Online-Only Supplemental Materials

e-Appendix 1: Electronic search strategy, coding-scheme for the systematic review and data analysis

The search strategy was designed by a PhD-trained clinical investigator with relevant domain expertise in neurology (DNT), an expert in emergency medicine (JAE), and the director of our Evidence-based Practice Center with extensive experience in systematic reviews (KAR).

We searched MEDLINE and Embase for English-language articles, using the following strategies with the following components: (1) diagnosis, (2) emergency, and (3) cerebrovascular (ischemic stroke, transient ischemic attack [TIA], or subarachnoid hemorrhage [SAH]). We did not expressly search for intracerebral hemorrhages (ICH) because these do not generally present diagnostic difficulty in the ED given the ready availability and high sensitivity of CT scans.[1] We also performed a manual search of reference lists from eligible articles, and contacted corresponding authors where necessary. We did not seek to identify research abstracts from meeting proceedings or unpublished studies.

MEDLINE Search *(accessed via PubMed at www.ncbi.nlm.nih.gov/pubmed)*

((diagnostic errors[mh] OR misdiagnosis[tiab] OR misdiagnosed[tiab] OR ((missed[tiab] OR delayed[tiab]) AND diagnosis[tiab])) OR ((population-based[tiab] OR population based[tiab] OR prospective[tiab] OR cohort[tiab] OR cross-sectional[tiab] OR cross sectional[tiab] OR systematic review[tiab] OR meta-analysis[tiab] OR meta analysis[tiab] OR random*[tiab]) AND (diagnosis[tiab] OR diagnosed[tiab] OR diagnostic[tiab]))) AND (emergency services,hospital[mh] OR emergency treatment[mh] OR emergency department*[tiab] OR emergency service*[tiab] OR emergency physician*[tiab] OR casualty[tiab] OR ambulance*[tiab] OR initial diagnosis[tiab] OR initial contact[tiab] OR warning[tiab]) AND (cerebrovascular disorders[mh] OR stroke[tiab] OR transient ischaemic attack[tiab] OR transient ischemic attack[tiab] OR TIA[tiab] OR subarachnoid hemorrhage[tiab] OR subarachnoid haemorrhage[tiab])

Embase Search *(accessed via embase.com)*

#1 'diagnostic error'/exp OR misdiagnosis:ti,ab OR misdiagnosed:ti,ab OR ((missed:ti,ab OR delayed:ti,ab) AND diagnosis:ti,ab) OR (('population-based':ti,ab OR 'population based':ti,ab OR prospective:ti,ab OR cohort:ti,ab OR 'cross-sectional':ti,ab OR 'cross sectional':ti,ab OR 'systematic review':ti,ab OR 'meta-analysis':ti,ab OR 'meta analysis':ti,ab OR random*:ti,ab) AND (diagnosis:ti,ab OR diagnosed:ti,ab OR diagnostic:ti,ab))

#2 'emergency health service'/exp OR 'emergency treatment'/exp OR 'emergency department*':ti,ab OR 'emergency service*':ti,ab OR 'emergency physician':ti,ab OR casualty:ti,ab OR ambulance*:ti,ab OR 'initial diagnosis':ti,ab OR 'initial contact':ti,ab OR warning:ti,ab

#3 'cerebrovascular disease'/exp OR stroke:ti,ab OR 'transient ischaemic attach':ti,ab OR 'transient ischemic attack':ti,ab OR TIA:ti,ab OR 'subarachnoid hemorrhage':ti,ab OR 'subarachnoid haemorrhage':ti,ab

#4 #1 AND #2 AND #3

Search Results

Our search identified 1693 unique citations, of which 1479 (87.4%) were excluded at the abstract level (Figure 1, main manuscript). We did not demand concordance on reason for abstract exclusion, but, among those abstracts with concordant reasons for exclusion (68.2%, n=1009), the distribution was as follows: 34.5% were not about misdiagnosis; 30.3% were not about a neurologic condition; 16.1% were not about cerebrovascular events; 10.0% had no original data; 4.1% were not about ED care; 2.3% had fewer than 5 subjects studied and 1.7% were not in English.

We sought to examine 214 full manuscripts (this included 19 articles identified by hand-search), one of which was unretrievable (Steiger (2000)).[2] After initial screening, there were a total of 16 disagreements about study inclusion for the two pairs of reviewers (period 1995-2009: reviewers DNT and JE, 8 disagreements, kappa=0.70 [95% CI: 0.50 – 0.90]; period 2009-2016: reviewers SL and AAT, 8 disagreements, kappa=0.80 [95% CI 0.66 – 0.93]). These differences were resolved by discussion and adjudication by a third reviewer. Overall agreement on reason for exclusion was 86.0% (DNT and JE, 73.7% agreement with 15 disagreements; SL and AAT, 92.5% agreement with 8 disagreements). We demanded concordance on reason for full-text exclusion and resolved differences by discussion.

At the end of our full-text review, 191 were excluded and 23 were considered eligible (Figure 1, main manuscript). These eligible studies represented 1.4% of the total (n=1693). Among all full-text manuscripts excluded (89.3%), the distribution of reason for exclusion was as follows: not about misdiagnosis (33.2%), abstract only (26.6%), not about the ED (11.7%), no original data (4.2%), not in English, despite English abstract (3.3%), reports only errors (3.3%), fewer

than 5 subjects (1.9%), low diagnostic reference standard (1.9%),[3-6] not about cerebrovascular disease (1.9%), study data collected prior to 1995 (0.9%), duplicate data (0.5%).[7]

For ten studies we attempted to contact the first or corresponding author for additional study information.[8-17] Seven authors responded and five provided additional information.

Data synthesis and analysis

We report the accuracy of ED diagnosis overall and by cerebrovascular condition. While several studies reported on intracerebral hemorrhage (ICH) patients as well, we did not attempt to analyze those patients separately since no numbers on false negatives and false positives could be retrieved. We calculated sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) for ED diagnoses for any cerebrovascular condition. Although properly considered *proportions* rather than *rates,*[18] we use here the more common terminology, defining the *false negative rate* as 1-sensitivity, *false positive rate* as 1-specificity, *false discovery rate* as 1-NPV, and *false omission rate* as 1-PPV.[18-20] When only partial results from a particular study were reported (e.g., no data on non-stroke patients or false positives), we excluded that study only from the relevant calculations (e.g., specificity, PPV), but not the remaining calculations (e.g., sensitivity, NPV) (e-Appendix 4). Where studies included data both from an ED population and other populations (e.g., primary care or pre-hospital), we included only results from the ED population. We present proportions and, where appropriate, 95% confidence intervals (95% CI). Tests of heterogeneity[21, 22] were conducted based on Cochran's Q test.[23] We assessed for possible trends in diagnostic accuracy parameters over time using weighted linear regression with weights equal to 1 over the estimated standard error of the accuracy parameters (equivalent to the χ^2 test for trend). Cohen's kappa, confidence intervals, heterogeneity statistics, trend analyses, and OR were calculated using R v3.2.4 (Foundation for statistical computing, Vienna, Austria) by a PhD biostatistician. This study is reported in accordance with PRISMA guidelines.[24]

Coding schema for abstract and full-text reviews

All gathered literature was subject to title/abstract screening by two independent reviewers (DNT/JAE or AAT/SL). Full-text screening was then applied to all citations considered eligible or possibly eligible by at least one reviewer. Two independent reviewers (DNT/JAE or AAT/SL) determined whether full-text manuscripts were eligible and, if not, provided a reason for exclusion. Differences were resolved by discussion and consensus. A third reviewer (KAR or DNT) verified the eligibility of selected articles. DNT and AAT completed a hand search of the reference lists of selected articles for additional citations. For citations identified by hand search, the full process was repeated iteratively until no additional manuscripts were found for inclusion. We calculated inter-rater agreement on full-text inclusion using Cohen's kappa.[25] A formal review protocol was not registered or posted.

<u>Abstract Review Coding Rules</u>
1) Coding status options are "Yes", "No", "Maybe". We will review full text of "Yes" and "Maybe". The purpose of "Yes" vs. "Maybe" is to look at kappa values agreement on "Yes" vs. "Maybe".
2) Err on the side of "Maybe" if there is doubt about a "No"; this is more conservative.
3) If there is only a title, exclude it only if you feel confident; otherwise code it as "Maybe".
4) Each "No" or "Maybe" should be coded with a reason for exclusion.
5) Reasons for exclusion are listed below 0-7. Go through them in order from 0 to 7 for each abstract, coding the first reason for exclusion only, not multiple reasons for exclusion. Only code "0" for "not English" if you are sure it is "not English".
6) Two independent raters will code reason for exclusion, but we will not mandate agreement on exclusion reason at the abstract level.
7) Occasionally an abstract seems inappropriate for another reason. In such cases, code as "other". There should be few "other" codings.

Abstract Reasons for Exclusion

0	not English	manuscript is not in English
1	no data	review paper; no original patient data
2	not neuro	no reasonable prospect that the study includes data about a cerebrovascular problem (e.g., study is about myocardial infarction, cardiac arrest, dementia)
3	not stroke	no reasonable prospect that the study includes data specifically about ischemic stroke, TIA, or subarachnoid hemorrhage (e.g., study is about intraparenchymal hemorrhage, traumatic subdural, epidural hematoma, giant cell arteritis)
4	not misDx	no reasonable prospect that the study includes data about clinical diagnostic accuracy or misdiagnosis of stroke (e.g., entry population is everyone with confirmed stroke or about radiographic accuracy)
5	not ED	no reasonable prospect that the study includes data about ED physician diagnostic accuracy (e.g., about pre-hospital accuracy, resident accuracy, reliability study of a diagnostic screening tool for stroke)
6	<5	fewer than 5 subjects (total participants reported, including cases and controls)
7	other	any other reason abstract is not included

Full-Text Review Coding Rules

1) Coding status options are “Yes” or “No”.
2) Each "No" should be coded with a reason for exclusion.
3) Reasons for exclusion are listed below 0-9. Go through them in order from 0 to 9 for each full manuscript, coding the first reason for exclusion only, not multiple reasons for exclusion.
4) Two independent raters will code reason for exclusion, and we will mandate agreement on exclusion reason at the manuscript level.
5) Coding differences will be adjudicated or consensus will be developed through dialogue.

Full-Text Reasons for Exclusion

0	not English	manuscript is not in English
1	abstract only	no full-text manuscript available
2	no data	review paper; no original patient data
3	not neuro	the study does not include data about a cerebrovascular problem (e.g., study is about myocardial infarction, cardiac arrest, dementia)
4	not stroke	the study does not include data specifically about ischemic stroke, TIA, or subarachnoid hemorrhage (e.g., study is about intraparenchymal hemorrhage, traumatic subdural, epidural hematoma, giant cell arteritis)
5	not misDx	the study does not include data about clinical diagnostic accuracy or misdiagnosis of stroke (e.g., entry population is everyone with confirmed stroke or about radiographic accuracy)
6	not ED	the study does not include data about ED physician diagnostic accuracy (e.g., about pre-hospital accuracy, resident accuracy, reliability study of a diagnostic screening tool for stroke)
7	<5	fewer than 5 subjects (total participants reported, including cases and controls)
8	only errors	the manuscript only has data on errors (numerator) without data on population from which patients were drawn (denominator) (e.g., non-consecutive case series; medico-legal cases not drawn from a known population)
9	pre-1995 data	study data were collected before 1995

e-Appendix 2: Included studies used in qualitative synthesis and quantitative meta-analysis (n=23)

The included studies were conducted in eight countries across four continents. Derivation of the study populations varied, but was split about evenly between disease cohorts (which included only patients with the target disorder [i.e., acute cerebrovascular event], n=12) and clinical referral cohorts (which included patients with suspicion of the target disorder who were sent for further evaluation either by consultation (n=4) or admission (n=5)). The remaining two studies were therapeutic (n=2) cohorts. More than half (n=13) were conducted at a single hospital, but eight were conducted at multiple hospitals (five of these representing population-based, regional samples). Data collection was prospective in six and, in the others, either retrospective or hybrid (e.g., prospective patient enrollment of an admission cohort where initial diagnoses were recorded retrospectively based on chart review after the patient was enrolled). Although several studies analyzed events over time (e.g., prior visits where diagnoses were missed), all studies took a cross-sectional approach to data analysis on the primary meta-analytic variables related to diagnostic accuracy.

e-Table 1. Meta-data of included studies stratified by primary cerebrovascular population and listed in alphabetical order

Author, Year (Citation)	Cerebrovascular Study Population (Derivation)	Setting (Country)	Data Collection (Analysis)	Reference Standard (Level, Provider)	ED Subjects (% Female)	Mean Age * (Standard Deviation, Range)	Special Comments
SUBARARCHNOID HEMORRHAGE (SAH)							
Kowalski et al (2004) [14] †, ‡	atraumatic SAH (disease cohort)	NICU, 1 tertiary care center (USA)	Hybrid (cross-sectional)	Imaging (CT) or LP in correct clinical context (high, specialist)	207 (68%)	NR (NR, 58% > 50)	Mix of patients initially seen in ED as well as outpatient clinics.
Mayer et al (1996)[15] †, ‡	aneurysmal SAH (disease cohort)	Neurosurgical service, 4 tertiary care centers (USA)	Retrospective (cross-sectional)	Imaging (CT) or LP in correct clinical context (high, specialist)	89 (61%)	52 (NR, 20-96)	Mix of patients initially seen in ED as well as outpatient clinics.
Vermeulen and Schull (2007)[26]	atraumatic SAH (disease cohort from regional EHR)	176 hospital EDs in a single province (Canada)	Retrospective (cross-sectional)	Discharge diagnosis (moderate, NR)	1507 (62%)	58 (~15, NR)	Administrative data study from a large regional health information exchange in Ontario, Canada.
ISCHEMIC STROKE (IS) or TRANSIENT ISCHEMIC ATTACK (TIA) §							
Jeng et al (1998)[27] †	IS >> ICH >> SAH > TIA (admission cohort)	Multiple units, 1 university hospital (Taiwan)	Retrospective (cross-sectional)	Imaging (CT) in correct clinical context (moderate, specialist)	2533 (41%)	63 (16, NR)	
Kerber et al (2006)[8] †, ¶	IS > TIA >> ICH with dizziness (disease cohort)	7 community hospital EDs in a single county (USA)	Retrospective (cross-sectional)	Validated medical record audit (moderate, specialist)	1629 (64%)	69 (12, 45-75+)	Study reported 1 hemorrhage not further specified (SAH or ICH).
Morgenstern et al (2004)[9] ¶, ‖	IS > TIA >> ICH > SAH (disease cohort)	7 community hospital EDs in a single county (USA)	Retrospective (cross-sectional)	Validated medical record audit (moderate, specialist)	2059 (58%)	NR (NR, 20% 45-59; 35% 60-74; 45% 75+)	There were incomplete data on FN, therefore specificity could not be determined. The study reported 211 hemorrhages that were not further specified (SAH or ICH).
Moulin et al (2003)[28]	IS or TIA (consultative cohort)	Neurology consult service, 1 university hospital (France)	Prospective (cross-sectional)	Consultant diagnosis (moderate, specialist)	1679 (48%)	57 (21, NR)	No breakdown between IS and TIA provided.
Nakajima et al (2008)[13] †	IS or TIA (disease cohort)	Neurology ward, 1 university hospital (Japan)	Retrospective (cross-sectional)	Imaging (CT) in correct clinical context (moderate, specialist)	299 (42%)	74 (11, NR)	No breakdown between IS and TIA provided.
Whiteley et al (2011)[29]	IS>> TIA > ICH > SAH (consultative cohort)	University hospital ED (UK)	Prospective (cross-sectional)	Panel of experts (moderate, specialists with independent review)	356 (51%)	72 (14, NR)	

ISCHEMIC STROKE (IS) §							
Arch et al (2016)[17]	IS (disease cohort)	1 community hospital (USA)	Retrospective (cross-sectional)	Imaging (CT or MRI) in correct clinical context (moderate, specialist)	185 (55%)	72 (FN) and 74 (TP) (NR, NR)	
Bhattacharya et al (2013)[12]	IS (disease cohort)	Multiple community hospitals and 1 university hospital (USA)	Prospective (cross-sectional)	Imaging (CT or MRI) in correct clinical context (high, specialist)	77 (57%)	38 (NR, NR)	
Broadley and Thompson (2003)[30]	IS >> ICH > TIA (admission cohort)	Stroke unit, 1 tertiary care center (Australia)	Hybrid (cross-sectional)	Imaging (CT or MRI) in correct clinical context (moderate, specialist)	284 (44%)	72 (NR, 20-100)	Reported age only for confirmed strokes (n=245).
Ferro et al (1998)[31] †	IS >> ICH (consultative cohort)	Neurology consult service, 3 large hospital EDs (Portugal)	Prospective (cross-sectional)	Consultant diagnosis (moderate, specialist with non-independent 2nd review)	185 (56%)	67 (17, 27-88)	Reported only the demographics of those misdiagnosed.
Harbison et al (2003)[32] †	IS >TIA> ICH (admission cohort)	Stroke unit, 1 secondary care center (UK)	Hybrid (cross-sectional)	Imaging (CT most) in correct clinical context (moderate, specialist)	93 (52%)	72 (NR, 22-98)	
Lever et al (2013)[33]	IS (disease cohort)	Single teaching hospital ED (USA)	Retrospective (cross-sectional)	Hospital discharge diagnosis (high, specialist)	189 (50%)	70 (16, 20-99)	
Leys et al (1997)[34]	IS >> ICH ~ TIA >> SAH (admission cohort)	Stroke unit, 1 university hospital (France)	Hybrid (cross-sectional)	Discharge diagnosis (moderate, specialist)	1245 (51%)	66 (NR, 15-101)	
Mohamed et al (2013)[35]	IS (disease cohort)	Numerous community hospitals and one academic medical center (USA)	Retrospective (cross-sectional)	Consultant diagnosis including imaging in most (moderate, specialist)	61 (NR)	38 (8, NR)	
Richoz et al (2015)[16] **	IS (disease cohort)	University hospital ED (CH)	Retrospective (cross-sectional)	Imaging (CT or MRI) in correct clinical context (moderate, specialist)	2200 (43.8%)	72.6 [median] (21.2 [interquartile range])	In this study only median age and "interquartile range" (not further defined) were reported.
Scott and Silbergleit (2003)[36]	IS >> TIA treated with thrombolytics (tPA) (therapeutic cohort)	4 mixed university/ community hospital EDs in a single region (USA)	Hybrid (cross-sectional)	Discharge diagnosis, hospital record (moderate, emergency physician)	151 (NR)	66 (~2, NR)	
Uchino et al (2010)[10] †	IS >> TIA treated with thrombolytics (tPA) (therapeutic cohort)	Stroke service, 1 tertiary care center (USA)	Retrospective (cross-sectional)	Imaging (CT or MRI) in correct clinical context (moderate, specialist)	133 (NR)	71 (NR, NR)	
TRANSIENT ISCHEMIC ATTACK (TIA) §							
Ferro et al (1996)[37] †	TIA (consultative cohort)	Neurology consult service, 3 large hospital EDs (Portugal)	Prospective (cross-sectional)	Consultant diagnosis (moderate, specialist with non-independent review)	31 (35%)	NR (NR, 65% ≤ 70)	10 cases with initial ED diagnosis of TIA were later confirmed to be IS. In the study, these were considered misdiagnoses, but we gave credit to ED physicians for these 10 cases and counted them as true positive cases in addition to the 4 confirmed TIA cases.
Prabhakaran et al (2008)[38]	TIA (admission cohort)	Stroke service, 1 tertiary care center (USA)	Prospective (cross-sectional)	Imaging (MRI most) in correct clinical context OR consultant diagnosis (moderate, specialist x 2)	100 (60%)	61 (~16, NR)	
Schrock et al (2012)[11]	TIA (disease cohort)	Stroke service, 1 county-owned hospital (USA)	Retrospective (cross-sectional)	Consultant diagnosis (moderate, specialist with non-independent review)	429 (62%)	~59 (NR, NR)	

ABBREVIATIONS: ED – emergency department; EHR – electronic health record; FN – false negatives; ICH – intracerebral hemorrhage; IS – ischemic stroke; LP – lumbar puncture; NICU – neurological intensive care unit; NR – not reported; SAH – subarachnoid hemorrhage; TIA – transient ischemic attack; TP – true positives.

* Patient age was not uniformly reported across studies. For studies only reporting ages separately for disease subgroups, we report a combined weighted meta-mean and approximate (~) standard deviation. For studies without other measures of age distribution, we report the proportion in each age bin as available. Additional clarifications about age reporting are noted (e.g., use of median rather than mean age).

† Referrals from general practitioners or other providers were admixed with those from emergency physicians in these original reports. Based upon data from the manuscript or direct communication with the corresponding authors, we were able to entirely segregate ED diagnoses from other frontline provider misdiagnoses, except as noted below.

‡ In these studies (Kowalski et al (2004);[14] Mayer et al (1996)),[15] we were unable to determine a precise number of SAH patients seen in the ED because studies reported only the total number of patients and the fraction misdiagnosed in the ED, but not the number seen in the ED or rate of ED misdiagnosis. Values here are imputed based on the assumption that the rate of misdiagnosis was equivalent across sites (ED, physician office, etc.). Subgroup analyses (including proportion female and SAH grade) employ the same assumption.

§ Studies were assigned to the "IS or TIA" section if no separate numbers for TIA and stroke were provided, or if both diagnoses were of similar frequency within the study population. Studies were assigned to the "IS" section or the "TIA" section if one diagnosis (IS or TIA) was dominating (e.g., IS >> TIA), otherwise the section "IS or TIA" was selected.

¶ The BASIC stroke surveillance project employs a combined active and passive surveillance strategy using trigger-symptom words to prompt structured medical record review. Although the emphasis of surveillance is a population-based disease cohort (all stroke patients), this approach enables ED symptom cohorts to be defined as done in Kerber et al (2006) [8] for all patients presenting with dizziness or vertigo.

|| After contacting the corresponding author, 51 patients were removed because of overlap with a different study from the same study group (Kerber et al (2006)).[8]

** In this study median age and inter-quartile range was reported.[16]

e-Appendix 3: Assessment of quality of diagnostic reference standard and QUADAS-2 assessment

Quality of evidence was assessed with respect to our primary outcome measures for diagnostic accuracy. The diagnostic reference standard for cerebrovascular disease was assessed as high, medium, or low (eBox 1) by a single rater (AAT) and confirmed by a second (DNT). Studies considered to have a low diagnostic reference standard were excluded during full-text assessment (Figure 1, main document).

e-Box 1. Definitions for quality of diagnostic reference standard

- High: Objective data (e.g., CT or LP or conventional catheter angiography for SAH; MRI-DWI for ischemic stroke, CT for intracranial hemorrhage) available for >80% of patients.
- Medium: Final diagnosis based on aggregate of clinical findings but a standardized process and specific definitions were applied.
- Low: All studies falling short of high or medium diagnostic reference standard.

ABBREVIATIONS: CT – computed tomography; LP – lumbar puncture; MRI-DWI – magnetic resonance imaging with diffusion weighted images; SAH – subarachnoid hemorrhage

QUADAS-2 assessment of included studies

For included studies, two independent raters (DNT/AAT) further assessed the risk of bias or applicability concerns using QUADAS-2[39] tailored study criteria, resolving disagreements by discussion. Inclusion was not restricted further based upon QUADAS-2 results. We chose this approach because we suspected from the outset that studies available to test our research hypothesis were of relatively limited number.

The QUADAS-2 tool for quality rating of diagnostic accuracy studies consists of four core domains (patient selection, index test, reference standard, and flow and timing).[39] Risk of bias is assessed for all four domains, and applicability is assessed for the first three domains. Thus, seven items per study are assessed to rate quality. For each item, pre-specified conditions must be met to qualify for "low risk" of bias. The tool's authors recommend review-specific tailoring of the rating criteria and process,[39] which we did for three specific items (e-Box 2).

e-Box 2. Tailored rating criteria for key domains of bias or applicability concerns

- Reference standard (bias): Studies were rated as *high* risk if all of the following criteria were met: (1) more than 25% of misdiagnoses had no independent, objective gold standard test (e.g., stroke without neuroimaging or any TIA); (2) the team was neurology or stroke specialist-based; (3) the focus or emphasis of the study was misdiagnosis; and (4) the team providing the reference standard was not masked to the index test result (e.g., ED physician diagnosis).
- Patient selection (applicability concerns): Studies were rated as *high* risk if a clinical or demographic subgroup that was clearly at higher (dizziness presentation or young patients only) or lower (rtPA treatment) risk of misdiagnosis than average was specifically studied. Studies were rated as *low* risk if a disease subgroup (e.g., TIA or SAH) was studied.
- Index test (applicability concerns): Studies were rated as *high* risk if ED diagnoses were known to consistently include input from a provider of lesser stroke training presumed to reduce accuracy (ED resident rather than attending; nurse, physician's assistant, paramedic) or of greater stroke training presumed to increase accuracy (neurology resident or attending). Studies were rated *unclear* if it was not reported or "most" diagnoses were ED physician diagnoses.

ABBREVIATIONS: rtPA – recombinant tissue plasminogen activator; SAH – subarachnoid hemorrhage; TIA – transient ischemic attack

QUADAS-2 results are shown in e-Table 2 using a minor modification of the recommended presentation technique that highlights the directional risk of bias or applicability concerns with respect to the primary meta-analytic outcome (i.e., diagnostic accuracy / misdiagnosis).[39]

e-Table 2. QUADAS-2 quality ratings for included studies (n=23)

		Risk of Bias				Applicability Concerns		
Citation	**CVD (n)**	**Patient Selection**	**Index Test**	**Reference Standard**	**Flow & Timing**	**Patient Selection**	**Index Test**	**Reference Standard**
Arch et al (2016)[17]	185	-	-	-	-	-	?	
Bhattacharya et al (2013)[12]	77	-	-	-	-	↑ (missed stroke)	-	-
Broadley and Thompson (2003)[30]	253	-	-	-	-	-	-	-
Ferro et al (1996)[37]	14	-	-	↑ (mis-Dx stroke)	-	-	-	-
Ferro et al (1998)[31]	169	-	-	↑ (mis-Dx stroke)	-	-	-	-
Harbison et al (2003)[32]	66	-	-	-	-	-	-	-
Jeng et al (1998)[27]	2226	-	-	-	-	-	?	-
Kerber et al (2006)[8]	46	-	-	-	-	↑ (missed stroke)	-	-
Kowalski et al (2004)[14]	207	-	-	-	-	-	-	-
Lever et al (2013)[33]	189	-	-	-	-	-	?	-
Leys et al (1997)[34]	1071	-	-	-	-	-	↓ (mis-Dx stroke)	-
Mayer et al (1996)[15]	89	-	-	-	-	-	-	-
Mohamed et al (2013)[35]	61	-	-	-	-	↑ (missed stroke)	?	-
Morgenstern et al (2004)[9]	1800	-	-	-	-	-	-	-
Moulin et al (2003)[28]	567	-	-	?	-	-	-	-
Nakajima et al (2008)[13]	299	-	-	-	-	-	-	-
Prabhakaran et al (2008)[38]	40	-	-	-	-	-	↓ (mis-Dx stroke)	-
Richoz et al (2015)[16]	2200	-	-	-	-	-	?	-
Scott and Silbergleit (2003)[36]	145	-	-	-	-	↓ (missed stroke)	-	-
Schrock et al (2012)[11]	273	-	-	-	-	-	-	-
Uchino et al (2009)[10]	125	-	-	-	-	↓ (missed stroke)	-	-
Vermeulen and Schull (2007)[26]	1507	-	-	-	-	-	-	-
Whiteley et al (2011)[29]	246	-	-	?	-	-	?	-

ABBREVIATIONS: mis-Dx – misdiagnosis

KEY: Possible ratings for risk of bias or applicability concerns were *low* ("-"), *high* (either resulting in over-estimation (↑) or under-estimation (↓) of the missed / misdiagnosed stroke rate) or *unclear* ("?").

Among 23 included studies, the risk of bias or applicability concerns was *low* for all seven criteria in eight studies. In the remaining 15 studies one (n=14) or two (n=1) of seven items were rated as either *high* or *unclear*. There were 9 studies for which one item was rated as *high* (risk of bias for the reference standard in 2; applicability concerns regarding patient selection in 5; applicability concerns regarding index test in 2). There were 7 studies for which one or more items were rated as *unclear* (risk of bias for reference standard in 2; applicability concerns regarding index test in 6).

Two studies scored *high* risk of bias for the reference standard because of review of cases by consensus stroke rounds, leading to potential over-estimation of misdiagnosis in stroke rate.[31, 37] One study scored *unclear* because likely more than 25% of stroke misdiagnoses had no firm, independent gold standard (e.g., TIA, or stroke without neuroimaging) (Moulin et al (2003));[28] another because numbers on imaging on patients with misdiagnosis were missing (Whiteley et al (2011)).[29] Among five studies scored *high* risk for applicability concerns for patient selection, three focused on special patient subgroups that were likely at increased risk for missed stroke because of clinical presentation (dizziness only – Kerber et al (2006))[8] or young age (Bhattacharya et al (2013);[12] Mohammed et al (2013));[35] two included patients that were likely at decreased risk for missed stroke because of presumably more obvious clinical presentations leading to treatment with tPA (Scott, and Silbergleit (2003),[36] Uchino et al (2009)).[10] Two studies scored *high* risk for applicability concerns for the index test due to involvement of neurology residents in all ED diagnoses (Leys et al (1997);[34] Prabhakaran et al (2008));[38] six scored *unclear* because neurology residents (Arch et al (2016);[17] Jeng et al (1998);[27] Mohamed et al (2013))[35] or nurses (Lever et al (2013);[33] Richoz et al (2015);[16] Whiteley et al (2011))[29] were involved in "some" cases.

e-Appendix 4: Misdiagnosis rates and diagnostic accuracy by primary study population

e-Table 3: ED diagnostic accuracy by primary study population and in alphabetical order

Citation	ED Subjects	Misdiagnosis 'n' (Proportion [%])	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
SUBARARCHNOID HEMORRHAGE (SAH)						
Kowalski et al (2004)[14] *	207	24 (11.6)	88.4	-	-	-
Mayer et al (1996)[15] *	89	22 (24.7)	75.3	-	-	-
Vermeulen and Schull (2007)[26]	1507	81 (5.4)	94.6	-	-	-
ISCHEMIC STROKE (IS) or TRANSIENT ISCHEMIC ATTACK (TIA) †						
Jeng et al (1998)[27] ‡	2533	307 (12.1)	-	-	87.9	-
Kerber et al (2006)[8]	1629	52 (3.2)	65.2	97.7	45.5	99.0
Morgenstern et al (2004)[9] §	2059	354 (17.2)	91.5	-	89.1	-
Moulin et al (2003)[28]	1679	259 (15.4)	76.2	88.8	77.7	88.0
Nakajima et al (2008)[13] *	299	16 (5.4)	94.6	-	-	-
Whiteley et al (2011)[29]	356	123 (34.6)	76.3	59.1	80.6	52.8
ISCHEMIC STROKE (IS) †						
Arch et al (2016)[17]	185	48 (25.9)	74.1	-	-	-
Bhattacharya et al (2013)[12]	77	11 (14.3)	85.7	-	-	-
Broadley and Thompson (2003)[30]	284	31 (10.9)	-	-	89.1	-
Ferro et al (1998)[31] *	185	16 (8.6)	-	-	91.4	-
Harbison et al (2003)[32] *	93	27 (29.0)	-	-	71.0	-
Lever et al (2013)[33]	189	29 (15.3)	84.7	-	-	-
Leys et al (1997)[34]	1245	174 (14.0)	-	-	86.0	-
Mohamed et al (2013)[35] *	61	11 (18.0)	82.0	-	-	-
Richoz et al (2015)[16]	2200	43 (2.0)	98.0	-	-	-
Scott and Silbergleit (2003)[36]	151	6 (4.0)	-	-	96.0	-
Uchino et al (2009)[10] *	133	8 (6.0)	-	-	94.0	-
TRANSIENT ISCHEMIC ATTACK (TIA) †						
Ferro et al (1996)[37] *,¶	31	17 (54.8)	-	-	45.2	-
Prabhakaran et al (2008)[38]	100	60 (60.0)	-	-	40.0	-
Schrock et al (2012)[11]	429	156 (36.4)	-	-	63.6	-

ABBREVIATIONS: ED – emergency department; IS – ischemic stroke; NPV – negative predictive value; PPV – positive predictive value; SAH – subarachnoid hemorrhage; TIA – transient ischemic attack

* Referrals from general practitioners or other providers were admixed with those from emergency physicians in these original reports. Based upon data from the manuscript or direct communication with the corresponding authors, we were able to entirely segregate ED diagnoses from other frontline provider misdiagnoses, except as noted in eTable 1.

† Studies were assigned to the "IS or TIA" section if no separate numbers for TIA and stroke were provided, or if both diagnoses were of similar frequency within the study population. Studies were assigned to the "IS" section or the "TIA" section if one diagnosis (IS or TIA) was dominating (e.g., IS >> TIA), otherwise the section "IS or TIA" was selected.

‡ Examined initially by both ED physician and "neurologist in training" for some[27] or all[34, 38] study subjects. The frequency of ED misdiagnoses might have been influenced (up or down) by the neurology trainees.

§ After contacting the corresponding author, 51 patients were removed because of overlap with a different study from the same study group (Kerber et al (2006)).[8]

¶ Ten cases with an initial ED diagnosis of TIA were later confirmed to be IS. In the study, these were considered misdiagnoses, but here we give credit to ED physicians for correctly diagnosing these 10 cases as cerebrovascular. We count them as true positive cases in addition to the 4 confirmed TIA cases noted in the study report.

e-Appendix 5: Initial misdiagnoses in missed strokes and overcalled stroke mimics

e-Table 4: Initial ED diagnoses in cases of missed cerebrovascular events (false negatives) *, †

Initial Diagnosis	Misdiagnosis 'n' (Proportion [%])	True SAH 'n' (Proportion [%])	True stroke / TIA 'n' (Proportion [%])
migraine and other headaches ‡	92 (26.1)	75 (81.5)	17 (18.5)
vertigo / dizziness	51 (14.5)	0 (0.0)	51 (100.0)
seizures	23 (6.5)	0 (0.0)	23 (100.0)
cognitive dysfunction / dementia	22 (6.3)	0 (0.0)	22 (100.0)
peripheral nerve disorders	22 (6.3)	0 (0.0)	22 (100.0)
sepsis/infection or confusion	22 (6.3)	0 (0.0)	22 (100.0)
meningitis or encephalitis	11 (3.1)	5 (45.5)	6 (54.5)
neck pain	11 (3.1)	11 (100.0)	0 (0.0)
hypertensive attack	7 (2.0)	7 (100.0)	0 (0.0)
sinusitis	6 (1.7)	6 (100.0)	0 (0.0)
viral syndrome	6 (1.7)	6 (100.0)	0 (0.0)
syncope	6 (1.7)	3 (50.0)	3 (50.0)
psychiatric disorders	6 (1.7)	0 (0.0)	6 (100.0)
subdural hemorrhage / SAH	5 (1.4)	0 (0.0)	5 (100.0)
musculoskeletal disorders	4 (1.1)	4 (100.0)	0 (0.0)
head trauma	3 (0.9)	0 (0.0)	3 (100.0)
arteriovenous malformation	2 (0.6)	2 (100.0)	0 (0.0)
inflammatory brain disease (including MS)	2 (0.6)	0 (0.0)	2 (100.0)
cerebral venous sinus thrombosis	2 (0.6)	0 (0.0)	2 (100.0)
brain tumor	2 (0.6)	0 (0.0)	2 (100.0)
coma of unclear origin	2 (0.6)	0 (0.0)	2 (100.0)
substance abuse	2 (0.6)	0 (0.0)	2 (100.0)
gastrointestinal disease	2 (0.6)	0 (0.0)	2 (100.0)
decompensated diabetes	1 (0.3)	0 (0.0)	1 (100.0)
aortic dissection	1 (0.3)	0 (0.0)	1 (100.0)
porphyria	1 (0.3)	0 (0.0)	1 (100.0)
other §	38 (10.8)	15 (39.5)	23 (60.5)
Total	**352 (100)**	**134 (38.1)**	**218 (61.9)**

ABBREVIATIONS: MS – multiple sclerosis; SAH – subarachnoid hemorrhage; TIA – transient ischemic attack

* Numbers were available from a total of 5 studies[8, 14, 16, 26, 28] reporting on 332 patients. In one study reporting on missed stroke cases, numbers from cases misdiagnosed by the emergency department physician (n=43) could not be separated from numbers of cases misdiagnosed by the neurologist (n=4), therefore we included these four cases here as well (Richoz et al (2015)).[16] Notably, in this same study, one or more initial diagnoses were allowed.

† In individual studies, other atypical or non-specific stroke presentations such as headache, generalized weakness, confusion, altered gait, and involuntary movements were linked to increased risk of stroke misdiagnosis.[11, 33, 38]

‡ Migraine headaches as opposed to other headache types were not separately reported in 3 studies reporting on 87 patients.[14, 26, 28] Headache diagnoses included migraine (n=3), hypertensive headache (n=2), cluster headache (n=1), headache of uncertain etiology (n=3), sinus headache (n=2), and tension-type headache (n=2).

§ In the source studies, there was no specific diagnosis provided in these 'other' cases.

e-Table 5: Final diagnosis in ED stroke / TIA mimics (false positives) *

Final Diagnosis	Misdiagnosis 'n' (Proportion [%])
seizures	161 (16.7)
vertigo / dizziness	91 (9.4)
migraine	78 (8.1)
metabolic encephalopathy †	77 (8.0)
brain tumor	73 (7.6)
non-vascular acute headache	53 (5.5)
psychiatric disorders ‡	44 (4.6)
cognitive dysfunction / dementia	44 (4.6)
syncope	43 (4.5)
sepsis/infection or confusion	26 (2.7)
meningitis or encephalitis	17 (1.8)
peripheral nerve disorders	11 (1.1)
transient global amnesia	4 (0.4)
acute peripheral artery occlusion	3 (0.3)
deep venous thrombosis	2 (0.2)
hypertensive attack	2 (0.2)
glaucoma	1 (0.1)
posttraumatic amnesia	1 (0.1)
other §	235 (24.3)
Total	**966 (100)**

* Numbers derive from a total of 12 studies reporting on 961 patients.[8, 10, 11, 27, 28, 30-32, 34, 36-38] None of the studies reported on SAH mimics; results represent only stroke and TIA mimics. In one study more than one diagnosis was allowed, resulting in 179 diagnoses in 174 patients (Leys et al (1997)).[34] One study provided only aggregate numbers of diagnoses on false negative and true negative subjects admixed; numbers from this study were not considered here (Whiteley et al (2011).[29]

† Underlying causes of metabolic encephalopathy included drug-induced encephalopathy (n=19), hypoglycemia (n=11), hyponatremia (n=11), encephalopathy of undetermined origin (n=9), post-anoxic encephalopathy (n=6), hyperglycemia (n=4), Wernicke encephalopathy (n=4), hypercalcemia (n=3), uremic encephalopathy (n=3), alcohol intoxication (n=2), carbon monoxide intoxication (n=1), dehydration (n=1). In the remaining cases (n=3) the cause was not further specified.

‡ Conversion disorder was the largest fraction of psychiatric final diagnoses (n=36).

§ No specific diagnoses were provided in these cases.

References

1. Chalela JA, Kidwell CS, Nentwich LM, et al. Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. Lancet 2007;369:293-298.
2. Steiger HJ. Subarachnoid hemorrhage. A common misdiagnosis! Remember that emergency treatment and immediate hospitalization saves lives. MMW-Fortschr Med 2000;142:40-64.
3. Grewal K, Austin PC, Kapral MK, Lu H, Atzema CL. Missed strokes using computed tomography imaging in patients with vertigo: population-based cohort study. Stroke 2015;46:108-113.
4. Fahimi J, Navi BB, Kamel H. Potential misdiagnoses of Bell's palsy in the emergency department. Ann Emerg Med 2014;63:428-434.
5. Kose A, Inal T, Armagan E, Kiyak R, Demir AB. Conditions that mimic stroke in elderly patients admitted to the emergency department. J Stroke Cerebrovasc Dis 2013;22:e522-527.
6. Newman-Toker DE, Moy E, Valente E, Coffey R, Hines AL. Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. Diagnosis 2014;1:155-166.
7. Kuruvilla A, Bhattacharya P, Rajamani K, Chaturvedi S. Factors associated with misdiagnosis of acute stroke in young adults. J Stroke Cerebrovasc Dis 2011;20:523-527.
8. Kerber KA, Brown DL, Lisabeth LD, Smith MA, Morgenstern LB. Stroke among patients with dizziness, vertigo, and imbalance in the emergency department: a population-based study. Stroke 2006;37:2484-2487.
9. Morgenstern LB, Lisabeth LD, Mecozzi AC, et al. A population-based study of acute stroke and TIA diagnosis. Neurology 2004;62:895-900.
10. Uchino K, Massaro L, Hammer MD. Transient ischemic attack after tissue plasminogen activator: aborted stroke or unnecessary stroke therapy? Cerebrovasc Dis 2010;29:57-61.
11. Schrock JW, Glasenapp M, Victor A, Losey T, Cydulka RK. Variables associated with discordance between emergency physician and neurologist diagnoses of transient ischemic attacks in the emergency department. Ann Emerg Med 2012;59:19-26.
12. Bhattacharya P, Nagaraja N, Rajamani K, Madhavan R, Santhakumar S, Chaturvedi S. Early use of MRI improves diagnostic accuracy in young adults with stroke. J Neurol Sci 2013;324:62-64.
13. Nakajima M, Hirano T, Uchino M. Patients with acute stroke admitted on the second visit. J Stroke Cerebrovasc Dis 2008;17:382-387.
14. Kowalski RG, Claassen J, Kreiter KT, et al. Initial misdiagnosis and outcome after subarachnoid hemorrhage. JAMA 2004;291:866-869.
15. Mayer PL, Awad IA, Todor R, et al. Misdiagnosis of symptomatic cerebral aneurysm. Prevalence and correlation with outcome at four institutions. Stroke 1996;27:1558-1563.
16. Richoz B, Hugli O, Dami F, Carron PN, Faouzi M, Michel P. Acute stroke chameleons in a university hospital: Risk factors, circumstances, and outcomes. Neurology 2015;85:505-511.
17. Arch AE, Weisman DC, Coca S, Nystrom KV, Wira CR, 3rd, Schindler JL. Missed Ischemic Stroke Diagnosis in the Emergency Department by Emergency Medicine and Neurology Services. Stroke 2016;47:668-673.
18. Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford: Oxford University Press, 2003.
19. Westover MB, Westover KD, Bianchi MT. Significance testing as perverse probabilistic reasoning. BMC Med 2011;9:20.
20. Positive and negative predictive values [online]. Available at: https://en.wikipedia.org/wiki/Positive_and_negative_predictive_values. Accessed June 20th 2016.
21. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539-1558.
22. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-560.
23. Cochran WG. The combination of estimates from different experiments. Biometrics 1954;10:101-129.
24. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009;6:e1000100.
25. Cohen J. A coefficient for agreement for nominal scales. Educ Psychol Meas 1960;20:37-46.
26. Vermeulen MJ, Schull MJ. Missed diagnosis of subarachnoid hemorrhage in the emergency department. Stroke 2007;38:1216-1221.
27. Jeng J-S, Huang Z-S, Chang Y-C, et al. Misdiagnosis of acute cerebrovascular disease: Experience from a hospital-based stroke registry in Taiwan (SCAN-V). Acta Neurol Taiwan 1998;7:185-192.
28. Moulin T, Sablot D, Vidry E, et al. Impact of emergency room neurologists on patient management and outcome. Eur Neurol 2003;50:207-214.
29. Whiteley WN, Wardlaw JM, Dennis MS, Sandercock PA. Clinical scores for the identification of stroke and transient ischaemic attack in the emergency department: a cross-sectional study. J Neurol Neurosurg Psychiatry 2011;82:1006-1010.
30. Broadley SA, Thompson PD. Time to hospital admission for acute stroke: an observational study. Med J Aust 2003;178:329-331.

31. Ferro JM, Pinto AN, Falcao I, et al. Diagnosis of stroke by the nonneurologist. A validation study. Stroke 1998;29:1106-1109.
32. Harbison J, Hossain O, Jenkinson D, Davis J, Louw SJ, Ford GA. Diagnostic accuracy of stroke referrals from primary care, emergency room physicians, and ambulance staff using the face arm speech test. Stroke 2003;34:71-76.
33. Lever NM, Nystrom KV, Schindler JL, Halliday J, Wira C, 3rd, Funk M. Missed opportunities for recognition of ischemic stroke in the emergency department. J Emerg Nurs 2013;39:434-439.
34. Leys D, Lucas C, Devos D, et al. Misdiagnoses in 1250 Consecutive Patients Admitted to an Acute Stroke Unit. Cerebrovasc Dis 1997;7:284-288.
35. Mohamed W, Bhattacharya P, Chaturvedi S. Early access to a neurologist reduces the rate of missed diagnosis in young strokes. J Stroke Cerebrovasc Dis 2013;22:e332-337.
36. Scott PA, Silbergleit R. Misdiagnosis of stroke in tissue plasminogen activator-treated patients: characteristics and outcomes. Ann Emerg Med 2003;42:611-618.
37. Ferro JM, Falcao I, Rodrigues G, et al. Diagnosis of transient ischemic attack by the nonneurologist. A validation study. Stroke 1996;27:2225-2229.
38. Prabhakaran S, Silver AJ, Warrior L, McClenathan B, Lee VH. Misdiagnosis of transient ischemic attacks in the emergency room. Cerebrovasc Dis 2008;26:630-635.
39. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529-536.