CAA ·Hypertension ·Undetermined	Y	Y	100 Caucasian 58	Random selection of hospital cases	66±13	R	Several years	2 years	100% CT and/or MRI 25% CTA,MRA or DSA	3	3	Experts	Y	N	Single	n/a	Y	Y	Full medical records, neurology opinion, imaging, all tests
Palm[25] Germany 2013	·Structural vascular pathology ·OAK related ·CAA ·Hypertension ·Undetermined	Y	Modified locally	142 Caucasian 51	Population based stroke registry	71.3± 13.5	R	6 months – 5 years	n/a	100% CT; 20% MRI; 26% CTA, MRA and/or DSA	2	n/a	Experts	Y	N	Single	n/a	?	Y	Anatomical classification, neuroimaging, age, medications on admission, INR, cardiovascular risk factors
Yeh[22] Taiwan 2014	·Structural lesion ·Systemic/other disease ·Medication ·CAA ·Hypertension ·Undetermined	Y	Modified locally	268 Chinese 66	Cases from National Taiwan University Hospital Stroke Registry database	60.9 ± 16.0	R	7-19 months	15 months	100% CT; 15% MRI; 30% CTA, MRA and/or DSA	2	1	Experts	Y	N	Single	n/a	Y	Y	Past medical hx and medication hx, imaging results,clinical information, blood tests

*For definitions of classification system categories see Supplemental.Table IV

†number of ICH cases classified for reliability assessment.

‡Method: R (retrospective): cases classified after initial presentation, usually through retrieving and reviewing medical records; P (prospective): patients classified at the time of/shortly after presenting and being recruited.

¶Time from symptom onset to classification (inter-rater reliability) or between two ratings (intra-rater reliability).

§Expertise (predefined categories): Expert=neuroradiology/neurosurgery/neurology trainee or consultant, stroke research fellow; Less expert=physicians 1-4 years post registration, general practitioners, nurses or medical students.

#Blinding: 1: rater unaware of other raters' decisions &/or their own previous decision; 2: rater unaware of being compared to other raters.

**Institutions: raters from single or multiple institutions.

††Blinding: raters blind to patient's clinical history (Hx) and/or outcome.

‡‡Same: same information available to each rater.

¶¶Information available: information available to the rater for classification purposes.
?: unknown; Y=yes; N=no;

Comments about specific studies: [8]mean age applies to a larger sample of 131 cases (including cases with subarachnoid/subdural/intraventricular hemorrhage); [24]only supratentorial cases; 50 cases rated by 3 raters to assess inter-rater agreement, 25/50 cases rated twice by one rater to assess intra-rater agreement, not specified how these 25 cases chosen from amongst the 50; [23]only supratentorial ICH cases; classification assumed retrospective, though not specifically mentioned; [21]not stated explicitly in the abstract but assumed to include only supratentorial ICH cases and to use expert rater(s). [25] % male, mean age and investigations performed applies to a larger sample of 152 non-traumatic ICH cases; [7]each case classified by 2 of 3 raters, with 2 raters classifying 50 cases each and one rater classifying 100 cases; Investigations performed apply to larger sample of 1013 ICH cases; 100 cases for reliability selected at random from the whole sample of 1013 patients, but weighted to ensure appropriate representation of all SMASH-U classifications.

Figure titles and legends:

Figure 1.

Title: Selection of included studies.

Legend:."n": overall number of ICH cases included

Figure 2.

Title: Inter- and intra-rater reliability of existing classification systems.

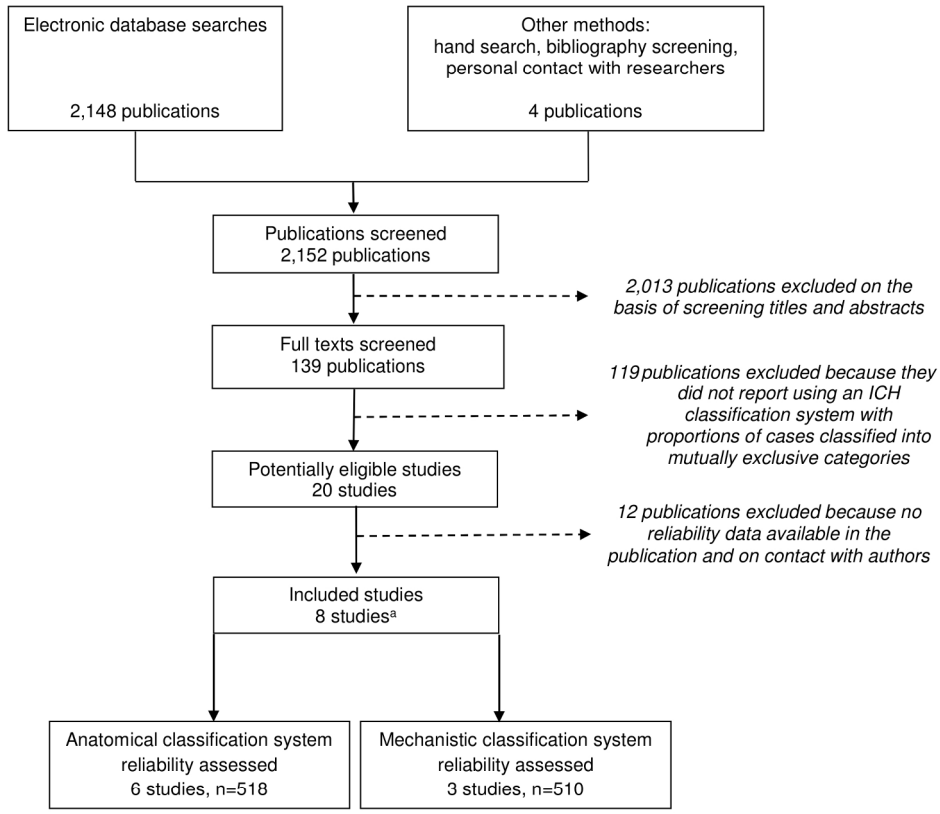
Legend: Squares represent individual study reliability estimates, and associated horizontal lines represent 95% confidence intervals.

Figure 3.

Title: Proportion of ICH cases in (A) anatomical and (B) mechanistic categories.

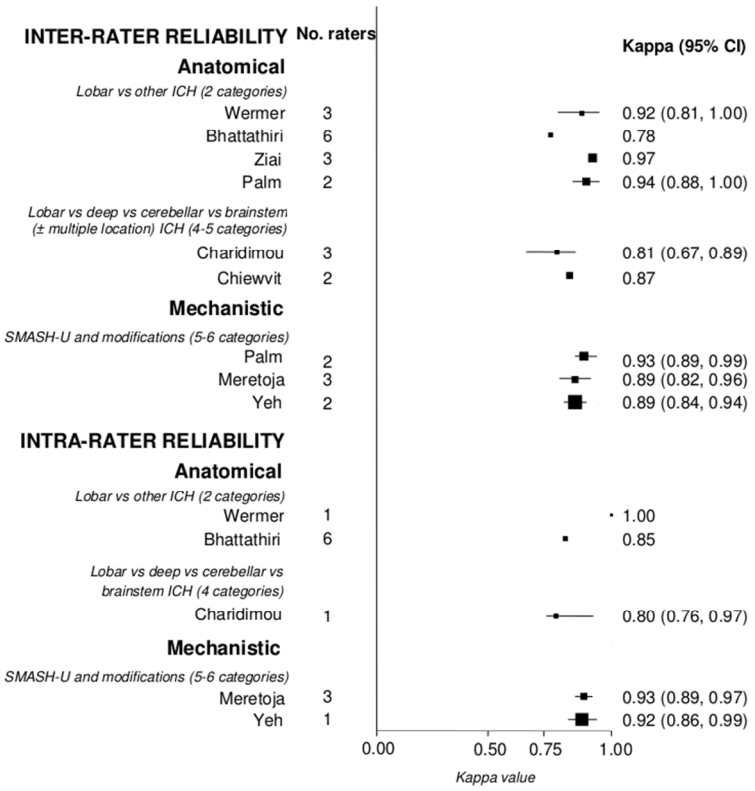
Legend: Notation as Figure2. Unfilled diamonds: pooled proportion estimates. Population-based (P), hospital-based (H), Yes (Y), No (N), Europe (E), Asia (A).

3A: *Chiewvit, Charidimou:* 7 multiple ICH/isolated intraventricular haemorrhage cases excluded



[a]one study used both anatomical and mechanistic classification systems[25],
one study was in the form of an abstract only[21], and one study was published in Spanish[20]

Figure 1
189x193mm (300 x 300 DPI)



All kappas are unweighted.
Wermer, Bhattathiri, Ziai: only supratentorial ICH cases included.
Wermer: mean kappa based on 3 sets of ratings between 2 raters (different combination of raters from amongst 3 raters).
Bhattathiri and Ziai: subtype kappa for lobar location vs any other location.
Bhattathiri: intra-rater reliability based on mean kappa for 6 raters all classifying each case twice.
Meretoja: each case was classified by 2 raters, but there were 3 raters in total with 2 raters classifying 50 cases each, and one rater classifying 100 cases.
Bhattathiri: various other comparisons were reported in the study but data was clearest for lobar vs any other category reliability reported on this figure.

Figure 2
69x98mm (300 x 300 DPI)

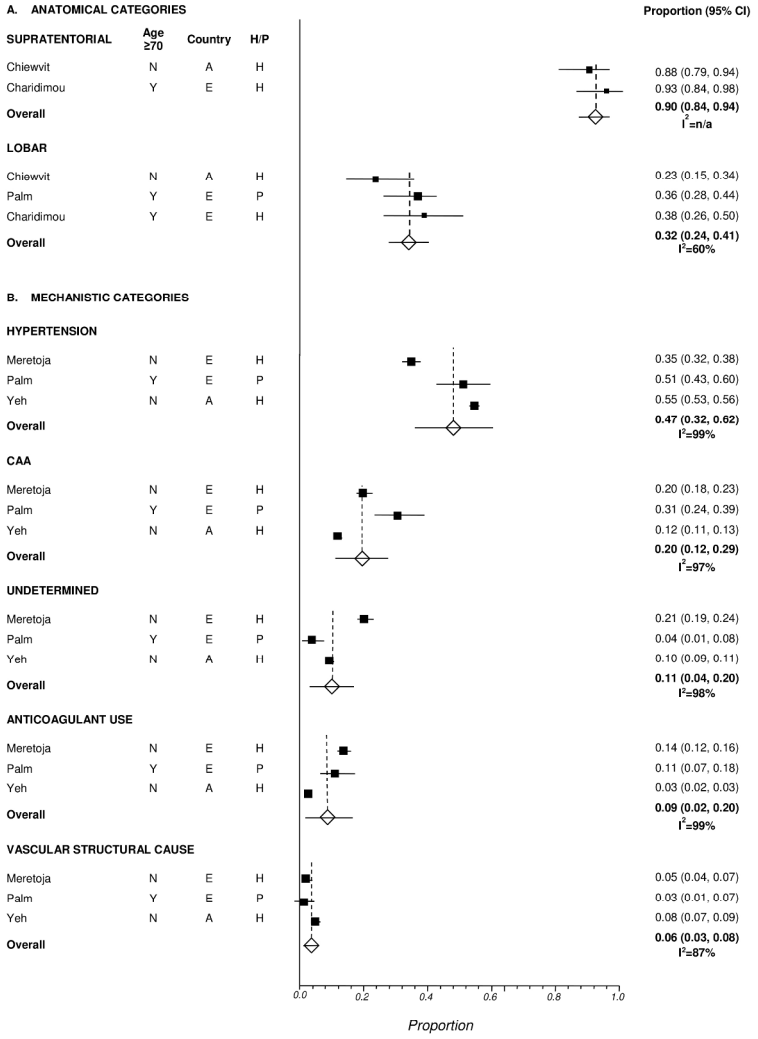


Figure 3
209x297mm (300 x 300 DPI)