Cochrane Library

Cochrane Database of Systematic Reviews

Prehospital stroke scales as screening tools for early identification of stroke and transient ischemic attack (Review)

Zhelev Z, Walker G, Henschke N, Fridhandler J, Yip S

Zhelev Z, Walker G, Henschke N, Fridhandler J, Yip S.
Prehospital stroke scales as screening tools for early identification of stroke and transient ischemic attack.
Cochrane Database of Systematic Reviews 2019, Issue 4. Art. No.: CD011427.
DOI: 10.1002/14651858.CD011427.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Prehospital stroke scales as screening tools for early identification of stroke and transient ischemic attack

Zhivko Zhelev[1], Greg Walker[2], Nicholas Henschke[3], Jonathan Fridhandler[2], Samuel Yip[2]

[1]NIHR CLAHRC South West Peninsula (PenCLAHRC), University of Exeter Medical School, University of Exeter, Exeter, UK. [2]Department of Neurology, University of British Columbia, Vancouver, Canada. [3]Cochrane Response, Cochrane, London, UK

Contact address: Greg Walker, Department of Neurology, University of British Columbia, Vancouver General Hospital, Vancouver, BC, Canada. gregorywalkerubc@gmail.com.

Editorial group: Cochrane Stroke Group.
Publication status and date: New, published in Issue 4, 2019.

Citation: Zhelev Z, Walker G, Henschke N, Fridhandler J, Yip S. Prehospital stroke scales as screening tools for early identification of stroke and transient ischemic attack. *Cochrane Database of Systematic Reviews* 2019, Issue 4. Art. No.: CD011427. DOI: 10.1002/14651858.CD011427.pub2.

ABSTRACT

Background

Rapid and accurate detection of stroke by paramedics or other emergency clinicians at the time of first contact is crucial for timely initiation of appropriate treatment. Several stroke recognition scales have been developed to support the initial triage. However, their accuracy remains uncertain and there is no agreement which of the scales perform better.

Objectives

To systematically identify and review the evidence pertaining to the test accuracy of validated stroke recognition scales, as used in a prehospital or emergency room (ER) setting to screen people suspected of having stroke.

Search methods

We searched CENTRAL, MEDLINE (Ovid), Embase (Ovid) and the Science Citation Index to 30 January 2018. We handsearched the reference lists of all included studies and other relevant publications and contacted experts in the field to identify additional studies or unpublished data.

Selection criteria

We included studies evaluating the accuracy of stroke recognition scales used in a prehospital or ER setting to identify stroke and transient Ischemic attack (TIA) in people suspected of stroke. The scales had to be applied to actual people and the results compared to a final diagnosis of stroke or TIA. We excluded studies that applied scales to patient records; enrolled only screen-positive participants and without complete 2 × 2 data.

Data collection and analysis

Two review authors independently conducted a two-stage screening of all publications identified by the searches, extracted data and assessed the methodologic quality of the included studies using a tailored version of QUADAS-2. A third review author acted as an arbiter. We recalculated study-level sensitivity and specificity with 95% confidence intervals (CI), and presented them in forest plots and in the receiver operating characteristics (ROC) space. When a sufficient number of studies reported the accuracy of the test in the same setting (prehospital or ER) and the level of heterogeneity was relatively low, we pooled the results using the bivariate random-effects model. We plotted the results in the summary ROC (SROC) space presenting an estimate point (mean sensitivity and specificity) with 95% CI and prediction regions. Because of the small number of studies, we did not conduct meta-regression to investigate between-study

heterogeneity and the relative accuracy of the scales. Instead, we summarized the results in tables and diagrams, and presented our findings narratively.

Main results

We selected 23 studies for inclusion (22 journal articles and one conference abstract). We evaluated the following scales: Cincinnati Prehospital Stroke Scale (CPSS; 11 studies), Recognition of Stroke in the Emergency Room (ROSIER; eight studies), Face Arm Speech Time (FAST; five studies), Los Angeles Prehospital Stroke Scale (LAPSS; five studies), Melbourne Ambulance Stroke Scale (MASS; three studies), Ontario Prehospital Stroke Screening Tool (OPSST; one study), Medic Prehospital Assessment for Code Stroke (MedPACS; one study) and PreHospital Ambulance Stroke Test (PreHAST; one study). Nine studies compared the accuracy of two or more scales. We considered 12 studies at high risk of bias and one with applicability concerns in the patient selection domain; 14 at unclear risk of bias and one with applicability concerns in the reference standard domain; and the risk of bias in the flow and timing domain was high in one study and unclear in another 16.

We pooled the results from five studies evaluating ROSIER in the ER and five studies evaluating LAPSS in a prehospital setting. The studies included in the meta-analysis of ROSIER were of relatively good methodologic quality and produced a summary sensitivity of 0.88 (95% CI 0.84 to 0.91), with the prediction interval ranging from approximately 0.75 to 0.95. This means that the test will miss on average 12% of people with stroke/TIA which, depending on the circumstances, could range from 5% to 25%. We could not obtain a reliable summary estimate of specificity due to extreme heterogeneity in study-level results. The summary sensitivity of LAPSS was 0.83 (95% CI 0.75 to 0.89) and summary specificity 0.93 (95% CI 0.88 to 0.96). However, we were uncertain in the validity of these results as four of the studies were at high and one at uncertain risk of bias. We did not report summary estimates for the rest of the scales, as the number of studies per test per setting was small, the risk of bias was high or uncertain, the results were highly heterogenous, or a combination of these.

Studies comparing two or more scales in the same participants reported that ROSIER and FAST had similar accuracy when used in the ER. In the field, CPSS was more sensitive than MedPACS and LAPSS, but had similar sensitivity to that of MASS; and MASS was more sensitive than LAPSS. In contrast, MASS, ROSIER and MedPACS were more specific than CPSS; and the difference in the specificities of MASS and LAPSS was not statistically significant.

Authors' conclusions

In the field, CPSS had consistently the highest sensitivity and, therefore, should be preferred to other scales. Further evidence is needed to determine its absolute accuracy and whether alternatives scales, such as MASS and ROSIER, which might have comparable sensitivity but higher specificity, should be used instead, to achieve better overall accuracy. In the ER, ROSIER should be the test of choice, as it was evaluated in more studies than FAST and showed consistently high sensitivity. In a cohort of 100 people of whom 62 have stroke/TIA, the test will miss on average seven people with stroke/TIA (ranging from three to 16). We were unable to obtain an estimate of its summary specificity. Because of the small number of studies per test per setting, high risk of bias, substantial differences in study characteristics and large between-study heterogeneity, these findings should be treated as provisional hypotheses that need further verification in better-designed studies.

PLAIN LANGUAGE SUMMARY

Accuracy of prehospital stroke scales to identify people with stroke or transient ischemic attack (TIA)

Background

Stroke is a life-threatening medical condition in which brain tissue is damaged. This could be caused by a clot blocking the blood supply to part of the brain or bleeding in the brain. If symptoms resolve within 24 hours without lasting consequences, the condition is called TIA (mini stroke). Effective treatment depends on early identification of stroke and any delays may result in brain damage or death.

Emergency medical services are the first point of contact for people experiencing symptoms suggestive of stroke. Medical responders could identify people with stroke more accurately if they use checklists called stroke recognition scales. Such scales include symptoms and other readily-available information. A positive result on the scale indicates high risk of stroke and the need of urgent specialist assessment. The scales do not differentiate between stroke and TIA; this is done in hospital by a neurologist or stroke physician.

Our objective was to review the research evidence on how accurately stroke recognition scales can detect stroke or TIA when used by paramedics or other prehospital clinicians, who are the first point of contact for people suspected of stroke.

Study characteristics

The evidence is current to 30 January 2018. We included studies assessing the accuracy of stroke recognition scales when applied to adults suspected of stroke out of hospital.

We included 23 studies evaluating the following scales: Cincinnati Prehospital Stroke Scale (CPSS; 11 studies), Recognition of Stroke in the Emergency Room (ROSIER; eight studies), Face Arm Speech Time (FAST; five studies), Los Angeles Prehospital Stroke Scale (LAPSS; five studies), Melbourne Ambulance Stroke Scale (MASS; three studies), Ontario Prehospital Stroke Screening Tool (OPSST; one study), Medic

Prehospital Assessment for Code Stroke (MedPACS; one study) and PreHospital Ambulance Stroke Test (PreHAST; one study). Nine studies compared two or more scales in the same people. The results from five studies were combined to estimate the accuracy of ROSIER in the emergency room (ER) and five studies to estimate the accuracy of LAPSS when used by ambulance clinicians.

Quality of the evidence

Many of the studies were of poor or unclear quality and we could not be sure that their results were valid.

Key results of the accuracy of the evaluated prehospital stroke scales

Studies differed considerably in terms of included participants and other characteristics. As a consequence, studies evaluating the same scale reported variable results.

We combined five studies evaluating ROSIER in the ER and obtained average sensitivity of 88% (88 out of 100 people with stroke/TIA will test positive on ROSIER). We were unable to obtain an estimate of specificity (how many people without stroke/TIA will test negative).

We also combined the results for LAPSS, but the included studies were of poor quality and the results may not be valid. The rest of the scales were evaluated in a smaller number of studies or the results were too variable to be combined statistically.

A small number of studies compared two or more scales when applied to the same participants. Such studies are more likely to produce valid results as the scales are used in the same circumstances. They reported that in the ER, ROSIER and FAST had similar accuracy, but ROSIER was evaluated in more studies. When used by ambulance staff, CPSS identified more people with stroke/TIA in all studies, but also more people without stroke/TIA tested positive.

Conclusion

Current evidence suggests that CPSS should be used by ambulance clinicians in the field. Further research is needed to estimate the proportion of wrong results and whether alternatives scales, such as MASS and ROSIER, which might have comparable sensitivity but higher specificity, should be used instead to achieve better overall accuracy. In the ER, ROSIER should be the test of choice. In a group of 100 people of whom 62 have stroke/TIA, the test will miss on average seven people with stroke/TIA (ranging from three to 16). Because of the small number of studies evaluating the tests in a specific setting, poor quality, substantial differences in study characteristics and variability in results, these findings should be treated with caution and need further verification in better-designed studies.

SUMMARY OF FINDINGS

Summary of findings 1. Prehospital stroke scales as screening tools for early identification of stroke and transient ischemic attack

OBJECTIVES AND METHODS

Review question: what is the absolute and relative (comparative) accuracy of stroke recognition scales used in a prehospital or ER setting to identify people with stroke and TIA?

Inclusion criteria: primary studies evaluating the test accuracy of stroke recognition scales in a prehospital or ER setting. The scales were used to identify stroke and TIA in people suspected of stroke, and the results were compared to a final diagnosis of stroke or TIA made by a neurologist or stroke physician (reference standard). Only studies reporting sufficient data to reconstruct the full 2 × 2 table were included. Studies in which the scales were applied to patient records, or including only scale-positive patients were excluded

Databases searched: CENTRAL, MEDLINE, Embase, Science Citation Index, plus hand-searches of reference lists

Search date: from earliest date possible to 30 January 2018

Methodologic quality assessment: QUADAS-2

Statistical analysis: if appropriate, the bivariate random-effects model was used to pool results

RESULTS

Number of studies included: 23 studies including 9230 participants, range 31–1130 participants, median 312 (IQR 154 to 554)

Number of scales evaluated: 8 scales, CPSS (11 studies), ROSIER (8 studies), FAST (5 studies), LAPSS (5 studies), MASS (3 studies), OPSST (1 study), MedPACS (1 study), PreHAST (1 study)

Setting: 6 studies evaluated the scales in the ER and 17 in a prehospital setting (16 evaluated the tests in the field and 1 in primary care)

Studies comparing scales in the same participants: 9 studies compared ≥ 2 scales in the same patients (3 studies each compared FAST vs ROSIER and CPSS vs MASS, 2 studies each compared ROSIER vs CPSS, LAPSS vs CPSS and LAPSS vs MASS, and 1 study each compared some of the remaining pairs)

Methodologic quality: 12 studies were at high risk of bias and 1 with applicability concerns in the patient selection domain; 14 at unclear risk of bias and 1 with applicability concerns in the reference standard domain; and 1 at high risk of bias and another 16 at unclear risk of bias in the flow and timing domain

CONCLUSIONS: CPSS should be preferred in the field as it had consistently high sensitivity in direct comparisons; further evidence is needed to determine its absolute accuracy and whether alternatives scales, such as MASS and ROSIER, which might have comparable sensitivity but higher specificity, should be used instead to achieve better overall accuracy. In the ER, ROSIER should be the test of choice. In a cohort of 100 people of whom 62 have stroke/TIA, the test will miss on average 7 people with stroke/TIA (range 3–16). We were unable to obtain an estimate of its summary specificity. Because of the small number of studies per test per setting, high risk of bias, substantial differences in study characteristics and large between-study heterogeneity, these findings should be treated as provisional hypotheses that need further verification in better-designed studies.

RESULTS: relative (comparative) accuracy

Considering only the results for which the statistical significance was reported or could be determined from the non-overlapping CIs of the accuracy estimates, the results of the comparative studies could be summarized as follows.

In the ER*:*

- ROSIER vs FAST: no statistically significant difference in sensitivities and specificities.

In the field:

- CPSS vs MASS: no statistically significant difference in sensitivities, but MASS was more specific;
- CPSS vs ROSIER: the specificity of ROSIER was higher (the result for sensitivity was uncertain);

- CPSS vs LAPSS: the difference in sensitivities was statistically significant in favor of CPSS (the difference in specificities was uncertain);
- CPSS vs MedPACS: both the differences in sensitivity and specificity were statistically significant, with CPSS being more sensitive but less specific;
- MASS vs LAPSS: the difference in sensitivities was statistically significant in favor of MASS, but no statistically significant difference in specificities was found.

Additional data from Purrucker 2015 (excluded from the main analysis) contradicted some of these results.

RESULTS: absolute accuracy

Index test	Number of studies	Number of studies at high risk of bias or applicability concerns	Results	Comments
ROSIER	8 (2 in the field, 1 in primary care and 5 in ER)	2 (1 in patient selection and 1 in flow and timing)	**Mean summary sensitivity** 0.88 (95% CI 0.84 to 0.91), prediction region 0.75 to 0.95 Specificity (study-level, range) 0.18 to 0.93	We report only a mean summary estimate for sensitivity, based on 5 studies of relatively good methodologic quality conducted in the ER. It means that in this setting the test will miss on average 12/100 people with stroke/TIA, but this could range from 5 to 25 people. Study-level specificities were extremely heterogeneous and the statistical uncertainty in the summary estimate was too great to allow meaningful clinical interpretation. Across the studies, between 7/100 and 82/100 people without stroke/TIA tested positive.
CPSS	11 (9 in the field, 1 in primary care and 1 in ER)	9 (8 in patient selection and 1 in the applicability of the reference standard)	Sensitivity 0.44 to 0.95 Specificity 0.21 to 0.79	High level of heterogeneity even when analysis restricted to use of CPSS in a prehospital setting by paramedics (7 studies). Across all studies, between 5/100 and 55/100 people with stroke/TIA were missed and between 21/100 and 79/100 without stroke/TIA tested positive
LAPSS	5 studies (prehospital)	4 (patient selection)	Summary sensitivity 0.83 (95% CI 0.75 to 0.89) Summary specificity 0.93 (95% CI 0.88 to 0.96)	According to the obtained summary estimates, the test will miss 17/100 people with stroke/TIA and 7/100 without stroke/TIA will test positive. However, these results should be treated with caution as 4/5 studies were at high risk of selection bias and for most the level of bias in the reference standard and the flow and timing domain could not be fully assessed.
FAST	5 studies (3 in prehospital and 2 in ER)	3 (2 in patient selection and 1 in flow and timing)	Sensitivity 0.64 to 0.97 Specificity 0.13 to 0.92	Heterogeneous results even when results analyzed separately by setting. Across studies the test missed between 3/100 and 36/100 people with stroke/TIA and between 8/100 and 87/100 people without stroke/TIA tested positive.
MASS	3 studies (prehospital)	3 (patient selection)	Sensitivity 0.74 to 0.90 Specificity 0.67 to 0.86	Heterogeneous results from studies at high risk of bias. Across studies, the test missed between 10/100 and 26/100 people with stroke/TIA and between 14/100 and 33/100 people without stroke/TIA tested positive.

OPSST	1 study	1 (patient selection)	Sensitivity 0.92 (95% CI 0.88 to 0.94) Specificity 0.86 (95% CI 0.80 to 0.90)	High risk of selection bias; the focus was on the positive predictive value which was 0.90 (95% CI 0.86 to 0.93). This means that 90/100 people with a positive test had stroke/TIA.
MedPACS	1 study	1 (patient selection)	Sensitivity 0.74 (95% CI 0.67 to 0.80) Specificity 0.33 (95% CI 0.27 to 0.39)	Retrospective data collection. The test missed 26/100 people with stroke/TIA and 67/100 people without stroke/TIA tested positive.
PreHAST	1 study	No quality issues	Sensitivity 1.00 (95% CI 0.87 to 1.00) Specificity 0.40 (95% CI 0.25 to 0.56)	PreHAST was designed for both recognition and severity assessment of stroke in the field; this was a pilot study focusing mainly on the accuracy of the scale to identify people with stroke/TIA. The test missed 0 people with stroke/TIA, but 60/100 people without stroke/TIA tested positive.

CI: confidence interval; CPSS: Cincinnati Prehospital Stroke Scale; ER: emergency room; FAST: Face Arm Speech Time; IQR: interquartile range; LAPSS: Los Angeles Prehospital Stroke Scale; MASS: Melbourne Ambulance Stroke Scale; MedPACS: Medic Prehospital Assessment for Code Stroke; OPSST: Ontario Prehospital Stroke Screening Tool; PreHAST: PreHospital Ambulance Stroke Test; ROSIER: Recognition of Stroke in the Emergency Room; TIA: transient ischemic attack.

BACKGROUND

Worldwide, stroke is the leading cause of death. By 2020, 19 million out of 25 million annual stroke deaths will occur in low- to middle-income countries. Some 88% of these events are ischemic strokes, with the remainder being hemorrhagic strokes. Stroke is also the leading cause of disability with 30% of stroke survivors requiring life-long assistance with their activities of daily living, 20% requiring assistance with ambulation and 16% requiring institutional levels of care (Daroff 2012). Ischemic stroke is caused by blockage of blood flow by thrombi, which are blood clots made of platelets, lipids, clotting factors and fibrin. Fibrin is the particular substrate of the thrombolytic, tissue plasminogen activator (tPA), which is a standard of care treatment for certain people with stroke. Failure to restore blood flow in a timely fashion results in an ischemic stroke, and infarction of brain tissue.

A transient ischemic attack (TIA) is an episode of neurologic deficit that reverses without any clinical evidence of neuronal damage. TIAs are prognosticators for future strokes and also require rapid identification, so that physicians can confirm whether or not the symptoms have resolved and then work towards early risk stratification, which has been shown to decrease recurrence (Amarenco 2008). The reference standard for diagnosis of stroke and TIA is the evaluation by a neurologist or stroke physician upon review of history, physical exam and a non-contrast brain computed tomography (CT) scan.

Intravenous tPA was the only approved treatment of acute ischemic stroke up until 2015. Utilization of intravenous tPA is limited by its time sensitivity and this medication can only be provided within a window of 0 to 4.5 hours after onset of symptoms. Data from multicenter randomized controlled trials (RCTs) of intravenous tPA treatment in acute stroke have shown that the odds of a favorable outcome at three months increased as onset-to-treatment time decreased (Hacke 2008; NINDS 1995; Sandercock 2012). Pooled analysis supported these findings (Wardlaw 2009).

In an effort to deliver thrombolytics at the earliest time point possible, various streamlined 'stroke code' systems have been developed to decrease the door-to-treatment time. The current American Heart Association guideline recommends that the target door-to-treatment time be less than 60 minutes. However, this is rarely achieved (Lyden 1994; Marler 2000; O'Connor 1999; Saver 2013). In addition to tPA, the American Heart and Stroke Association (Powers 2018), and the European Stroke Organisation (ESO 2018), now recommend that for selected people with acute ischemic stroke, mechanical thrombectomy be considered up to 24 hours of onset of symptoms.

An ideal system for rapid thrombolytic delivery and, now, consideration for revascularization, begins with rapid and accurate stroke detection at the time of first contact with medical personnel (in most cases paramedics). Stroke pathways that include prehospital notification have been demonstrated to reduce door-to-treatment time and improve outcomes. Furthermore, hemorrhagic strokes require rapid assessment and it is believed that early identification and intervention is associated with signals toward decreased end volume size of hemorrhage (Anderson 2008). Optimal time of intervention and specific therapy including blood pressure agents and targets (Butcher 2013; Hill 2013), and use of clotting factors (Flaherty 2014), are currently under investigation.

Notably, all studies are focused on initiating therapies as soon as possible. However, the diagnostic accuracy of paramedics' diagnosis of stroke based on unstructured clinical assessment is poor (Harbison 1999). Better results could be achieved if validated stroke recognition tools are used to support the initial triage. Several such instruments have been developed and implemented in different countries worldwide, but the question about their absolute and relative accuracy remains unanswered.

Target condition being diagnosed

We included all suspected acute strokes (ischemic, hemorrhagic or TIAs) in people assessed by prehospital or emergency room (ER) staff including paramedics, emergency medicine technicians (EMTs), nurses, emergency physicians or general practitioners (GPs). Trauma must not be a primary disease mechanism, but we considered eligible studies including people with secondary trauma (a fall due to stroke). In the case of TIA, it is impossible to know if the neurologic deficit has resolved until the person has been assessed by a neurologist or stroke physician and has had adequate imaging. The presentation of TIA is analogous on a spectrum of disease that cannot be separated into categories at the time of first contact by the responding healthcare staff.

We included TIA in the target condition as the scales are not intended to differentiate between stroke and TIA. Therefore, if a person with relevant symptoms at presentation and a final diagnosis of TIA tests positive on the scale, the result will be treated as true positive rather than false positive. However, we appreciate that a lack of clear guidance on whether or not to apply the scales on people who are no longer symptomatic at the time of first contact is likely to introduce variation in the spectrum of included participants and, as a result, in test accuracy. We took this into consideration when interpreting the results from the included studies.

Index test(s)

The index tests are prehospital scales used to determine whether the person is having stroke. They are based on the National Institutes of Health Stroke Scale (NIHSS) and the first such tools, Cincinnati Prehospital Stroke Scale (CPSS) and Los Angeles Prehospital Stroke Scale (LAPSS), were developed and introduced in the USA in the mid-1990s (Nor 2004). The use of prehospital stroke scales by emergency medical responders is recommended by the American Heart and Stroke Association (Powers 2018), the European Academy of Neurology and the European Stroke Organisation (Kobayashi 2018). However, they make no recommendations about the use of specific instruments. The scales are in wide circulation worldwide and emergency medical responders receive training on how to use them as part of their professional education.

The scales are screening tools intended for use by prehospital and ER staff, and are not meant for diagnosis of any neurologic condition. Furthermore, they are not for determining the severity of stroke (unless they have a dual purpose) or the type of stroke (ischemic versus hemorrhagic versus TIA, or any subtypes). Due to the urgency to act on any type of stroke, the prehospital environment is not the appropriate setting in the decision tree to separate ischemic from hemorrhagic stroke or stroke from TIA. This is done by the attending neurologist or stroke physician.

The following stroke recognition scales were evaluated in the studies eligible for inclusion in the current review.

- Cincinnati Prehospital Stroke Scale (CPSS; Kothari 1999).
- Los Angeles Prehospital Stroke Scale (LAPSS; Kidwell 2000).
- Melbourne Ambulance Stroke Scale (MASS; Bray 2005a).
- Ontario Prehospital Stroke Screening Tool (OPSST; Chenkin 2009).
- Face Arm Speech Time (FAST; Harbison 2003).
- Recognition of Stroke in the Emergency Room (ROSIER; Nor 2005).
- Medic Prehospital Assessment for Code Stroke (MedPACS; Studnek 2013).
- PreHospital Ambulance Stroke Test (PreHAST; Andsberg 2017).

We summarized the characteristics of the evaluated scales in Table 1. Each scale consists of a list of checkbox items from the patient's history of presenting illness, past medical history, physical exam and basic laboratory values. The presence of any of the symptoms listed on the scale indicates high probability of stroke and should trigger an emergency stroke protocol. If none of the listed symptoms are present, diagnosis of stroke is less likely but not completely ruled out. Each symptom is scored '+1' when present and '0' when absent. Because the number of symptoms that could be scored '+1' is different for different scales, the total score varies. However, for all scales included in our review, a total score '+1 or greater' indicates high probability of stroke and warrants a referral for specialist assessment.

Some of the scales include additional criteria which determine whether the person is eligible for assessment with the respective scale. These criteria have been added to improve specificity by excluding people with common stroke mimics. However, the eligibility criteria of OPSST aim not only to reduce unnecessary triage of people with stroke mimics, but also of people who would be ineligible for fibrinolysis, regardless of whether stroke is present or not (e.g. people who could not be transported on time to an acute stroke care center) (Chenkin 2009).

ROSIER and PreHAST use slightly different scoring systems. ROSIER comprises five physical symptoms, each scored '+1' and two additional items, seizure activity and abnormal blood sugar, each scored '-1'. The presence of any of these additional items makes the diagnosis of stroke less likely even when some of the five listed symptoms are present. The total score could range from '-2' (none of the five physical symptoms and both additional items are present) to '+5' (all physical symptoms are present and neither of the two additional items). The positivity threshold is the same as for the other scales '+1 or greater'.

PreHAST is a tool "designed to screen for common stroke symptoms and grade severity, similarly to the NIHSS." (p. 2) It includes stroke symptoms that could predict main arterial vessel occlusion in addition to recognizing people with stroke in the field (Andsberg 2017). It comprises eight items that are scored differently (e.g. 0 or 1; 0, 1 or 2; 0 or 2), with 0 indicating absence of the symptom and 1 and 2, different levels of severity of a present symptom. The total score ranges from 0 to 19 points and '+1 or greater' is used as a positivity threshold to identify potential stroke. One study eligible for inclusion in this review evaluated PreHAST as the only prehospital stroke scale combining recognition of stroke and assessment of severity. We identified studies evaluating similar 'dual purpose' scales, but none of them met our inclusion criteria, mainly because the scales were applied to patient records (e.g. Purrucker 2015; Purrucker 2017).

Clinical pathway

The clinical pathway is very simple. When the paramedic, ambulance worker or medical attendant who is first on the scene is suspicious that the person may be having stroke, they are to implement a stroke scale in their evaluation of the person. Thus, the point of first contact between emergency medical responders and the person is where the index tests are to be implemented. The people are then brought to an ER for further evaluation and clinical workup. The triage of people who present directly to the ER and are suspected of stroke could also involve a stroke recognition scale.

Alternative test(s)

As discussed earlier, in addition to PreHAST we identified other prehospital stroke scales that combine stroke identification and severity assessment. Most of them were repurposed stroke severity scales (e.g. Kurashiki Prehospital Stroke Scale (KPSS), Los Angeles Motor Scale (LAMS), eight-item National Institutes of Health Stroke Scale (sNIHSS-8) and five-item National Institutes of Health Stroke Scale (sNIHSS-5)) initially designed to identify people with large vessel occlusion (LVO), who might be candidates for thrombectomy. They were evaluated in a small number of studies none of which met our inclusion criteria. In addition, one of the included studies compared CPSS to a panel of blood biomarkers but, as far as we are aware, these are not routinely used in clinical practice and have not been recommended for prehospital triage of people suspected of stroke (Vanni 2011).

Rationale

Despite the fact that prehospital stroke recognition scales are widely used in clinical practice, there has been little effort to systematically identify and review the evidence pertaining to their accuracy. Two non-Cochrane systematic reviews with objectives similar to ours have been published (Brandler 2014; Rudd 2016). The first review, Brandler 2014, included only studies in which the scales were used by paramedics, in agreement with the usual practice in the USA emergency medical services (EMS). The authors noted the heterogeneity in test accuracy estimates and concluded that "LAPSS and CPSS had similar diagnostic capabilities" (p. 1). This was questioned by the authors of the second review, Rudd 2016, which had a broader scope and concluded that "Available data do not allow a strong recommendation to be made about the superiority of a stroke recognition instrument." (p. 1). Given the contradicting outcomes from these two investigations, we decided to review the evidence pertaining to the absolute and relative accuracy of prehospital stroke recognition scales using well-defined inclusion criteria and established Cochrane Review methods, in order to make recommendations for future research and, if appropriate, for clinical practice.

OBJECTIVES

To systematically identify and review the evidence pertaining to the test accuracy of validated stroke recognition scales used in a prehospital or emergency room (ER) setting to screen people suspected of having stroke.

Secondary objectives

To investigate the effect of potential sources of heterogeneity on test accuracy estimates.

METHODS

Criteria for considering studies for this review

Types of studies

We considered all primary test accuracy studies if they evaluated a stroke recognition scale (index test) used in a prehospital or ER setting, against a final diagnosis of stroke/TIA. We included only those studies reporting sufficient data to determine test accuracy parameters (2 × 2 table). We included retrospective studies using stroke and EMS registry data, if the scales had been applied directly, face-to-face, to eligible patients. We excluded studies in which the scales were applied to patient records rather than to actual patients. We also excluded studies that enrolled only screen-positive patients.

Participants

We defined the target population as non-comatose, non-trauma patients suspected of stroke, with symptom duration under 24 hours at the time of presentation. Participants had to be over 18 years of age as this is a criterion for thrombolytic use. We included studies that had a subpopulation of people with previous history of stroke. The stroke recognition scales had to be applied in a prehospital or emergency setting.

We defined comatose patients as people who presented in the field with a Glasgow Coma Scale (GCS) score less than 8 and, therefore, required intubation and life-saving airway management. We included studies on people with a depressed level of consciousness who were protecting their airway, as their exam was not confounded by medications used for the induction and maintenance of an artificial airway.

Index tests

The Index tests were prehospital scales for the determination of whether the person was having stroke or not. We included all such scales if they were evaluated in eligible studies. The index tests could have been administered by a paramedic, an emergency medical responder, a nurse, an emergency physician or a GP. There were no limitations on the amount of training the scale administrator had received with the particular stroke scale. However, we acknowledge that differences in knowledge, experience and training could contribute to heterogeneity in test accuracy results. Here we used 'prehospital' as an umbrella term referring to the use of the scales in any prehospital setting including in the field (i.e. people attended by the ambulance), the ER or primary care. We specified setting (prehospital versus ER versus primary care) when discussing the use of specific scales in the studies.

Target conditions

The target condition was stroke, regardless of its type or severity, including ischemic stroke, hemorrhagic stroke or TIA. We defined ischemic stroke as irreversible neurologic damage due to obstruction of a blood vessel, corresponding to the parenchymal territory responsible for the neurologic function that was lost. We defined intracerebral hemorrhage as a stroke due to a bleed within the brain parenchyma. TIA is, by definition, transient and the neurologic deficit reverses without any clinical evidence of neuronal damage. To be included, studies could have used either the tissue-based definition of TIA (a negative diffusion-weighted imaging study) or the time-based definition of TIA (resolution of symptoms in less than 24 hours (but may be diffusion-weighted imaging positive).

Reference standards

There is no single, 'gold standard' diagnostic test to determine stroke. Therefore, we used the following criteria to define an acceptable reference standard.

- The initial inhospital diagnosis of stroke must have been done by a physician (neurologist, stroke physician, internist, emergency physician) who performed the history, physical exam and interpretation of the non-contrast CT head scan and any other imaging. It could alternatively be done by an internist, family physician or an emergency physician with the assistance of a consulting radiologist, neurologist, stroke physician, or a combination of these, available in person or by telephone.
- The person must have a documented discharge diagnosis of stroke or 'other', where 'other' could have been a neurologic or non-neurologic diagnosis. 'Other' could have been any medical condition that was determined by a physician, where the symptoms that mimicked stroke were accounted for.
- The person's chart must have been reviewed by a neurologist or stroke physician and the final diagnosis signed off by a neurologist or stroke physician, once the evolution of the person's condition had occurred to the point where they were discharged. For the purpose of the review, we considered a neurologist and a stroke physician equivalent. Non-neurologic discharge diagnoses made by non-neurologists were considered valid.
- Every participant who was assessed with the index test by prehospital staff/emergency responders was then to be assessed by a neurologist or stroke physician at some point prior to having a neurologic discharge diagnosis. This applied even to people with a 'negative' score on the index test. The path at which they arrived at a non-stroke diagnosis was beyond the scope of this review.

Search methods for identification of studies

We searched relevant computerized databases (listed below) from the earliest year possible to 30 January 2018. We applied no restrictions on language of publication.

Electronic searches

We searched the following electronic bibliographic databases:

- Cochrane Central Register of Controlled Trials (CENTRAL; 2018, Issue 1) in the Cochrane Library (searched 30 January 2018; Appendix 1);
- MEDLINE (Ovid) (1946 to 30 January 2018; Appendix 2);
- Embase (Ovid) (1974 to 30 January 2018; Appendix 3);
- Science Citation Index Cited Reference Search for forward tracking of important articles (up to 13 February 2018).

We developed the MEDLINE search strategy with the help of the Cochrane Stroke Group Information Specialist and adapted it for the other databases (Appendix 2).

Searching other resources

We searched the reference lists of all included studies and other relevant publications to identify additional studies. We contacted authors of the known prehospital stroke scales and asked them to provide information regarding unpublished studies.

Data collection and analysis

Selection of studies

Due to the large volume of initial titles produced by our database searches, we divided the references into two groups. We screened each title twice independently; ZZ and NH screened half of all titles, and GW and JF the other half. We retrieved the full texts of potentially relevant papers and JF and GW assessed their eligibility against the inclusion criteria. We resolved discrepancies by discussion or arbitration by a third review author (ZZ). We coded the studies excluded at full-text screening with a particular reason for exclusion.

Data extraction and management

To collect data from studies, we used a prespecified data extraction form, which included information on study characteristics, participant population and relevant outcomes (Appendix 4). Two review authors (NH and JF) independently extracted the data to ensure adequate reliability and quality, and a third review author (ZZ) adjudicated any disagreements. If reported in the paper, we extracted 2 × 2 data directly (true positives, false positives, true negatives and false negatives) for each index test. Alternatively, we reconstructed 2 × 2 tables by entering data on sensitivity, specificity, total number of participants and the proportion of diseased participants in the Review Manager 5 diagnostic accuracy calculator (Review Manager 2014). We sent data requests to study authors before excluding a study due to insufficient data.

Assessment of methodological quality

We assessed the methodologic quality of each study using the Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) tool (Whiting 2011). The tool consists of four domains: patient selection, index test, reference standard, and flow and timing. The first three domains are assessed in terms of risk of bias and concerns regarding applicability, and rated as 'high', 'low' or 'unclear'. The fourth domain, flow and timing, is assessed only in terms of risk of bias using the same rating categories. The tailored version of the tool including a set of operational definitions is provided in Appendix 5.

We added to the patient selection domain an additional signaling question to check if data were collected prospectively or retrieved from EMS and stroke registries. Retrospective data are prone to selective and incomplete recording, and matching patient records across different databases is not always possible. Therefore, we considered all studies using retrospective data collection to be at high risk of bias. We included only studies that applied the scales to actual patients and not to patient records, regardless of whether the patients were enrolled prospectively (prospective design) or data were retrieved from registries (retrospective design).

In the current review, the index tests were prehospital scales used to screen people suspected of having stroke at the first point of contact. The reference standard was a combination of tests performed once the person had already been admitted to hospital. Therefore, the question "Were the index test results interpreted without knowledge of the results of the reference standard?" would always be answered 'Yes'. This question was initially removed from the checklist, but included again during the editorial process upon advice from the Diagnostic Test Accuracy Editorial team.

It is unlikely that awareness of the index test results will affect the final diagnosis of stroke, if made by a neurologist/stroke physician using the results from imaging and other objective tests. However, it is possible to affect the diagnosis of TIA, which is based on the patient's presenting symptoms and assessment of their resolution within 24 hours. To capture this, we included in the reference standard domain a signaling question asking whether the clinicians making the final diagnosis were blinded to the results from the index test. However, the presenting symptoms were both part of the index test and the reference standard for TIA and, therefore, complete independence was not possible. Clinicians making the final diagnosis of TIA will always have access to this information, regardless of whether the results from the stroke scale are available to them or not. Therefore, we acknowledge that there could be risk of incorporation bias even in studies in which stroke adjudicators were blinded to the index test results.

Two review authors (GW, SY) independently assessed the methodologic quality of the studies and resolved any disagreements through discussion or arbitration by a third review author (ZZ or NH). If any of the signaling questions in the domain was rated 'high risk of bias', the overall domain was also categorized as 'high risk of bias'.

Statistical analysis and data synthesis

The index tests being reviewed are each made up of a set of criteria that are individually assessed and then combined to assign each participant a particular score. All scales use the same positivity threshold, '1 or greater', which indicates that the person may have been having a stroke. For each index test, we generated a diagnostic 2 × 2 table (true positives, false positives, true negatives and false negatives) from which we calculated sensitivity and specificity with 95% confidence intervals (95% CI). We also created forest plots and receiver operating characteristics (ROC) plots to show the variation in test accuracy estimates across studies.

When at least four studies evaluating the same index test were conducted in the same setting and reported consistent test accuracy estimates, we pooled sensitivity and specificity using the bivariate random-effects method. This method is recommended for studies using the same positivity threshold; it preserves the two-dimensional nature of the data; accounts for between-study variability by using a random-effects approach, and allows for the possibility of a negative correlation that may exist between sensitivity and specificity across studies (Reitsma 2005). We presented the summary estimates with a 95% confidence ellipse (i.e. a bivariate CI) and a 95% prediction region in the summary ROC space.

When only a small number of studies are included in a meta-analysis, the prediction regions generated by the Review Manager 5 are excessively conservative (Review Manager 2014). They may

appear inconsistent with the estimated CIs, as they depend on the number of included studies as well as on the standard errors and the covariance of the estimated mean logit sensitivity and specificity. To mitigate this, we followed the practice suggested in Gurusamy 2015. It recommends that when fewer than 10 studies are included in a meta-analysis, the number of studies entered into the Review Manager's analysis panel should be 10 (rather than the actual number of pooled studies). According to the authors, this provides a better approximation of the prediction region than using the actual (smaller) number of studies.

We calculated positive and negative likelihood ratios from the summary sensitivity and specificity, and plotted the results from comparative studies in the ROC plane to illustrate the relative accuracy of the tests. All statistical analyses were carried out using the analysis functions of Review Manager 5 (Review Manager 2014) and STATA statistical software version 15 (StataCorp 2011).

Investigations of heterogeneity

In the protocol, we listed the following variables as potential contributors to between-study variation in test accuracy estimates:

- participant demographics (e.g. age, gender);
- proportion of different types of stroke (ischemic, hemorrhagic or TIA);
- level of training;
- methodologic quality of included studies.

While working on the review, we identified additional potential sources of variation, the most important of which were:

- different triggers for applying the tool (prespecified criteria versus general suspicion of stroke);
- different procedures to obtain test scores, when more than one stroke scale was performed (e.g. consecutive application of both scales versus deriving the score of the simpler scale from the more complex scale);
- differences in the reference standard (e.g. hospital discharge diagnosis versus independent panel of clinicians).

Statistical investigation of the influence of the above sources of heterogeneity was not feasible, because of the small number of studies per test conducted in the same setting. Instead, we conducted a visual inspection of the ROC and forest plots, and provided a narrative description of the observed heterogeneity.

Sensitivity analyses

The small number of studies included in the meta-analyses precluded quantitative sensitivity analysis based on the methodologic quality of included studies. However, when reporting the results, we considered the methodologic quality of studies evaluating specific tests and highlighted the results reported by better-quality studies.

Assessment of reporting bias

Following the recommendations of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*, we did not investigate publication bias because of the low power of the recommended test for funnel plot asymmetry, when there is heterogeneity in the diagnostic odds ratios and, more generally, because of the limited research in this area (Macaskill 2010). However, publication bias might be present and might affect the results from the review. In order to mitigate this, we conducted comprehensive searches of the published literature and contacted experts in the field to identify any additional or unpublished studies. We also interpreted our results with caution, acknowledging the possibility of publication bias.

RESULTS

Results of the search

Figure 1 illustrates the selection process and Appendix 6 shows the number of records per database. From the initial electronic searches conducted in January 2015, we identified 8481 unique references. After screening titles and abstracts, we selected 162 publications for full-text assessment. Of those, we excluded 71 conference abstracts that did not report sufficient data and for which there were no full-text articles or additional data. Two review authors (JF, GW or SY) independently assessed the eligibility of the remaining 91 titles and selected 19 studies for inclusion in the review. We last updated the searches on 30 January 2018 and identified 3526 additional unique references. We screened 33 at full-text level and considered four of them to be inclusions. We also searched the reference lists of all included studies and other relevant publications, but found no additional inclusions. The final number of studies included in the review was 23.

Figure 1. Study flow diagram. ER: emergency room; LAPSS: Los Angeles Prehospital Stroke Scale; ROSIER: Recognition of Stroke in the Emergency Room.

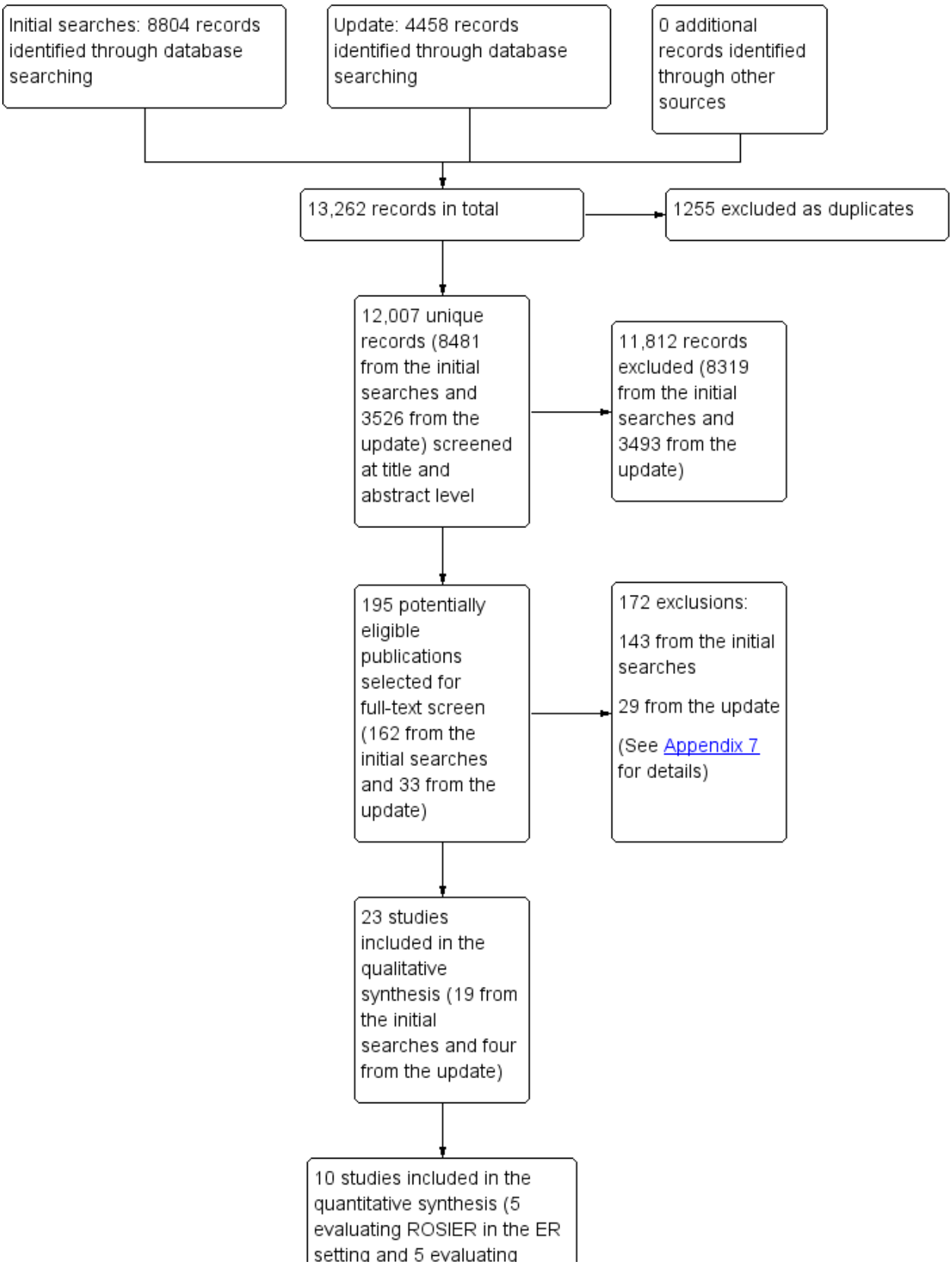
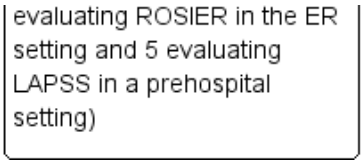


Figure 1. (Continued)



Characteristics of the included studies

The characteristics of the included studies are summarized in Table 2, Table 3, and Table 4. Twenty-two studies were journal articles and one was a conference abstract (Kim 2017). They were published between 2000 and 2017. Five studies were conducted in China (Chen 2013; Ding 2009; Jiang 2014; Mingfeng 2012; Mingfeng 2017); five in the USA (English 2018; Frendl 2009; Kidwell 2000; Ramanujam 2008; Studnek 2013); four in the UK (Fothergill 2013; Jackson 2008; Nor 2005; Whiteley 2011); two each in Australia (Bray 2005a; Bray 2010); Sweden (Andsberg 2017; Berglund 2014); and the Republic of Korea (Kim 2017; Lee 2015); and one each in Belgium (Bergs 2010), Canada (Chenkin 2009), and Italy (Vanni 2011).

Twenty-one studies were published in English, one in Korean (Lee 2015), and one in Chinese (Ding 2009). The data extraction and methodologic quality assessment of the two non-English language studies were done by stroke neurologists fluent in the respective language: the translation from Chinese was done by a member of our team (SY); and the translation from Korean was done by Dr Sang Min Sung from the Pusan National University Hospital in South Korea. Additional data or answers to specific queries, or both, were very kindly provided by the authors of the following included papers: Berglund 2014, Jiang 2014, and Lee 2015.

The studies evaluated eight prehospital stroke scales:

- CPSS (11 studies; Bergs 2010; Bray 2005a; Bray 2010; English 2018; Frendl 2009; Kim 2017; Mingfeng 2012; Mingfeng 2017; Ramanujam 2008; Studnek 2013; Vanni 2011);
- ROSIER (eight studies; Fothergill 2013; Jackson 2008; Jiang 2014; Lee 2015; Mingfeng 2012; Mingfeng 2017; Nor 2005; Whiteley 2011);
- FAST (five studies; Berglund 2014; Bergs 2010; Fothergill 2013; Lee 2015; Whiteley 2011);
- LAPSS (five studies; Bergs 2010; Bray 2005a; Chen 2013; Ding 2009; Kidwell 2000);
- MASS (three studies; Bergs 2010; Bray 2005a; Bray 2010);
- OPSST (one study; Chenkin 2009);
- MedPACS (one study; Studnek 2013);
- PreHAST (one study; Andsberg 2017).

Nine of the included studies (39%) evaluated more than one stroke scale in the same participants (Bergs 2010; Bray 2005a; Bray 2010; Fothergill 2013; Mingfeng 2012; Mingfeng 2017; Lee 2015; Studnek 2013; Whiteley 2011), and one study compared CPSS to a panel of blood biomarkers used to identify people with stroke in the ER (Vanni 2011). All studies obtained the scores directly, by face-to-face application of the scales to people suspected of stroke.

Nor 2005 also compared the accuracy of ROSIER with that of FAST, CPSS and LAPSS, but the scores for the latter three scales were derived post hoc from neurologist-recorded signs. An additional analysis from the same study compared the accuracy of ROSIER (completed by ER physicians) with that of FAST (completed by paramedics) in a subgroup of 49 participants. We included the data for ROSIER, which was the main focus of the study, but excluded the two comparative data sets: the first one because the scores for FAST, CPSS and LAPSS were derived from patient records, and the second one because it was a post-hoc analysis of a small convenience sample and the tests were performed in different setting by different clinicians.

The total number of participants in the included studies was 9230 and ranged from 31 (Bergs 2010) to 1130 (Chen 2013), median 312 (interquartile range (IQR) 154 to 554). The prevalence of the target condition (stroke and TIA) ranged from 16% (Ding 2009) to 92% (Jackson 2008), mean 54% (standard deviation (SD) 20%). The index tests were used in an ER setting in six studies (Jackson 2008; Jiang 2014; Lee 2015; Nor 2005; Vanni 2011; Whiteley 2011); in three of them they were applied by ER physicians, in two by ER physicians or nurses, and by nurses in one study. The rest of the studies were conducted in a prehospital setting and the scales were applied by paramedics, with the exception of Ding 2009 and Mingfeng 2012 (ER physicians as part of an ambulance crew), Andsberg 2017 and Bergs 2010 (nurses), Berglund 2014 (nurses or paramedics), and Mingfeng 2017 (GPs).

The amount of training and the trigger for applying the scales also varied across studies. Some studies applied the stroke recognition tool to all participants suspected of stroke by the attending clinician (Fothergill 2013; Frendl 2009; Nor 2005; Studnek 2013). Other studies required the participants to meet specific eligibility criteria to be tested (Table 2). This most likely led to differences between study cohorts and contributed to the observed between-study heterogeneity.

Methodological quality of included studies

The methodologic quality of the included studies is summarized in Figure 2 and Figure 3. We considered 12 studies (52%) at high risk of bias in the patient selection domain: seven because they were retrospective analyses of stroke registry data (Bray 2010; Chenkin 2009; English 2018; Frendl 2009; Kim 2017; Ramanujam 2008; Studnek 2013), and five prospective studies that failed to include all eligible consecutive participants (Bergs 2010; Bray 2005a; Chen 2013; Fothergill 2013; Kidwell 2000). Retrospective studies depend on routinely collected data, which are susceptible to selective and incomplete recording. For instance, Bray 2010 and Studnek 2013 excluded over 10% of all eligible patient records because they were missing relevant data. Therefore, we considered all retrospective studies at high risk of selection bias.

Figure 2. Risk of bias and applicability concerns graph: review authors' judgments about each domain presented as percentages across included studies.

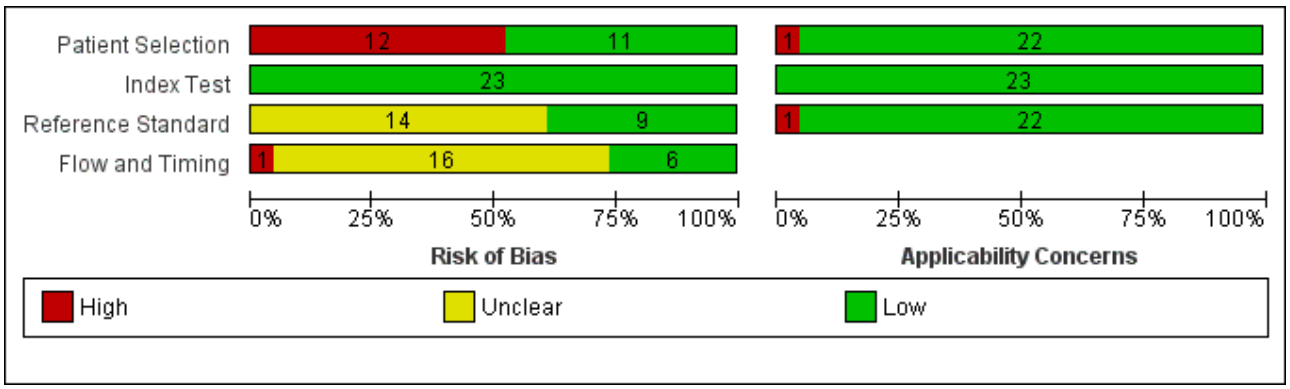


Figure 3. Risk of bias and applicability concerns summary: review authors' judgments about each domain for each included study.

Another potential source of bias in this domain was the failure to include all people with a negative screen. The reason was that in some studies participants who tested positive on the scale were transported to an acute stroke care center, while those with a negative screen were taken to other hospitals and could not always be included in the study sample. Such non-consecutive selection is likely to affect both sensitivity and specificity by missing false negative and true negative cases. The authors of some papers acknowledged this issue (Chenkin 2009; Ramanujam 2008), but it is entirely possible that more studies were affected by such non-consecutive selection.

The retrospective analysis conducted by English 2018 included only participants identified by the EMS dispatchers as potential stroke cases. Most likely, this led to a significant proportion of the true stroke patients and those without stroke, but with relevant clinical presentation, to be missed. The effect of such selection has been demonstrated by Berglund and colleagues (Berglund 2014); in their study, EMS dispatchers missed about 30% of the people with stroke/TIA. In addition, we had applicability concerns about the selection of participants in one study, as 41% of the included participants were assessed more than 24 hours after the onset of symptoms (Jiang 2014).

All studies used a prespecified positivity threshold for the index tests and, as the stroke scales were always performed before the reference standard, the clinician administering the test was unaware of the reference standard results. We could not determine the risk of bias in the reference standard domain in 14 studies (61%), as they failed to report sufficient detail. The reference standard was hospital discharge diagnosis with no information on the actual tests and procedures, and the blinding of the clinicians to the results from the index tests. We noted applicability concerns in the reference standard domain for one study that excluded TIA from the target condition (Vanni 2011).

For 16 studies (70%), we could not exclude the possibility of bias in the flow and timing domain. Some of the studies failed to report the time to diagnosis (within 14 days of presentation or longer), whether all participants received (the same) reference standard, or whether they included all participants in the analysis. Also, one study excluded 7% of the participants and was rated as high risk of bias (Lee 2015).

With regards to comparative accuracy, not included in the Methodological Quality diagrams, we identified two main issues. First, only a few studies reported the statistical significance of their results; even when they did, they did not report whether the study was adequately powered to detect clinically meaningful differences in the accuracy of the compared tests. Second, in the index test domain, one potential source of bias and applicability concerned the method by which the scores of individual tests were derived. In the case of 'nested tests', that is where one scale contained all items of the other scales, such as CPSS and LAPSS nested in MedPACS, the scores for the nested scales were derived from the more complex one (Bray 2005a; Bray 2010; Fothergill 2013; Studnek 2013). Other studies applied the tests individually, but by the same test administrator and no random order of application was reported (Lee 2015; Mingfeng 2012; Whiteley 2011). In the remaining study, all compared tests (CPSS, FAST, LAPSS and MASS) were combined in a 10-item questionnaire completed by the attending EMS nurses (Bergs 2010). In theory, these different methods of application could lead to biased results and variability in the performance of the scales.

Findings

Face Arm Speech Test (FAST)

Five studies evaluated the FAST tool (Berglund 2014; Bergs 2010; Fothergill 2013; Lee 2015; Whiteley 2011). The total number of participants was 1894 and ranged from 31 to 900, median 312. The mean prevalence of stroke/TIA was 56% (SD 12%) and ranged from 36% to 69%. The proportion of TIA in people with the target condition was also variable and ranged from 5% (Bergs 2010) to 27% (Berglund 2014), suggesting different spectrum of included participants. This could be explained, at least partly, with differences in the inclusion criteria (e.g. Berglund 2014 included people with symptom onset less than six hours; Bergs 2010 kept the inclusion criteria broad to avoid missing cases).

All studies had prospective design, but we considered two of them at high risk of selection bias (Bergs 2010; Fothergill 2013), as they failed to include all eligible consecutive participants. We determined the risk of bias in the flow and timing domain to be high for Lee 2015, as they excluded from the analysis 7% of all participants due to incomplete records; and for Bergs 2010, we could not fully assess the risk of bias in the reference standard, and flow and timing domains.

Three studies evaluated the accuracy of FAST in a prehospital setting with the test being performed by paramedics or nurses (Berglund 2014; Bergs 2010; Fothergill 2013). The reported sensitivities were 0.64 (Berglund 2014), 0.95 (Bergs 2010), and 0.97 (Fothergill 2013), and the reported specificities 0.75 (Berglund 2014), 0.33 (Bergs 2010), and 0.13 (Fothergill 2013). This suggests potential presence of a threshold effect, which could be related to differences in selection criteria, as all three studies used the same positivity threshold ('1 or greater').

In the two studies conducted in the ER, the test was administered by ER physicians or ER physicians and nurses (Lee 2015; Whiteley 2011). The reported sensitivities were 0.86 (Lee 2015) and 0.81 (Whiteley 2011), and the specificities were 0.92 (Lee 2015) and 0.39 (Whiteley 2011) (Figure 4; Figure 5). We could not find an obvious explanation for the large difference in specificities, but noted that the two cohorts differed in important aspects, such as prevalence of the target condition (36% (Lee 2015) versus 69% (Whiteley 2011)) and mean age (60 years (Lee 2015) versus 72 years (Whiteley 2011)).

Figure 4. Forest plot of 2 Face Arm Speech Time (FAST).

Study	TP	FP	FN	TN	Test administrator	Setting	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
Fothergill 2013	171	103	6	15	Paramedic	Prehospital	0.6	0.97 [0.93, 0.99]	0.13 [0.07, 0.20]
Bergs 2010	18	8	1	4	Nurse	Prehospital	0.61	0.95 [0.74, 1.00]	0.33 [0.10, 0.65]
Berglund 2014	301	107	171	321	Nurse or paramedic	Prehospital	0.52	0.64 [0.59, 0.68]	0.75 [0.71, 0.79]
Lee 2015	97	15	16	184	Emergency physician	ER	0.36	0.86 [0.78, 0.92]	0.92 [0.88, 0.96]
Whiteley 2011	199	67	47	43	Emergency physician or nurse	ER	0.69	0.81 [0.75, 0.86]	0.39 [0.30, 0.49]

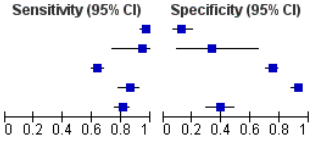
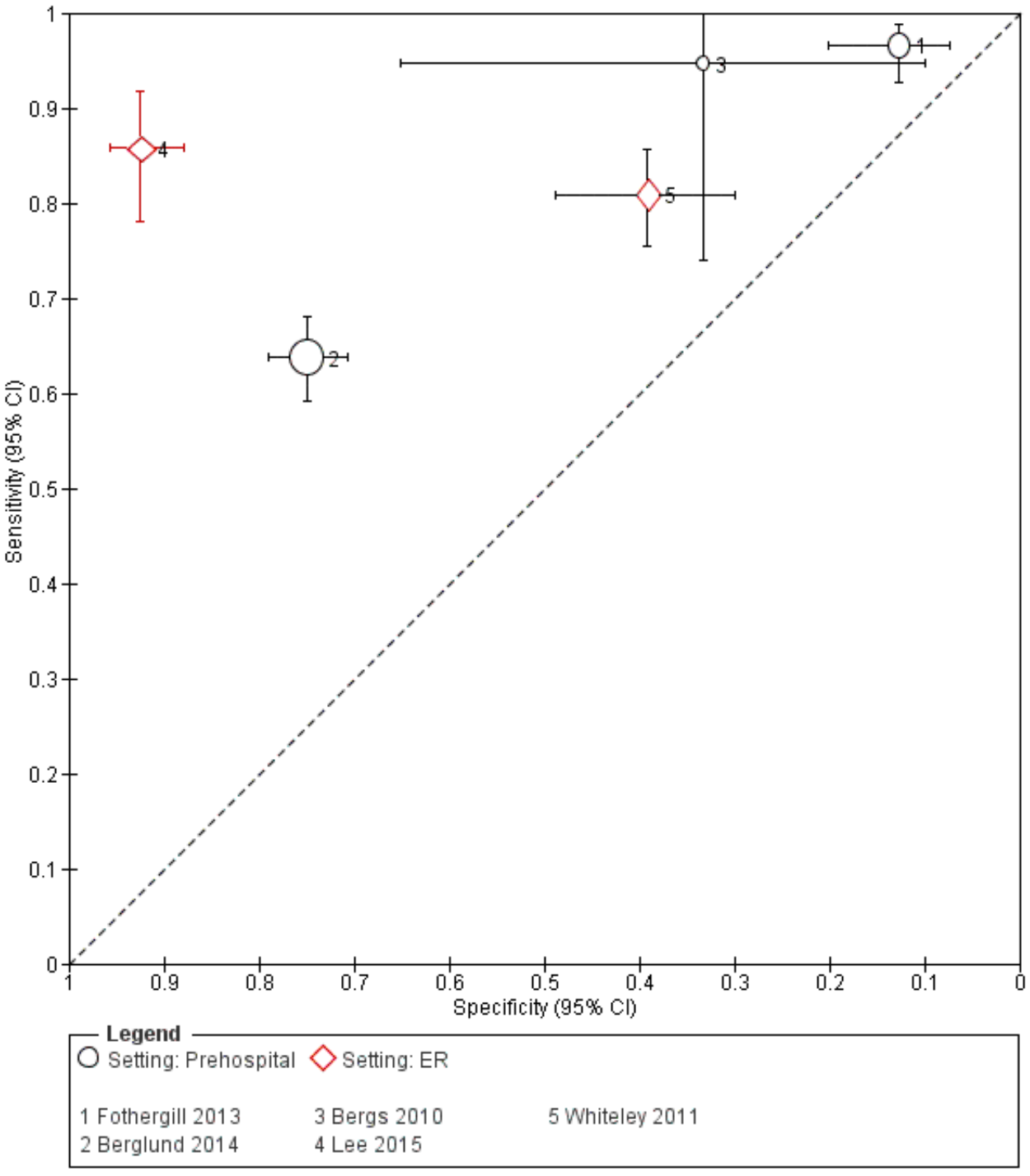


Figure 5. Summary receiver operating characteristics plot of 2 Face Arm Speech Time (FAST). ER: emergency room.



Los Angeles Prehospital Stroke Scale (LAPSS)

Five studies evaluated LAPSS (Bergs 2010; Bray 2005a; Chen 2013; Ding 2009; Kidwell 2000). The total number of included participants was 1794, with median 206 and range 31 to 1130. The mean prevalence of stroke/TIA was 51% (SD 33%) and ranged from 16% to 88%. All studies had prospective design but, with the exception of Ding 2009, were at high risk of selection bias because of non-consecutive sampling. We could not determine the risk of bias in the reference standard domain for three studies (Bergs 2010; Bray

2005a; Ding 2009), and considered all studies at unclear risk of bias in the flow and timing domain.

All studies used the index test in a prehospital setting, performed by paramedics in three studies (Bray 2005a; Chen 2013; Kidwell 2000), by emergency nurses in one (Bergs 2010), and by ER physicians in one (Ding 2009). The sensitivity of LAPSS ranged from 0.74 to 0.92 and specificity from 0.83 to 0.97 (Figure 6; Figure 7). Kidwell 2000 and Ding 2009 reported much higher sensitivity and specificity compared to the other three studies (sensitivity: 0.91 to 0.92 (Ding 2009; Kidwell 2000) versus 0.74 to 0.78 (other three studies), and specificity: 0.96 to 0.97 (Ding 2009; Kidwell 2000) versus 0.83 to 0.90 (other three studies)).

Figure 6. Forest plot of 3 Los Angeles Prehospital Stroke Scale (LAPSS).

Study	TP	FP	FN	TN	Test administrator	Setting	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
Bergs 2010	14	2	5	10	Nurse	Prehospital	0.61	0.74 [0.49, 0.91]	0.83 [0.52, 0.98]
Bray 2005a	57	4	16	23	Paramedic	Prehospital	0.73	0.78 [0.67, 0.87]	0.85 [0.66, 0.96]
Chen 2013	782	13	215	120	Paramedic	Prehospital	0.88	0.78 [0.76, 0.81]	0.90 [0.84, 0.95]
Ding 2009	47	12	4	264	Emergency physician	Prehospital	0.16	0.92 [0.81, 0.98]	0.96 [0.93, 0.98]
Kidwell 2000	31	5	3	167	Paramedic	Prehospital	0.17	0.91 [0.76, 0.98]	0.97 [0.93, 0.99]

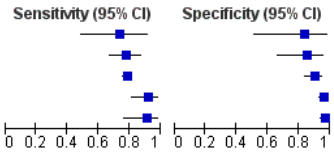
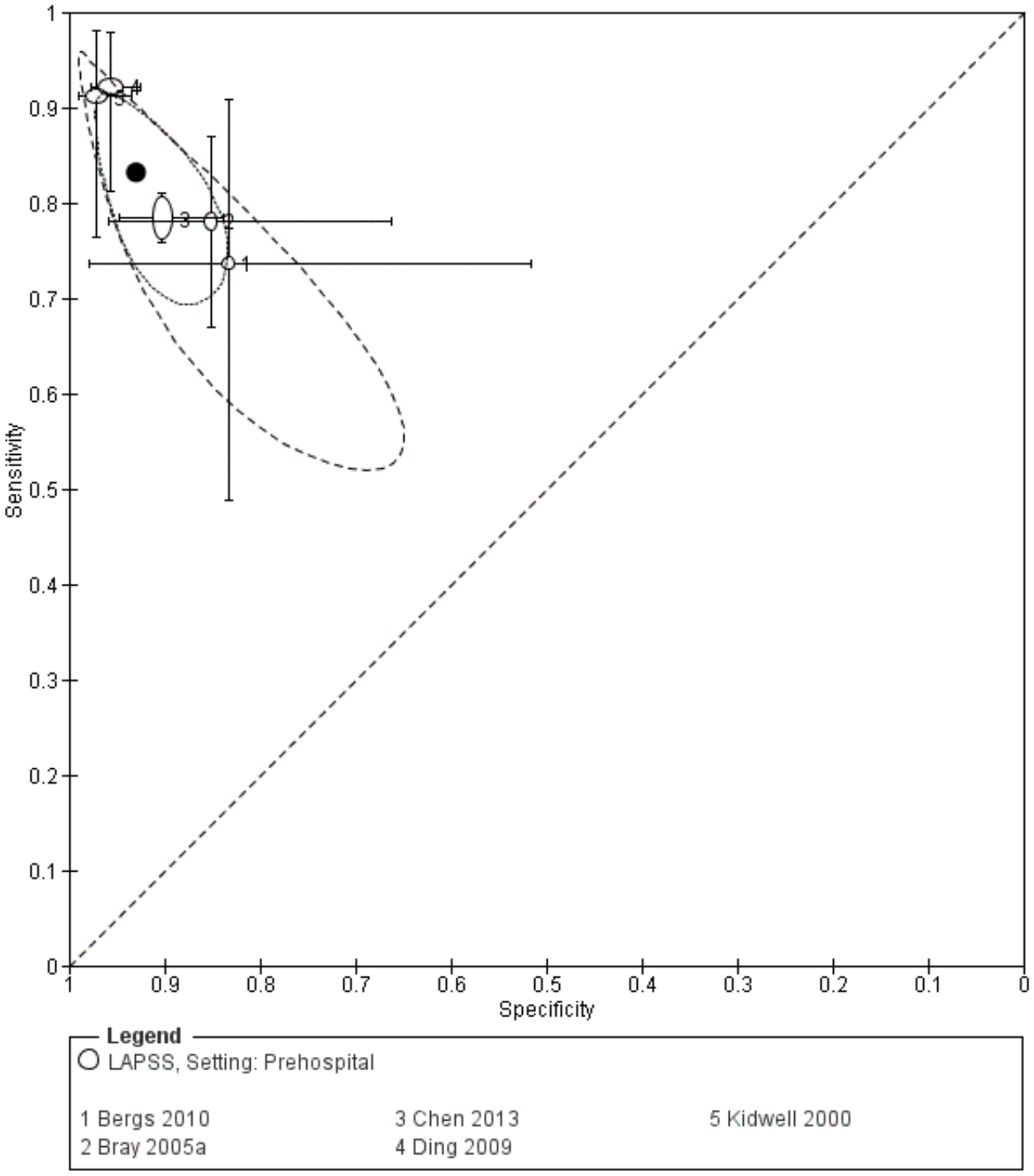


Figure 7. Summary receiver operating characteristics plot of 3 Los Angeles Prehospital Stroke Scale (LAPSS).



One possible explanation of these differences was the level of training and expertise in the two groups of studies. Ding 2009 did not report training, but in this study ER physicians (as part of an ambulance crew) used the test, so we can assume much higher level of expertise compared to even trained paramedics. In Kidwell 2000, the paramedics received extensive training on stroke and the use of LAPSS, including video vignettes of people with stroke and stroke mimics. They also had to pass an exam which, if failed, was followed by further training. The test administrators in the other three studies included nurses and paramedics with far less intensive training (Table 4).

The two groups of studies differed in other important ways, which may also have contributed to the observed differences in the

reported accuracy estimates. The prevalence of stroke/TIA in Ding 2009 and Kidwell 2000 was much lower (16% to 17% (Ding 2009; Kidwell 2000) versus 61% to 88% (in other three studies)), and the proportion of eligible participants out of all emergency runs was much higher (16% to 34% (Ding 2009; Kidwell 2000) versus 2.1% to 7.6% (in other three studies). Also, in comparison to the other three studies, the cohorts in Ding 2009 and Kidwell 2000 were much younger (mean age: 58 to 63 years (Ding 2009; Kidwell 2000) versus 72 to 77 years (other three studies)), and the proportion of women was higher (48% (Ding 2009; Kidwell 2000) versus 39% (Bergs 2010; Chen 2013 (Bray 2005a did not report sex distribution))).

The authors of Chen 2013 tried to explain the difference between their study and that of Kidwell 2000 by pointing to differences in study populations, level of training, and different EMS systems in the USA and China. However, Ding 2009 was also conducted in China, suggesting that the differences in the healthcare systems might have had less impact than other factors, such as training and selection of participants.

As the studies reported relatively consistent sensitivity and specificity estimates, we decided to pool the results of all five studies. They applied the same positivity threshold ('1 or greater'), so we used the random-effects bivariate model (Reitsma 2005). This produced a mean summary sensitivity 0.83 (95% CI 0.75 to 0.89) and a mean summary specificity 0.93 (95% CI 0.88 to 0.96), with fairly wide prediction region reflecting the small number of studies and the presence of between-study heterogeneity (Figure 7). The respective mean positive likelihood ratio was 12 (95% CI 6 to 23) and the mean negative likelihood ratio was 0.18 (95% CI 0.11 to 0.29). However, these estimates should be treated with caution because of the methodologic limitations noted above.

Melbourne Ambulance Stroke Scale (MASS)

Three studies evaluated the MASS tool in a prehospital setting (Bray 2005a; Bray 2010; Bergs 2010). The number of participants included in the studies was 100 (Bray 2005a), 850 (Bray 2010), and 31 (Bergs 2010); the prevalence of stroke/TIA was 73% (Bray 2005a), 23% (Bray 2010), and 61% (Bergs 2010), and the proportion of eligible participants out of all EMS runs was 2.1% (Bray 2005a), 19.0% (Bray 2010), and 7.6% (Bergs 2010), suggesting differences in the selection and composition of study cohorts.

The test was administered by nurses in Bergs 2010, and by paramedics in Bray 2005a and Bray 2010. We considered all three studies to be at high risk of selection bias and to have 'unclear' risk of bias in the reference standard, and flow and timing domains. The sensitivity of MASS was 0.90 (Bray 2005a), 0.83 (Bray 2010), and 0.74 (Bergs 2010), and the specificity was 0.74 (Bray 2005a), 0.86 (Bray 2010), and 0.67 (Bergs 2010) (Figure 8; Figure 9).

Figure 8. Forest plot of 5 Melbourne Ambulance Stroke Scale (MASS).

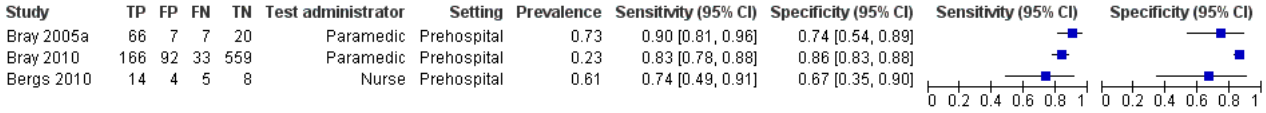
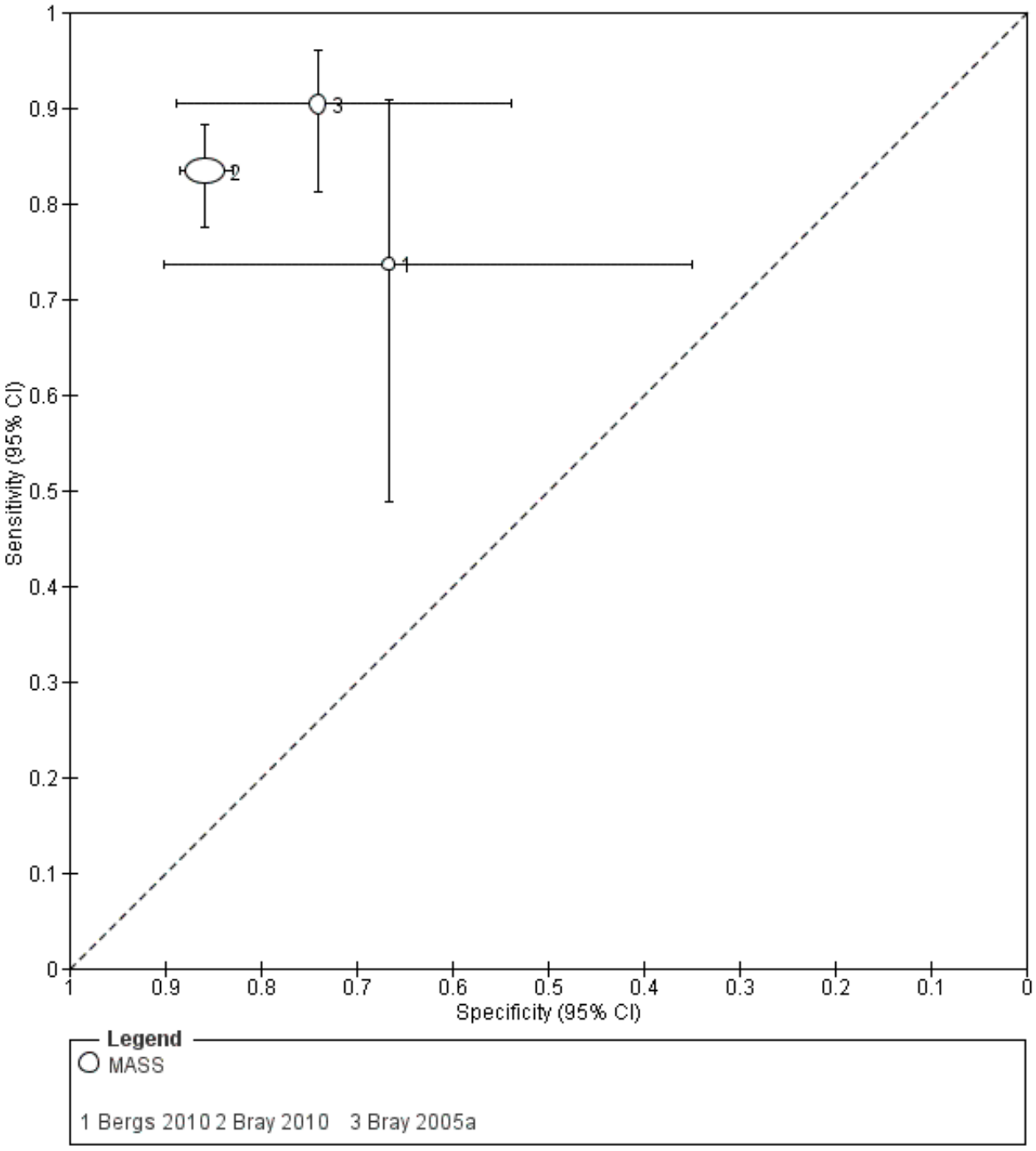


Figure 9. Summary receiver operating characteristics plot of 5 Melbourne Ambulance Stroke Scale (MASS; prehospital setting)



Cincinnati Prehospital Stroke Scale (CPSS)

Eleven studies evaluated CPSS (Bergs 2010; Bray 2005a; Bray 2010; English 2018; Frendl 2009; Kim 2017; Mingfeng 2012; Mingfeng 2017; Ramanujam 2008; Studnek 2013; Vanni 2011). The total number of participants was 4157 with median 268 and range 31 to 1045. The mean prevalence of stroke/TIA was 56% (SD 17%) and ranged from 23% to 74%.

We determined the risk of bias in the patient selection domain to be high in eight studies, of which six had retrospective design and two because of non-consecutive sampling (Bergs 2010; Bray 2005a). We could not assess the risk of bias in the reference standard domain in eight studies and in the flow and timing domain in all but two studies (Mingfeng 2017; Vanni 2011). Also, Vanni 2011 excluded TIA

from their definition of target condition and was at high level of applicability concerns.

One study used the test in the ER and reported sensitivity and specificity estimates 0.75 and 0.78, respectively (Vanni 2011). In another study, the test was used by GPs to decide whether to transfer people suspected of stroke from primary care to a hospital with an acute stroke center (Mingfeng 2017). The reported sensitivity and specificity of the test were 0.78 and 0.71, respectively.

The remaining nine studies used CPSS administered by ambulance staff in the field: by paramedics in seven studies, nurses in one study, and ER physicians in one study. However, professional expertise and training could not explain the differences in study estimates. The two studies in which the test was applied by nurses (Bergs 2010), and ER physicians (Mingfeng 2012), reported relatively high estimates of sensitivity (0.95 (Bergs 2010) and 0.89 (Mingfeng 2012)), but variable specificity (0.33 (Bergs 2010) and 0.69 (Mingfeng 2012)). Across studies, the level of between-study heterogeneity was very high, with sensitivity ranging from 0.44 to 0.95 and specificity from 0.21 to 0.79 (Figure 10; Figure 11).

Figure 10. Forest plot of 1 Cincinnati Prehospital Stroke Scale (CPSS). ER: emergency room; GP: general practitioner.

Study	TP	FP	FN	TN	Test administrator	Setting	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
Ramanujam 2008	193	284	247	321	Paramedic	Prehospital	0.42	0.44 [0.39, 0.49]	0.53 [0.49, 0.57]
Frendl 2009	43	45	18	48	Paramedic	Prehospital	0.4	0.70 [0.57, 0.81]	0.52 [0.41, 0.62]
English 2018	72	27	24	7	Paramedic	Prehospital	0.74	0.75 [0.65, 0.83]	0.21 [0.09, 0.38]
Studnek 2013	147	175	39	55	Paramedic	Prehospital	0.45	0.79 [0.72, 0.85]	0.24 [0.19, 0.30]
Bray 2010	176	138	23	513	Paramedic	Prehospital	0.23	0.88 [0.83, 0.93]	0.79 [0.75, 0.82]
Kim 2017	142	31	10	85	Paramedic	Prehospital	0.57	0.93 [0.88, 0.97]	0.73 [0.64, 0.81]
Bray 2005a	69	12	4	15	Paramedic	Prehospital	0.73	0.95 [0.87, 0.98]	0.56 [0.35, 0.75]
Bergs 2010	18	8	1	4	Nurse	Prehospital	0.61	0.95 [0.74, 1.00]	0.33 [0.10, 0.65]
Mingfeng 2017	259	40	73	96	GP	Prehospital	0.71	0.78 [0.73, 0.82]	0.71 [0.62, 0.78]
Mingfeng 2012	340	49	43	108	Emergency physician	Prehospital	0.7	0.89 [0.85, 0.92]	0.69 [0.61, 0.76]
Vanni 2011	65	15	22	53	Nurse	ER	0.56	0.75 [0.64, 0.83]	0.78 [0.66, 0.87]

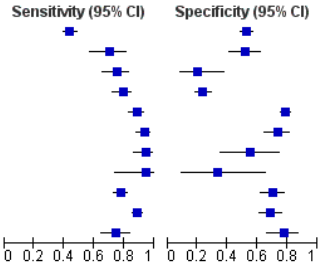
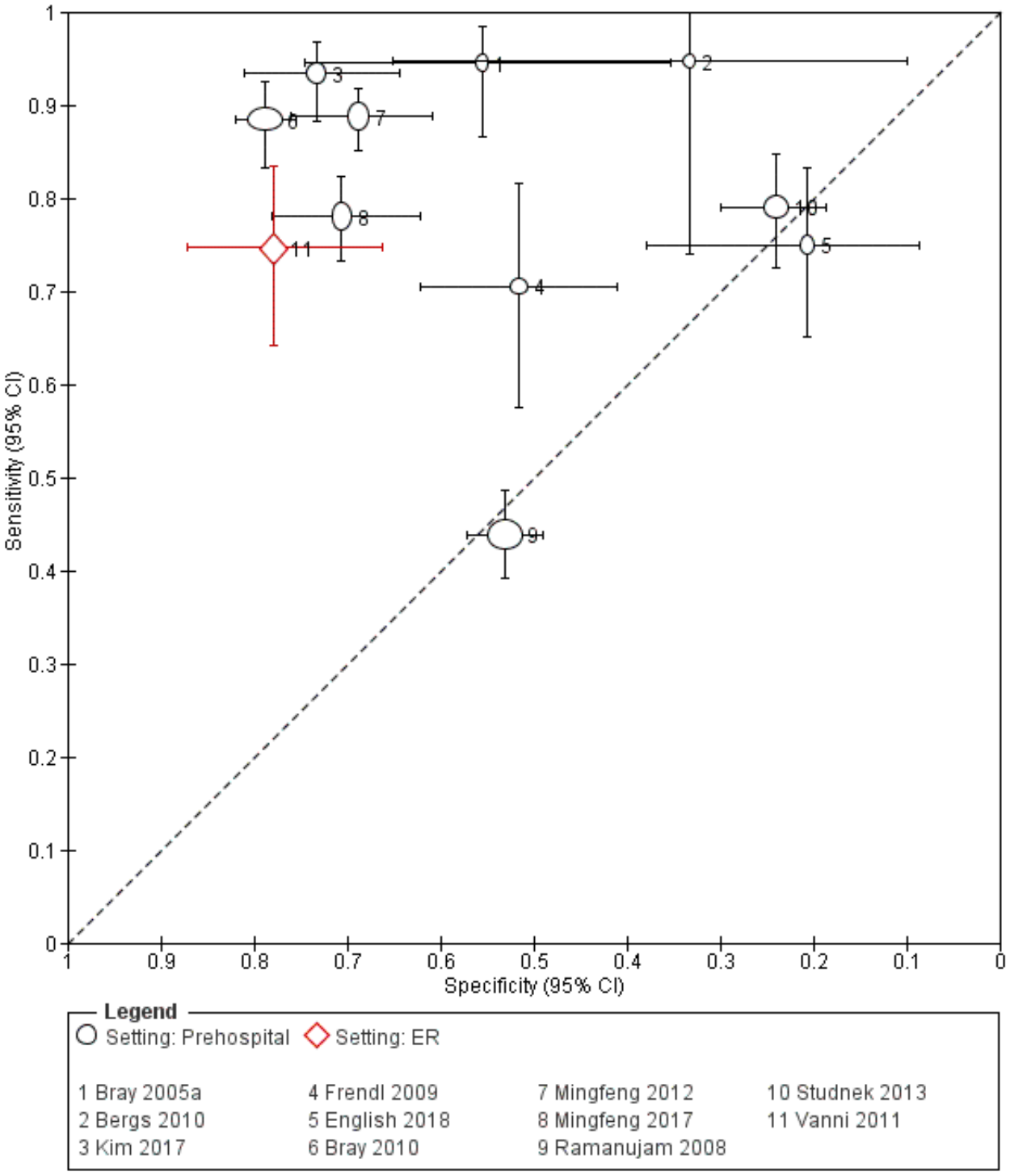


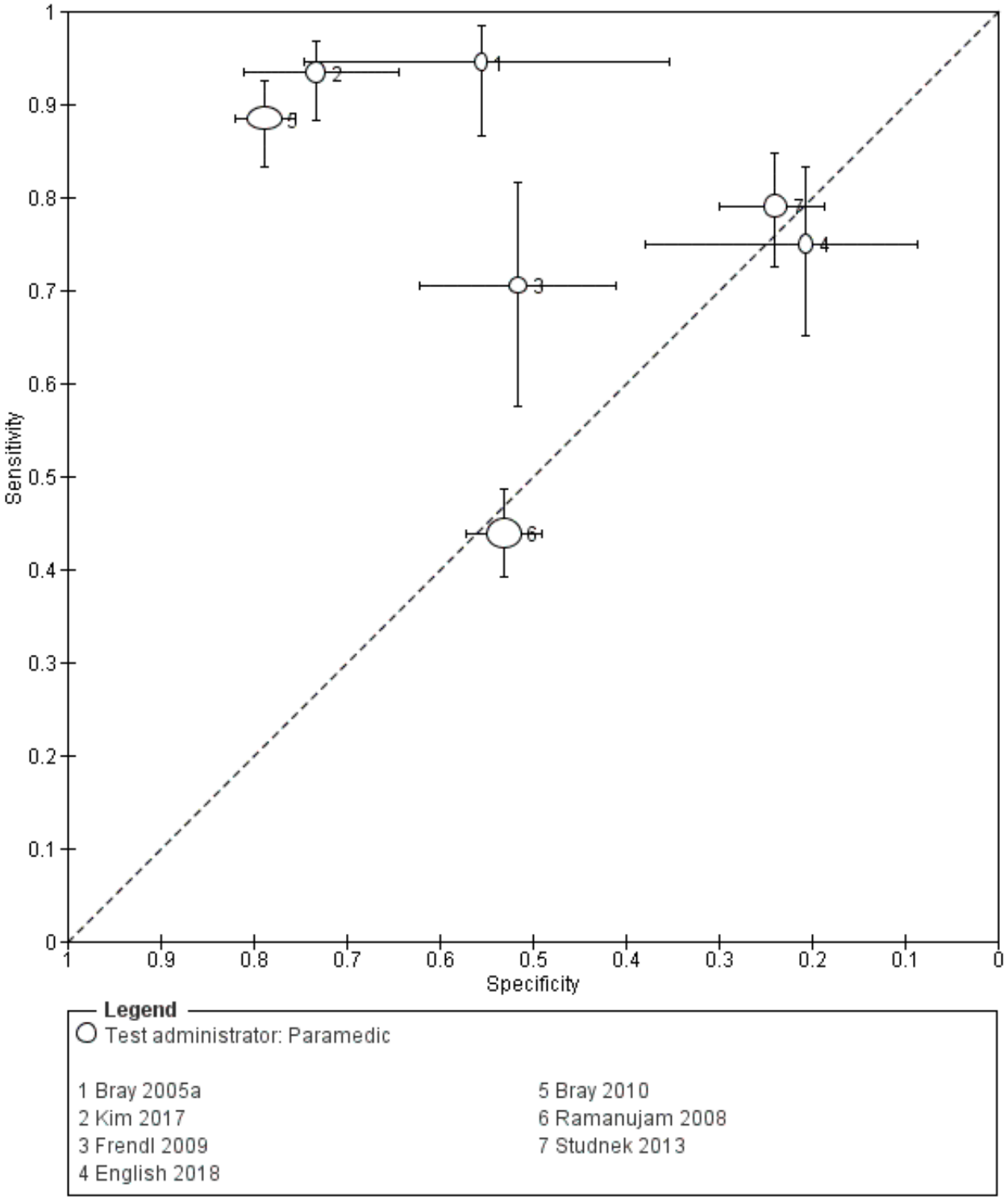
Figure 11. Summary receiver operating characteristics plot of 1 Cincinnati Prehospital Stroke Scale. ER: emergency room.



Considering only the seven studies where paramedics performed the test in a prehospital setting did not result in more consistent study-level estimates (Figure 12). English 2018, Ramanujam 2008, and Studnek 2013 reported accuracy estimates lying very close to the line of no-discrimination. This indicated that, in these studies, the test performed no better than a random guess. In contrast, Bray 2005a, Bray 2010, and Kim 2017 reported very high sensitivity (greater than 85%) and specificity higher than in the other four studies (greater than 50%). We could not find an obvious explanation of this extreme variation and assumed that the most

likely reasons were differences in the inclusion criteria and the presence of selection bias.

Figure 12. Summary receiver operating characteristics plot of 1 Cincinnati Prehospital Stroke Scale (CPSS) including only studies evaluating the test when used by paramedics in a prehospital setting.



Indeed, all studies conducted in a prehospital setting varied considerably in terms of inclusion criteria and selection of participants (Table 2). For instance, English 2018 included only people suspected of stroke by emergency dispatchers, while Frendl 2009 included all people transported by EMS and coded as having possible stroke or TIA (Table 2). As a consequence, the composition of study cohorts was very different: the prevalence of stroke/TIA in this group of studies ranged from 23% to 74%, median 57% (IQR 41% to 72%); the proportion of TIA in people with the target condition ranged from 3% to 23%; mean age ranged from 63 to 77 years; the proportion of women ranged from 32% to 56%; and the proportion of eligible participants out of all EMS runs ranged from 1.3% to 34.5% (Table 3). Studies also differed in terms of training and reference standard, and most of them were at high risk of bias in at least one domain (especially selection bias).

Given the high level of between-study heterogeneity, which could not be reduced through stratification, and the high risk of bias in most of the studies, we decided not to pool the results. We considered three of the studies at low risk of bias (Mingfeng 2012; Mingfeng 2017; Vanni 2011), and, therefore, more likely to provide unbiased test accuracy estimates. They were conducted in different countries (China and Italy), different settings (ambulance, primary care and the ER) and CPSS was used by emergency physicians, GPs, and nurses (Figure 10). As expected, specificity was higher relative to most of the CPSS studies in which paramedics used the test in the field. Vanni 2011 and Mingfeng 2017 reported moderate sensitivities (0.75 (Vanni 2011) and 0.78 (Mingfeng 2017)), while the sensitivity in Mingfeng 2012 was relatively high (0.89). Vanni 2011 excluded TIA from their definition of target condition: people with positive CPSS screen diagnosed with TIA were considered false positives rather than true positives, and people with negative screen diagnosed with TIA were considered true negatives instead of false negatives. However, the effect of TIA exclusion on the reported accuracy estimates remains unclear.

Recognition Of Stroke In the Emergency Room tool (ROSIER)

Eight studies evaluated the ROSIER tool (Fothergill 2013; Jackson 2008; Jiang 2014; Lee 2015; Mingfeng 2012; Mingfeng 2017; Nor 2005; Whiteley 2011). Across all studies, the total number of participants was 2895 and ranged from 50 to 714, median 334. The mean prevalence was 64% (SD 16%) and ranged from 36% to 92%. Three studies used the test in a prehospital setting: it was administered by paramedics in Fothergill 2013, by ER physicians in Mingfeng 2012, and by GPs at a primary healthcare center in Mingfeng 2017. In the remaining five studies, the setting was the ER and the test was administered by ER physicians in three studies and ER physicians or nurses in two studies.

The methodologic quality of the studies was better compared with the studies evaluating the other tests: we considered Fothergill 2013 at high risk of selection bias because of non-consecutive sampling; Lee 2015 at high risk of bias in the flow and timing domain as they excluded 7% of the participants due to incomplete records (data provided by the authors); we could not fully assess the risk of bias in the reference standard domain in Jackson 2008; and for Jiang 2014, there were applicability concerns in the patient selection domain as 41% of the included participants were assessed more than 24 hours after the onset of symptoms (Figure 3).

The sensitivity of ROSIER was consistently high and ranged from 0.83 to 0.97, while specificity was extremely heterogeneous and ranged from 0.18 to 0.93 (Figure 13). This did not change even when studies conducted in different settings were considered separately (Figure 14). In two of the five studies conducted in the ER, the test was applied by ER physicians and nurses (versus ER physicians only); these studies reported relatively low specificity estimates. However, in Jackson 2008, ROSIER was used by ER physicians only and the reported specificity was even lower (Figure 15).

Figure 13. Forest plot of 4 Recognition of Stroke in the Emergency Room (ROSIER). ER: emergency room; GP: general practitioner.

Study	TP	FP	FN	TN	Test administrator	Setting	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
Mingfeng 2017	276	26	56	110	GP	Prehospital	0.71	0.83 [0.79, 0.87]	0.81 [0.73, 0.87]
Mingfeng 2012	341	27	38	134	Emergency physician	Prehospital	0.7	0.90 [0.86, 0.93]	0.83 [0.77, 0.89]
Fothergill 2013	171	97	6	21	Paramedic	Prehospital	0.6	0.97 [0.93, 0.99]	0.18 [0.11, 0.26]
Whiteley 2011	203	62	43	48	Emergency physician or nurse	ER	0.69	0.83 [0.77, 0.87]	0.44 [0.34, 0.53]
Lee 2015	98	14	15	185	Emergency physician	ER	0.36	0.87 [0.79, 0.92]	0.93 [0.88, 0.96]
Jiang 2014	323	202	48	141	Emergency physician or nurse	ER	0.52	0.87 [0.83, 0.90]	0.41 [0.36, 0.47]
Nor 2005	94	10	7	49	Emergency physician	ER	0.63	0.93 [0.86, 0.97]	0.83 [0.71, 0.92]
Jackson 2008	44	3	2	1	Emergency physician	ER	0.92	0.96 [0.85, 0.99]	0.25 [0.01, 0.81]

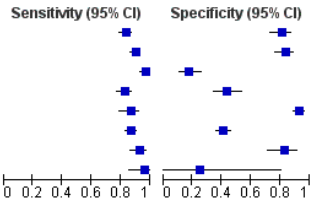


Figure 14. Summary receiver operating characteristics plot of 4 Recognition of Stroke in the Emergency Room (ROSIER). ER: emergency room.

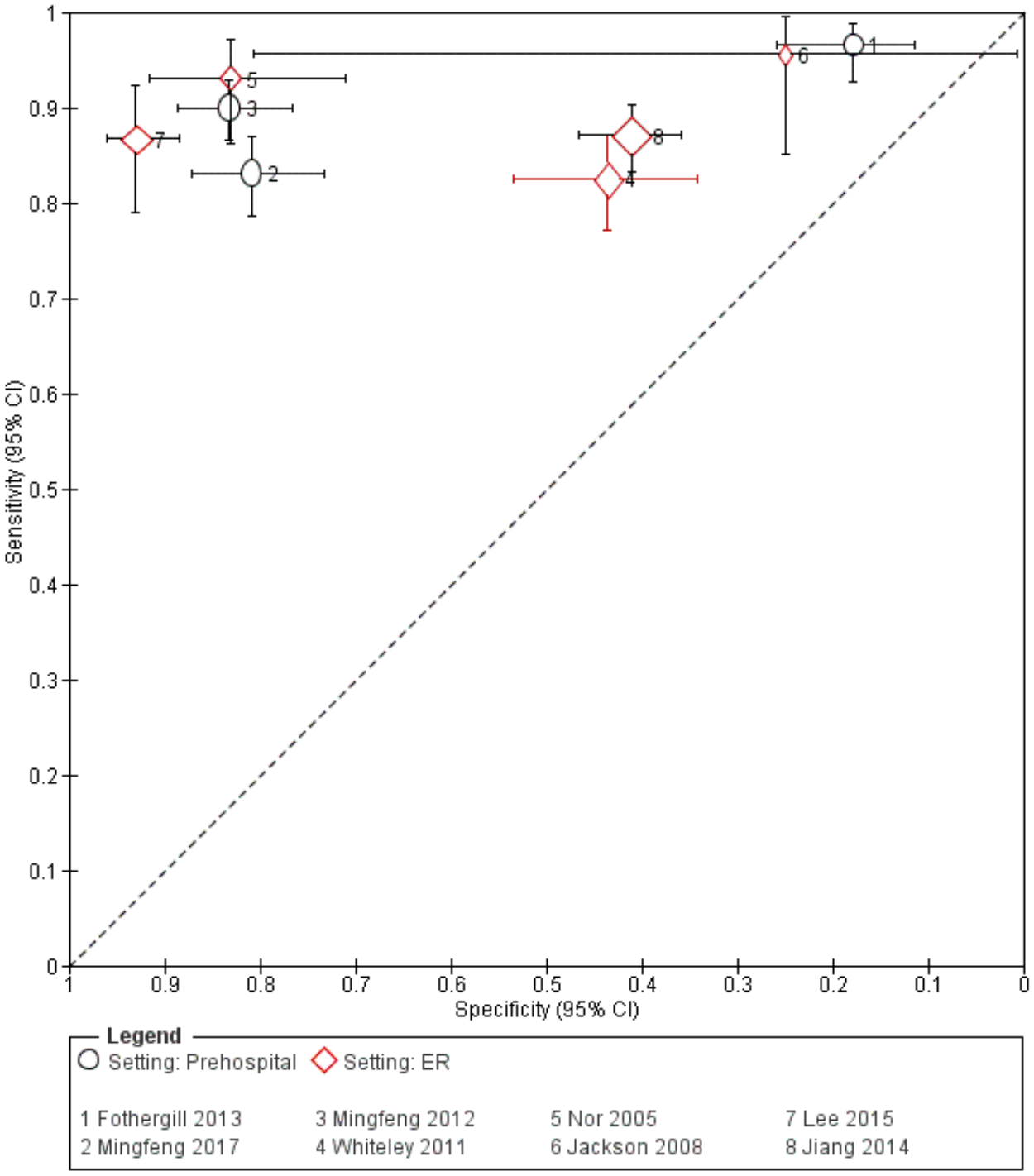
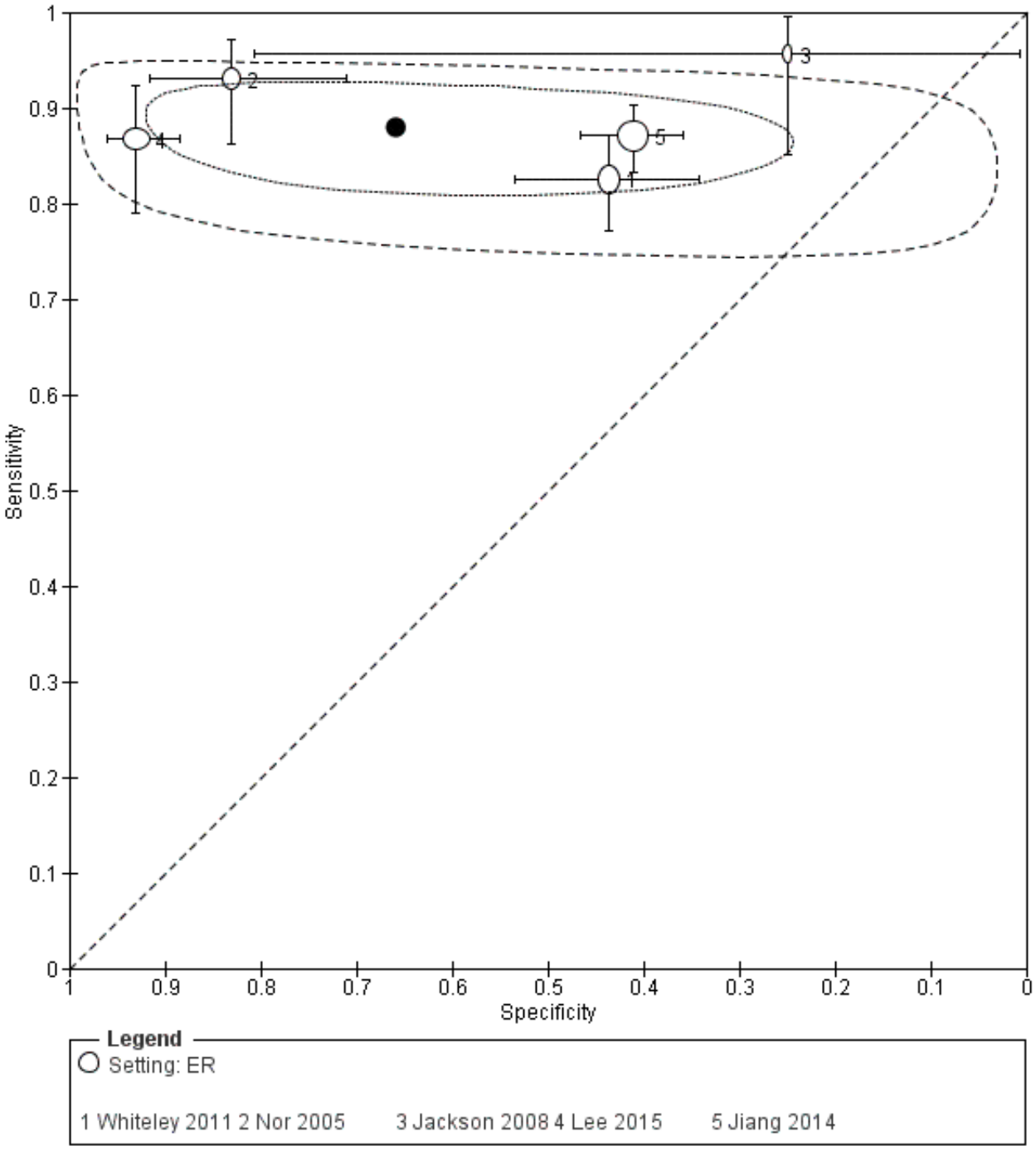


Figure 15. Summary receiver operating characteristics plot of Recognition of Stroke in the Emergency Room (ROSIER) including only studies conducted in the emergency room (ER).



Also, there was some variability in study cohorts, even when analysis was restricted to studies conducted in the ER. Thus, the prevalence ranged from 36% to 92%; the proportion of people with ischemic stroke ranged from 57% to 81%, hemorrhagic stroke from 5% to 31%, and TIA from 7% to 15%. However, we could not find obvious determinants of the difference in specificities and assumed that this was the cumulative effect of a range of factors, including differences in populations, patient selection, test administrators and reference standard. In addition, we could not rule out the presence of bias in some of the included studies.

Given the relatively consistent sensitivity estimates and better methodologic quality, we decided to pool the results from the five studies conducted in the ER. Since all studies applied the same

positivity threshold ('1 or greater') we used the random-effects bivariate model (Reitsma 2005). The mean summary sensitivity was 0.88 (95% CI 0.84 to 0.91) and the mean summary specificity was 0.66 (95% CI 0.37 to 0.86) (Figure 15). The corresponding mean positive likelihood ratio was 2.58 (95% CI 1.18 to 5.66) and the mean negative likelihood ratio was 0.18 (95% CI 0.10 to 0.32).

Figure 15 shows the extreme between-study heterogeneity and the associated statistical uncertainty around the specificity estimate. The between-study heterogeneity in sensitivity was less pronounced but still substantial, with the prediction region (the outer dashed line) indicating that the sensitivity of similar (interchangeable) future studies could range approximately from 0.75 to 0.95. Four of the five studies included in the meta-analysis reported details about the type of stroke missed by the test (Appendix 8), with the most frequent diagnosis, as reported by Nor 2005 and Whiteley 2011, being posterior circulation stroke.

Ontario Prehospital Stroke Screening Tool (OPSST)

This tool should be considered separately from the other stroke recognition scales, as the exclusion criteria in OPSST were added "for reducing the unnecessary triage of patients with stroke mimics and patients who are ineligible for fibrinolysis" (Chenkin 2009, p. 154). Therefore, the intended use of the scale was not only to identify people with stroke/TIA, but also to assess their suitability for treatment.

Only one included study evaluated OPSST and provided diagnostic accuracy data (Chenkin 2009). The study was a retrospective analysis of consecutive participants with symptoms suggesting an acute neurologic problem. Paramedics applied the tool in a prehospital setting and the participants were transported to a single stroke center. Data included 554 participants transported over a one-year period and identified by reviewing the database of the Registry of the Canadian Stroke Network.

The prevalence of stroke/TIA in the sample was 57%. Data were unavailable for participants who were screened negative and were not taken to the regional stroke center. The main focus of the study was the positive predictive value (PPV) of the tool which was 0.90 (95% CI 0.86 to 0.93). The paper reported full 2 × 2 data, but the authors cautioned that the reported estimates of sensitivity, specificity, and negative predictive value (NPV) might be biased. Sensitivity was 0.92 (95% CI 0.88 to 0.94) and specificity 0.86 (95% CI 0.80 to 0.90) (Figure 16).

Figure 16. Forest plot of 6 Ontario Prehospital Stroke Screening Tool (OPSST).

Medic Prehospital Assessment for Code Stroke (MedPACS)

Only one included study evaluated the MedPACS tool (Studnek 2013). The study was a retrospective analysis of patient records and included 416 participants. Data were obtained from the EMS electronic patient care reports and the local stroke registries. People were included in the study if they received a prehospital MedPACS screen and were transported to one of the seven local hospitals. The test was performed by paramedics and the prevalence of stroke/TIA in the study sample was 45%. The reported sensitivity was 0.74 (95% CI 0.67 to 0.80) and specificity was 0.33 (95% CI 0.27 to 0.39) (Figure 17).

Figure 17. Forest plot of 7 Medic Prehospital Assessment for Code Stroke (MedPACS).

PreHospital Ambulance Stroke Test (PreHAST)

One pilot study, with 69 participants, evaluated the accuracy of PreHAST(Andsberg 2017). This prehospital stroke scale was designed to combine stroke identification and severity assessment, and to be used by ambulance staff in the field. The study was of good methodologic quality. It had a prospective design and included unselected participants (greater than 18 years old) suspected of stroke. The test was performed by ambulance nurses who received extensive training in using the tool. Two stroke neurologists, blinded to the PreHAST results, independently made the final diagnosis, using all available information including history, clinical and imaging results.

Of the 78 participants assessed with the PreHAST tool, nine were excluded: five did not give informed consent; it was not possible to perform PreHAST in two due to agitation and an ongoing epileptic seizure, and two had no symptoms at ambulance arrival. Using a cutoff of '1 or greater' to identify people with stroke/TIA produced sensitivity estimate of 1.00 (95% CI 0.87 to 1.00) and a specificity estimate of 0.40 (95% CI 0.25 to 0.56) (Figure 18). Higher thresholds resulted in better specificity but lower sensitivity (e.g. using "≥ 2" produced sensitivity 0.81).

Figure 18. Forest plot of 8 PreHospital Ambulance Stroke Test (PreHAST).

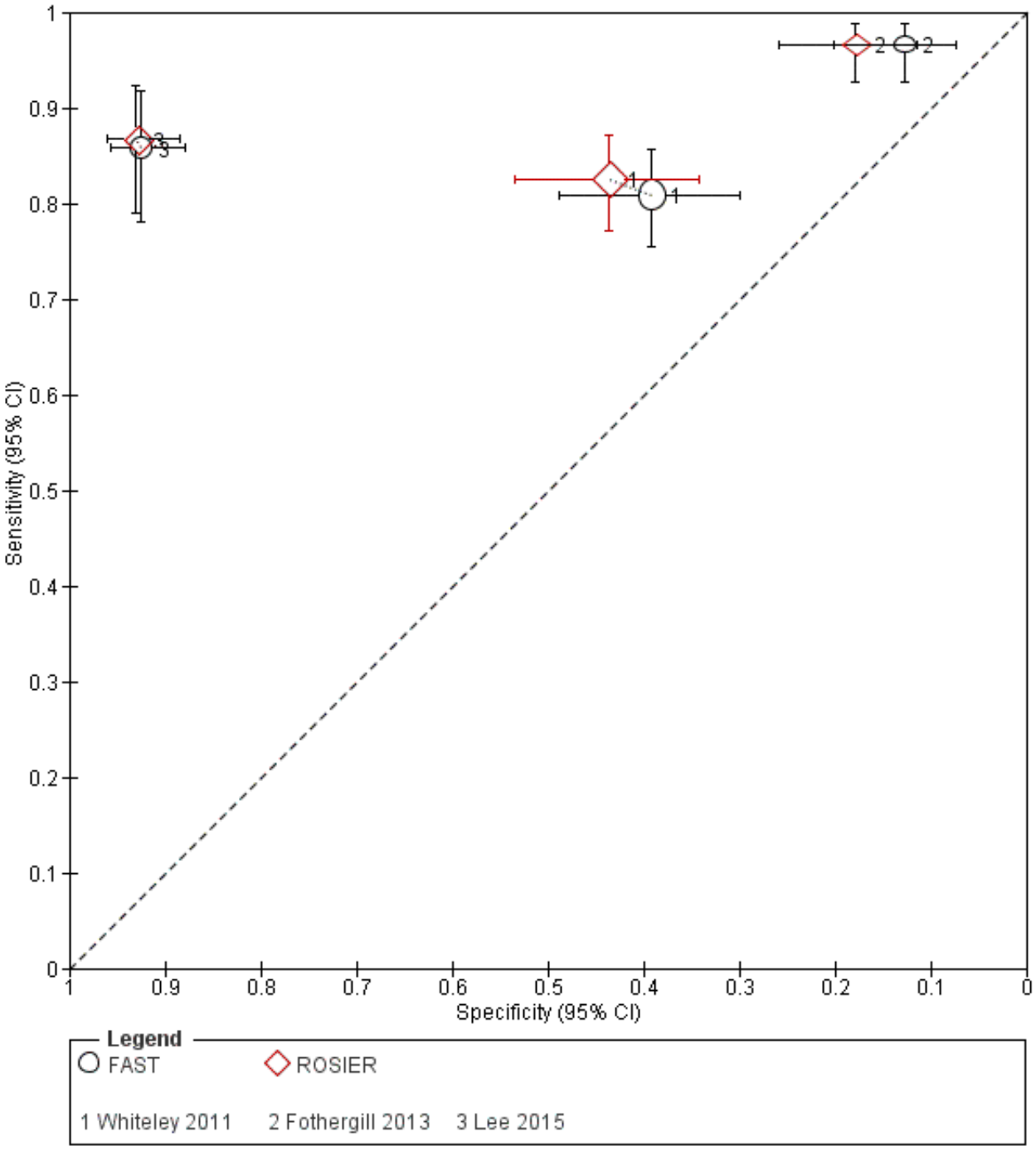
Paired test data

In this section, we report the results from studies comparing directly, in the same participants, two or more prehospital stroke scales. The small number of studies comparing the same scales precluded statistical comparison of the summary estimates of sensitivity and specificity through meta-regression. Instead, we tried to summarize the results by focusing on the statistical significance of the differences in sensitivities and specificities, as reported in the individual studies.

FAST versus ROSIER

Three studies compared directly, in the same cohort of participants, FAST and ROSIER: two in the ER (Lee 2015; Whiteley 2011), and one in a prehospital setting (Fothergill 2013). As Figure 19 shows, there was no obvious difference in the performance of the two tools. The two studies conducted in the ER reported the statistical significance of their results: Lee 2015 found no difference in the area under the curve of the two tools (0.918 for ROSIER and 0.910 for FAST, P = 0.376), and Whiteley 2011 reported that the differences in both sensitivities and specificities of the tests were non-significant (sensitivity P = 0.39 and specificity P = 0.30).

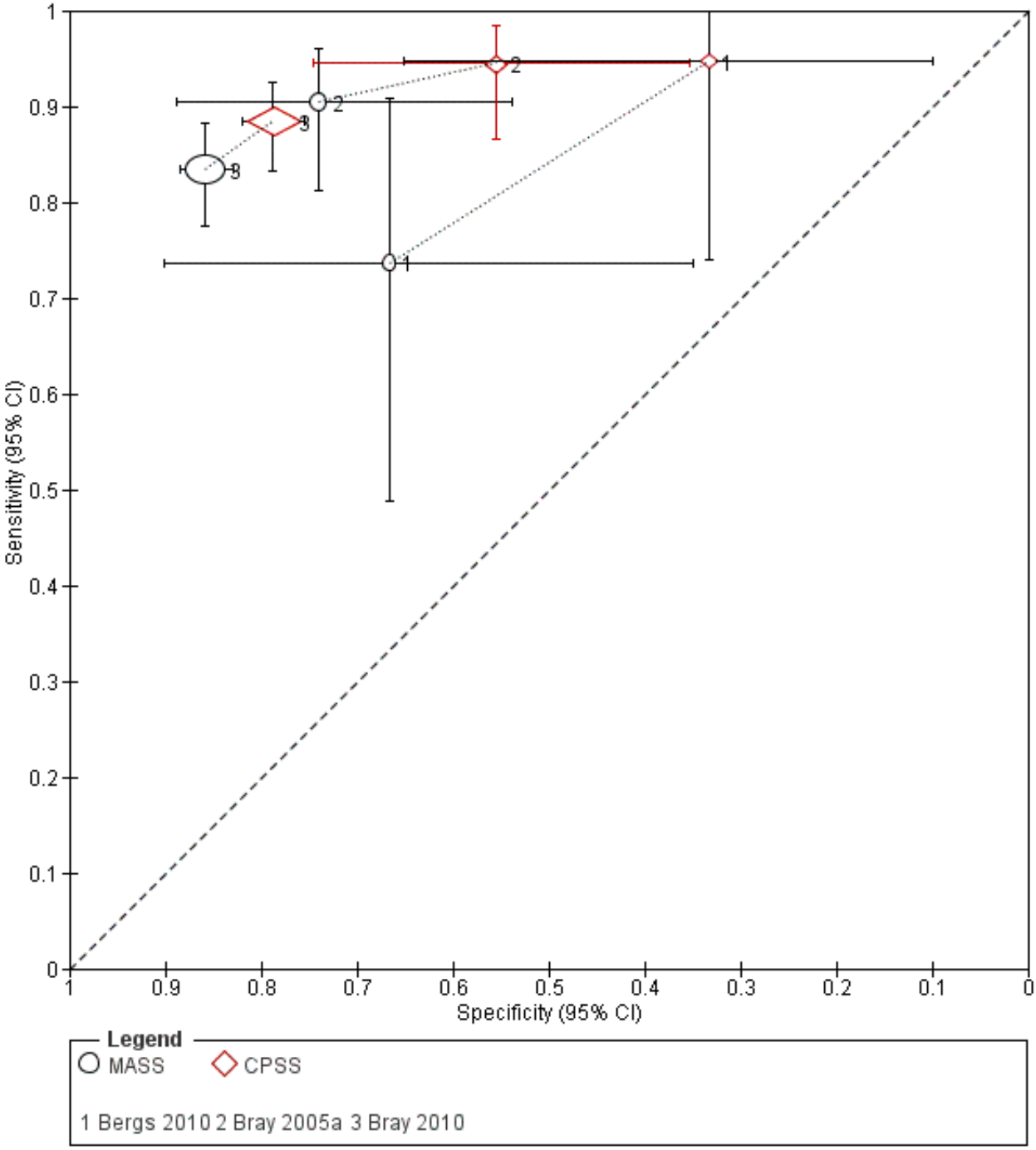
Figure 19. Summary receiver operating characteristics plot of tests: 1 Face Arm Speech Time (FAST), 5 Recognition of Stroke in the Emergency Room (ROSIER), paired data.



CPSS versus MASS

Three studies directly compared MASS and CPSS in a prehospital setting (Bergs 2010; Bray 2005a; Bray 2010) (Figure 20). In both Bray 2005a and Bray 2010, the difference in sensitivities was not statistically significant (P = 0.45 (Bray 2005a) and P = 0.149 (Bray 2010)), but MASS was more specific than CPSS (P = 0.007 (Bray 2005a) and P = 0.001 (Bray 2010)). In Bergs 2010, the differences in sensitivities (0.95 versus 0.74 in favor of CPSS) and specificities (0.67 versus 0.33 in favor of MASS) were considerable. However, this was a small study (31 participants) at high risk of selection bias and there was a significant overlap in the CIs.

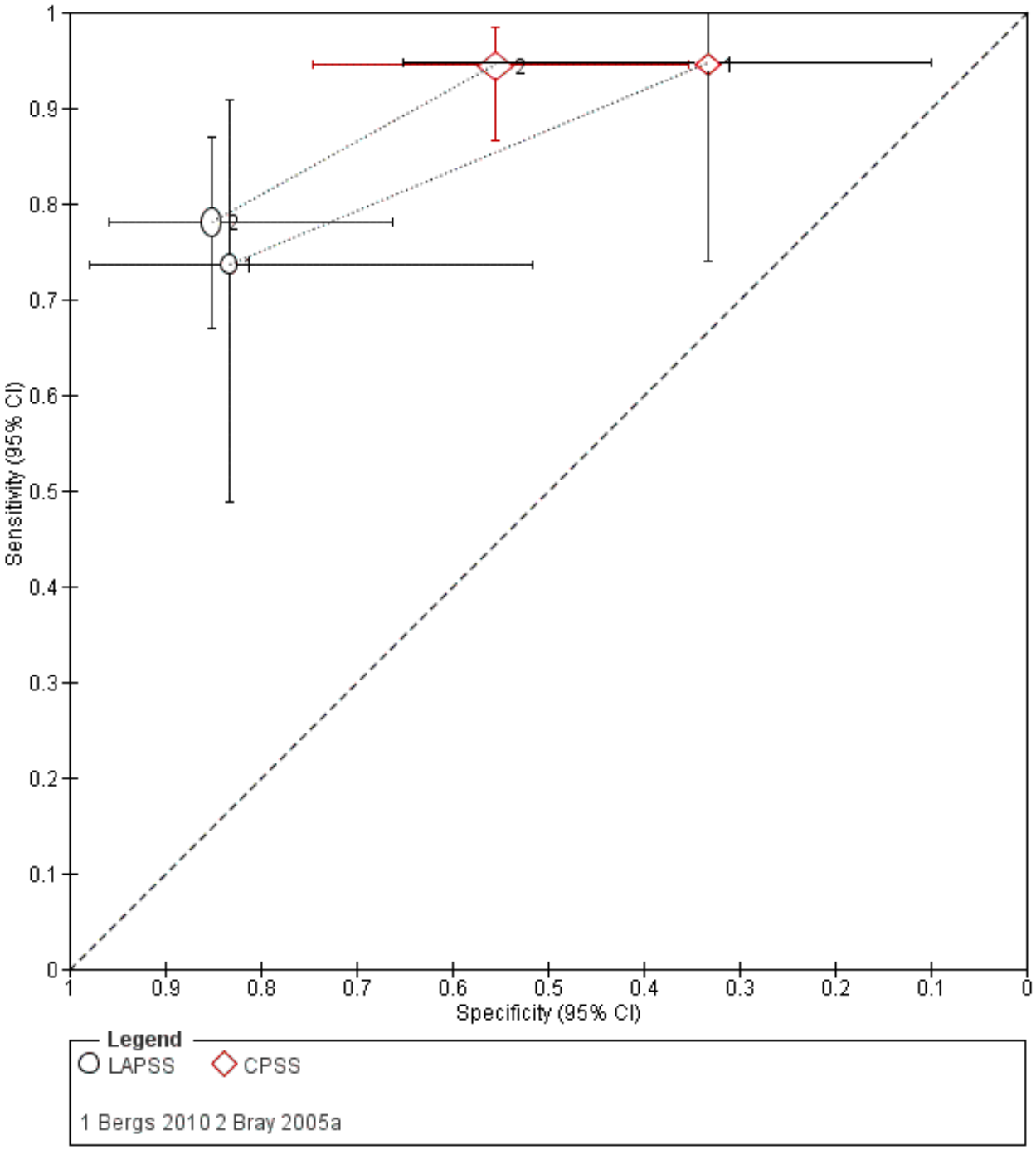
Figure 20. Summary receiver operating characteristics plot of tests: 3 Melbourne Ambulance Stroke Scale (MASS), 4 Cincinnati Prehospital Stroke Scale (CPSS), paired data.



CPSS versus LAPSS

Two studies directly compared CPSS and LAPSS (Bergs 2010; Bray 2005a). Both studies were conducted in a prehospital setting and in both studies CPSS was more sensitive but less specific than LAPSS (Figure 21). Neither of the studies reported the statistical significance of their results, but the CIs around the sensitivity estimates reported by Bray 2005a were almost non-overlapping: 0.78 (95% CI 0.67 to 0.87) versus 0.95 (95% CI 0.87 to 0.98) in favor of CPSS, suggesting statistical significance.

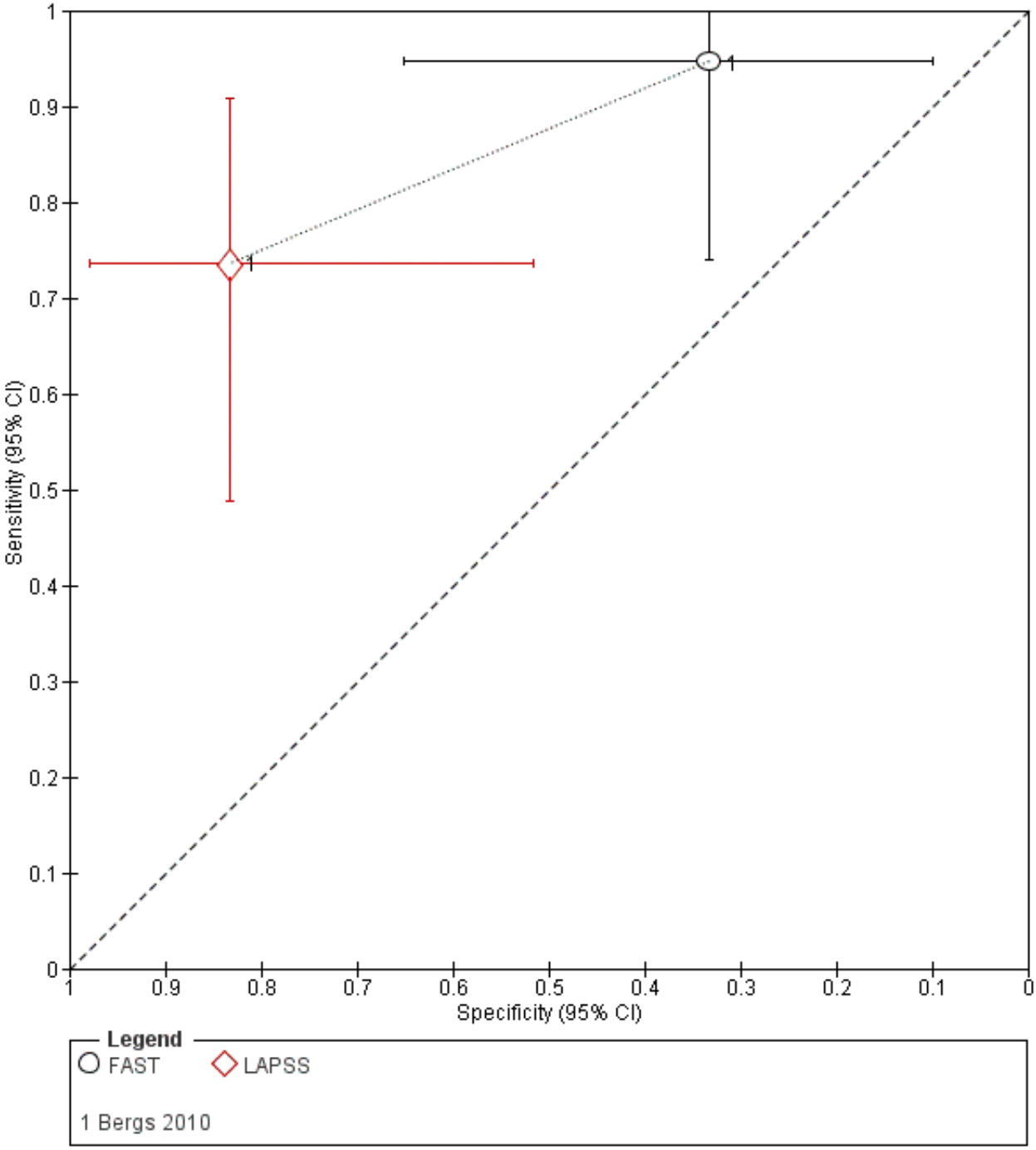
Figure 21. Summary receiver operating characteristics plot of tests: 2 Los Angeles Prehospital Stroke Scale (LAPSS), 4 Cincinnati Prehospital Stroke Scale (CPSS), paired data.



FAST versus LAPSS

One small study directly compared FAST and LAPSS (Bergs 2010; 31 participants). It reported that FAST was more sensitive but less specific than LAPSS (sensitivity: 0.95 (95% CI 0.74 to 1.00) with FAST versus 0.74 (95% CI 0.49 to 0.91) with LAPSS; specificity: 0.33 (95% CI 0.10 to 0.65) with FAST versus 0.83 (95% CI 0.52 to 0.98) with LAPSS) (Figure 22). However, this was a small study at high risk of bias and the reported estimates had overlapping CIs.

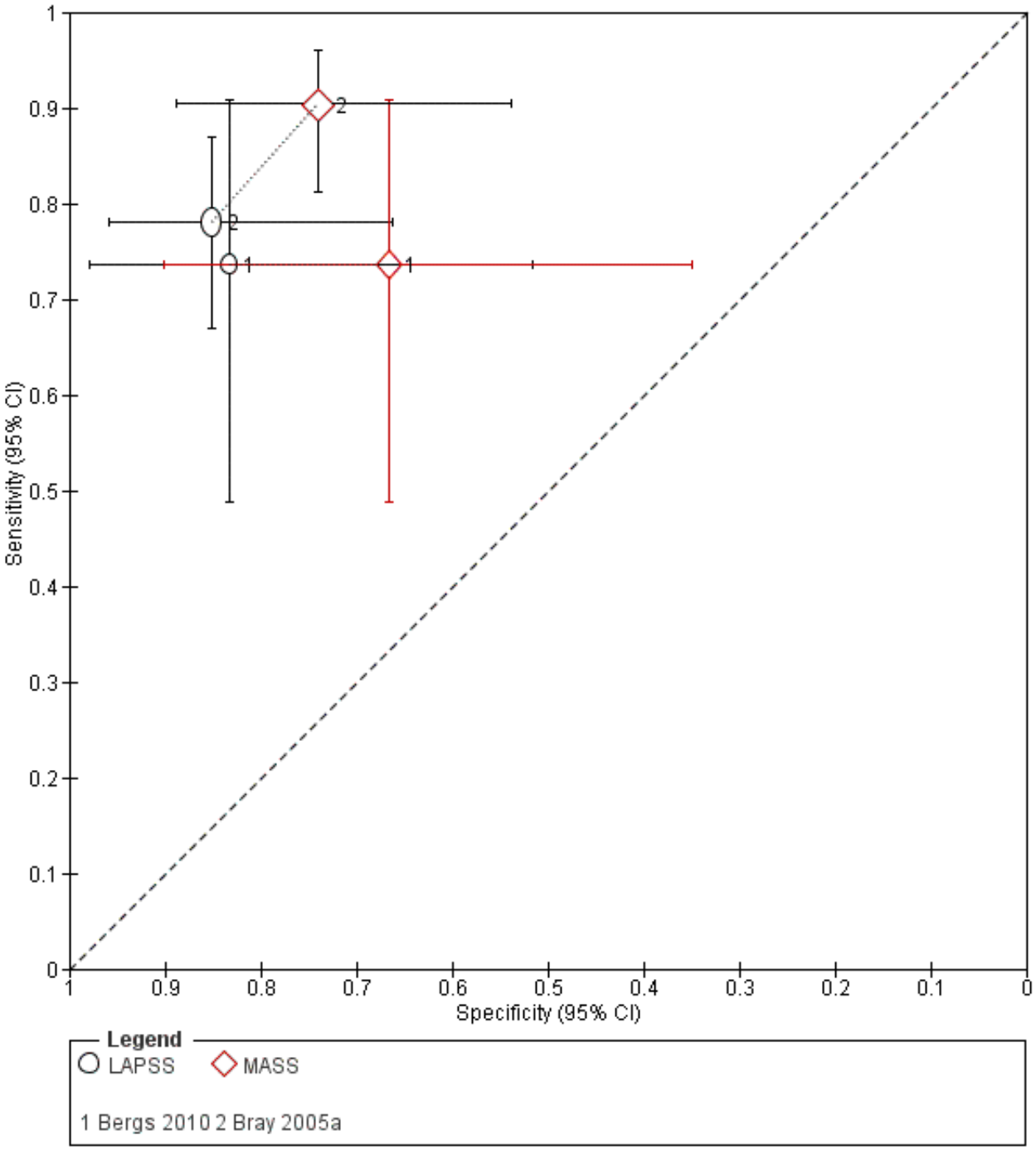
Figure 22. Summary receiver operating characteristics plot of tests: 1 Face Arm Speech Time (FAST), 2 Los Angeles Prehospital Stroke Scale (LAPSS), paired data.



MASS versus LAPSS

Two studies, both conducted in a prehospital setting, compared MASS and LAPSS (Bray 2005a; Bergs 2010). The smaller study, Bergs 2010 (31 participants), reported equal sensitivities (0.74) but LAPSS was more specific than MASS (0.83 (95% CI 0.52 to 0.98) with LAPSS versus 0.67 (95% CI 0.35 to 0.90) with MASS). However, the CIs were very wide and almost completely overlapping. In Bray 2005a (100 participants), MASS was more sensitive than LAPSS (P = 0.008), but the difference in specificities was not statistically significant (P = 0.25) (Figure 23).

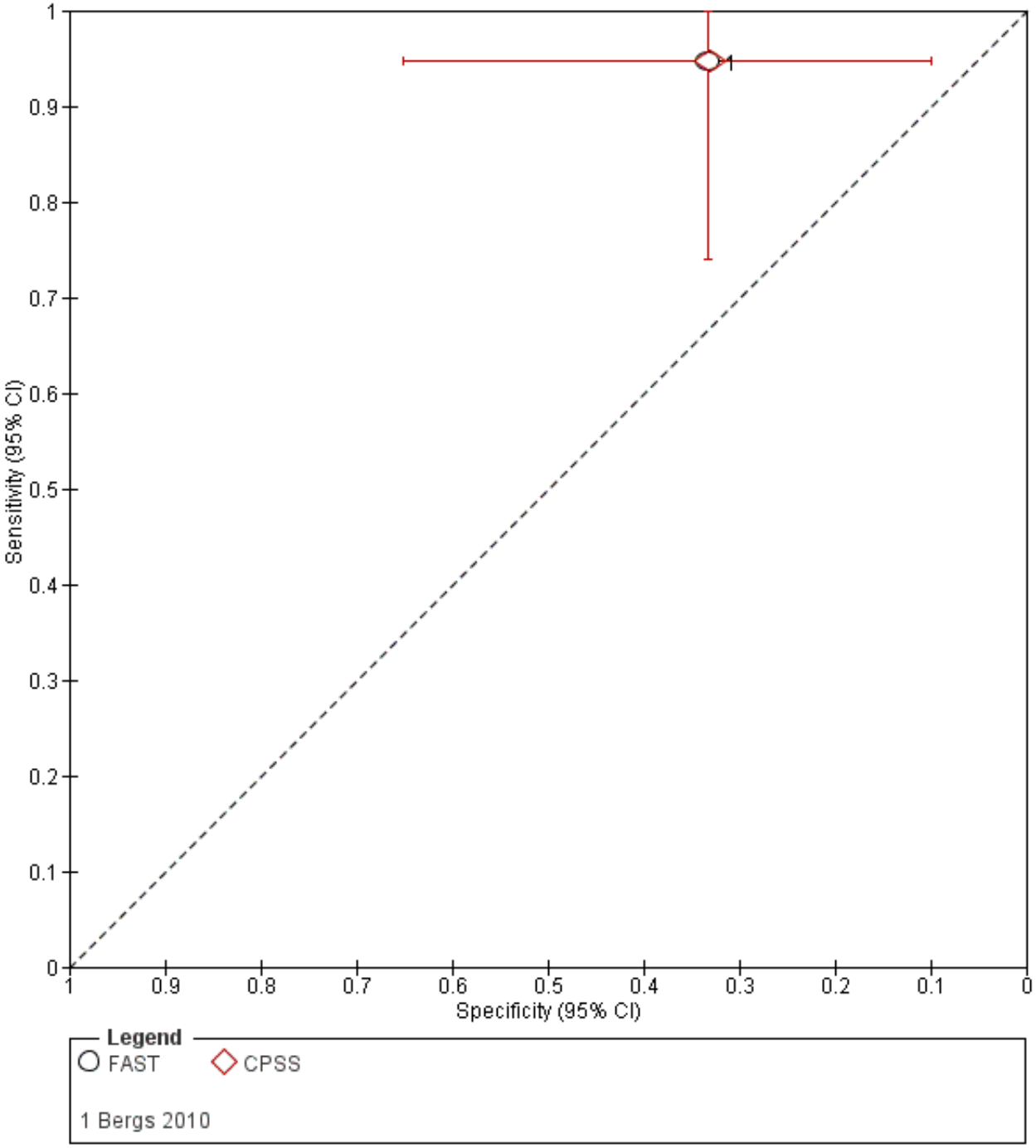
Figure 23. Summary receiver operating characteristics plot of tests: 2 Los Angeles Prehospital Stroke Scale (LAPSS), 3 Melbourne Ambulance Stroke Scale (MASS), paired data.



CPSS versus FAST

Only Bergs 2010 (31 participants) reported data on the comparison of FAST and CPSS. The estimates of sensitivity and specificity for both tests were the same (sensitivity: 0.95; specificity: 0.33) (Figure 24).

Figure 24. Summary receiver operating characteristics plot of tests: 1 Face Arm Speech Time (FAST), 4 Cincinnati Prehospital Stroke Scale (CPSS), paired data.



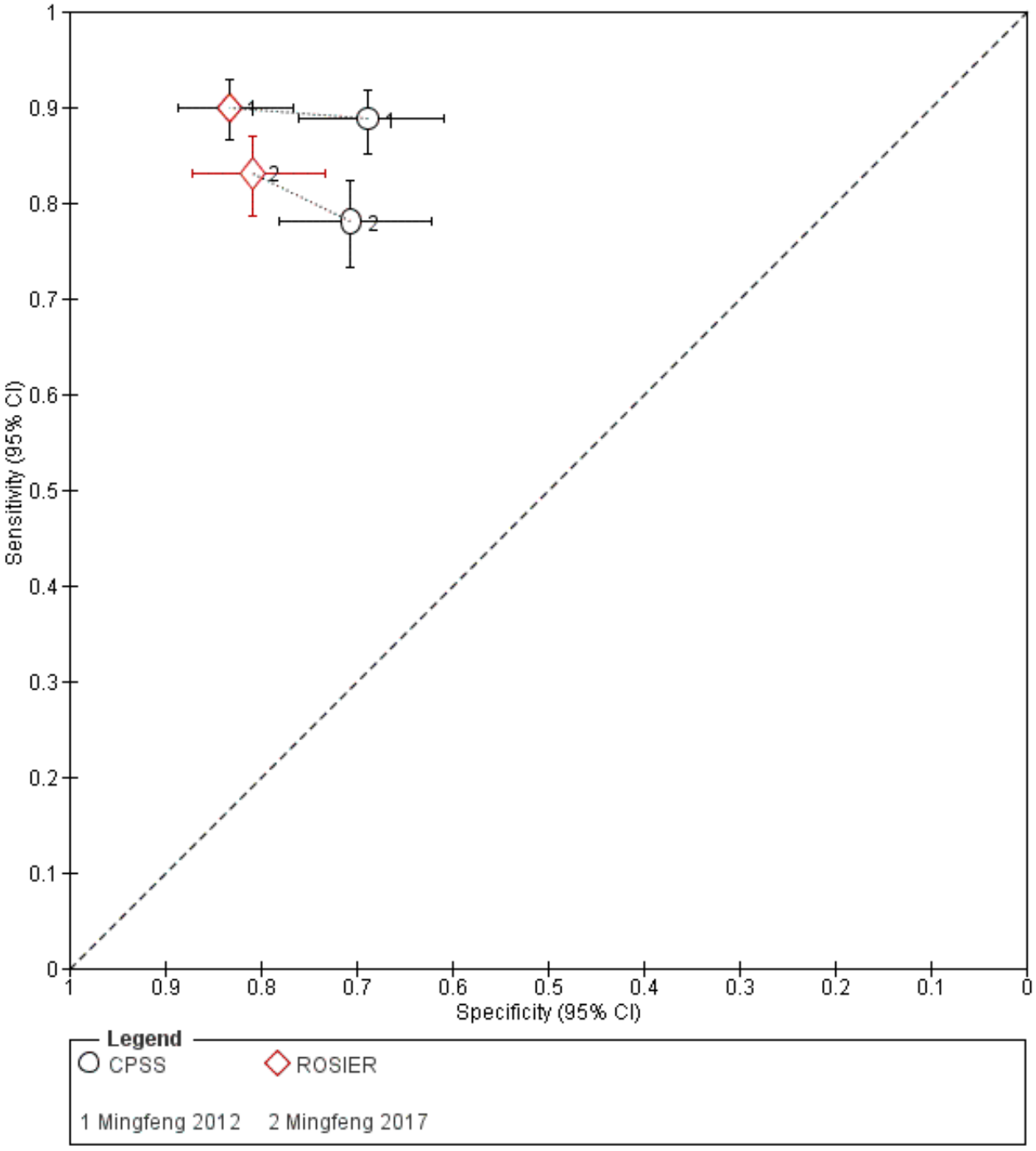
CPSS versus ROSIER

Two studies directly compared CPSS and ROSIER (Mingfeng 2012: 540 participants; Mingfeng 2017: 468 participants). The studies were conducted in China by the same research team. Mingfeng 2012 had emergency physicians in the field use the scales, whereas Mingfeng 2017 had GPs in primary healthcare centers use the scales to decide whether people suspected of stroke should be transferred to a hospital with acute stroke care facilities.

In Mingfeng 2012, the sensitivity of the two scales was equivalent, with almost completely overlapping CIs. In Mingfeng 2017, ROSIER was more sensitive than CPSS, but the CIs overlapped (0.83 (95% CI 0.79 to 0.87) with ROSIER versus 0.78 (95% CI 0.73 to 0.82) with

CPSS). In terms of specificity, ROSIER was more specific in both studies: in Mingfeng 2012 the difference appeared to be significant as the CIs did not overlap (0.83 (95% CI 0.77 to 0.89) with ROSIER versus 0.69 (95% CI 0.61 to 0.76) with CPSS). In Mingfeng 2017, the difference in specificities was less pronounced and the CIs overlapped (Figure 25).

Figure 25. Summary receiver operating characteristics plot of tests: 1 Cincinnati Prehospital Stroke Scale (CPSS), 4 Recognition of Stroke in the Emergency Room (ROSIER).

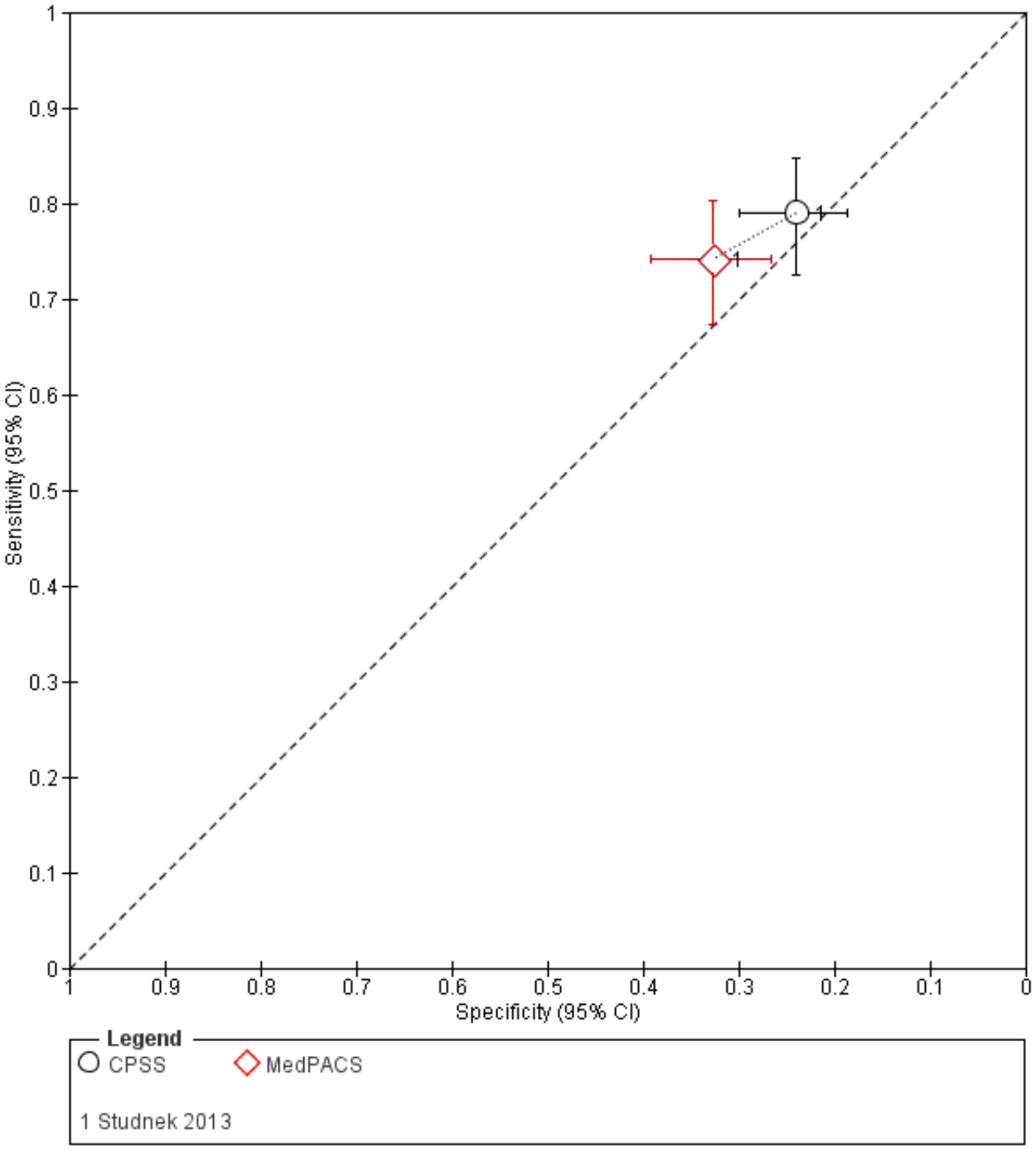


The authors of the studies reported that ROSIER was superior to CPSS both in terms of sensitivity and specificity, but that there was no difference in the "positivity rate" of the two tests. However, they did not report the statistical significance of these results and we were unable to clarify what they meant by "positivity rate".

CPSS versus MedPACS

The retrospective analysis performed by Studnek 2013 (416 participants) reported slightly higher sensitivity, but lower specificity for CPSS compared with MedPACS (sensitivity: 0.79 with CPSS versus 0.74 with MedPACS; specificity: 0.24 with CPSS versus 0.33 with MedPACS) (Figure 26). Both differences, in sensitivities and specificities, were statistically significant when compared using the McNemar's test (sensitivity: 0.048 (95% CI 0.009 to 0.088), P = 0.011; specificity: 0.086 (95% CI 0.042 to 0.131), P < 0.001).

Figure 26. Summary receiver operating characteristics plot of tests: 1 Cincinnati Prehospital Stroke Scale (CPSS), 7 Medic Prehospital Assessment for Code Stroke (MedPACS).



Summary of results for which statistical significance could be determined

We made the following observations considering only the results for which the statistical significance was reported or could be determined from the non-overlapping CIs of the accuracy estimates.

In the ER:

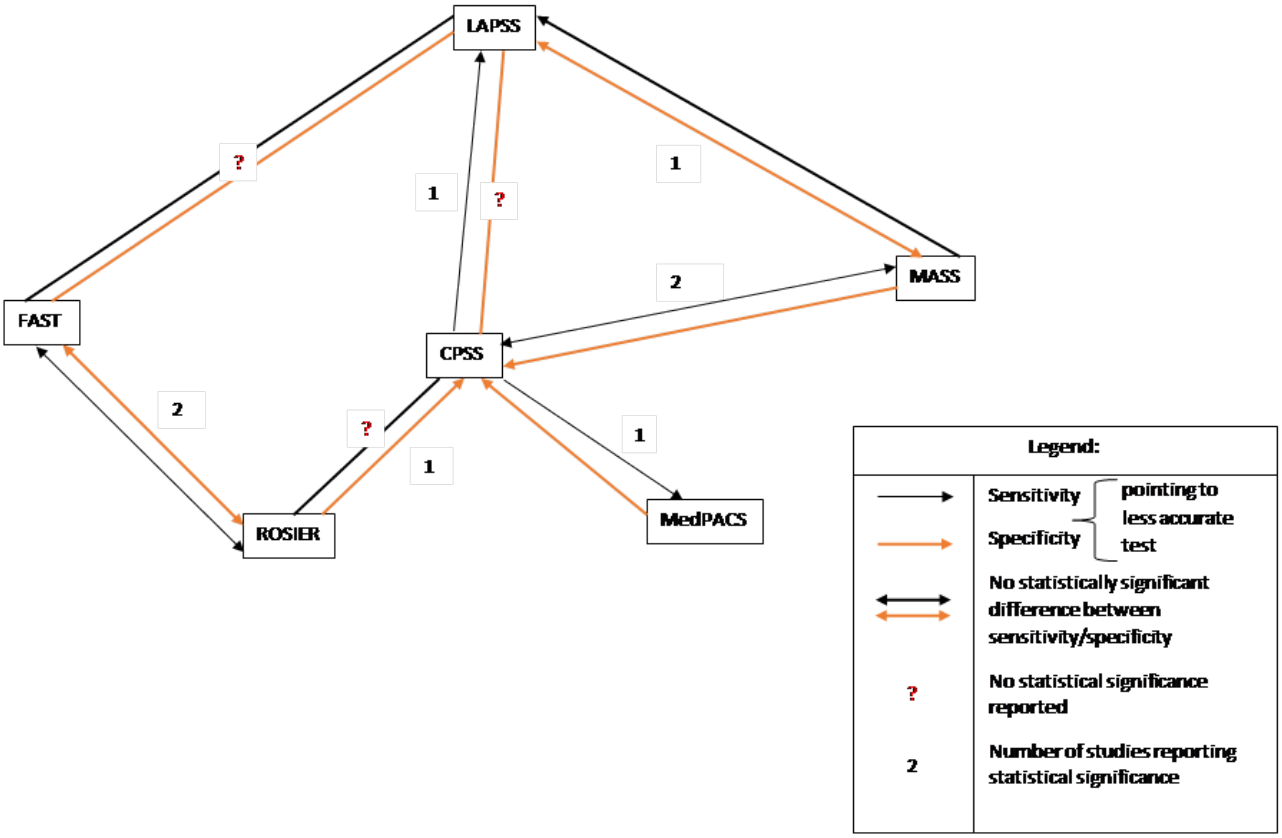
- ROSIER versus FAST: there was no statistically significant difference in sensitivities and specificities (Lee 2015; Whiteley 2011).

In the field:

- CPSS versus MASS: there was no statistically significant difference in sensitivities, but MASS was more specific (Bray 2005a; Bray 2010);
- CPSS versus ROSIER: the specificity of ROSIER was higher (the result for sensitivity was uncertain) (Mingfeng 2012);
- CPSS versus LAPSS: the difference in sensitivities was statistically significant in favor of CPSS (the difference in specificities was uncertain) (Bray 2005a);
- CPSS versus MedPACS: both the differences in sensitivity and specificity were statistically significant, with CPSS being more sensitive but less specific (Studnek 2013);
- MASS versus LAPSS: the difference in sensitivities was statistically significant in favor of MASS, but there was no statistically significant difference in specificities (Bray 2005a).

We summarized the above results in Table 5 and Figure 27. These results should be treated with caution as none of the studies reported whether they had the statistical power to detect clinically meaningful differences in sensitivity and specificity estimates. Also, many of the studies were at high or unclear risk of bias in at least one of the QUADAS-2 domains and the compared scales were not applied independently with blinding to the results of the comparator.

Figure 27. Summary diagram of the results of studies comparing directly (in the same cohort of participants) two or more scales and reporting the statistical significance of their results. The number next to the arrows indicates the number of studies comparing the tests. CPSS: Cincinnati Prehospital Stroke Scale; FAST: Face Arm Speech Time; LAPSS: Los Angeles Prehospital Stroke Scale; MASS: Melbourne Ambulance Stroke Scale; MedPACS: Medic Prehospital Assessment for Code Stroke; ROSIER: Recognition of Stroke in the Emergency Room.



Additional comparative data

In Appendix 9, we provide additional data on the comparative accuracy of the scales extracted from Purrucker 2015. We excluded this study from the main analysis as it did not meet our inclusion criteria: the test scores were derived by applying the scales to patient records rather than to actual participants. However, considering the paucity of comparative data, this study provided unique information by comparing multiple scales in the same

participants. Also, unlike most of the comparative studies discussed earlier, it included a large cohort of 689 participants and, hence, was more likely to have the statistical power to detect small differences in the accuracy estimates. Besides, indirect data collection is less likely to affect the relative accuracy of the tests, as all scores are derived using the same method. Therefore, we decided to include these data as additional evidence, but to report them separately.

The study included consecutive participants attended by EMS paramedics and emergency physicians using a prospective database (DATAPEC GmbH, Germany). All participants allocated to the database category 'suspected central nervous system disorder' were included with the exception of 33 cases with missing discharge diagnosis. The reference standard was the hospital discharge diagnosis reviewed by two of the authors (neurologists), who had access to all clinical information including imaging. Both neurologists were blinded to the results of the individual scales.

The paper reported the method of sample size calculation: the study was powered to ensure maximum marginal error of estimate of 5% with 95% confidence level of the true value of sensitivity (or specificity). The statistical significance of the differences in the accuracy of the individual scales was not reported, but for some of these comparisons they could be determined as the 95% CIs did not overlap.

Briefly, CPSS, FAST and ROSIER were the most sensitive and LAPSS, MASS and MedPACS were the most specific, with many of the differences in sensitivity and specificity being statistically significant (non-overlapping CIs) (Appendix 9). In Appendix 10, we provide an amended version of the comparative accuracy table including the results from Purrucker 2015. The main differences between the results from the included studies and Purrucker 2015 could be summarized as follows:

- FAST versus ROSIER: Purrucker 2015 detected difference in specificities, with ROSIER being more specific;
- CPSS versus MASS: Purrucker 2015 detected difference in sensitivities, with CPSS being more sensitive;
- MASS versus LAPSS: Purrucker 2015 detected a trend towards difference in specificities, with LAPSS being more specific.

The rest of the results were in agreement with those reported in the included studies. These results are consistent with the expectation that simpler scales, such as CPSS and FAST, would be more sensitive but less specific than more complex scales with eligibility criteria intended to identify stroke mimics. One possible explanation of the discrepancies in the results reported by the included studies and Purrucker 2015 was that the latter had more statistical power to detect small differences in sensitivity and specificity estimates (of the included studies only Bray 2010 had a larger sample size, 850 participants). However, this could not explain the discrepancies for CPSS versus MASS, as one of the included studies was Bray 2010.

DISCUSSION

Summary of main results

We systematically identified and reviewed studies reporting on the accuracy of prehospital stroke recognition scales. The scales had to be applied to actual people suspected of stroke, to identify those with a final diagnosis of stroke or TIA. Twenty-three studies conducted in nine different countries and reporting on eight scales met our inclusion criteria. Nine of them compared the accuracy of two or more scales in the same participants. As an additional source of evidence, we reported the results from one large study that compared multiple scales by applying them to patient records (hence, it was excluded from the main analysis; Appendix 9; Appendix 10) (Purrucker 2015). Six studies were conducted in the ER and one in primary care; in the remaining studies the scales were used in the field by ambulance crew clinicians. We pooled the results from five studies evaluating ROSIER in the ER and five studies evaluating LAPSS in the field.

Limitations of the available evidence

We could summarize the main limitations as follows:

- small number of studies per test (or comparison) conducted in the same setting;
- high or unclear risk of bias in most studies;
- significant clinical and methodologic differences between studies;
- large between-study heterogeneity in the reported estimates of test accuracy.

Paucity of evidence

Only CPSS, ROSIER and LAPSS were evaluated in more than five studies in the same setting: CPSS in nine (ambulance), ROSIER in five (the ER) and LAPSS in five (ambulance). The rest of the scales were evaluated in fewer studies at high risk of bias, precluding pooling of results or statistical investigation of heterogeneity. The number of direct comparisons was also small: only ROSIER versus FAST and CPSS versus MASS were evaluated in three studies each. All other test pairs were evaluated in one, two or none.

Methodologic quality of studies

Most of the studies had serious methodologic issues, which means that the reported estimates could be biased and not reflect the actual performance of the test in the study-specific conditions. The most important of these limitations concerned the selection of participants and the quality of the reference standard. More than half of the studies (12/23) were at high risk of selection bias because they failed to include all eligible consecutive participants. The reasons included: retrospective data collection, failure to apply the tool to all eligible participants and failure to include all people with a negative screen.

With respect to the reference standard, some of the studies, especially those using inhospital discharge diagnosis, failed to provide sufficient information on the tests and procedures, to allow full assessment of the risk of bias. Also, in the diagnosis of TIA, the index test and the reference standard are not independent by default, as the neurologist or stroke physician making the final diagnosis usually relies on the initial assessment of presenting symptoms. As Brandler 2014 pointed out, this is particularly the case where the presenting symptoms have resolved by the time of neurologic exam and the clinician had to rely on the patient's record. Differences in the reference standard and the presence of incorporation bias may explain, at least to some extent, the substantial variation in the proportion of TIA (out of all stroke/TIA diagnoses) which, across studies, ranged from 3% to 27%.

In addition, studies comparing two or more scales were at high risk of bias and applicability concerns specific to comparative studies.

The main issue was that the scales were not applied independently of each other, by test administrators blinded to the results of the alternative scale. Also, many of the studies failed to report the statistical and clinical significance of their results, and none reported whether the study had the statistical power to detect such differences.

Sources of heterogeneity

As evident from the forest and ROC plots, there was considerable between-study heterogeneity in test accuracy estimates, even when studies conducted in different settings were considered separately. Due to the small number of studies, we were unable to investigate statistically the effect of these variables. Instead, we summarized them in tables and suggested hypotheses based on visual inspection of the forest and ROC plots. The most important between-study differences that might have contributed to the observed variation included differences in study cohorts, qualification and training of test administrators, and differences in the reference standard.

All studies included people suspected of stroke, but the variation in study cohorts was considerable:

- the mean age ranged from 58 to 77 years (19 studies);
- the proportion of women ranged from 31% to 59% (18 studies);
- the prevalence of stroke/TIA ranged from 16% (Ding 2009) to 92% (Jackson 2008), mean 54% (SD 20%);
- the proportion of people with ischemic stroke ranged from 44% to 89% (15 studies), hemorrhagic stroke ranged from 4% to 41% (16 studies) and TIA ranged from 3% to 27% (14 studies) (Table 3); and
- the proportion of participants eligible for assessment with the evaluated stroke scale, out of all EMS runs, ranged from 1.3% to 34.4% (based on eight studies that reported such data).

In theory, prevalence should not affect sensitivity and specificity if the spectrum of diseased and non-diseased patients remains consistent. However, large variation in prevalence may reflect underlying differences in populations and patient selection likely to affect accuracy (Leeflang 2009). As the above summary shows, study samples varied considerably not only in terms of prevalence, but also in terms of important patient characteristics, suggesting differences in the spectrum of included participants.

One particularly important source of variation was the eligibility criteria for applying the scales. These included criteria incorporated into the scales and additional study-specific criteria. Different scales had different eligibility criteria (Table 1), which would inevitably affect the accuracy of the scale in identifying people with stroke/TIA and ruling out stroke mimics. Eligibility criteria have been added to scales, such as LAPSS, MASS, MedPACS and OPSST, to increase specificity by excluding participants with symptoms most likely to be caused by stroke mimics. Given the inverse relation between sensitivity and specificity, we can expect that increasing specificity in this way would result in lower sensitivity, as some of the ineligible participants may in fact have stroke.

Even before making a decision whether or not to apply a stroke scale, a medical responder needs to decide whether the suspected condition is likely to be stroke. A range of factors may influence this decision. We can assume that if stroke is suspected by the EMS dispatchers, the ambulance crew will more readily consider stroke in the differential diagnosis and apply the scale. Indeed, in some studies (e.g. Berglund 2014; Ramanujam 2008), all participants suspected of stroke by the EMS dispatchers were included in the study. Variation in the definition of 'suspected stroke' is also likely to contribute to variation in study samples. In some of the included studies (Table 2), 'suspicion of stroke' was defined very loosely and the decision was left to the clinical judgment of the medical responder. We can assume that in this case, qualification, training, experience and awareness that one participates in a study are all likely to affect the decision and cause variation in sample composition.

In other studies, the initial eligibility criteria were defined more tightly, acting as a prior screen that determined whether or not the stroke scale should be applied. One example of such a criterion was the time from onset of symptoms: in Berglund 2014 it was six hours whereas Kidwell 2000 used a 24-hour threshold. Another example is the rule for dealing with people who were no longer symptomatic at the time of first contact. Several studies excluded such people from their cohorts (Andsberg 2017; Bray 2005a; Bray 2010; Jiang 2014; Kidwell 2000; Mingfeng 2012; Mingfeng 2017; Nor 2005; Vanni 2011; Whiteley 2011), but the rest did not explicitly define the rule. Such cases might have been dealt with differently by different responders, especially in studies with retrospective data collection, thus contributing to the observed variation in the proportion of people with TIA. Although the available evidence did not allow us to investigate the impact of these differences, we can assume that their cumulative effect on the reported test accuracy estimates has been significant.

Even when used in the same setting, the stroke scales were administered by clinicians with different qualifications and training. For instance, in four of the 16 studies conducted in a prehospital setting, nurses, nurses or paramedics, and ER physicians administered the scales, while paramedics administered them in the rest of the studies. Training also varied, from a two-hour continuing education lecture about neurologic emergencies (Studnek 2013), to more intensive instrument-specific training followed by an exam and additional training (Kidwell 2000). In the same way, there was variation in the reference standard, with many of the studies using inhospital discharge diagnosis, while in others an independent review by blinded experts was performed (Andsberg 2017; Chen 2013; Kidwell 2000; Mingfeng 2017; Vanni 2011; Whiteley 2011) (Table 4).

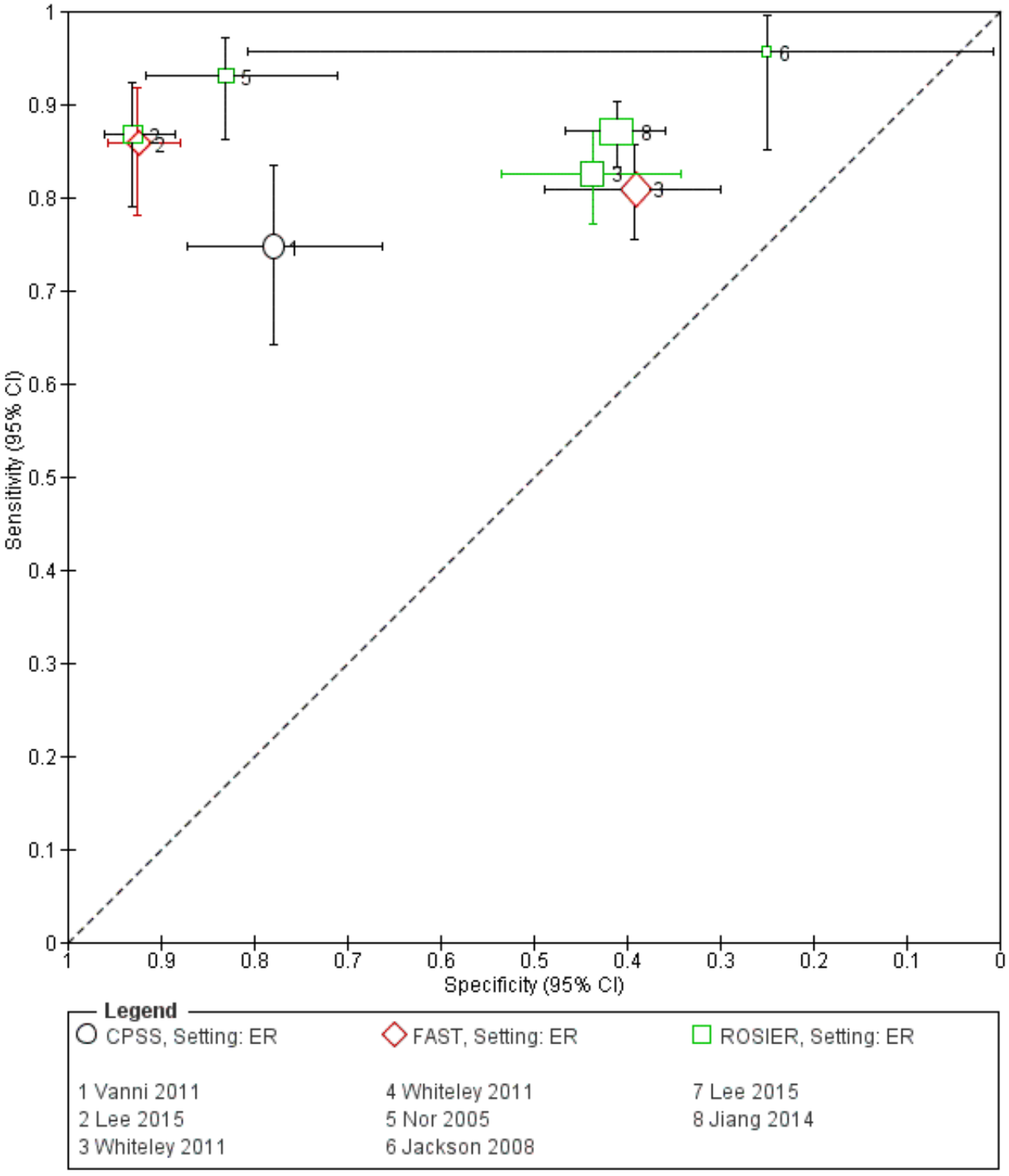
Summary of findings

Considering the above limitations, the results from the review could be summarized as follows.

Accuracy of the scales in the emergency room

Of the six studies conducted in the ER, one evaluated the accuracy of CPSS, five evaluated ROSIER, and two (out of the latter five) compared ROSIER and FAST in the same participants (Figure 28). Across all studies and scales, sensitivity was relatively high, ranging from 0.75 (CPSS, Vanni 2011) to 0.96 (ROSIER, Jackson 2008). In contrast, specificity was extremely variable, ranging from 0.25 (ROSIER, Jackson 2008) to 0.93 (ROSIER, Lee 2015). The methodologic quality of these studies was better compared with the studies conducted in the field.

Figure 28. Summary receiver operating characteristics plot of all studies conducted in the emergency room (ER). CPSS: Cincinnati Prehospital Stroke Scale; FAST: Face Arm Speech Tim; ROSIER: Recognition of Stroke in the Emergency Room.



Pooling the results of the five studies evaluating ROSIER produced a mean summary sensitivity of 0.88 (95% CI 0.84 to 0.91) and a mean summary specificity of 0.66 (95% CI 0.37 to 0.86). The prediction interval for sensitivity ranged from approximately 0.75 to 0.95. This means that the test, when used in an ER setting, will miss on average 12% of all people with stroke/TIA, although this may range from 5% to 25%, depending on the specific conditions. In a cohort of 100 participants of whom 62 have stroke/TIA (62%

mean prevalence across the five studies) the test will miss on average seven participants with stroke/TIA (ranging from three to 16). Due to the high level of between-study heterogeneity, the statistical uncertainty in summary specificity was too extreme to allow meaningful clinical interpretation. The fact that the test was performed by ER physicians in three of the studies and ER physicians or nurses in the other two did not seem to have bearing on the accuracy of the test (based on visual inspection of the ROC plot).

Although the study evaluating CPSS was of good methodologic quality, it did not include TIA in the definition of the target condition (Vanni 2011) (Figure 3). Therefore, we were unable to compare the accuracy of CPSS (sensitivity 0.75 and specificity 0.78) to that of FAST and ROSIER. Both studies comparing FAST and ROSIER in the same participants reported that there was no statistically significant difference in the area under the curve of the two tests (Lee 2015), or between their sensitivities and specificities (Whiteley 2011) (Figure 28). The high sensitivity of ROSIER could be explained by its scoring system. It contains five 'positive' items aiming to capture a broad range of stroke/TIA presentations, and two items intended to increase specificity ("no seizure at onset" and "blood glucose > 3.5 mmol/L"; both scored '–1'). The latter are subtracted from the total score, rather than making the patient ineligible for assessment (as in the other complex scales).

However, these results should be treated with caution as Purrucker 2015, which was a much larger study, reported that ROSIER was more specific than FAST (as the CIs of specificity estimates did not overlap), but probably less sensitive (although the statistical significance of this result was not reported). Purrucker 2015 applied the scales to patient records of a mixed prehospital population (ambulance and the ER), so this hypothesis needs further investigation.

Accuracy of the scales in primary care

Only one study, conducted in China, evaluated the accuracy of ROSIER and CPSS when used by GPs in a primary healthcare center (Mingfeng 2017). The test was used to decide whether people suspected of stroke should be transferred to a hospital with acute stroke care facilities. The study was of good methodologic quality (Figure 3), and reported that ROSIER was both more sensitive and more specific than CPSS (Figure 25). However, the authors did not report the statistical significance of these results. If we assume that the reported estimates were unbiased and the prevalence of stroke/TIA was 71%, in the study-specific cohort of 100 patients suspected of stroke, ROSIER will miss 12 (95% CI 9 to 15) out of 71 people with stroke/TIA and will misclassify as positive 6 (95% CI 4 to 8) out of 29 people without stroke/TIA.

Accuracy of the scales when used by ambulance staff in the field

Although most of the included studies (16) evaluated the accuracy of the scales in the field, the evidence for this setting was least reliable due to the small number of studies per test, high or unclear risk of bias, and high level of between-study heterogeneity. In this group, only the five studies evaluating LAPSS reported relatively consistent accuracy estimates that allowed pooling of results. The mean summary sensitivity was 0.83 (95% CI 0.75 to 0.89) and the mean summary specificity was 0.93 (95% CI 0.88 to 0.96). The prediction region extending down to 0.55 for sensitivity and 0.65 for specificity (Figure 7). This means that in a cohort of 100 people of whom 51 have stroke/TIA (51% mean prevalence across the five studies), on average the test will miss 9 (and up to 23) people with stroke/TIA and will misclassify as positive 3 (up to 17) people without stroke/TIA. We considered four of the five studies at high risk of selection bias and could not establish the risk of bias in the reference standard or flow and timing domain (or both) for all five studies. Therefore, we could not be sure that the summary estimates reflected the actual performance of LAPSS in this patient population.

The nine studies evaluating CPSS in the field reported extremely variable accuracy estimates, even after we stratified the results by test administrator (Figure 12). Since the risk of bias was high or unclear in most of these studies, we decided that pooling of results would be inappropriate. However, even if we consider only the better-quality studies reporting the highest accuracy estimates, the proportion of people with stroke that will be missed by the scale is still considerable. For instance, if we assume that the 0.89 (95% CI 0.85 to 0.92) sensitivity of CPSS reported by Mingfeng 2012 was unbiased, this still means that in the specific study conditions (China, prehospital, ER physicians) the test will miss between 8% and 15% of people with stroke/TIA, in the same time misclassifying as positive almost one third of the people without stroke.

The rest of the scales, FAST, MASS, OPSST, MedPACS and PreHAST, were evaluated in a small number of heterogeneous studies, most of which were at high risk of bias. Therefore, we were unable to report reliable estimates of their absolute accuracy or to compare their accuracy indirectly (across studies).

The number of studies comparing the scales in the same participants was also small and most of them were at high risk of bias or had other methodologic issues. The latter included the fact that only a few studies reported the statistical significance of their results and none reported whether the study had the statistical power to detect clinically significant differences in sensitivity and specificity. In summary, CPSS was more sensitive than MedPACS and LAPSS, but had similar sensitivity to that of MASS; and MASS was more sensitive than LAPSS. In contrast, MASS, ROSIER and MedPACS were more specific than CPSS; and the difference in the specificities of MASS and LAPSS was not statistically significant (Figure 28).

These findings should be considered with caution, as the data reported by Purrucker 2015 question some of the results (Appendix 9; Appendix 10). In particular, this study found that CPSS was more sensitive but less specific than MASS. This result could not be explained by lack of statistical power in the initial comparison, Bray 2010, as the latter study was much larger than Purrucker 2015 (sample size: 850 with Bray 2010 versus 689 with Purrucker 2015).

Despite the limitations of the available evidence, one consistent finding across all studies conducted in this setting was that CPSS was more or equally sensitive than the other scales, but has a lower specificity. This has been demonstrated in studies conducted independently from the development team, which provides further reassurance that the finding is robust. Theoretically, this makes sense as the higher sensitivity and lower specificity of CPSS, relative to the more complex scales, could be explained by the additional eligibility criteria in the latter, which increase specificity, but at the expense of sensitivity (threshold effect). We can expect FAST to have similar accuracy (high sensitivity and low specificity), as the two tests are very similar (they only differ in the assessment of

speech, Purrucker 2015), but further evidence is needed to confirm (or reject) this hypothesis.

Strengths and weaknesses of the review

In conducting the review, we followed the recommendations of the Cochrane Collaboration. We published a review protocol detailing the objectives, inclusion and exclusion criteria, and the methods for conducting the review. We ran comprehensive searches with no language or publication date restrictions. Our searches identified all studies included in the previous two reviews (Brandler 2014; Rudd 2016), plus two non-English language studies (published in Korean and Chinese) that have not been previously reviewed. When uncertain, we contacted study authors to clarify the eligibility of a study, or to obtain non-published data for included studies. Two review authors independently selected studies for inclusion, extracted data and assessed the methodologic quality of the included studies using QUADAS-2, as recommended by the Cochrane Diagnostic Test Accuracy Group.

However, we should also acknowledge some limitations of the review. First, despite our efforts, we failed to ascertain the eligibility of all studies published as conference abstracts only and to obtain additional data for some of the included studies.

Second, the focus of our review was on the accuracy of the scales and not their impact on patient outcomes. The main advantage of early identification of stroke is the opportunity for treatment of eligible patients with intravenous tPA, and now with thrombectomy. Therefore, the impact of the scales will depend first on the type of patients with stroke that are missed (false negatives) and, second, on the proportion of false positives which, if too high, could burden the emergency services and block resources needed for patients who actually have stroke. Future studies should be designed as end-to-end studies (studies that examine the impact of accuracy on patient outcomes) and include an economic evaluation, as such a design is more likely to identify the instruments that maximize the use of these interventions and lead to better patient outcomes.

Third, we focused on prehospital stroke scales designed to identify people with stroke and not to assess its severity or the person's eligibility for specific interventions. With the advances in thrombectomy, the assessment of stroke severity in a prehospital setting is likely to become more relevant and instruments combining recognition and severity assessment will attract interest. We included only one study that evaluated a 'dual purpose' scale (PreHAST), but the range of available tools is increasing. For instance, Purrucker 2015 compared several such instruments and the results suggested that the repurposed stroke severity scales could be as sensitive as CPSS and FAST in identifying stroke in the field, with the added advantage of being able to assess its severity. Well-designed end-to-end studies comparing alternative triage strategies, including those that make use of the new communication technologies, are needed to show which one is the most effective and cost-effective in a specific context.

Our review compared to previous reviews

We identified two systematic reviews, Brandler 2014 and Rudd 2016, with similar objectives to ours.

Brandler 2014, included only studies in which the scales were used by paramedics or EMTs because "physicians are not present in most EMS systems in the United States" (p. 2). All but one of the eight studies included in the review were also included in ours; we excluded Wojner-Alexandrov 2005 because it did not meet our inclusion criteria. Four of the included studies evaluated LAPSS; three evaluated CPSS; two evaluated MASS; and one each evaluated MedPACS, OPSST, ROSIER and FAST. The authors concluded that LAPSS was superior to the other scales, despite the fact that only a single study compared directly (within the same patient cohort) LAPSS to other scales (Bray 2005a), and reported that both CPSS and MASS had higher sensitivity than LAPSS (0.95 (CPSS) and 0.90 (MASS) versus 0.78 (LAPSS)). Also, two of the four studies evaluating LAPSS reported sensitivity of 0.78 (Bray 2005a; Chen 2013). The other two studies that reported higher sensitivities had serious methodologic issues: Kidwell 2000 was at high risk of selection bias because in this study paramedics completed LAPSS in only approximately half of all eligible participants; the accuracy estimates reported by Wojner-Alexandrov 2005 related to the paramedic's diagnosis of stroke in unselected ambulance patients, the majority of whom would not have been suspected of stroke in the first place.

In our review, five studies evaluated LAPSS: in addition to Bray 2005a, Chen 2013, and Kidwell 2000, we included Bergs 2010, in which the scales were completed by nurses, and Ding 2009, in which test administrators were emergency physicians (we excluded Wojner-Alexandrov 2005 for the above reason). Although we pooled the results from the five included studies, we could not be certain that the reported summary estimates were unbiased. The high risk of bias and between-study variation in test accuracy means that indirect (between-study) comparison of the scales was unlikely to produce valid results. Instead, we focused on direct (within-study) comparisons which, as stated earlier, suggested that LAPSS was less sensitive than CPSS and MASS, but more specific than CPSS, a conclusion very different from that made in Brandler 2014.

The second review conducted by Rudd and colleagues included all studies in which the scales were administered face-to-face by any prehospital or hospital clinicians to identify adults suspected of stroke (Rudd 2016). In comparison to our review, they did not include studies published in languages other than English or German, but included studies without complete 2 × 2 data, such as studies that included only screen-positive patients. They included 21 studies (18 papers and three abstracts) evaluating the same stroke scales as in our review (with the exception of PreHAST). The authors of the review discussed various sources of bias and between-study heterogeneity, and concluded that: "Available data do not allow a strong recommendation to be made about the superiority of a stroke recognition instrument." (Rudd 2016, p.1).

While we agree with this general conclusion, we feel that we could be more specific. We included six studies (two published in Chinese and Korean, and four published in 2017) that were not included in Rudd 2016. We pooled the results for two scales: we obtained setting-specific summary sensitivity for ROSIER based on five studies of better methodologic quality; and summary sensitivity and specificity for LAPSS which, unfortunately, were based on studies at high risk of bias. In addition, by focusing on comparative studies, we were able to arrive at more specific conclusions about the relative accuracy of the scales that agree with what would be theoretically expected, considering the tradeoff between sensitivity and specificity.

Applicability of findings to the review question

Our findings were relevant to the review question we tried to answer.

AUTHORS' CONCLUSIONS

Implications for practice

The primary goal of a prehospital stroke scale is to identify all people with stroke or transient ischemic attack (TIA), so they can be transported to hospital and undergo diagnostic assessment without delay. Some of these people will experience irreversible brain damage if they do not receive treatment in the first few hours of the onset of symptoms. We do not want to miss any people with stroke, even at the expense of some reasonable overcapturing of stroke mimickers. Therefore, in terms of accuracy, we consider sensitivity the prime metric for determining a 'superior' stroke scale.

Our findings suggest that in the field Cincinnati Prehospital Stroke Scale (CPSS) had consistently the highest sensitivity, but was less specific than most of the scales. Unfortunately, the high risk of bias and extreme between-study heterogeneity did not allow us to obtain summary estimates of the absolute accuracy of CPSS. Also, further evidence is needed to determine whether alternative scales that might have comparable sensitivity but higher specificity, such as Melbourne Ambulance Stroke Scale (MASS) and Recognition of Stroke in the Emergency Room (ROSIER), should be used instead, to achieve better overall accuracy.

The reviewed evidence suggests that in the emergency room (ER), ROSIER should be the test of choice, as it was evaluated in the largest number of studies, its sensitivity was consistently high and had similar accuracy to Face Arm Speech Time (FAST), the only alternative scale it was directly compared against in this setting. When used in the ER, ROSIER will miss on average 12% (range 5% to 25%) of all people with stroke/TIA. This means that in a cohort of 100 people of whom 62 have stroke/TIA the test will miss on average seven people with stroke/TIA (range three to 16). We are unable to provide an estimate of the summary specificity, as the study-level results were too heterogeneous.

Because of the small number of studies per test per setting, high risk of bias, substantial differences in study characteristics and large between-study heterogeneity in the reported accuracy estimates, these findings should be treated with caution. They are only provisional hypotheses that need further verification in better-designed studies.

Implications for research

Future studies should try to address some of the methodologic issues identified here. Given the significant variation in clinical practice and populations, indirect comparisons are unlikely to produce valid results. Also, the narrow focus on test accuracy does not allow more important questions to be answered, such as the impact of alternative triage strategies on patient outcomes and their cost effectiveness. Scales that have the same accuracy may have very different clinical effectiveness if they identify a different proportion of the patients that would benefit from early treatment (as opposed to those who would not, e.g. people with TIA).

Unfortunately, the impact of different triage protocols on patient outcomes is not a simple question: first, because it is difficult to predict who will benefit from early identification; and, second, because subsequent clinical decisions, such as diagnostic assessment and treatment, will also contribute to the final outcome. Such questions could not be answered by diagnostic accuracy studies and, therefore, future research should be designed as end-to-end studies. They should incorporate health economic evaluations and compare directly alternative triage strategies, including alternative stroke scales intended for the same use in a specific setting; stroke recognition scales and repurposed stroke severity scales; or the use of new communication technologies, to determine which one is the most effective and cost-effective in a specific context.

Researchers should try to maximize the external validity of such studies, bearing in mind that the introduction of extra elements (e.g. intensive training) that will not be implemented in practice, may jeopardize the applicability of results. Care should be taken to include all consecutive patients suspected of stroke, including those with a negative screen. The trigger for applying the test (e.g. clinical suspicion versus prespecified criteria) is also likely to affect the estimated accuracy of the scales and future studies should try to investigate this in order to identify the most effective strategy.

It is unclear if using different methods to obtain scores when comparing two or more scales (as described earlier) has bearing on the accuracy of the scales. If randomizing people to assessment with alternative instruments is not feasible, the impact of using one or another method should be considered when collecting data (e.g. through a qualitative investigation) and in interpreting the results. Using the hospital discharge diagnosis as a reference standard could introduce bias and contribute to heterogeneity. Therefore, a chart review by independent neurologists/stroke physicians using all available information should be preferred.

ACKNOWLEDGEMENTS

We acknowledge the Cochrane Stroke Group for their input and participation.

We thank Dr Sang Min Sung for translating from Korean, and Dr Huiqin Yang for translating from Chinese.

REFERENCES

References to studies included in this review

Andsberg 2017 *{published data only}*

Andsberg G, Esbjörnsson M, Olofsson A, Lindgren A, Norrving B, von Euler M. PreHospital Ambulance Stroke Test – pilot study of a novel stroke test. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 2017;**25**:37.

Berglund 2014 *{published data only}*

Berglund A, Svensson L, Wahlgren N, von Euler M. Face Arm Speech Time Test use in the prehospital setting, better in the ambulance than in the emergency medical communication center. *Cerebrovascular Diseases* 2014;**37**(3):212-6.

Bergs 2010 *{published data only}*

Bergs J, Sabbe M, Moons P. Prehospital stroke scales in a Belgian pre-hospital setting: a pilot study. *European Journal of Emergency Medicine* 2010;**17**(1):2-6.

Bray 2005a *{published data only}*

Bray JE, Martin J, Cooper G, Barger B, Bernard S, Bladin C. Paramedic identification of stroke: community validation of the Melbourne Ambulance Stroke Screen. *Cerebrovascular Diseases* 2005;**20**(1):28-33.

Bray 2010 *{published data only}*

Bray JE, Coughlan K, Barger B, Bladin C. Paramedic diagnosis of stroke: examining long-term use of the Melbourne Ambulance Stroke Screen (MASS) in the field. *Stroke* 2010;**41**(7):1363-6.

Chen 2013 *{published data only}*

Chen S, Sun H, Lei Y, Gao D, Wang Y, Wang Y, et al. Validation of the Los Angeles Pre-hospital Stroke Screen (LAPSS) in a Chinese urban emergency medical service population. *PloS One* 2013;**8**(8):e70742.

Chenkin 2009 *{published data only}*

Chenkin J, Gladstone DJ, Verbeek PR, Lindsay P, Fang J, Black SE, et al. Predictive value of the Ontario Prehospital Stroke Screening Tool for the identification of patients with acute stroke. *Prehospital Emergency Care* 2009;**13**(2):153-9.

Ding 2009 *{published data only}*

Ding HX, Chen H, Yin ST, Wang WP, Shen XY, Shi SF, et al. Effect of the Los Angeles Prehospital Stroke Screen in prehospital diagnosis of stroke. *Chinese Journal of Cerebrovascular Diseases* 2009;**6**(6):285-8.

English 2018 *{published data only}*

English SW, Rabinstein AA, Mandrekar J, Klaas JP. Rethinking prehospital stroke notification: assessing utility of Emergency Medical Services Impression and Cincinnati Prehospital Stroke Scale. *Journal of Stroke and Cerebrovascular Diseases* 2018;**27**(4):919-25.

Fothergill 2013 *{published data only}*

Fothergill RT, Williams J, Edwards MJ, Russell IT, Gompertz P. Does use of the recognition of stroke in the emergency room stroke assessment tool enhance stroke recognition by ambulance clinicians?. Stroke 2013; Vol. 44, issue 11:3007-12.

Frendl 2009 *{published data only}*

Frendl DM, Strauss DG, Underhill BK, Goldstein LB. Lack of impact of paramedic training and use of the Cincinnati Prehospital Stroke Scale on stroke patient identification and on-scene time. Stroke 2009; Vol. 40, issue 3:754-6.

Jackson 2008 *{published data only}*

Jackson A, Deasy C, Geary UM, Plunkett PK, Harbison J. Validation of the use of the ROSIER stroke recognition instrument in an Irish emergency department. *Irish Journal of Medical Science* 2008;**177**:189-92.

Jiang 2014 *{published data only}*

Jiang HL, Chan CPY, Leung YK, Li YM, Graham CA, Rainer TH. Evaluation of the Recognition of Stroke in the Emergency Room (ROSIER) scale in Chinese patients in Hong Kong. PloS One 2014; Vol. 9, issue 10:e109762.

Kidwell 2000 *{published data only}*

Kidwell CS, Starkman S, Eckstein M, Weems K, Saver JL. Identifying stroke in the field. Prospective validation of the Los Angeles Prehospital Stroke Screen (LAPSS). *Stroke* 2000;**31**:71-6.

Kim 2017 *{published data only}*

Kim DH, Kim SW, Jun SM, Kim JK. Accuracy of prehospital stroke recognition by paramedics and TPA therapy rate after transportation. *Journal of the Neurological Sciences* 2017;**381 Suppl 1**:867.

Lee 2015 *{published data only}*

Lee S, Seo JS, Lee SC, Lee JH, Doh H. Prospective evaluation of the Recognition Of Stroke In the Emergency Room (ROSIER) Scale in emergency department. *Journal of the Korean Society of Emergency Medicine* 2015;**26**(5):466-73.

Mingfeng 2012 *{published data only}*

Mingfeng H, Zhixin W, Qihong G, Lianda L, Yanbin Y, Jinfang F. Validation of the use of the ROSIER scale in prehospital assessment of stroke. *Annals of Indian Academy of Neurology* 2012;**15**(3):191-5.

Mingfeng 2017 *{published data only}*

Mingfeng H, Zhixin W, Jianyi Zh, Gai Zh, Yingying L, Wenyuan Ch, et al. ROSIER scale is useful in an emergency medical service transfer protocol for acute stroke patients in primary care center: a southern China study. *Neurology Asia* 2017;**22**(2):93-8.

Nor 2005 *{published data only}*

Nor AM, Davis J, Sen B, Shipsey D, Louw SJ, Dyker AG, et al. The Recognition of Stroke in the Emergency Room (ROSIER) scale: development and validation of a stroke recognition instrument. Lancet Neurology 2005; Vol. 4, issue 11:727-34.

Ramanujam 2008 *{published data only}*

Ramanujam P, Guluma KZ, Castillo EM, Chacon M, Jensen MB, Patel E, et al. Accuracy of stroke recognition by emergency

medical dispatchers and paramedics – San Diego experience. *Prehospital Emergency Care* 2008;**12**:307-13.

Studnek 2013 *{published data only}*

Studnek JR, Asimos A, Dodds J, Swanson D. Assessing the validity of the Cincinnati Prehospital Stroke Scale and the medic prehospital assessment for code stroke in an urban emergency medical services agency. Prehospital Emergency Care 2013; Vol. 17, issue 3:348-53.

Vanni 2011 *{published data only}*

Vanni S, Polidori G, Pepe G, Chiarlone M, Albani A, Pagnanelli A, et al. Use of biomarkers in triage of patients with suspected stroke. Journal of Emergency Medicine 2011; Vol. 40, issue 5:499-505.

Whiteley 2011 *{published data only}*

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. *Journal of Neurology, Neurosurgery, and Psychiatry* 2011;**82**:1006-10.

References to studies excluded from this review

Alhanati 2014 *{published data only}*

Alhanati L, Dubourdieu S, Hoffmann C, Beguec F, Travers S, Lefort H, et al. Stroke: prospective evaluation of a prehospital management process based on rescuers under medical direction. *American Journal of Emergency Medicine* 2014;**32**(5):438-42.

Asimos 2014 *{published data only}*

Asimos AW, Ward S, Brice JH, Rosamond WD, Goldstein LB, Studnek J. Out-of-hospital stroke screen accuracy in a state with an emergency medical services protocol for routing patients to acute stroke centers. *Annals of Emergency Medicine* 2014;**64**(5):509-15.

Belvis 2005 *{published data only}*

Belvis R, Cocho D, Marti-Fabregas J, Pagonabarraga J, Aleu Ai, Garcia-Bargo MD, et al. Benefits of a prehospital stroke code system. Feasibility and efficacy in the first year of clinical practice in Barcelona, Spain. *Cerebrovascular Diseases* 2005;**19**(2):96-101.

Benjamin 2013 *{published data only}*

Benjamin LA, Joekes El, Das K, Beeching NJ, Wilkins E, Solomon T. Diagnostic accuracy of the Recognition of Stroke in the Emergency Room (ROSIER) score and CT brain in an HIV population. *Journal of Infection* 2013;**67**(6):619-22.

Birnbaum 2008 *{published data only}*

Birnbaum A, Esses D, Bijur P, Wollowitz A, Gallagher EJ. Failure to validate the San Francisco Syncope Rule in an independent emergency department population. *Annals of Emergency Medicine* 2008;**52**(2):151-9.

Blomberg 2014 *{published data only}*

Blomberg H, Lundstrom E, Toss H, Gedeborg R, Johansson J. Agreement between ambulance nurses and physicians in assessing stroke patients. *Acta Neurologica Scandinavica* 2014;**129**(1):49-55.

Bray 2005b *{published data only}*

Bray JE, Martin J, Cooper G, Barger B, Bernard S, Bladin C. An interventional study to improve paramedic diagnosis of stroke. *Prehospital Emergency Care* 2005;**9**(3):297-302.

Brott 1989 *{published data only}*

Brott T, Adams HP Jr, Olinger CP, Marler JR, Barsan WG, Biller J, et al. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* 1989;**20**(7):864-70.

Buck 2009 *{published data only}*

Buck BH, Starkman S, Eckstein M, Kidwell CS, Haines J, Huang R, et al. Dispatcher recognition of stroke using the National Academy Medical Priority Dispatch System. *Stroke* 2009;**40**(6):2027-30.

Caceres 2013 *{published data only}*

Caceres JA, Adil MM, Jadhav V, Chaudhry SA, Pawar S, Rodriguez GJ, et al. Diagnosis of stroke by emergency medical dispatchers and its impact on the prehospital care of patients. *Journal of Stroke and Cerebrovascular Diseases* 2013;**22**(8):e610-4.

Camerlingo 2002 *{published data only}*

Camerlingo M, Casto L, Censori B, Ferraro B, Gazzaniga G, Partziguian T, et al. Experience with a questionnaire administered by emergency medical service for pre-hospital identification of patients with acute stroke. *Neurological Sciences* 2002;**22**(5):357-61.

Casolla 2013 *{published data only}*

Casolla B, Bodenant M, Girot M, Cordonnier C, Pruvo J P, Wiel E, et al. Intra-hospital delays in stroke patients treated with rt-PA: impact of preadmission notification.. *Journal of Neurology* 2013;**260**(2):635-9.

Deakin 2009 *{published data only}*

Deakin C D, Alasaad M, King P, Thompson F. Is ambulance telephone triage using advanced medical priority dispatch protocols able to identify patients with acute stroke correctly?. *Emergency Medicine Journal* 2009;**26**(6):442-5.

De La Ossa 2014 *{published data only}*

De La Ossa NP, Carrera D, Gorchs M, Querol M, Millan M, Gomis M, et al. Design and validation of a prehospital stroke scale to predict large arterial occlusion: the Rapid Arterial oCclusion Evaluation Scale. *Stroke* 2014;**45**(1):87-91.

Demeestere 2017 *{published data only}*

Demeestere J, Garcia-Esperon C, Lin L, Bivard A, Ang T, Smoll NR, et al. Validation of the National Institutes of Health Stroke Scale-8 to detect large vessel occlusion in ischemic stroke. *Journal of Stroke and Cerebrovascular Diseases* 2017;**26**(7):1419-26.

Ellison 2004 *{published data only}*

Ellison SR, Gratton MC, Schwab RA, Ma OJ. Prehospital dispatch assessment of stroke. *Missouri Medicine* 2004;**101**(1):64-6.

Ferri 2005 *{published data only}*

Ferri M, Luca A, Rossi PG, Lori G, Guasticchi G. Does a pre-hospital emergency pathway improve early diagnosis and referral in suspect stroke patients? – Study protocol of a cluster randomised trial [ISRCTN41456865]. *BMC Health Services Research* 2005;**66**(5):1-9.

Ferro 1996 *{published data only}*

Ferro JM, Falcao I, Rodrigues G, Canhao P, Melo T P, Oliveira V, et al. Diagnosis of transient ischemic attack by the non-neurologist. A validation study. *Stroke* 1996;**27**(12):2225-9.

Ferro 1998 *{published data only}*

Ferro JM, Pinto AN, Falcao I, Rodrigues G, Ferreira J, Falcao F, et al. Diagnosis of stroke by the non-neurologist. A validation study. *Stroke* 1998;**29**(6):1106-9.

Fischer 2008 *{published data only}*

Fischer CE, Barnung S, Nielsen SL, Rasmussen LS. Prehospital identification of stroke – room for improvement. *European Journal of Neurology* 2008;**15**(8):792-6.

Garnett 2010 *{published data only}*

Garnett AR, Marsden DL, Parsons MW, Quain DA, Spratt NJ, Loudfoot AR, et al. The rural Prehospital Acute Stroke Triage (PAST) trial protocol: a controlled trial for rapid facilitated transport of rural acute stroke patients to a regional stroke centre. *International Journal of Stroke* 2010;**5**(6):506-13.

Garrett 2013 *{published data only}*

Garrett JS, Zarghouni M, Layton KF, Graybeal D, Daoud YA. Validation of clinical prediction scores in patients with primary intracerebral hemorrhage. *Neurocritical Care* 2013;**19**(3):329-35.

Govindarajan 2011 *{published data only}*

Govindarajan P, Ghilarducci D, McCulloch C, Pierog J, Bloom E, Johnston C. Comparative evaluation of stroke triage algorithms for emergency medical dispatchers (MeDS): prospective cohort study protocol. *BMC Neurology* 2011;**11**:14.

Govindarajan 2012 *{published data only}*

Govindarajan P, Desouza NT, Pierog J, Ghilarducci D, Johnston SC. Feasibility study to assess the use of the Cincinnati stroke scale by emergency medical dispatchers: a pilot study. *Emergency Medicine Journal* 2012;**29**(10):848-50.

Gropen 2018 *{published data only}*

Gropen TI, Boehme A, Martin-Schild S, Albright K, Samai A, Pishanidar S, et al. Derivation and validation of the Emergency Medical Stroke Assessment and comparison of large vessel occlusion scales. *Journal of Stroke and Cerebrovascular Diseases* 2018;**27**(3):806-15.

Grossman 2011 *{published data only}*

Grossman AW, Navi BB, Kamel H, Shah MP, Wong CS, Poisson SN, et al. Initial validation of the ABCD2 score for predicting cerebrovascular diagnoses in emergency department (ED) patients with acute dizziness. *Stroke* 2011;**42**(3):e85.

Hand 2006 *{published data only}*

Hand PJ, Kwan J, Lindley RI, Dennis MS, Wardlaw JM. Distinguishing between stroke and mimic at the bedside: the brain attack study. *Stroke* 2006;**37**(3):769-75.

Harbison 2003 *{published data only}*

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**(1):71-6.

Hasegawa 2013 *{published data only}*

Hasegawa Y, Sasaki N, Yamada K, Ono H, Kumai J, Tsumura K, et al. Prediction of thrombolytic therapy after stroke-bypass transportation: the Maria Prehospital Stroke Scale score. *Journal of Stroke and Cerebrovascular Diseases* 2013;**22**(4):514-9.

Henry-Morrow 2017 *{published data only}*

Henry-Morrow TK, Nelson BD, Conahan E, Mathiesen C, Glenn-Porter B, Niehaus MT, et al. An educational intervention allows for greater prehospital recognition of acute stroke. *American Journal of Emergency Medicine* 2017;**35**(12):1959-61.

Herzberg 2014 *{published data only}*

Herzberg M, Boy S, Holscher T, Ertl M, Zimmermann M, Ittner KP, et al. Prehospital stroke diagnostics based on neurological examination and transcranial ultrasound. *Critical Ultrasound Journal* 2014;**6**(1):1-13.

Heuer 2012 *{published data only}*

Heuer JF, Gruschka D, Crozier TA, Bleckmann A, Plock E, Moerer O, et al. Accuracy of prehospital diagnoses by emergency physicians: comparison with discharge diagnosis. *European Journal of Emergency Medicine* 2012;**19**(5):292-6.

Huang 1994 *{published data only}*

Huang JA, Wang PY, Chang MC, Chia LG, Yang DY, Wu TC. Allen score in clinical diagnosis of intracranial hemorrhage. *Zhonghua Yi Xue za Zhi* 1994;**54**(6):407-11.

Hurwitz 2005 *{published data only}*

Hurwitz AS, Brice JH, Overby BA, Evenson KR. Directed use of the Cincinnati Prehospital Stroke Scale by laypersons. *Prehospital Emergency Care* 2005;**9**(3):292-6.

Iguchi 2011 *{published data only}*

Iguchi Y, Kimura K, Watanabe M, Shibazaki K, Aoki J. Utility of the Kurashiki Prehospital Stroke Scale for hyperacute stroke. *Cerebrovascular Diseases* 2011;**31**(1):51-6.

Jang 2014 *{published data only}*

Jang J, Chung SP, Park I, You JS, Lee HS, Park JW, et al. The usefulness of the Kurashiki prehospital stroke scale in identifying thrombolytic candidates in acute ischemic stroke. *Yonsei Medical Journal* 2014;**55**(2):410-6.

Jia 2017 *{published data only}*

Jia J, Band R, Abboud ME, Pajerowski W, Guo M, David G, et al. Accuracy of emergency medical services dispatcher and crew

diagnosis of stroke in clinical practice. *Frontiers in Neurology* 2017;**8**:466.

Josephson 2008 *{published data only}*

Josephson SA, Sidney S, Pham TN, Bernstein AL, Johnston SC. Higher ABCD2 score predicts patients most likely to have true transient ischemic attack. *Stroke* 2008;**39**(11):3096-8.

Kidwell 1998 *{published data only}*

Kidwell CS, Saver JL, Schubert GB, Eckstein M, Starkman S. Design and retrospective analysis of the Los Angeles Prehospital Stroke Screen (LAPSS). *Prehospital Emergency Care* 1998;**2**(4):267-73.

Kimura 2008 *{published data only}*

Kimura K, Inoue T, Iguchi Y, Shibazaki K. Kurashiki prehospital stroke scale. *Cerebrovascular Diseases* 2008;**25**(1-2):189-91.

Kothari 1995 *{published data only}*

Kothari R, Barsan W, Brott T, Broderick J, Ashbrock S. Frequency and accuracy of prehospital diagnosis of acute stroke. *Stroke* 1995;**26**(6):937-41.

Kothari 1997 *{published data only}*

Kothari R, Hall K, Brott T, Broderick J. Early stroke recognition: developing an out-of-hospital NIH Stroke Scale. *Academic Emergency Medicine* 1997;**4**(10):986-90.

Kothari 1999 *{published data only}*

Kothari RU, Pancioli A, Liu T, Brott T, Broderick J. Cincinnati Prehospital Stroke Scale: reproducibility and validity. *Annals of Emergency Medicine* 1999;**33**(4):373-8.

Krebes 2012 *{published data only}*

Krebes Se, Ebinger M, Baumann AM, Kellner PA, Rozanski M, Doepp F, et al. Development and validation of a dispatcher identification algorithm for stroke emergencies. *Stroke* 2012;**43**(3):776-81.

Kwiatkowski 2006 *{published data only}*

Kwiatkowski T. The Recognition of Stroke in the Emergency Room (ROSIER) scale: a useful diagnostic tool?. *Nature Clinical Practice Neurology* 2006;**2**(5):244-5.

Lange 2011 *{published data only}*

Lange MC, Zetola VF, Parolin MF, Zamproni LN, Fernandes AF, Piovesan EJ, et al. Curitiba acute ischemic stroke protocol: a university hospital and EMS initiative in a large Brazilian city. *Arquivos de Neuro-psiquiatria* 2011;**69**(3):441-5.

Lavin 2014 *{published data only}*

Lavin TL, Ranta A. Transient ischaemic attack/stroke electronic decision support: a 14-month safety audit. *Journal of Stroke and Cerebrovascular Diseases* 2014;**23**(2):267-70.

Levine 2016 *{published data only}*

Levine SR, Switzer JA. Acute stroke in the field. *Neurology* 2016;**87**(1):13-4.

Liferidge 2004 *{published data only}*

Liferidge AT, Brice JH, Overby BA, Evenson KR. Ability of laypersons to use the Cincinnati Prehospital Stroke Scale. *Prehospital Emergency Care* 2004;**8**(4):384-7.

Lin 2012 *{published data only}*

Lin CB, Peterson ED, Smith EE, Saver JL, Liang L, Xian Y, et al. Emergency medical service hospital prenotification is associated with improved evaluation and treatment of acute ischemic stroke. *Circulation: Cardiovascular Quality and Outcomes* 2012;**5**(4):514-22.

Llanes 2004 *{published data only}*

Llanes JN, Kidwell CS, Starkman S, Leary MC, Eckstein M, Saver JL. The Los Angeles Motor Scale (LAMS): a new measure to characterize stroke severity in the field. *Prehospital Emergency Care* 2004;**8**(1):46-50.

Malekzadeh 2015 *{published data only}*

Malekzadeh J, Shafaee H, Behnam H, Mirhaghi A. The effect of Cincinnati Prehospital Stroke Scale on telephone triage of stroke patients: evidence-based practice in emergency medical services. International Journal of Evidence-based Healthcare 2015; Vol. 13, issue 2:87-92.

Mao 2016 *{published data only}*

Mao H, Lin P, Mo J, Li Y, Chen X, Rainer TH, et al. Development of a new stroke scale in an emergency setting. *BMC Neurology* 2016;**16**:168. [PUBMED: 27608839]

Middleton 2016 *{published data only}*

Middleton S, Levi C, Dale S, Cheung NW, McInnes E, Considine J, et al. Triage, treatment and transfer of patients with stroke in emergency department trial (the T(3) Trial): a cluster randomised trial protocol. *Implementation Science* 2016;**11**(1):139.

Mohd 2004 *{published data only}*

Mohd NA, McAllister C, Louw SJ, Dyker AG, Davis M, Jenkinson D, et al. Agreement between ambulance paramedic- and physician-recorded neurological signs with Face Arm Speech Test (FAST) in acute stroke patients. *Stroke* 2004;**35**(6):1355-9.

Mosley 2013 *{published data only}*

Mosley I, Morphet J, Innes K, Braitberg G. Triage assessments and the activation of rapid care protocols for acute stroke patients. *Australasian Emergency Nursing Journal* 2013;**16**(1):4-9.

Nam 2014 *{published data only}*

Nam HS, Heo JN, Kim J, Kim YD, Song TJ, Park E, et al. Development of smartphone application that aids stroke screening and identifying nearby acute stroke care hospitals. *Yonsei Medical Journal* 2014;**55**(1):25-9.

Nazliel 2008 *{published data only}*

Nazliel B, Starkman S, Liebeskind DS, Ovbiagele B, Kim D, Sanossian N, et al. A brief prehospital stroke severity scale identifies ischemic stroke patients harboring persisting large arterial occlusions. *Stroke* 2008;**39**(8):2264-7.

Newman-Toker 2013 *{published data only}*

Newman-Toker DE, Kerber KA, Hsieh Y, Pula JH, Omron R, Saber Tehrani AS, et al. HINTS outperforms ABCD2 to screen for stroke in acute continuous vertigo and dizziness. *Academic Emergency Medicine* 2013;**20**(10):986-96.

Noorian 2016 *{published data only}*

Noorian Al, Sanossian N, Liebeskind DS. Field validation of Prehospital LAMSScore to identify large vessel occlusion ischemic stroke patients for direct routing to Emergency Neuroendovascular Centers. Stroke 2016; Vol. 47 Suppl 1.

Nor 2004 *{published data only}*

Nor AM, McAllister C, Louw SJ, Dyker AG, Davis M, Jenkinson D, et al. Agreement between ambulance paramedic- and physician-recorded neurological signs with Face Arm Speech Test (FAST) in acute stroke patients. *Stroke* 2004;**35**(6):1355-9.

O'Brien 2012 *{published data only}*

O'Brien W, Crimmins D, Donaldson W, Risti R, Clarke TA, Whyte S, et al. FASTER (Face, Arm, Speech, Time, Emergency Response): experience of Central Coast Stroke Services implementation of a pre-hospital notification system for expedient management of acute stroke. *Journal of Clinical Neuroscience* 2012;**19**(2):241-5.

Ollikainen 2018 *{published data only}*

Ollikainen JP, Janhunen HV, Tynkkynen JA, Mattila KM, Halinen MM, Oksala NK, et al. The Finnish Prehospital Stroke Scale Detects Thrombectomy and Thrombolysis Candidates – a propensity score-matched study. *Journal of Stroke and Cerebrovascular Diseases* 2018;**27**(3):771-7.

Oostema 2015 *{published data only}*

Oostema JA, Nasiri M, Chassee T, Reeves MJ. Use of a prehospital stroke scale increases accuracy of emergency medical services stroke recognition. Stroke 2015; Vol. 46:Abstract W P220.

Oostema 2016 *{published data only}*

Oostema JA, Carle T, Talia N, Reeves M. Dispatcher stroke recognition using a stroke screening tool: a systematic review. *Cerebrovascular Diseases* 2016;**42**(5-6):370-7.

Perez de la Ossa 2014 *{published data only}*

Perez de la Ossa N, Carrera D, Gorchs M, Querol M, Millan M, Gomis M, et al. Design and validation of a prehospital stroke scale to predict large arterial occlusion: the rapid arterial occlusion evaluation scale. *Stroke* 2014;**45**(1):87-91.

Purrucker 2015 *{published data only}*

Purrucker JC, Hametner C, Engelbrecht A, Bruckner T, Popp E, Poli S. Comparison of stroke recognition and stroke severity scores for stroke detection in a single cohort. *Journal of Neurology, Neurosurgery, and Psychiatry* 2015;**86**(9):1021-8. [PUBMED: 25466259]

Purrucker 2017 *{published data only}*

Purrucker JC, Hartig F, Richter H, Engelbrecht A, Hartmann J, Auer J, et al. Design and validation of a clinical scale for prehospital stroke recognition, severity grading and prediction of large vessel occlusion: the shortened NIH Stroke Scale for emergency medical services. *BMJ Open* 2017;**7**(9):e016893.

Quenardelle 2016 *{published data only}*

Quenardelle V, Lauer-Ober V, Zinchenko I, Bataillard M, Rouyer O, Beaujeux R, et al. Stroke mimics in a stroke care pathway based on MRI Screening. *Cerebrovascular Diseases* 2016;**42**(3-4):205-12.

Qureshi 2016 *{published data only}*

Qureshi S, Janjua AU, Raza SA, Brasher C, Anderson A, Belagaje S, et al. Stroke screening tools have high specificity for detecting large vessel occlusion in a Southeastern US prospective cohort study. *Stroke* 2016;**47 Suppl 1**:Abstract WMP67.

Richards 2016 *{published data only}*

Richards CT, Markul E, Stein-Spencer L, Prabhakaran S. A dichotomized Cincinnati Prehospital Stroke Score is correlated with revascularization therapy in acute stroke. Stroke 2016; Vol. 47:Abstract WMP65.

Robinson 2013 *{published data only}*

Robinson TG, Reid A, Haunton VJ, Wilson A, Naylor AR. The face arm speech test: does it encourage rapid recognition of important stroke warning symptoms?. *Emergency Medicine Journal* 2013;**30**(6):467-71.

Rodriguez-Pardo 2017 *{published data only}*

Rodriguez-Pardo J, Fuentes B, Alonso de Lecinana M, Ximenez-Carrillo A, Zapata-Wainberg G, Alvarez-Fraga J, et al. The Direct Referral to Endovascular Center criteria: a proposal for pre-hospital evaluation of acute stroke in the Madrid Stroke Network. *European Journal of Neurology* 2017;**24**(3):509-15.

Ross 2007 *{published data only}*

Ross MA, Compton S, Medado P, Fitzgerald M, Kilanowski P, O'Neil BJ. An emergency department diagnostic protocol for patients with transient ischemic attack: a randomized controlled trial. *Annals of Emergency Medicine* 2007;**50**(2):109-19.

Rudd 2015 *{published data only}*

Rudd MP, Price CI, Ford GA. Prehospital stroke scales in urban environments: a systematic review. *Neurology* 2015;**84**(9):962.

Rudd 2016 *{published data only}*

Rudd M, Buck D, Ford GA, Price CI. A systematic review of stroke recognition instruments in hospital and prehospital settings. *Emergency Medicine Journal* 2016;**33**(11):818-22.

Schilling 2012 *{published data only}*

Schilling M, Kros M, Ritter M, Ohms M, Schabitz WR, Kusch W, et al. Concept for allocation of acute stroke patients. Evaluation of the quality of diagnosis reached by the emergency medical services of Munster. *Der Nervenarzt* 2012;**83**(6):759-65.

Schrock 2009 *{published data only}*

Schrock JW, Victor A, Losey T. Can the ABCD2 risk score predict positive diagnostic testing for emergency department

patients admitted for transient ischemic attack?. *Stroke* 2009;**40**(10):3202-5.

Sequeira 2015 *{published data only}*

Sequeira D, Martin-Gill C, Guyette FX, Jadhav AP. Comparison of prehospital stroke scales. Circulation 2015; Vol. 132:Abstract 12493.

Shapiro 2003 *{published data only}*

Shapiro J, Bessette M, Levine SR, Baumlin K. HandiStroke: a handheld tool for the emergent evaluation of acute stroke patients. *Academic Emergency Medicine* 2003;**10**(12):1325-8.

Sheppard 2015 *{published data only}*

Sheppard JP, Mellor RM, Greenfield S, Mant J, Quinn T, Sandler D, et al. The association between prehospital care and in-hospital treatment decisions in acute stroke: a cohort study. *Emergency Medicine Journal* 2015;**32**(2):93-9.

Silva 2015 *{published data only}*

Silva GS, Miranda RC, Massaud RM, Vacari AM, Neto MC. False negative results on acute stroke recognition tool has no impact on stroke prognosis. Stroke 2015; Vol. 46:Abstract WMP82.

Smith 1998 *{published data only}*

Smith WS, Isaacs M, Corry MD. Accuracy of paramedic identification of stroke and transient ischemic attack in the field. *Prehospital Emergency Care* 1998;**2**(3):170-5.

Smith 1999 *{published data only}*

Smith WS, Corry MD, Fazackerley J, Isaacs SM. Improved paramedic sensitivity in identifying stroke victims in the prehospital setting. *Prehospital Emergency Care* 1999;**3**(3):207-10.

Soda 2016 *{published data only}*

Soda H, Shammas L, Griewing B, Rashid A. 4-Item Stroke Scale (4ISS) – prehospital notification scale for stroke patients. *Cerebrovascular Diseases* 2016;**41**:195.

Timerding 1989 *{published data only}*

Timerding BL, Barsan WG, Hedges JR, Brott TG, VanLigten PF, Spilker JA, et al. Stroke patient evaluation in the emergency department before pharmacologic therapy. *American Journal of Emergency Medicine* 1989;**7**(1):11-5.

Tirschwell 2002 *{published data only}*

Tirschwell DL, Longstreth WT, Becker KJ, Gammans RE Sr, Sabounjian LA, Hamilton S, et al. Shortening the NIH Stroke scale for use in the prehospital setting. *Stroke* 2002;**33**(12):2801-6.

Tonomura 2015 *{published data only}*

Tonomura S. Clinical characteristics of pseudo negative cases in prehospital triage for stroke. Stroke 2015; Vol. 46:Abstract W P211.

Trivedi 2015 *{published data only}*

Trivedi T, Heidari K, Merchant A, Jauch E, Venkatesh S, Sen S. Accuracy and clinical implications of Cincinnati pre-hospital stroke scale and Los Angeles pre-hospital stroke scale use by emergency medical services. Stroke 2015; Vol. 46.

Turc 2016 *{published data only}*

Turc G, Maïer B, Naggara O, Seners P, Isabel C, Tisserand M, et al. Clinical scales do not reliably identify acute ischemic stroke patients with large-artery occlusion. *Stroke* 2016;**47**(6):1466.

Van Hooff 2013 *{published data only}*

Van Hooff RJ, Cambron M, Dyck R, Smedt A, Moens M, Espinoza AV, et al. Prehospital unassisted assessment of stroke severity using telemedicine: a feasibility study. *Stroke* 2013;**44**(10):2907-9.

Verma 2010 *{published data only}*

Verma A, Gladstone DJ, Fang J, Chenkin J, Black SE, Verbeek PR. Effect of online medical control on prehospital Code Stroke triage. *Canadian Journal of Emergency Medicine* 2010;**12**(2):103-10.

Wennman 2012 *{published data only}*

Wennman I, Klittermark P, Herlitz J, Lernfelt B, Kihlgren M, Gustafsson C, et al. The clinical consequences of a pre-hospital diagnosis of stroke by the emergency medical service system. A pilot study. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 2012;**20**:48.

Wesley 2016 *{published data only}*

Wesley K, Wesley K. Stroke recognition. Lower the threshold for stroke alert activation. *Journal of Emergency Medical Services* 2016;**41**(2):22.

Williams 2015 *{published data only}*

Williams TA, Blacker D, Arendts G, Patrick E, Finn J. Accuracy of stroke identification by paramedics in a metropolitan prehospital setting. *International Journal of Stroke* 2015;**10**:83.

Williams 2017 *{published data only}*

Williams TA, Blacker D, Arendts G, Patrick E, Brink D, Finn J. Accuracy of stroke identification by paramedics in a metropolitan pre-hospital setting: a cohort study. *Australasian Journal of Paramedicine* 2017;**14**(2):ajp.paramedics.org/index.php/ajp/article/view/521.

Wojner-Alexandrov 2005 *{published data only}*

Wojner-Alexandrov AW, Alexandrov AV, Rodriguez D, Persse D, Grotta JC. Houston paramedic and emergency stroke treatment and outcomes study (HoPSTO). *Stroke* 2005;**36**(7):1512-8. [PUBMED: 15961712]

Yamashita 2011 *{published data only}*

Yamashita S, Kimura K, Iguchi Y, Shibazaki K, Watanabe M, Iwanaga T. Kurashiki Prehospital Stroke Subtyping Score (KP3S) as a means of distinguishing ischemic from hemorrhagic stroke in emergency medical services. *European Neurology* 2011;**65**(4):233-8.

Yilmaz 2014 *{published data only}*

Yilmaz BK, Cevik E, Celik Y, Dogan H, Uzun O, Sayin T, et al. A new scoring system predicting acute ischemic stroke: NIHSS +ABCD2. *Acta Medica Mediterranea* 2014;**30**(3):621-6.

Yock-Corrales 2011 *{published data only}*

Yock-Corrales A, Babl FE, Mosley IT, Mackay MT. Can the FAST and ROSIER adult stroke recognition tools be applied to confirmed childhood arterial ischemic stroke?. *BMC Pediatrics* 2011;**11**:93.

You 2013 *{published data only}*

You JS, Chung SP, Chung HS, Park JW, Kim HJ, Lee SH, et al. Predictive value of the Cincinnati Prehospital Stroke Scale for identifying thrombolytic candidates in acute ischemic stroke. *American Journal of Emergency Medicine* 2013;**31**(12):1699-702.

Zamora 2013 *{published data only}*

Zamora SG, Elias R, Schipper NP, Palameta F, Parodi R, Gallo R, et al. Awareness of initial stoke evaluation scales among emergency room physicians in three Argentinian Provinces. *Revista Medica de Rosario* 2013;**79**(2):62-72.

Ziegler 2008 *{published data only}*

Ziegler V, Rashid A, Muller-Gorchs M, Kippnich U, Hiermann E, Kogerl C, et al. Use of mobile computing systems in preclinical stroke care [Einsatz mobiler Computing-Systeme in der praklinischen Schlaganfallversorgung]. *Der Anaesthesist* 2008;**57**(7):677-85.

Zohrevandi 2015 *{published data only}*

Zohrevandi B, Monsef KV, Asadi P, Tajik H, Azizzade RN. Diagnostic accuracy of Cincinnati Pre-Hospital Stroke Scale. *Emergency* 2015;**3**(3):95-8.

Additional references

Amarenco 2008

Amarenco P, Benavente O. EXPRESS transient ischemic attack study: speed the process. *Stroke* 2008;**39**(8):2400-1.

Anderson 2008

Anderson CS, Huang Y, Wang JG, Arima H, Neal B, Peng B, INTERACT Investigators. Intensive blood pressure reduction in acute cerebral haemorrhage trial (INTERACT): a randomised pilot trial. *Lancet Neurology* 2008;**7**(5):391-9.

Brandler 2014

Brandler ES, Sharma M, Sinert RH, Levine SR. Prehospital stroke scales in urban environments: a systematic review. *Neurology* 2014;**82**(24):2241-9. [PUBMED: 24850487]

Butcher 2013

Butcher KS, Jeerakathil T, Hill M, Demchuk AM, Dowlatshahi D, Coutts SB, ICH ADAPT Investigators. The Intracerebral Hemorrhage Acutely Decreasing Arterial Pressure Trial. Stroke 2013; Vol. 44, issue 3:620-6.

Daroff 2012

Daroff RB, Fenichel GM, Jankovic J, Mazziotta JC. Bradley's Neurology in Clinical Practice. 6th Edition. Vol. **2**, Philadelphia (PA): Elsevier, 2012.

ESO 2018

European Stroke Organisation. ESO Guidelines Session 2018. eso-stroke.org/esoc/esoc-2018-eso-guidelines-session-2018/ (accessed 24 August 2018).

Flaherty 2014

Flaherty M. STOP-IT Study. International Stroke Conference; 2014 Feb 12-14; San Diego (CA). 2014.

Gurusamy 2015

Gurusamy KS, Giljaca V, Takwoingi Y, Higgie D, Poropat G, Štimac D, et al. Endoscopic retrograde cholangiopancreatography versus intraoperative cholangiography for diagnosis of common bile duct stones. *Cochrane Database of Systematic Reviews* 2015, Issue 2. [DOI: 10.1002/14651858.CD010339.pub2]

Hacke 2008

Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Guidetti D, ECASS Investigators. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *New England Journal of Medicine* 2008;**359**(13):1317-29.

Harbison 1999

Harbison J, Massey A, Barnett L, Hodge D, Ford GA. Rapid ambulance protocol for acute stroke. Lancet 1999; Vol. 353, issue 9168:1935. [PUBMED: 10371574]

Hill 2013

Hill MD, Muir KW. Should blood pressure be aggressively lowered acutely after intracerebral hemorrhage?. *Stroke* 2013;**44**(10):2951-2.

Kleindorfer 2007

Kleindorfer DO, Miller R, Moomaw CJ, Alwell K, Broderick JP, Khoury J, et al. Designing a message for public education regarding stroke: does FAST capture enough stroke?. *Stroke* 2007;**38**(10):2864-8. [PUBMED: 17761926]

Kobayashi 2018

Kobayashi A, Czlonkowska A, Ford GA, Fonseca AC, Luijckx GJ, Korv J, et al. European Academy of Neurology and European Stroke Organization consensus statement and practical guidance for pre-hospital management of stroke. *European Journal of Neurology* 2018;**25**(3):425-33. [PUBMED: 29218822]

Leeflang 2009

Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *Journal of Clinical Epidemiology* 2009;**62**(1):5-12. [PUBMED: 18778913]

Lyden 1994

Lyden PD, Rapp K, Babcock T, Rothrock R. Ultra-rapid identification and enrollment of stroke patients into clinical trials. *Journal of Stroke and Cerebrovascular Disease* 1994;**4**(2):106-13.

Macaskill 2010

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and presenting results. In: Deeks JJ,

Bossuyt PM, Gatsonis C, editor(s). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: srdta.cochrane.org.

Marler 2000

Marler JR, Tilley BC, Lu M, Brott TG, Lyden PC, Grotta JC, et al. Early stroke treatment associated with better outcome: the NINDS rt-PA Stroke Study. *Neurology* 2000;**55**(11):1649-55.

NINDS 1995

National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *New England Journal of Medicine* 1995;**333**(24):1581-7.

O'Connor 1999

O'Connor RE, McGraw P, Edelsohn L. Thrombolytic therapy for acute ischemic stroke: why the majority of patients remain ineligible for treatment. *Annals of Emergency Medicine* 1999;**33**(1):9-14.

Powers 2018

Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. 2018 Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2018;**49**:e46-e99.

Reitsma 2005

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005;**58**(10):982-90.

Review Manager 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Sandercock 2012

Sandercock P, Wardlaw JM, Lindley RI, Dennis M, Cohen G, Murray G, IST-3 Collaborative Group. The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 hours of acute ischemic stroke (the third international stroke trial [IST-3]): a randomised controlled trial. *Lancet* 2012;**379**(9834):2352-63.

Saver 2013

Saver JL, Fonarow GC, Smith EE, Reeves MJ, Grau-Sepulveda MV, Pan W, et al. Time to treatment with intravenous tissue plasminogen activator and outcome from acute ischemic stroke. *JAMA* 2013;**23**(309):2480-8.

StataCorp 2011 [Computer program]

StataCorp LP. Stata Statistical Software: Release 12. College Station (TX): StataCorp LP, 2011.

Wardlaw 2009

Wardlaw JM, Murray V, Berge E, del Zoppo GJ. Thrombolysis for acute ischaemic stroke. *Cochrane Database of Systematic Reviews* 2009, Issue 4. [DOI: 10.1002/14651858.CD000213.pub2]

Warlow 2001

Warlow CP, Dennis MS, van Gijn J, Hankey GJ, Sandercock PAG, Bamford JM, et al. Stroke: a Practical Guide to Management. 2nd Edition. Oxford: Blackwell, 2001.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.

References to other published versions of this review

Walker 2014

Walker G, Yip S, Zhelev Z, Henschke N. Prehospital stroke scales as screening tools for early identification of stroke and transient ischemic attack. *Cochrane Database of Systematic Reviews* 2014, Issue 12. [DOI: 10.1002/14651858.CD011427]

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Andsberg 2017

Study characteristics	
Patient sampling	Pilot study conducted in an ambulance district staffed by 43 ambulance nurses for 24-hour service for the community hospital of Hässleholm, Sweden. **Sampling:** during the study period (9 January to 23 May 2014) neurologic assessment with PreHAST was done if stroke was suspected in conscious people > 18 years of age.
Patient characteristics and setting	**Inclusion and exclusion criteria:** neurologic assessment with PreHAST was done if stroke was suspected, defined as sudden onset of focal neurologic symptoms/signs, in conscious people > 18 years of age.

Andsberg 2017 *(Continued)*

	Participant characteristics: not reported
Index tests	**Index test:** PreHAST (score > 0 considered positive) **Test administrator:** ambulance nurses **Training:** before the study period all ambulance nurses received a 4-hour education program, covering basic stroke knowledge and assessment and grading of stroke symptoms according to PreHAST. The education program included practical PreHAST training in pairs, where each ambulance nurse performed the PreHAST items under supervision and proper execution. During the study an instruction video for PreHAST was available on YouTube.
Target condition and reference standard(s)	**Target condition:** stroke/TIA (people with TIA included in the study were required to have ongoing symptoms when evaluated by the ambulance staff; used time-based definition of TIA). **Reference standard:** 2 stroke physicians, blinded to the PreHAST scores, independently reviewed the medical records of the participants, including evaluation of history, and clinical and radiologic findings. In case of disagreement, a third evaluator adjudicated the final diagnosis.
Flow and timing	78 participants were assessed with PreHAST; 9 were excluded (5 informed consent not obtained; 2 PreHAST was not possible to perform due to agitation and ongoing epileptic seizure; 2 did not have symptoms at ambulance arrival).
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 26/69 (37.7%) participants had stroke/TIA PreHAST: TP = 26, FP = 26, FN = 0, TN = 17
Notes	**Non-stroke diagnoses:** 8 epilepsy, 7 late effect after stroke, 5 migraine, 3 Bell's palsy, 3 fatigue, 1 SDH, 1 dementia, 5 vertigo, 3 syncope, 2 infection, 2 delirium, 1 transitory global amnesia, 1 opsoclonus syndrome **Ease of use and time:** in the poststudy survey, the ambulance staff reported PreHAST easy to execute and estimated the test time to be 2–3 minutes. **Funding:** financed by departmental funds

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		Low	**Low**
DOMAIN 2: Index Test Index tests			

Andsberg 2017 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
		Low	

Berglund 2014

Study characteristics	
Patient sampling	Sweden, Stockholm county, for 6 months in 2008 **Sampling:** people calling the Swedish equivalent of 911 in Sweden, Stockholm county were sampled. (EMCC procedure prior to ambulance assessment: when the nurse at the EMCC suspected stroke with onset within 6 hours the inclusion criteria were checked. If the criteria were met, the nurse opened a sealed envelope of the priority level. The FAST test was then an optional tool for the nurse at the EMCC to use for identifying symptoms of stroke. The nurse was free to include the participant on her/his own suspicion of stroke. As the study aimed to analyze priority code and delay, the FAST test was not mandatory for the EMCC as it might have caused a delay). Almost one-third of the participants in the study were identified and included from the ambulance, when missed from the EMCC. The ambulance contacted the EMCC to have a code for randomization

Berglund 2014 *(Continued)*

	and the priority level. All participants included in the study were tested for FAST by the ambulance personnel (hence, the sampling considered consecutive; this was confirmed by authors).
Patient characteristics and setting	**Inclusion and exclusion criteria:** suspected stroke with symptom onset within 6 hours; ages 18–85 years; previous independence in activities of daily living; and no other acute condition requiring a priority level 1 (see HASTA study). **Participant characteristics:** ages 22–93 years; 55.5% men
Index tests	**Index test:** FAST (prespecified threshold, positive if ≥ 1) **Test administrator:** registered specialist nurse and a registered general nurse or an ambulance educated staff, a paramedic. **Training:** all personnel at the ambulance companies supplying the Stockholm area with ambulance transport participated in 1 lecture about stroke and the FAST test, prior to start of the study. The personnel at the EMCC were also given education of stroke and FAST, adjusted for testing FAST by telephone, prior to study start. All emergency calls concerning medical issues were connected to and evaluated by a registered nurse.
Target condition and reference standard(s)	**Target condition:** stroke or TIA **Reference standard:** CT brain scan and in some cases CTA or MRI, neurologic exam, if necessary, EEG (differential diagnosis), laboratory tests. People with obvious signs of other conditions than stroke, such as diabetes or alcohol intoxication might not have been evaluated by a CT scan. All participants received a final diagnosis by a neurologist or stroke specialist.
Flow and timing	Withdrawals: 39 (4.3%) left at home or deviated from the ER before seeing a doctor, considered as non-stroke diagnoses.
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 472/900 (52%) **FAST:** TP = 387; FP = 255; FN = 85; TN = 173 Combined data from EMCC (nurse) and ambulance (paramedic)
Notes	**Categorization of alternate diagnoses:** 472 (52%) stroke/TIA, 166 (18%) neurology, 223 (25%) non-neurology; 39 (4%) unknown **Funding:** declared, provided by the authors as supplementary data.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		

Berglund 2014 *(Continued)*

		Low	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
		Low	

Bergs 2010

Study characteristics	
Patient sampling	EMS of University Hospitals Leuven, Belgium, 11 December 2005 to 30 April 2006 **Sampling:** prospective but not consecutive (only 31/124 eligible patients were included in the study). Inclusion criteria kept broad to avoid missing cases; nurses asked to complete a questionnaire (combining all evaluated scales) for every patient transported with relevant neurologic complaints.
Patient characteristics and setting	**Inclusion and exclusion criteria:**

Bergs 2010 *(Continued)*

	Inclusion: all patients transported with relevant neurologic complaints: an acute neurologic event without clear origin, altered LOC, convulsions, syncope, headache, symptoms of weakness, dizziness or decreased well-being, aphasia, visual impairment, weakness in arms and or legs and facial paralysis. Exclusion: ages < 18 years, GCS < 9, transported to alternate hospital, trauma **Participant characteristics:** mean age 77 years; 61% men; seizure history not reported
Index tests	**Index tests:** FAST (prespecified threshold, positive if ≥ 1), CPSS (prespecified threshold ≥ 1), LAPSS, MASS **Test administrator:** emergency nurses **Training:** all nurses were briefed on purpose of study, stroke scales and guidelines.
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** unspecified, diagnosis at ER discharge
Flow and timing	Of 132 eligible participants, 70 were transported to the study hospital, but the ambulance nurses completed the questionnaire only in 31 (39 participants did not have a completed questionnaire by EMS despite meeting study criteria).
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 19/31 (61.3%) **LAPSS:** TP = 14; FP = 2; FN = 5; TN = 10 **MASS:** TP = 14; FP = 4; FN = 5; TN = 8 **FAST:** TP = 18; FP = 8; FN = 1; TN = 4 **CPSS:** TP = 18; FP = 8; FN = 1; TN = 4
Notes	**Categorization of alternate diagnosis for people who did not have an ischemic stroke:** not reported **Funding:** not reported

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Prospective design	Yes		
		High	**Low**
DOMAIN 2: Index Test Index tests			

Bergs 2010 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Unclear		
		Unclear	

Bray 2005a

Study characteristics	
Patient sampling	Box Hill Hospital, Melbourne, Australia, September 2002 to September 2003 **Sampling:** prospective but not consecutive sampling: 3327/5957 consecutive calls attended by the study paramedics were transported to the regional university teaching hospital (a single research site); 19 patients were excluded due to incomplete diagnosis prior to discharge; only 100/127 eligible transports had completed MASS forms; of the 27 cases with incomplete forms 10 (37%) were strokes and 17 (63%) were 'stroke mimics'.
Patient characteristics and setting	**Inclusion and exclusion criteria:** possible stroke based on triage (advanced MPDS) or focal neurologic deficit on initial assessment. Specific inclusion and exclusion criteria not reported. **Participant characteristics:** not reported
Index tests	**Index test:** LAPSS, CPSS, MASS

Bray 2005a *(Continued)*

	Test administrator: paramedics **Training:** 1-hour educational session on pathogenesis and management of acute stroke, and instruction in assessment and documentation of items used in a prehospital stroke scale.
Target condition and reference standard(s)	**Target condition:** stroke and TIA **Reference standard:** standard criteria for diagnosis of stroke or TIA (Warlow 2001); review of discharge diagnosis
Flow and timing	27 incomplete/no documentation
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 73/100 (73%) **MASS:** TP = 66; FP = 7; FN = 7; TN = 20 **LAPSS:** TP = 57; FP = 4; FN = 16; TN = 23 **CPSS:** TP = 69; FP = 12; FN = 4; TN = 15 Data recalculated from sensitivity, specificity and likelihood ratio in Table 3 in the paper.
Notes	**Categorization of alternate diagnosis for people who did not have an ischemic stroke:** 27 total: 7 cardiac, 5 seizure, 3 hypoglycemia, 3 SDH, 3 fracture, 2 tumor, 1 sepsis, 1 migraine, 1 vertigo, 1 Parkinson disease **Exclusions:** 19 people excluded due to incomplete diagnosis; 27 eligible cases were excluded due to incomplete MASS sheets: 10 (37%) were strokes and 17 (63%) were 'stroke mimics'. **Funding:** not reported

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		High	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Bray 2005a *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Unclear		
		Unclear	

Bray 2010

Study characteristics	
Patient sampling	Box Hill Hospital, Melbourne, Australia, January to May 2008 **Sampling:** retrospective, consecutive; 2 groups of people admitted to the study hospital were used in this study: 1. people transported by EMS with documented MASS assessments of hand grip, speech and facial weakness; and 2. people with a discharge diagnosis of stroke or TIA included in the stroke/TIA registry.
Patient characteristics and setting	**Inclusion and exclusion criteria:** conscious but neurologically compromised patients of no obvious cause such as drug overdose or trauma. If these assessments were positive for stroke, they obtained the remaining MASS history items and performed a blood sugar level to rule out stroke mimics and suitability for thrombolysis. People who were unconscious or asymptomatic at the time of paramedic assessment were excluded. **Participant characteristics:** not reported
Index tests	**Index test:** MASS, CPSS **Test administrator:** paramedics

Bray 2010 *(Continued)*

	Training: 1-hour stroke education program and instruction on use of the MASS
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** data were cross-referenced against the hospital stroke/TIA registry (name, date, gender and age) to determine if the discharge diagnosis was stroke or TIA. For 8 people with stroke and TIA with no MASS documentation, MASS and CPSS were retrospectively applied based on the paramedic assessment.
Flow and timing	154 participants did not have completed questionnaire; in 8 cases, based on final discharge diagnosis MASS was calculated retrospectively based on EMS report (not included in our analysis).
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 199/850 (23.4%) **MASS:** TP = 166; FP = 92; FN = 33; TN = 559 **CPSS:** TP = 176; FP = 138; FN = 23; TN = 513 Data recalculated from sensitivity, specificity and likelihood ratio in Table 3.
Notes	**Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** not reported **Funding:** not reported

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	No		
		High	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it prespecified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			

Bray 2010 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Unclear		
		Unclear	

Chen 2013

Study characteristics	
Patient sampling	Beijing Tiantan Hospital, China **Sampling:** prospective but non-consecutive as 400/1550 "target stroke" runs did not complete LAPSS.
Patient characteristics and setting	**Inclusion and exclusion criteria:** ages > 18 years, absence of trauma, absence of coma, neurologically relevant complaint (altered LOC, local neurologic signs, seizure syncope, head pain, weak + dizzy + sick) **Participant characteristics:** age range 20–101 years, median 72 years; 60.5% men; seizure history included
Index tests	**Index test:** LAPSS **Test administrator:** paramedics **Training:** 3 hours' LAPSS-based stroke training session with 3 experts from study team.
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** 2 blinded neurologists reviewed the ER charts, recorded final ER discharge diagnoses, and verified absence or presence of potential stroke symptoms. The medical documents and neuroimaging records were reviewed before the final diagnoses were verified.

Chen 2013 *(Continued)*

Flow and timing	420 participants did not have questionnaire completed.		
Comparative			
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 997/1130 (88.2%) **LAPSS:** TP = 782; FP = 13; FN = 215; TN = 120		
Notes	**Categorization of alternate diagnosis for participants who did not have stroke:** not reported **Funding:** Ministry of Science and Technology and the Ministry of Health of the People's Republic of China		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		High	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Chen 2013 *(Continued)*

Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Did all patients receive a reference standard?	Yes
	Unclear

Chenkin 2009

Study characteristics	
Patient sampling	Sunnybrook Health Sciences Centre, Toronto, Canada; 1 March 2005 to 28 February 2006 **Sampling:** retrospective analysis of consecutive participants. Consecutive participants transported to the stroke center by ambulance under the acute stroke protocol over a 1-year period. Participants were identified by reviewing the database of the Registry of the Canadian Stroke Network. Paramedics applied this tool in the field to any person with symptoms suggesting an acute neurologic problem.
Patient characteristics and setting	**Inclusion and exclusion criteria:** consecutive patients with symptoms suggesting an acute neurologic problem transported to the stroke center by ambulance under the acute stroke protocol over 1-year period **Participant characteristics:** mean age 73.7 (SD 13.5) years; 69.1% men; positive screen only
Index tests	**Index test:** OPSST **Test administrator:** paramedics **Training:** 90-minute training session on stroke screening tool prior to implementation
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** final inhospital diagnosis of acute stroke defined as either ischemic stroke, ICH or TIA according to the consulting neurologist. No other details provided.
Flow and timing	Interval between index test and final diagnosis not reported.
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 214/554 (39%) **OPSST:** TP = 187; FP = 138; FN = 27; TN = 202
Notes	**Categorization of alternate diagnosis for participants who did not have stroke:** not reported **Funding:** not reported
Methodological quality	

Chenkin 2009 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	No		
		High	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Unclear		
		Unclear	

Ding 2009

Study characteristics	
Patient sampling	3 local hospitals, Yantian District, Shenzhen, China; June 2007 to June 2008

Ding 2009 *(Continued)*

	Sampling: prospective consecutive sample; 2035 people with non-traumatic, non-comatose, non-obstetrics presentation and 327 had acute neurologic symptoms and were included in the study.
Patient characteristics and setting	**Inclusion and exclusion criteria:** non-traumatic, non-comatose, non-obstetrics people with acute neurologic symptoms (no further details) **Participant characteristics:** mean age 58 years; 48% women
Index tests	**Index test:** LAPSS **Test administrator:** emergency physicians applied LAPSS in a prehospital setting. **Training:** not reported
Target condition and reference standard(s)	**Target condition:** stroke or TIA **Reference standard:** final diagnosis made by a specialist group including 1 neurologist, 1 radiologist and 1 hospital doctor.
Flow and timing	Unclear if diagnosis was done within 14 days of presentation.
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** ischemic stroke = 44.1%, hemorrhagic stroke = 35.6%, TIA = 20.3% **LAPSS:** TP = 47, FP = 12, FN = 4, TN = 264
Notes	**Categorization of alternate diagnosis for participants who did not have stroke:** not reported **Funding:** not reported

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Prospective design	Yes		
		Low	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Ding 2009 *(Continued)*

		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
		Unclear	

English 2018

Study characteristics	
Patient sampling	Mayo Clinic Hospital, St. Mary's Campus Emergency Department, Rochester, MI, USA **Sampling:** people identified with potential stroke by emergency dispatchers; 1 January 2014 to 31 December 2015
Patient characteristics and setting	**Inclusion criteria:** any 1 of: positive CPSS in field; EMS impression of cerebrovascular accident or TIA; acute stroke pager activation in the ER; or discharge diagnosis of cerebrovascular accident or TIA **Exclusion criteria:** any 1 of: hospital arrival via helicopter; outside hospital transfer; direct admission without ER evaluation or last known well time > 6 hours **Participant characteristics:** mean age: stroke 76.6 (SD 13.5), no stroke 72.1 (SD 14.6); 50% men
Index tests	**Index test:** CPSS **Test administrator:** paramedics **Training:** 1-hour online module annually on stroke recognition and assessment in the field as part of their required job training.
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** final diagnosis at discharge documented following a review of hospital admission note and discharge summary.

English 2018 *(Continued)*

Flow and timing	377 people identified by EMS dispatchers with possible stroke were transported to the clinic; of these, 185 met inclusion criteria; of the 185 people, 55 were excluded according to the prespecified exclusion criteria and 5 did not have CPSS documentation; the latter 5 people were included in the analysis but it was unclear how the CPSS scores for these people were determined.
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke:** 96/130 (73.8%); 64.5% were ischemic strokes, 20.8% were TIA and 14.6% were ICH **CPSS:** TP = 72, FP = 27, FN = 24, TN = 7
Notes	**Discharge diagnosis of participants falsely identified as stroke by EMS in the field, number:** 9 (26.5%) seizure, 7 (20.6%) infection, 6 (17.6%) encephalopathy, 3 (8.8%) syncope, 2 (5.9%) migraine, 2 (5.9%) peripheral nerve injury, 2 (5.9%) electrolyte disturbance, 3 (8.8%) other **Funding:** none declared

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	No		
		High	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it prespecified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

English 2018 *(Continued)*

		Unclear	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
		Unclear	

Fothergill 2013

Study characteristics	
Patient sampling	Royal London Hospital, UK **Sampling:** prospective but not consecutive as only people assessed with ROSIER and conveyed to the Royal London Hospital were included (32 people did not receive ROSIER and were not included in the study).
Patient characteristics and setting	**Inclusion criteria:** ages > 18 years presenting with symptoms of stroke **Exclusion criteria:** ages < 18 years, not assessed using the ROSIER or transferred to another hospital **Participant characteristics:** mean age 65 years, range 20–95 years; 53% men; people with seizure history included
Index tests	**Index test:** FAST, ROSIER **Test administrator:** paramedics **Training:** 1-hour stroke educational program, scenario-based demonstration of ROSIER, 15-minute educational DVD
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** final diagnosis made by a stroke consultant or other senior medical physician caring for the person within 72 hours of the person's admission to hospital. Routine tests used, including CT and MRI scans, undertaken by the clinical team to confirm whether the person had a stroke. Final diagnosis was confirmed by a senior stroke consultant.
Flow and timing	32 people not assessed by index tests; 17 without reference standard results.
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 177/295 (60%)

Fothergill 2013 *(Continued)*

	ROSIER: TP = 171; FP = 97; FN = 6; TN = 21 **FAST:** TP = 171; FP = 103; FN = 6; TN = 15
Notes	**Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** not reported **Funding:** not reported

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Prospective design	Yes		
		High	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		

Fothergill 2013 *(Continued)*

	Low

Frendl 2009

Study characteristics	
Patient sampling	Duke University Medical Center, USA **Sampling:** retrospective chart review
Patient characteristics and setting	**Inclusion and exclusion criteria:** all participants transported by EMS and coded as having possible stroke or TIA. Any people identified by EMS as "unresponsive" were excluded. **Participant characteristics:** mean age 67 (SD 16) years; 44% men
Index tests	**Index test:** CPSS **Test administrator:** paramedics **Training:** 1-hour interactive educational presentation on stroke recognition and use of the CPSS.
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** participants' final diagnosis in the hospital stroke registry, which reflects the results of routine clinical, laboratory and radiographic evaluations.
Flow and timing	30 people were excluded from study as they were noted to be "unresponsive".
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 61/154 (40%) **CPSS:** TP = 43; FP = 45; FN = 18; TN = 48
Notes	**Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** not reported **Funding:** supported by an American Stroke Association student scholarship award

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		

Frendl 2009 *(Continued)*

Prospective design	No		
		High	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Unclear		
		Unclear	

Jackson 2008

Study characteristics	
Patient sampling	St James's Hospital, Newcastle-upon-Tyne, UK **Sampling:** prospective, consecutive
Patient characteristics and setting	**Inclusion and exclusion criteria:** suspected stroke on routine triage **Participant characteristics:** mean age 73 years, range 29–41 years; 48% men
Index tests	**Index test:** ROSIER **Test administrator:** emergency physicians **Training:** none reported
Target condition and reference standard(s)	**Target condition:** stroke (TIA not mentioned in the paper)

Jackson 2008 *(Continued)*

	Reference standard: participants' records were later followed up to determine accuracy of initial diagnosis; stroke confirmed on investigation (no further details reported).
Flow and timing	Time interval between index test and discharge diagnosis not reported.
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke:** 46/50 (92%) **ROSIER:** TP = 44; FP = 3; FN = 2, TN = 1
Notes	**Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** not reported **Funding:** not reported

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		Low	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Jackson 2008 *(Continued)*

Did all patients receive the same reference standard?	Unclear	
Were all patients included in the analysis?	Yes	
Did all patients receive a reference standard?	Unclear	
		Unclear

Jiang 2014

Study characteristics	
Patient sampling	Prince of Wales hospital, Chinese University of Hong Kong; 1 June 2011 to 31 December 2011 **Sampling:** prospective cohort, consecutive participants presenting to ER
Patient characteristics and setting	**Inclusion and exclusion criteria:** Inclusion: > 18 years old, presenting to ER with symptoms/signs suggestive of stroke/TIA Exclusion: trauma brain injury with an external cause, incomplete medical records, direct admission to ward, SAH, SDH, TIA without symptoms/signs during the assessment **Participant characteristics:** mean age in stroke/TIA group 72 (SD 13) years, mean age in stroke mimics group 69 (SD 14) years; 53.4% men; seizure history: 11 people with seizure history; 295/715 (41%) participants had an onset time > 24 hours prior to assessment, hence the high concerns regarding the applicability of results.
Index tests	**Index test:** ROSIER **Test administrator:** specialist stroke nurses or consultant in emergency medicine **Training:** research staff received the specific training by stroke nurse and by a test provided by the NIHSS website. All the criteria for the scale followed the rules of the NIHSS.
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** stroke defined as a focal or global neurologic deficit with symptoms lasting for 24 hours, or resulting in death within 24 hours, which after investigation was thought to be due to a vascular cause; TIAs were defined as clinical syndromes characterized by an acute loss of focal cerebral or monocular function with symptoms lasting < 24 hours and thought to be caused by inadequate blood supply as a result of thrombosis or embolism. All people suspected of stroke were reviewed by the stroke team which included 4 stroke nurses and 2 specialist doctors. The final diagnoses were made after their assessment and after review of clinical symptoms and the acute neuroimaging (CT and MRI), and this was used as the reference standard for diagnosis in the study.
Flow and timing	Cases excluded from analysis: 51 in total of which: 4 incomplete records, 2 not accessible, 45 did not meet original ROSIER scale criteria (people with SAH, SDH, TIA). Time interval between the use of the scales and the final diagnosis, mean: 4.96 (SD 0.23) days.
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 371/715 (52%) **ROSIER:** TP = 323; FP = 202; FN = 48; TN = 142
Notes	**Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** 34 spinal neuropathy, 27 dementia, 27 labyrinthitis, 27 sepsis, 24 musculoskeletal, 24 syncope, 21 hy-

Jiang 2014 *(Continued)*

pertension, 20 somatization, 18 metabolic, 17 uncertain, 16 brain tumor, 16 peripheral neuropathy, 14 encephalopathy, 14 numbness, 13 TGA

Additional outcomes: provided by the authors

Funding: Direct Grant for research of the Chinese University of Hong Kong

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		Low	**High**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Jiang 2014 *(Continued)*

Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Did all patients receive a reference standard?	Yes
	Unclear

Kidwell 2000

Study characteristics	
Patient sampling	UCLA Medical Center, USA **Sampling:** prospective but not consecutive as the paramedics completed LAPSS forms on 206/446 people with neurologically relevant symptoms (analysis based on all runs, including those without completed forms, also reported but not included in the review).

Patient characteristics and setting	**Inclusion and exclusion criteria:** sole inclusion criterion was transport by a paramedic vehicle involved in the study during the enrolment period. Target stroke population was non-comatose, non-trauma people with symptom duration < 24 hours with suspected ischemic stroke, ICH or TIA if person was still symptomatic at the time of initial paramedic examination. 2-stage LAPSS screening process:
	
	• stage 1: paramedics were asked to identify all non-comatose, non-trauma people having neurologic complaints, i.e. people with potential stroke or stroke mimic:
	  * altered LOC,
	  * local neurologic signs,
	  * seizure,
	  * syncope,
	  * head pain, and
	  * cluster category of weak/dizzy/sick.
	
	Examples of categories that were not neurologically relevant included chest pain, allergic reaction, abdominal pain, and shortness of breath;
	
	• stage 2: criteria for LAPSS form completion were:
	  * age ≥ 18 years,
	  * neurologically relevant complaint,
	  * absence of coma, and
	  * non-traumatic presentation.
	
	Participant characteristics: mean age 67 years; 52% men

Index tests	**Index test:** LAPSS **Test administrator:** paramedics **Training:** a brief certification tape was used that consisted of 5 video vignettes of paramedics performing the LAPSS examination on 3 people with stroke, 1 stroke mimic (alcohol intoxication) and 1 healthy person. After the LAPSS-based education session, trainees watched the certification tape and completed the LAPSS exam on each vignette. Certification for use of the LAPSS required correct completion of the LAPSS exam on all 5 people. If certification was not achieved on the first trial, paramedics underwent further education in a small group setting focused on typical exam errors and then repeated the certification exam until all 5 people were correctly identified. To both reinforce and determine the impact of the training sessions, a

Kidwell 2000 *(Continued)*

	19-item stroke knowledge test was administered before and after the education session. The test included 8 items regarding stroke symptoms and diagnosis, 7 questions regarding acute stroke care and 4 items regarding stroke pathophysiology.
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** for all runs, 1 blinded author (KW) reviewed ER charts, recorded final ER discharge diagnoses and confirmed absence or presence of potential stroke symptoms. On all potential target stroke runs (people meeting LAPSS form completion criteria), 1 blinded author (CSK) additionally examined all inpatient medical records to confirm hospital discharge diagnoses of ischemic stroke, ICH and TIA by review of reports from imaging studies and attending physician notes. For people with the diagnosis of TIA, a consensus on final diagnosis was reached after complete medical record review and case discussion with a second stroke neurologist. In all people with cerebral infarct and ICH, the diagnosis of the blinded reviewer agreed with the charted diagnosis of the attending neurologist.
Flow and timing	446 people had neurologically relevant symptoms, of these paramedics applied LAPSS on 206.
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 34/206 (16.5%) **LAPSS:** TP = 31; FP = 5; FN = 3; TN = 167
Notes	**Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** not reported **Funding:** Grant-in-aid from the American Heart Association, Greater Los Angeles Affiliate

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		High	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Kidwell 2000 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
		Unclear	

Kim 2017

Study characteristics	
Patient sampling	Study conducted in Busan, Republic of Korea **Sampling:** people with suspected stroke transported to a single hospital by EMS paramedics and people with true stroke without stroke recognition by EMS for 12 months; data extracted from emergency care records including CPSS documented by EMS paramedics and hospital medical records.
Patient characteristics and setting	**Inclusion and exclusion criteria:** people with suspected stroke referred by paramedics and people with true stroke admitted during the same period (no further details). **Participant characteristics:** not reported

Kim 2017 *(Continued)*

Index tests	**Index test:** CPSS **Test administrator:** paramedics **Training:** not reported
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** hospital medical records (no further details)
Flow and timing	268 people with suspected stroke referred, of whom 152 had confirmed stroke/TIA; time between CPSS and final diagnosis not reported.
Comparative	
Diagnostic test accuracy data	**Prevalence:** 152/268 (56.7%) participants had stroke/TIA **CPSS:** TP = 142, FP = 31, FN = 10, TN = 85
Notes	**Proportion of stroke/TIA:** stroke = 149, TIA = 3 **Alternative diagnoses:** not reported **Additional information:** people with ischemic stroke with stroke recognition by EMS using CPSS had a higher NIHSS score (10 with CPSS vs 6 with negative or no documented CPSS; P = 0.001) and shorter on-scene to door time (22 with CPSS vs 25 minutes with negative or no documented CPSS; P = 0.009) and were more likely to be treated with tissue plasminogen activator (49.7% with CPSS vs 32.2% with negative or no documented CPSS; P = 0.007) than these with negative or no documented CPSS.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Prospective design	No		
		High	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**

Kim 2017 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Did all patients receive a reference standard?	Unclear		
		Unclear	

Lee 2015

Study characteristics	
Patient sampling	Dongguk University Ilsan Hospital, Korea, from August 2013 to February 2014 **Sampling:** prospective, consecutive sample (people who had come to the emergency care center directly).
Patient characteristics and setting	**Inclusion criteria:** age ≥ 18 years; presentation symptoms: weakness of extremities, dysarthria, sensory changes, alternation of consciousness, speech disturbance, visual disturbance, gait disturbance, dizziness, severe headache, syncope; hospital arrival within 12 hours from symptom onset. **Exclusion criteria:** age < 18 years; symptoms caused by trauma; transfer from other hospitals **Participant characteristics:** mean age 59.7 (SD 16.1) years in all included participants; 68.6 (SD 13.1) years in people with stroke/TIA; 54.6 (SD 15.5) years in people without stroke/TIA 45.2% men; 61.0% men in stroke or TIA group; 36.2% men in without stroke/TIA group
Index tests	**Index test:** ROSIER, FAST (applied consecutively, information supplied by the authors), usual cutoff ≥ 1 used for both scales **Test administrator:** emergency physicians (including residents and ER consultants) **Training:** 3 hours of training on theory of stroke and the acute stroke registration system from an emergency medicine specialist
Target condition and reference standard(s)	**Target condition:** stroke/TIA

Lee 2015 *(Continued)*

	Reference standard: final diagnosis of stroke/TIA signed off by a neurologist after reviewing clinical information and results MRI (confirmed by the authors). Classification of stroke was according to Oxford Community Stroke Project.
Flow and timing	Interval between index test and reference standard < 14 days. 23 (7%) participants excluded from analysis due to incomplete data (information supplied by the authors upon request)
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 113/312 (36.2%) **ROSIER:** TP = 98; FP = 14; FN = 15; TN = 185 **FAST:** TP = 97; FP = 15; FN = 16; TN = 184
Notes	**Proportion of ischemic, hemorrhagic and TIA** Ischemic 77/312 (24.7%) Hemorrhage 21/312 (6.7%) TIA 14/312 (4.5%) **Specific diagnosis in participants diagnosed with stroke or TIA:** 32 (28.3%) lacunar stroke, 28 (24.8%) partial anterior circulation stroke, 14 (12.4%) primary ICH, 14 (12.4%) TIA, 13 (11.5%) total anterior circulation stroke, 7 (6.2%) SAH, 4 (4.4%) posterior circulation stroke **Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** 59/312 (18.9%) non-specific dizziness, 39/312 (12.5%) primary headache, 30/312 (9.6%) BPPV, 15/312 (4.8%) vestibular neuritis, 12/312 (3.8%) syncope, 8/312 (2.6%) migraine, 7/312 (2.2%) drug intoxication, 6/312 (1.9%) facial palsy, 23/312 (7.4%) other **Comorbidities (% in patients with ischemic vs hemorrhagic vs TIA)** Hypertension 140 (44.9%) vs 74 (65.5%) vs 66 (33.2%); P < 0.001* Diabetes mellitus 51 (16.3%) vs 26 (23.0%) vs 25 (12.6%); P < 0.025* Previous stroke 33 (10.6%) vs 19 (16.8%) vs 14 (7.0%); P < 0.012* Hyperlipidemia 33 (10.6%) vs 16 (14.2%) vs 17 (8.5%); P < 0.129* Heart disease 21 (6.7%) vs 12 (10.6%) vs 9 (4.5%); P < 0.058* Smoking 53 (17.0%) vs 30 (26.5%) vs 23 (11.6%); P < 0.001* **Current medication (% in patients with ischemic vs hemorrhagic vs TIA)** Aspirin 35 (11.2%) vs 20 (17.7%) vs 15 (7.5%); P < 0.009* Plavix 15 (4.8%) vs 9 (8.0%) vs 6 (3.0%); P < 0.058* Warfarin 2 (0.6%) vs 0 (0%) vs 2 (1.0%); P < 0.537 *Statistically significant. **Funding:** not reported

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			

Lee 2015 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		Low	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		

Lee 2015 *(Continued)*

Did all patients receive a reference standard?	Yes
	High

Mingfeng 2012

Study characteristics	
Patient sampling	Guangzhou University of Traditional Chinese Medicine, China; April 2010 to November 2011 **Sampling:** prospective, consecutive sample
Patient characteristics and setting	**Inclusion and exclusion criteria:** Inclusion: aged > 18 years with suspected stroke or TIA with symptoms or signs seen by emergency physician in the prehospital setting Exclusion: head trauma, did not accept medical treatment in prehospital setting, refused CT or MRI; details of presenting symptoms/reason for query stroke: sudden weakness/numbness of face, arms, legs, sudden confusion, trouble speaking or understanding, sudden trouble seeing/walking, dizziness, loss of balance/co-ordination, sudden severe headache **Participant characteristics:** mean age 63 years, range 18–96 years; 67.6% men
Index tests	**Index test:** ROSIER, CPSS **Test administrator:** emergency physicians **Training:** 6-hour course on ROSIER and CPSS
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** CT or MRI; final discharge diagnosis of stroke or TIA made by the neurologists was used as the reference standard for diagnosis in this study.
Flow and timing	42 people with suspected stroke did not meet the study criteria.
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 380/540 (70.4%) **ROSIER:** TP = 341; FP = 27; FN = 38; TN = 134 **CPSS:** TP = 340; FP = 49; FN = 43; TN = 108
Notes	**Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** 40 vertigo, 27 seizure, 22 syncope, 20 cardiac, 15 sepsis, 10 hypoglycemia, 10 hysteria, 6 ethyl alcohol, 4 brain tumor, 3 demyelinating, 2 hypokalemia, 1 labyrinthitis **Funding:** not reported

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Mingfeng 2012 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		Low	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
		Unclear	

Mingfeng 2017

Study characteristics

Mingfeng 2017 *(Continued)*

Patient sampling	Luocun Community Health Service Center (LCHSC) of Nanhai District affiliated with the Nanhai Hi-Tech Industrial Zone Hospital (NHIZH) of Foshan, a Level-II hospital with the capability of managing acute stroke, and Zhangcha Community Health Service Center (ZCHSC) of Chancheng District, Foshan City, affiliated with the Foshan Hospital of Traditional Chinese Medicine (FSTCM), a tertiary care teaching hospital with acute stroke center, China **Sampling:** all people presenting to the 2 health centers; August 2012 to January 2016; ages > 18 years with suspected stroke or TIA and with symptoms or signs observed by GPs in LCHSC and ZCHSC were included in this study.
Patient characteristics and setting	**Inclusion criteria:** suspected stroke or TIA based on the following clinical signs: numbness or weakness in the face, arms or legs (especially on 1 side of the body); confusion, difficulty in speaking or understanding speech; vision disturbance in 1 or both eyes; dizziness, walking difficulties, loss of balance or co-ordination; severe headache without known cause **Exclusion criteria:** head trauma or surgery in recent months; previous stroke with neurologic deficits; incomplete medical testing was excluded from this study **Participant characteristics:** mean age 67.54 (SD 12.66) years; 192 (41.03%) women
Index tests	**Index test:** ROSIER, CPSS **Test administrator:** 16 GPs **Training:** GPs were trained by emergency physicians on the use of the ROSIER scale and CPSS for 10 hours before the study
Target condition and reference standard(s)	**Target condition:** stroke/TIA (only people with observed symptoms/signs by the GPs were included) **Reference standard:** final discharge diagnosis of stroke or TIA made by neurologists reviewing all diagnostic information including CT scan of the brain (immediately after transfer), blood tests and 12-lead ECG conducted in the ER; comprehensive neurologic assessment including additional tests, such as continuous ECG monitoring, 24-hour Holter ECG, duplex carotid and cardiac ultrasound, TCD, MRI or MRA, and conventional cerebral angiography were performed as requested by the neurologists once the participant was transferred to the neurology ward. The neurologists who made the final diagnosis were blinded to the results from the ROSIER and CPSS.
Flow and timing	512 people with suspected stroke assessed by GPs, of whom 468 met the study inclusion criteria; the CT scan of the brain was done immediately after the participant was transferred to the ER.
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 332/468 (70.94%) participants had final diagnosis stroke or TIA; 240 (72.29%) of the participants had ischemic stroke/TIA and 92 (27.71%) hemorrhagic stroke **ROSIER:** TP = 276, FP = 26, FN = 56, TN = 110 **CPSS:** TP = 259, FP = 40, FN = 73, TN = 96
Notes	**Stroke mimics:** 136 participants (45 syncope, 28 seizure, 26 vertigo, 10 hypoglycemia, 8 hypokalemia, 5 sepsis, 4 brain tumor, 3 hysteria, 2 alcohol intoxication, 2 hepatic encephalopathy, 2 Meniere's syndrome and 1 demyelination) **Funding:** funding provided by the Internal Grants from Science and Technology Foundation of Foshan City, China (no. 2014AB00328, no. 2014AG10002, and no. 2015AB00354), Guangdong Province Science and Technology Foundation (no. 2014A020212002) and the Municipal Clinical Key Specialty Construction Project Funds of Foshan City (no. Fspy2-2015004 and no. FSGSSPZD135025).
Methodological quality	

Mingfeng 2017 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		Low	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		

Mingfeng 2017 *(Continued)*

Were all patients included in the analysis?	Yes
Did all patients receive a reference standard?	Yes
	Low

Nor 2005

Study characteristics	
Patient sampling	Newcastle Hospital, UK; 1 November 2002 to 31 July 2003 **Sampling:** prospective, consecutive validation study
Patient characteristics and setting	**Inclusion and exclusion criteria:** age > 18 years, suspected stroke or TIA **Participant characteristics:** mean age: stroke 71 (SD 14) years, non-stroke 72 (SD 16) years; 41.3% men
Index tests	**Index test:** ROSIER (CPSS, LAPSS, FAST also calculated based on the neurologist-recorded signs from the prospective validation cohort) **Test administrator:** emergency physicians (ROSIER); retrospective calculation based on neurologist-recorded signs (CPSS, LAPSS, FAST) **Training:** regular educational program on how to use the instrument with twice monthly updates given to small groups of ER staff.
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** final diagnosis made by the consultant stroke physician, after assessment and review of clinical symptomatology and brain imaging findings, was used as the reference standard.
Flow and timing	Not reported
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 101/160 (63.1%) **ROSIER:** TP = 94; FP = 10; FN = 7; TN = 49 **FAST:** TP = 83; FP = 10; FN = 18; TN = 49 **LAPSS:** TP = 60; FP = 9; FN = 41; TN = 50 **CPSS:** TP = 86; FP = 12; FN = 15; TN = 47
Notes	**Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** 13 syncope, 8 seizure, 8 sepsis, 7 somatization, 4 brain tumor, 3 labyrinthitis, 3 SDH, 13 other **Funding:** Stroke Association UK
Methodological quality	

Nor 2005 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		Low	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
		Unclear	

Ramanujam 2008

Study characteristics	
Patient sampling	6 hospitals in San Diego, USA; 1 January 2005 to 31 December 2005

Ramanujam 2008 *(Continued)*

	Sampling: retrospective analysis of consecutive patient records
Patient characteristics and setting	**Inclusion and exclusion criteria:** Inclusion: ages ≥ 18 years, identified as having stroke in the prehospital phase using MPDS stroke protocol dispatchers Exclusion: people taken to other acute care hospitals not participating in the study; people with a dispatch determinant of stroke (card 28) who were not transported by City EMS agency (SDMSE) to participating hospitals; people in the stroke registry not transported by SDMSE; or people with no final outcome data **Participant characteristics:** not reported
Index tests	**Index test:** CPSS **Test administrator:** paramedics **Training:** annual 1-hour education session on recognizing stroke
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** discharge diagnosis for participants in stroke registry
Flow and timing	Missing data for 16 participants
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 440/1045 (42.1%) **CPSS:** TP = 193; FP = 284; FN = 247; TN = 321
Notes	**Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** not reported **Funding:** Stroke Center, University of California San Diego Medical Center

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Prospective design	No		
		High	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Ramanujam 2008 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Unclear		
		Unclear	

Studnek 2013

Study characteristics	
Patient sampling	7 hospitals under 2 healthcare systems, USA **Sampling:** retrospective analysis of patient records; people were included in this study if they received a prehospital MedPACS screen and were transported to 1 of the 7 local hospitals
Patient characteristics and setting	**Inclusion and exclusion criteria:** Inclusion: signs or symptoms of acute stroke or TIA Exclusion: no documented assessment, ages < 18 years, secondary transports **Participant characteristics:** mean age 66.8 (SD 16.7) years; 45.7% men
Index tests	**Index test:** CPSS, MedPACS **Test administrator:** paramedics **Training:** 2-hour continuing education lecture regarding neurologic emergencies
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** discharge diagnosis of stroke or TIA

Studnek 2013 *(Continued)*

Flow and timing	52 participants with no stroke screen performed or incomplete results
Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 186/416 (44.7%) **MedPACS:** TP = 138; FP = 155; FN = 48; TN = 75 **CPSS:** TP = 147; FP = 175; FN = 39; TN = 55
Notes	**Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** not reported **Funding:** not reported

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Prospective design	No		
		High	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Studnek 2013 *(Continued)*

Did all patients receive the same reference standard?	Unclear
Were all patients included in the analysis?	Yes
Did all patients receive a reference standard?	Unclear
	Unclear

Vanni 2011

Study characteristics	
Patient sampling	ERs of 3 hospitals located in 3 towns of the northern (Florence), middle (Rome), and southern (Pescara) regions of Italy; July 2006 to September 2006 **Sampling:** consecutive prospective sample
Patient characteristics and setting	**Inclusion and exclusion criteria:** Inclusion: presence at triage of acute focal neurologic deficit (including also signs of posterior circulation ischemia: vertigo, double vision, visual field defects or disorders of perception, balance, and co-ordination) or a local EMS dispatch of suspected stroke. Exclusion: major trauma and coma (GCS ≤ 8); people with terminal illnesses (life expectancy < 3 months) **Participant characteristics:** age in all participants 72 (SD 14) years, in people with stroke 74 (SD 11) years, in people without stroke 69 (SD 16) years; P = 0.027; men in all participants 59%, in people with stroke 61%, in people without stroke 56%; P = 0.622
Index tests	**Index test:** CPSS (compared to The Triage® Stroke Panel including quantitative measurement of B-type natriuretic peptide, fibrin degradation products containing D-dimer, matrix metalloproteinase-9 and S100β) **Test administrator:** trained nurses (at triage) **Training:** not reported
Target condition and reference standard(s)	**Target condition:** stroke, defined according to WHO criteria of "a focal or global neurologic deficit with symptoms lasting for 24 h or resulting in death before 24 hrs, which was considered to be due to a vascular cause after investigation" (reference standard positive). All other patients, including people with TIA (defined as clinical syndromes characterized by an acute loss of focal cerebral or monocular function with symptoms lasting < 24 hours and thought to be caused by inadequate blood supply as a result of thrombosis or embolism), were assigned to non-stroke diagnosis and considered free of the target condition (reference standard negatives). **Reference standard:** stroke diagnosis established by a consensus of 3 experts (1 emergency physician, 1 specialist in internal medicine and 1 neurologist), blinded to the index test results, after reviewing all clinical data and brain imaging results. Stroke diagnosis established when at least 2 of the 3 experts matched. Non-contrast-enhanced CT scanning of the brain by a 12-channel CT scanner was performed within 1 hour from emergency physician evaluation. MRI, MRA, CTA or conventional angiography was performed when appropriate.
Flow and timing	The mean time from symptom onset to presentation was 7 hours (range 1–24 hours), and only 32% of participants arrived at the hospital within 3 hours. Non-contrast-enhanced CT scanning of the brain by a 12-channel CT scanner was performed within 1 hour from the emergency physician's evaluation.

Vanni 2011 *(Continued)*

Comparative	
Diagnostic test accuracy data	**Prevalence of stroke:** 87/155 (56.1%)
Notes	**Specific diagnosis in participants diagnosed with stroke/TIA:** not reported **Categorization of alternate diagnosis for participants who did not have stroke:** 23/155 (15%) had TIA; other diagnoses not reported **Comorbidities:** Previous stroke or TIA: 21% Diabetes mellitus: 16% Atrial fibrillation: 12% SBP: 156 (SD 22) mmHg **Funding:** not declared

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		Low	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Vanni 2011 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**High**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
		Low	

Whiteley 2011

Study characteristics	
Patient sampling	Edinburgh Western General Hospital, UK; 21 March 2007 to 27 February 2009 **Sampling:** prospective consecutive people with suspected stroke, symptoms < 24 hours, symptomatic at time of assessment, GP/paramedic/member of ER staff made a diagnosis of "suspected stroke".
Patient characteristics and setting	**Inclusion and exclusion criteria:** symptoms began < 24 hours before admission; still symptomatic at the time of assessment; and in whom a GP, a paramedic or a member of the ER staff had made a diagnosis of 'suspected stroke' **Participant characteristics:** mean age 72 (SD 14) years; 51% women
Index tests	**Index test:** FAST, ROSIER **Test administrator:** emergency physician or nurse **Training:** not reported
Target condition and reference standard(s)	**Target condition:** stroke/TIA **Reference standard:** panel of experts (which included stroke physicians, neurologists and neuroradiologists), who had access to the clinical findings, imaging results and the participant's subsequent clinical course. Diagnostic criteria provided.
Flow and timing	50 participants excluded from analysis due to incomplete assessment.

Whiteley 2011 *(Continued)*

Comparative	
Diagnostic test accuracy data	**Prevalence of stroke/TIA:** 246/356 (69.1%) **FAST:** TP = 199; FP = 67; FN = 47; TN = 43 **ROSIER:** TP = 203; FP = 62; FN = 43; TN = 48
Notes	**Categorization of alternate diagnosis for participants who did not have an ischemic stroke:** TIA 37, ICH 10, SAH 2 **Funding:** none relevant

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Prospective design	Yes		
		Low	**Low**
DOMAIN 2: Index Test Index tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		

Whiteley 2011 *(Continued)*

Were all patients included in the analysis?	Yes
Did all patients receive a reference standard?	Yes
	Low

BPPV: benign paroxysmal positional vertigo; CPSS: Cincinnati Prehospital Stroke Scale; CT: computed tomography; CTA: computed tomography angiography; ECG: electrocardiogram; EEG: electroencephalogram; EMCC: Emergency Medical Communication Center; EMS: emergency medical service; ER: emergency room; FAST: Face Arm Speech Time; FN: false negative; FP: false positive; GCS: Glasgow Coma Scale; GP: general practitioner; ICH: intracerebral hemorrhage; LAPSS: Los Angeles Prehospital Stroke Scale; LOC: level of consciousness; MASS: Melbourne Ambulance Stroke Scale; MedPACS: Medic Prehospital Assessment for Code Stroke; MPDS: medical priority dispatch system; MRA: magnetic resonance angiography; MRI: magnetic resonance imaging; NIHSS: National Institutes of Health Stroke Scale; OPSST: Ontario Prehospital Stroke Screening Tool; PreHAST: PreHospital Ambulance Stroke Test; ROSIER: Recognition of Stroke in the Emergency Room; SAH: subarachnoid hemorrhage; SBP: systolic blood pressure; SD: standard deviation; SDH: subdural hematoma; SDMSE: San Diego Medical Services Enterprise; TCD: transcranial Doppler; TGA: transient global amnesia; TIA: transient ischemic attack; TN: true negative; TP: true positive; UCLA: University of California at Los Angeles; WHO: World Health Organization.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Alhanati 2014	No stroke scale specified.
Asimos 2014	Did not meet gold standard of diagnosis of stroke/diagnosis not confirmed by stroke physician, just used ICD 9/10 codes.
Belvis 2005	No stroke scale specified.
Benjamin 2013	Diagnosis not confirmed by neurologist.
Birnbaum 2008	Not a scale specific to stroke/inappropriate scale.
Blomberg 2014	No stroke scale specified.
Bray 2005b	Did not meet reference standard, author contacted requesting more information, did not receive adequate response.
Brott 1989	Inappropriate scale.
Buck 2009	Study of dispatchers, no stroke scale specified.
Caceres 2013	No stroke scale specified.
Camerlingo 2002	Questionnaire over the telephone, no stroke scale specified.
Casolla 2013	No stroke scale specified.
De La Ossa 2014	Scale only assessed large artery occlusions.
Deakin 2009	Test administered by dispatcher.
Demeestere 2017	Evaluation of NIHSS-8 for LVO using retrospectively derived scores.
Ellison 2004	No stroke scale specified.

Study	Reason for exclusion
Ferri 2005	Study protocol.
Ferro 1996	No stroke scale specified.
Ferro 1998	No stroke scale specified.
Fischer 2008	No stroke scale specified.
Garnett 2010	Protocol only, no scale.
Garrett 2013	Prognosis scale for ICH.
Govindarajan 2011	Protocol only.
Govindarajan 2012	Dispatchers only.
Gropen 2018	New stroke scale, EMSA, combining stroke recognition and severity assessment; paper described derivation and internal validation using patient records (inappropriate design).
Grossman 2011	Inappropriate scale.
Hand 2006	Not a prehospital stroke scale, scale administered by research fellow, neurologists and internists.
Harbison 2003	Did not meet reference standard.
Hasegawa 2013	Did not meet reference standard.
Henry-Morrow 2017	Evaluation of educational intervention for prehospital recognition of stroke (not an evaluation of a stroke scale).
Herzberg 2014	Use of prehospital TCD, no stroke scale specified.
Heuer 2012	Did not evaluate stroke scale (evaluation of diagnostic accuracy of ER physicians).
Huang 1994	Allen score in clinical diagnosis of intracranial hemorrhage.
Hurwitz 2005	Test administered by lay telephone caller.
Iguchi 2011	Evaluation of the Kurashiki prehospital severity scale in identifying thrombolytic candidates.
Jang 2014	Evaluation of the Kurashiki prehospital severity scale in identifying thrombolytic candidates.
Jia 2017	Accuracy of EMS diagnosis of stroke, no specific scale evaluated; retrospective data used.
Josephson 2008	ABCD score calculated post-hoc.
Kidwell 1998	Scale tested retrospectively.
Kimura 2008	Scale of severity.
Kothari 1995	No stroke scale specified.
Kothari 1997	Creation of a new stroke scale from NIHSS.
Kothari 1999	Scale done postadmission, after diagnosis have been made.

Study	Reason for exclusion
Krebes 2012	Evaluation of dispatcher's use.
Kwiatkowski 2006	Comment on another study.
Lange 2011	Only analyzed results of people who would be eligible for tPA, do not report data regarding accuracy of stroke scale.
Lavin 2014	Electronic triaging tool for family doctors.
Levine 2016	Review paper (not primary study).
Liferidge 2004	CPSS used by layperson.
Lin 2012	Prenotification study.
Llanes 2004	Scale of severity.
Malekzadeh 2015	CPSS used by dispatch nurses.
Mao 2016	Recruited people with suspected stroke presenting to the ER with symptoms or signs within 7 days.
Middleton 2016	Study protocol, general ER triage scale, not specific to stroke.
Mohd 2004	Only FAST-positive cases, focus on agreement between paramedics and physicians.
Mosley 2013	No stroke scale specified.
Nam 2014	Description of smart phone app, not accuracy study.
Nazliel 2008	Measure of stroke severity, indication of LVO.
Newman-Toker 2013	Only included people with vertigo/dizziness.
Noorian 2016	Stroke severity scale, target condition LVO.
Nor 2004	Objective of the study was an agreement between paramedics on the scene and by stroke physicians after admission in determining acute stroke signs using FAST. Analysis was confined to acute stroke cases; only FAST-positive participants identified as suspected stroke by paramedics were included. Nor 2005 reported on the same cohort of participants.
O'Brien 2012	Analysis of prehospital protocol, do not examine accuracy of FAST.
Ollikainen 2018	New stroke scale, FPSS, which combines stroke recognition and LVO identification. Study reported development and validation of scale using patient records (inappropriate design).
Oostema 2015	Focus on EMS accuracy to recognize stroke, not the accuracy of CPSS; also conference abstract, no full text.
Oostema 2016	Systematic review of the accuracy of emergency dispatchers stroke recognition when employing stroke screening tools.
Perez de la Ossa 2014	Predictor for LVO.
Purrucker 2015	Scales determined retrospectively from NIHSS; data collected from the ER neurological report.

Study	Reason for exclusion
Purrucker 2017	Reports the development of "an NIH Stroke Scale (NIHSS) compatible, all-in-one scale for rapid and comprehensive prehospital stroke assessment including stroke recognition, severity grading and progression monitoring as well as prediction of large vessel occlusion (LVO)." Validation using patient records (same cohort as in Purrucker 2015) (inappropriate design).
Quenardelle 2016	FAST calculated a posteriori, data collection to produce a FAST score not described in detail but very likely derived from the neurologist's initial assessment.
Qureshi 2016	Stroke severity scales evaluated; target condition LVO.
Richards 2016	Evaluated dichotomized CPSS to recognize people suitable for revascularization.
Robinson 2013	Public knowledge study.
Rodriguez-Pardo 2017	Evaluation of new criteria to identify people eligible for mechanical thrombectomy.
Ross 2007	Protocol for workup of TIA.
Rudd 2015	Review paper of Rudd 2016.
Rudd 2016	Systematic review.
Schilling 2012	No stroke scale specified.
Schrock 2009	ABCD score as a predictor of positive work up for TIA.
Sequeira 2015	Retrospective analysis of the accuracy of several scales but the reference standard was the NIHSS; conference abstract, no full data.
Shapiro 2003	Evaluation of electronic stroke tool, not a purely diagnostic scale, no outcomes reported for accuracy.
Sheppard 2015	Not a diagnostic accuracy study; only people with established stroke diagnosis included.
Silva 2015	End-to-end study of the impact of LAPSS results upon clinical outcome (mRS < 3) at discharge; conference abstract, no diagnostic accuracy data reported.
Smith 1998	No stroke scale specified.
Smith 1999	No stroke scale specified.
Soda 2016	"4iss" scale; probably a combined recognition/severity tool but only abstract available; "4iss" was used in people who had a positive score on FAST, as decided by paramedics.
Timerding 1989	No stroke scale specified.
Tirschwell 2002	NIHSS performed by a neurologist.
Tonomura 2015	Investigated the clinical characteristics of pseudonegative cases in prehospital triage for stroke/TIA by EMS; only people with established stroke diagnosis included; conference abstract.
Trivedi 2015	End-to-end retrospective analysis of a statewide database; only conference abstract with no sufficient test accuracy data to recreate 2 × 2 table; authors did not reply to data request.
Turc 2016	Target condition LVO.

Study	Reason for exclusion
Van Hooff 2013	Telemedicine, simulation.
Verma 2010	No stroke scale specified.
Wennman 2012	No stroke scale specified.
Wesley 2016	Comment, no primary study.
Williams 2015	FAST score obtained retrospectively, evaluates accuracy of paramedics' decision, not the instrument.
Williams 2017	Evaluation of paramedics' accuracy, not FAST (FAST was recorded only in half of the cases); ER discharge diagnosis (not hospital) was used as a reference standard (inappropriate intervention and reference standard).
Wojner-Alexandrov 2005	Evaluates accuracy of paramedics' decision, not the instrument; all EMS runs rather than those with neurologically relevant symptoms used to calculate false-negative rate.
Yamashita 2011	Inappropriate scale; attempts to differentiate hemorrhage from ischemic stroke.
Yilmaz 2014	Inappropriate reference standard, MRI only.
Yock-Corrales 2011	Participants were children.
You 2013	CPSS as a predictor of thrombolysis.
Zamora 2013	Awareness of scales in a population of medical doctors.
Ziegler 2008	No reference standard mentioned; unable to obtain additional information from authors.
Zohrevandi 2015	Retrospective analysis based on ER records.

ABCD: age, blood pressure, clinical features, duration of TIA, and presence of diabetes; CPSS: Cincinnati Prehospital Stroke Scale; ER: emergency room; EMS: emergency medical service; EMSA: Emergency Medical Stroke Assessment; FAST: Face Arm Speech Time; ICD: International Classification of Disease; ICH: intracerebral hemorrhage; LAPSS: Los Angeles Prehospital Stroke Scale; LVO: large vessel occlusion; MRI: magnetic resonance imaging; mRS: modified Rankin Scale; NIHSS: National Institutes of Health Stroke Scale; TCD: transcranial Doppler; TIA: transient ischemic attack; tPA: tissue plasminogen activator.

DATA

Presented below are all the data for all of the tests entered into the review.

Table Tests. Data tables by test

Test	No. of studies	No. of participants
1 CPSS	11	4157
2 FAST	5	1894
3 LAPSS	5	1794
4 ROSIER	8	2895

Test	No. of studies	No. of participants
5 MASS	3	981
6 OPSST	1	554
7 MedPACS	1	416
8 PreHAST	1	69

Test 1. CPSS.

Test 2. FAST.

Test 3. LAPSS.

Test 4. ROSIER.

Test 5. MASS.

Test 6. OPSST.

Test 7. MedPACS.

Test 8. PreHAST.

ADDITIONAL TABLES

Table 1. Characteristics of the evaluated stroke identification scales

—	CPSS	FAST	LAPSS	MASS	ROSIER	MedPACS	PreHAST	OPSST
	Eligibility criteria							**Exclusion criteria**
—	—	—	Age > 45 years History of seizures or epilepsy **absent** Symptom duration < 24 hours At baseline, patient not wheelchair bound or bedridden Blood glucose 60–400 mg/dL (3.3–22.2 mmol/L)	Age > 45 years History of seizures or epilepsy **absent** At baseline, patient not wheelchair bound or bedridden Blood glucose 2.8–22.2 mmol/L	—	History of seizures or epilepsy **absent** Symptom duration ≤ 25 hours Blood glucose 60–400 mg/dL (3.3–22.2 mmol/L)	Age > 18 years Intended for use only in conscious people, i.e. alert or aroused by minor stimulation	CTAS level 1; or uncorrected airway, breathing or circulatory problem (or both) Symptoms of the stroke have resolved Blood sugar < 4 mmol/L Seizure at onset of symptoms or observed by paramedic GCS < 10 Terminally ill or palliative care patient Could not arrive to a stroke center within 2 hours of a clearly determined time of symptom onset or the time the patients was "last seen in a usual state of health"
Screen items								
Facial palsy	+1	+1	+1	+1	+1	+1	+1	+1
Gaze preference	—	—	—	—	—	+1	+2	—
Vision	—	—	—	—	+ 1	—	+ 2	—
Speech disturbance	+1	+1	—	+1	+1	+1	0–2	+1
Hand grip	—	—	+1	+1	—	—	—	—

Table 1. Characteristics of the evaluated stroke identification scales *(Continued)*

Arm drift/weakness	+1	+1	+1	+1	+1	+1	0–2	+1
Leg drift/weakness	—	—	—	—	+1	+1	0–2	+1
No seizure at onset	—	—	—	—	–1	—	—	—
Blood glucose > 3 5 mmol/L	—	—	—	—	–1	—	—	—
Other	—	—	—	—	—	—	Verbal instructions +2 Sensory (pain) 0–2	—
Score range	0–3	0–3	0–3	0–4	–2 to 5	0–5	0–19	0–4
Positivity threshold	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1

CPSS: Cincinnati Prehospital Stroke Scale; Kothari 1997 and Kothari 1999.
CTAS: Canadian Triage and Acuity Scale.
FAST: Face Arm Speech Time; Kleindorfer 2007.
GCS: Glasgow Coma Scale.
LAPSS: Los Angeles Prehospital Stroke Scale. This refers to the LAPSS criteria published in Kidwell 2000, which differ slightly from an earlier version of the scale published in Kidwell 1998. In LAPSS 2000 an eligibility criterion is considered met even when the answer to the question is unknown (e.g. the person will be considered > 45 years old when this information is not available). In LAPSS 1998 when the answer to an eligibility question is unknown, the criterion is considered unmet and the person is not eligible for assessment with LAPSS; in addition, in the earlier version the symptom duration was 12 (and not 24) hours. Only LAPSS 2000 criteria are presented in the above table as no studies using LAPSS 1998 were included.
MASS: Melbourne Ambulance Stroke Scale; Bray 2005a.
MedPACS: Medic Prehospital Assessment for Code Stroke; Studnek 2013.
OPSST: Ontario Prehospital Stroke Screening Too; Chenkin 2009. The authors point out that "The addition of these exclusion criteria may be helpful for reducing the unnecessary triage of patients with stroke mimics and patients who are ineligible for fibrinolysis" (p. 154).
PreHAST: PreHospital Ambulance Stroke Test. Designed to screen for common stroke symptoms and grade severity, similarly to the National Institutes of Health Stroke Scale (NIHSS); simultaneous testing of right and left side for visual field and sensory items; only verbal instructions allowed, so it tests indirectly for sensory (Wernicke's) aphasia, Andsberg 2017.
ROSIER: Recognition of Stroke in the Emergency Room; Nor 2005.

Table 2. Inclusion criteria of the studies included in the review

Study ID	Country	Inclusion criteria
Prehospital setting		
Andsberg 2017	Sweden	Suspected stroke defined as sudden onset of focal neurologic symptoms/signs, in conscious people > 18 years of age.
Berglund 2014	Sweden	Suspected stroke with symptom onset within 6 hours; ages 18–85 years; previous independence in activities of daily living; and no other acute condition requiring a priority level 1.
Bergs 2010	Belgium	Acute neurologic event without clear origin, altered level of consciousness, convulsions, syncope, headache, and symptoms of weakness, dizziness or decreased well-being, aphasia, visual impairment, weakness in arms or legs (or both) and facial paralysis. People age < 18 years, trauma, unconsciousness (GCS ≤ 8), and people transported to another hospital were excluded.
Bray 2005a	Australia	Paramedics were instructed to complete a MASS assessment sheet on all designated EMS dispatches for 'stroke' that were symptomatic, conscious and to be transported to Box Hill Hospital. Paramedics were also asked to complete a MASS sheet for other people suspected of stroke where a focal neurologic deficit (i.e. unilateral limb weakness, speech disturbance) was noted during an initial exam.
Bray 2010	Australia	People transported by EMS with documented MASS assessments of hand grip, speech, and facial weakness; and people with a discharge diagnosis of stroke or TIA included in the stroke/TIA registry. People who were unconscious or asymptomatic at the time of paramedic assessment were excluded.
Chen 2013	China	Baseline screen criteria for the 'target stroke' population were referred from the original LAPSS study including age ≥ 18 years; neurologically relevant complaints; absence of coma and non-traumatic. The neurologically relevant complaints were identified with 6 categories, including altered level of consciousness; local neurologic signs; seizure; syncope; head pain and the cluster category of weak/dizzy/sick.
Chenkin 2009	Canada	People screened as positive by paramedics using OPSST and transported directly to a predesignated stroke center based on the person's current geographic location. Also all people with suspected stroke arriving by ambulance who did not have a positive screen were examined.
Ding 2009	China	People with acute neurologic problems and non-traumatic, non-comatose, non-obstetrics presentation transported to 3 local hospitals.
English 2018	USA	People identified by EMS dispatchers as potential stroke/TIA cases were included. Those who met any of the following inclusion criteria were selected: positive CPSS in field; EMS impression of cerebrovascular accident or TIA; acute stroke pager activation in the ER; discharge diagnosis of cerebrovascular accident or TIA. People were excluded if they met any of the following: hospital arrival via helicopter, outside hospital transfer, direct admission without ER evaluation or last known well time > 6 hours.
Fothergill 2013	UK	People aged > 18 years if they presented with symptoms of stroke, were assessed by participating ambulance clinicians using the ROSIER, and conveyed to the Royal London Hospital. Those who were ages < 18 years, not assessed using the ROSIER or transferred to another hospital were excluded.

Table 2. Inclusion criteria of the studies included in the review *(Continued)*

Frendl 2009	USA	All people transported to the Duke University Medical Center and coded by EMS as having a possible stroke or TIA were identified retrospectively by review of computerized and paper-based paramedic records for the year before and after training, regardless of whether or not an abnormality was noted for a CPSS item. These records were then compared with the hospital's prospective stroke registry for the same period. The stroke registry includes all patients admitted to the study hospital with a discharge diagnosis of stroke or TIA.
Kidwell 2000	USA	A 'target stroke' population was predefined as non-comatose, non-trauma patients with symptom duration < 24 hours with ischemic stroke, intracerebral hemorrhage, or TIA if the person was still symptomatic at the time of initial paramedic examination. These people constituted the population the LAPSS was designed to identify.
Kim 2017	Republic of Korea	People suspected of stroke and transported to a single hospital by EMS paramedics and people with true stroke without stroke recognition by EMS (retrospective sample), for a period of 12 months (data extracted from EMS records, including CPSS score).
Mingfeng 2012	China	All people > 18 years with suspected stroke or TIA with symptoms or signs seen by an emergency physician in the prehospital setting were included. According to the ASA guidelines, people who got ≥ 1 of these suggestive clinical elements as follows were defined as people with suspect acute stroke or TIA. The suggestive clinical elements included sudden weakness or numbness of the face, arm or leg, especially on 1 side of the body; sudden confusion, trouble speaking or understanding; sudden trouble seeing in 1 or both eyes; sudden trouble walking, dizziness, loss of balance or co-ordination; or sudden severe headache with unknown cause.
Mingfeng 2017	China	All people > 18 years with suspected stroke or TIA who presented to 2 primary care centers during the recruitment period. The following clinical signs were considered suggestive of stroke: numbness or weakness in the face, arms or legs (especially on 1 side of the body); confusion, difficulty in speaking or understanding speech; vision disturbance in 1 or both eyes; dizziness, walking difficulties, loss of balance or co-ordination; severe headache without known cause. Patients were excluded if they had head trauma or surgery in recent months; previous stroke with neurologic deficits or incomplete medical testing.
Ramanujam 2008	USA	People age ≥ 18 years identified as having stroke in the prehospital phase using the MPDS Stroke protocol by emergency medical dispatchers or by use of CPSS by paramedics. People taken to other acute care hospitals not participating in the study, people with a dispatch determinant of stroke who were not transported by City EMS agency (SDMSE) to participating hospitals, people in the stroke registry not transported by SDMSE or people with no final outcome data were excluded from the study.
Studnek 2013	USA	People were included if they received a prehospital MedPACS screen and were transported to 1 of 7 local hospitals. The EMS agency protocols stipulated that a MedPACS screen be performed on all people who had signs or symptoms of acute stroke or TIA. People with no documented MedPACS screen, who nevertheless ended up with a hospital diagnosis of stroke were excluded from the primary analysis. People were also excluded if they were < 18 years, if they were transported to any medical facility other than those in the inclusion criteria or if they were secondary transports from a regional facility.
ER setting		

Table 2. Inclusion criteria of the studies included in the review *(Continued)*

Jackson 2008	UK	Consecutive participants admitted to a single ER identified on routine initial triage as having possible or suspected stroke.
Jiang 2014	China	Consecutive participants ≥ 18 years old, presenting to the ER with symptoms or signs suggestive of stroke or TIA. The following people were excluded: traumatic brain injury with an external cause such as motor vehicle crashes and falls; incomplete medical records; people who did not present first to the ER (e.g. direct admission to a ward); and in accordance with the criteria for the original ROSIER scale, people with subarachnoid hemorrhage, subdural hematoma and TIA without symptoms and signs during this period.
Lee 2015	Korea	People with suspected acute stroke who were admitted to the ER.
Nor 2005	UK	People age > 18 years with suspected stroke or TIA with symptoms or signs seen by ER physicians in the ER were included.
Vanni 2011	Italy	Consecutive adults with suspected stroke who presented to the ERs of 3 hospitals. Inclusion criteria were the presence at triage of acute focal neurologic deficit (including also signs of posterior circulation ischemia: vertigo, double vision, visual field defects or disorders of perception, balance, and co-ordination) or a 118 (local EMS) dispatch of suspected stroke. Exclusion criteria were major trauma and coma (GCS score ≤ 8). People with terminal illnesses (life expectancy < 3 months) were also excluded.
Whiteley 2011	UK	Consecutive participants with suspected acute stroke who presented to the ER of the Western General Hospital, Edinburgh, while the study neurologist was available. Acute stroke was suspected in people: whose symptoms began < 24 hours before admission; who were still symptomatic at the time of assessment; and in whom a general practitioner, a paramedic or a member of the emergency-room staff had made a diagnosis of 'suspected stroke'.

ASA: American Stroke Association; CPSS: Cincinnati Prehospital Stroke Scale; EMS: emergency medical services; ER: emergency room; GCS: Glasgow Coma Scale; LAPSS: Los Angeles Prehospital Stroke Scale; MASS: Melbourne Ambulance Stroke Scale; MPDS: medical priority dispatch system; OPSST: Ontario Prehospital Stroke Screening Tool; ROSIER: Recognition of Stroke in the Emergency Room; SDMSE: San Diego Medical Services Enterprise; TIA: transient ischemic attack.

Table 3. Characteristics of study cohorts

Study ID	Country	Sample size	Prevalence (%)	Ischemic stroke (%)	Hemorrhagic stroke (%)	TIA (%)	Mean age (years)	Sex (% women)	Eligible participants out of all EMS runs or ER presentations (%)
Prehospital setting									
Andsberg 2017	Sweden	69	38	69	4	27	n/a	n/a	n/a
Berglund 2014	Sweden	900	52	64	9	27	71	44	n/a
Bergs 2010	Belgium	31	61	79	16	5	77	39	7.6
Bray 2005a	Australia	100	73	68	13	23	76	n/a	2.1
Bray 2010	Australia	850	23	n/a	n/a	n/a	n/a	n/a	19
Chen 2013	China	1130	88	61	25	3	72	39	3.1
Chenkin 2009	Canada	554	57	58	21	11	74	31	n/a
Ding 2009	China	327	16	44	36	20	58	48	16.1
English 2018	USA	130	74	64	15	21	72–77	50–52	34.5
Fothergill 2013	UK	295	60	71	23	6	64	47	n/a
Frendl 2009	USA	154	40	n/a	n/a	n/a	67	56	n/a
Kidwell 2000	USA	206	17	n/a	n/a	n/a	63	48	34.4
Kim 2017	Korea	268	57	n/a	n/a	3	n/a	n/a	n/a
Mingfeng 2012	China	540	70	n/a	41	n/a	63	32	n/a
Mingfeng 2017	China	468	71	n/a	28	n/a	71	51	n/a
Ramanujam 2008	USA	1045	42	n/a	n/a	n/a	n/a	n/a	1.3
Studnek 2013	USA	416	45	82	n/a	n/a	67	54	n/a

Table 3. Characteristics of study cohorts *(Continued)*

ER setting

Jackson 2008	UK	50	92	n/a	n/a	n/a	73	52	n/a
Jiang 2014	China	714	52	81	12	7	72	47	n/a
Lee 2015	Korea	312	36	57	31	12	60	55	n/a
Nor 2005	UK	160	63	76	14	10	70	59	n/a
Vanni 2011	Italy	155	56	89	11	n/a	72	41	6.8
Whiteley 2011	UK	356	69	80	5	15	72	51	n/a

EMS: emergency medical service; ER: emergency room; n/a: not applicable; TIA: transient ischemic attack.

Table 4. Index test and reference standard

Study ID	Setting	Index tests	Test administrator	Training	Reference standard
Andsberg 2017	Prehospital	PreHAST	Nurse	4-hour educational program, covering basic stroke knowledge and assessment and grading of stroke symptoms according to PreHAST; it included practical PreHAST training in pairs, where each ambulance nurse performed the PreHAST items under supervision and proper execution. During the study an instruction video for PreHAST was available on YouTube.	2 stroke physicians, blinded to the PreHAST scores, independently reviewed the medical records of the participants, including evaluation of history, clinical and radiologic findings. In case of disagreement, a third evaluator adjudicated the final diagnosis.
Berglund 2014	Prehospital	FAST	Nurse or paramedic	1 lecture about stroke and the FAST test, prior to start of the study.	CT brain scan and in some cases CTA or MRI, neurologic examination, if necessary, EEG (differential diagnosis), laboratory tests. All participants received a final diagnosis by a neurologist or stroke specialist.
Bergs 2010	Prehospital	CPSS, FAST, LAPSS, MASS	Nurse	All nurses were briefed on purpose of study, stroke scales and guidelines.	Diagnosis at ER discharge (unspecified).
Bray 2005a	Prehospital	CPSS, LAPSS, MASS	Paramedic	1-hour educational session on stroke and use of the prehospital stroke scale.	Standard criteria for diagnosis of stroke or TIA (Warlow 2001); review of discharge diagnosis (no further details).
Bray 2010	Prehospital	CPSS, MASS	Paramedic	1-hour educational program and instruction on use of the MASS.	Discharge diagnosis based on hospital stroke registry.
Chen 2013	Prehospital	LAPSS	Paramedic	3 hours' LAPSS-based stroke training session with 3 experts from study team.	2 blinded neurologists reviewed the ER charts, recorded final ER discharge diagnoses, and verified absence or presence of potential stroke symptoms. The medical documents and neuroimaging records were reviewed before the final diagnoses were verified.
Chenkin 2009	Prehospital	OPSST	Paramedic	90-minute training session on the stroke	Hospital discharge diagnosis (no further details).

Table 4. Index test and reference standard *(Continued)*

				screening tool prior to implementation.	
Ding 2009	Prehospital	LAPSS	Emergency physician	Not reported.	Hospital final diagnosis made by a specialist group including a neurologist, a radiologist and a generalist.
English 2018	Prehospital	CPSS	Paramedic	1-hour online module annually on stroke recognition and assessment in the field as part of their required job training.	Hospital discharge diagnosis (no further details).
Fothergill 2013	Prehospital	FAST, ROSIER	Paramedic	1-hour stroke educational program, scenario based demonstration of ROSIER, 15-minute educational DVD.	Final diagnosis made by a stroke consultant or other senior medical physician caring for the person within 72 hours of the person's admission to hospital, based on CT and MRI scans. The final diagnosis was confirmed by a senior stroke consultant.
Frendl 2009	Prehospital	CPSS	Paramedic	1-hour interactive educational presentation on stroke recognition and use of the CPSS.	Hospital discharge diagnosis based on the results of routine clinical, laboratory and radiographic evaluations.
Kidwell 2000	Prehospital	LAPSS	Paramedic	Video vignettes of paramedics performing the LAPSS examination on 3 people with stroke, 1 stroke mimic person, and 1 healthy person. Following a LAPSS-focused education session, trainees had to pass an exam which, if failed, was followed by further training.	For all runs, 1 blinded author reviewed ER charts, recorded final ER discharge diagnoses and confirmed absence or presence of potential stroke symptoms. On all potential target stroke runs, 1 blinded author additionally examined all inpatient medical records to confirm hospital discharge diagnoses of stroke/TIA by review of reports from imaging studies and attending physician notes. For people with the diagnosis of TIA, a consensus on final diagnosis was reached after complete medical record review and case discussion with a second stroke neurologist. In all people with cerebral infarct and intracerebral hemorrhage, the diagnosis of the blinded reviewer agreed with the charted diagnosis of the attending neurologist.
Kim 2017	Prehospital	CPSS	Paramedic	Not reported.	Hospital medical records.
Mingfeng 2012	Prehospital	CPSS, ROSIER	Emergency physician	6-hour course on ROSIER and CPSS.	The final discharge diagnosis of stroke/TIA made by neurologists and based on CT or MRI.
Mingfeng 2017	Prehospital	CPSS, ROSIER	GP	Trained by emergency physicians on the use of the	Final discharge diagnosis of stroke or TIA made by neurologists reviewing all diagnostic information includ-

Table 4. Index test and reference standard *(Continued)*

				ROSIER scale and CPSS for 10 hours before the study.	ing CT scan of the brain (immediately after transfer), blood tests and 12-lead ECG conducted in the ER; comprehensive neurologic assessment including additional tests, such as continuous ECG monitoring, 24-hour Holter ECG, duplex carotid and cardiac ultrasound, TCD, MRI or MRA, and conventional cerebral angiography were performed as requested by the neurologists once the person was transferred to the neurology ward. The neurologists who made the final diagnosis were blinded to the results from the ROSIER and CPSS.
Ramanujam 2008	Prehospital	CPSS	Paramedic	Annual 1-hour education session on recognizing stroke.	Discharge diagnosis for people in stroke registry.
Studnek 2013	Prehospital	CPSS, MedPACS	Paramedic	2-hour continuing education lecture regarding neurologic emergencies.	Discharge diagnosis of stroke/TIA.
Jackson 2008	ER	ROSIER	Emergency physician	No training reported.	Patients' records were later followed up to determine accuracy of initial diagnosis; stroke confirmed on investigation (no further details reported).
Jiang 2014	ER	ROSIER	Emergency physician or nurse	The research staff received the specific training by a stroke nurse and a module/exam provided by the NIHSS website.	The final diagnoses were made after people suspected of stroke were reviewed by the stroke team and after review of clinical symptoms and the acute neuroimaging (CT and MRI).
Lee 2015	ER	FAST, ROSIER	Emergency physician	3 hours of training on theory of stroke and the acute stroke registration system from an emergency medicine specialist.	Ischemic stroke and bleeding were determined in accordance with brain CT and MRI results. The final diagnosis was confirmed at the time through the electronic medical record.
Nor 2005	ER	ROSIER	Emergency physician	Regular educational program on the use of the instrument with twice monthly updates given to small groups of ER staff.	Final diagnosis made by the consultant stroke physician, after assessment and review of clinical symptomatology and brain imaging findings.
Vanni 2011	ER	CPSS	Nurse	No training reported.	TIA was excluded from the target condition. Stroke diagnosis established by a consensus of 3 experts, blinded to the index test results, after reviewing all clinical data and brain imaging results.

Table 4. Index test and reference standard *(Continued)*

Whiteley 2011	ER	FAST, ROSIER	Emergency physician or nurse	No training reported.	Diagnosis made by a panel of experts, who had access to the clinical findings, imaging results and the person's subsequent clinical course.

CPSS: Cincinnati Prehospital Stroke Scale; CT: computed tomography; CTA: computed tomography angiography; ECG: electrocardiogram; ER: emergency room; EEG: electroencephalogram; FAST: Face Arm Speech Time; LAPSS: Los Angeles Prehospital Stroke Scale; MASS: Melbourne Ambulance Stroke Scale; MRA: magnetic resonance angiography; MRI: magnetic resonance imaging; NIHSS: National Institutes of Health Stroke Scale; OPSST: Ontario Prehospital Stroke Screening Tool; PreHAST: PreHospital Ambulance Stroke Test; ROSIER: Recognition of Stroke in the Emergency Room; TCD: transcranial Doppler; TIA: transient ischemic attack.

Table 5. Comparative accuracy of the scales

Scales	Sensitivity	Specificity	Studies
FAST vs ROSIER	=	=	Lee 2015; Whiteley 2011
CPSS vs MASS	=	<	Bray 2005a; Bray 2010
CPSS vs ROSIER	?	<	Mingfeng 2012
CPSS vs LAPSS	>	?	Bray 2005a
CPSS vs MedPACS	>	<	Studnek 2013
MASS vs LAPSS	>	=	Bray 2005a

The table summarizes the results from studies reporting the statistical significance of the differences in sensitivity and specificity estimates of the scales and is equivalent to Figure 27. A version of the table including the results from Purrucker 2015 is given in Appendix 10.

APPENDICES

Appendix 1. Cochrane Library search strategy

The Cochrane Library databases

Cochrane Database of Systematic Reviews (CDSR)

Database of Reviews of Effects (DARE)

Cochrane Central Register of Controlled Trials (CENTRAL)

Health Technology Assessment database (HTA)

NHS Economic Evaluation Database (NHSEED)

#1 [mh ^"cerebrovascular disorders"] or [mh "basal ganglia cerebrovascular disease"] or [mh "brain ischemia"] or [mh "carotid artery diseases"] or [mh "cerebrovascular trauma"] or [mh "intracranial arterial diseases"] or [mh "intracranial arteriovenous malformations"] or [mh "intracranial embolism and thrombosis"] or [mh "intracranial hemorrhages"] or [mh ^stroke] or [mh "brain infarction"] or [mh ^"stroke, lacunar"] or [mh ^"vasospasm, intracranial"] or [mh ^"vertebral artery dissection"]

#2 (stroke* or apoplex* or cerebral next vasc* or cerebrovasc* or cva or SAH):ti,ab

#3 ((brain or cerebr* or cerebell* or vertebrobasil* or hemispher* or intracran* or intracerebral or infratentorial or supratentorial or middle next cerebr* or mca* or "anterior circulation" or "basilar artery" or "vertebral artery") near/5 (isch*emi* or infarct* or thrombo* or emboli* or occlus* or hypoxi*)):ti,ab

#4 ((brain* or cerebr* or cerebell* or intracerebral or intracran* or parenchymal or intraparenchymal or intraventricular or infratentorial or supratentorial or basal next gangli* or putaminal or putamen or "posterior fossa" or hemispher* or subarachnoid) near/5 (hemorrhag* or haemorrhag* or hematoma* or hematoma* or bleed*)):ti,ab

#5 [mh ^"ischemic attack, transient"] or ((transient near/3 isch*emi*) or TIA or TIAs):ti,ab

#6 #1 or #2 or #3 or #4 or #5

#7 ((Cincinnati or "Los Angeles" or Ontario or Maria or Kurashiki) near/10 (scale* or screen*)):ti,ab

#8 (("Melbourne Ambulance" near/5 (screen* or scale*)) or "face arm speech test"):ti,ab

#9 #7 or #8

#10 #6 and #9

#11 [mh ^"emergency medical services"]

#12 [mh ^"emergency medical service communication systems"]

#13 [mh ^"emergency service, hospital"] or [mh ^"emergency medicine"] or [mh ^"emergency treatment"]

#14 [mh ^ambulances] or [mh ^"emergency responders"] or [mh ^"allied health personnel"]

#15 (prehospital* or pre-hospital* or pre hospital* or ambulance* or paramedic* or EMS):ti,ab

#16 (Emergency near/3 (medical or health) near/3 (service* or system* or worker* or personnel* or responder* or dispatcher* or unit or units or technician* or vehicle*)):ti,ab

#17 (emergency near/5 (physician* or staff or room* or department*)):ti,ab

#18 #11 or #12 or #13 or #14 or #15 or #16 or #17

#19 (stroke* near/5 (scale* or screen* or checklist* or assess* or identif* or recogni* or evaluat* or diagnos* or detect*)):ti,ab

#20 #18 and #19

#21 [mh ^"cerebrovascular disorders"/DI,CL] or [mh "basal ganglia cerebrovascular disease"/DI,CL] or [mh "brain ischemia"/DI,CL] or [mh "carotid artery diseases"/DI,CL] or [mh "cerebrovascular trauma"/DI,CL] or [mh "intracranial arterial diseases"/DI,CL] or [mh "intracranial arteriovenous malformations"/DI,CL] or [mh "intracranial embolism and thrombosis"/DI,CL] or [mh "intracranial hemorrhages"/DI,CL] or [mh ^stroke/DI,CL] or [mh "brain infarction"/DI,CL] or [mh ^"stroke, lacunar"/DI,CL] or [mh ^"vasospasm, intracranial"/DI,CL] or [mh ^"vertebral artery dissection"/DI,CL]

#22 #18 and #21

#23 [mh ^diagnosis] or [mh ^"early diagnosis"]

#24 #6 and #18 and #23

#25 [mh "sensitivity and specificity"]

#26 (sensitiv* or specificity):ti,ab

#27 (predictive near/5 value*):ti,ab

#28 [mh "diagnostic errors"]

#29 ((false next positive*) or (false next negative*)):ti,ab

#30 (observer next variation*):ti,ab

#31 (roc next curve*):ti,ab

#32 (likelihood near/3 ratio*):ti,ab

#33 [mh ^"likelihood function"]

#34 #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33

#35 #6 and #18 and #34

#36 #10 or #20 or #22 or #24 or #35

Appendix 2. MEDLINE search strategy

MEDLINE (Ovid)

1. cerebrovascular disorders/ or exp basal ganglia cerebrovascular disease/ or exp brain ischemia/ or exp carotid artery diseases/ or exp cerebrovascular trauma/ or exp intracranial arterial diseases/ or exp intracranial arteriovenous malformations/ or exp "intracranial embolism and thrombosis"/ or exp intracranial hemorrhages/ or stroke/ or exp brain infarction/ or stroke, lacunar/ or vasospasm, intracranial/ or vertebral artery dissection/
2. (stroke$ or apoplex$ or cerebral vasc$ or cerebrovasc$ or cva or SAH).tw.
3. ((brain or cerebr$ or cerebell$ or vertebrobasil$ or hemispher$ or intracran$ or intracerebral or infratentorial or supratentorial or middle cerebr$ or mca$ or anterior circulation or basilar artery or vertebral artery) adj5 (isch?emi$ or infarct$ or thrombo$ or emboli$ or occlus$ or hypoxi$)).tw.
4. ((brain$ or cerebr$ or cerebell$ or intracerebral or intracran$ or parenchymal or intraparenchymal or intraventricular or infratentorial or supratentorial or basal ganglia$ or putaminal or putamen or posterior fossa or hemispher$ or subarachnoid) adj5 (h?emorrhag$ or h?ematoma$ or bleed$)).tw
5. ischemic attack, transient/ or ((transient adj3 isch?emi$) or TIA or TIAs).tw.
6. 1 or 2 or 3 or 4 or 5
7. ((Cincinnati or Los Angeles or Ontario or Maria or Kurashiki) adj10 (scale$ or screen$)).tw.
8. (Melbourne Ambulance adj5 (screen$ or scale$)).tw.
9. 7 or 8
10. 6 and 9
11. emergency medical services/
12. emergency medical service communication systems/
13. emergency service, hospital/ or emergency medicine/ or emergency treatment/
14. ambulances/ or emergency responders/ or allied health personnel/
15. (prehospital$ or pre-hospital$ or pre hospital$ or ambulance$ or paramedic$ or EMS).tw.
16. (Emergency adj3 (medical or health) adj3 (service$ or system$ or worker$ or personnel$ or responder$ or dispatcher$ or unit or units or technician$ or vehicle$)).tw.
17. (emergency adj5 (physician$ or staff or room$ or department$)).tw.
18. 11 or 12 or 13 or 14 or 15 or 16 or 17
19. (stroke$ adj5 (scale$ or screen$ or checklist$ or assess$ or identif$ or recogni$ or evaluat$ or diagnos$ or detect$)).tw.
20. 18 and 19
21. cerebrovascular disorders/di, cl or basal ganglia cerebrovascular disease/di, cl or brain ischemia/di, cl or exp brain infarction/di, cl or hypoxia-ischemia, brain/di, cl or carotid artery diseases/di, cl or carotid artery thrombosis/di, cl or carotid artery, internal, dissection/di, cl or intracranial arterial diseases/di, cl or cerebral arterial diseases/di, cl or infarction, anterior cerebral artery/di, cl or infarction, middle cerebral artery/di, cl or infarction, posterior cerebral artery/di, cl or exp "intracranial embolism and thrombosis"/di, cl or exp stroke/di, cl or vertebral artery dissection/di, cl
22. 18 and 21
23. diagnosis/ or early diagnosis/
24. 6 and 18 and 23
25. exp "sensitivity and specificity"/
26. (sensitiv$ or specificity).tw.
27. (predictive adj5 value$).tw.
28. exp diagnostic errors/
29. ((false adj positive$) or (false adj negative$)).tw.
30. (observer adj variation$).tw.
31. (roc adj curve$).tw.
32. (likelihood adj3 ratio$).tw.
33. likelihood function/
34. 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33
35. 6 and 18 and 34
36. 10 or 20 or 22 or 24 or 35

Appendix 3. Embase search strategy

1. cerebrovascular disease/ or exp basal ganglion hemorrhage/ or exp brain hematoma/ or exp brain hemorrhage/ or exp brain infarction/ or exp brain ischemia/ or exp carotid artery disease/ or cerebral artery disease/ or exp cerebrovascular accident/ or exp occlusive cerebrovascular disease/ or stroke patient/

2. (stroke$ or apoplex$ or cerebral vasc$ or cerebrovasc$ or cva or SAH).tw.

3. ((brain or cerebr$ or cerebell$ or vertebrobasil$ or hemispher$ or intracran$ or intracerebral or infratentorial or supratentorial or middle cerebr$ or mca$ or anterior circulation or basilar artery or vertebral artery) adj5 (isch?emi$ or infarct$ or thrombo$ or emboli$ or occlus$ or hypoxi$)).tw.

4. ((brain$ or cerebr$ or cerebell$ or intracerebral or intracran$ or parenchymal or intraparenchymal or intraventricular or infratentorial or supratentorial or basal gangli$ or putaminal or putamen or posterior fossa or hemispher$ or subarachnoid) adj5 (h?emorrhag$ or h?ematoma$ or bleed$)).tw.

5. ischemic attack, transient/ or ((transient adj3 isch?emi$) or TIA or TIAs).tw.

6. 1 or 2 or 3 or 4 or 5

7. ((Cincinnati or Los Angeles or Ontario or Maria or Kurashiki) adj10 (scale$ or screen$)).tw.

8. ((Melbourne Ambulance adj5 (screen$ or scale$)) or face arm speech test).tw.

9. 7 or 8

10. 6 and 9

11. emergency health service/

12. emergency/ or emergency call system/ or emergency care/ or emergency medicine/

13. emergency treatment/ or emergency patient/ or emergency nurse practitioner/ or emergency nursing/ or emergency physician/

14. ambulance/ or rescue personnel/ or rapid response team/ or paramedical personnel/ or paramedical profession/

15. (prehospital$ or pre-hospital$ or pre hospital$ or ambulance$ or paramedic$ or EMS).tw.

16. (Emergency adj3 (medical or health) adj3 (service$ or system$ or worker$ or personnel$ or responder$ or dispatcher$ or unit or units or technician$ or vehicle$)).tw.

17. (emergency adj5 (physician$ or staff or room$ or department$)).tw.

18. 11 or 12 or 13 or 14 or 15 or 16 or 17

19. (stroke$ adj5 (scale$ or screen$ or checklist$ or assess$ or identif$ or recogni$ or evaluat$ or diagnos$ or detect$)).tw.

20. 18 and 19

21. cerebrovascular disease/di or exp basal ganglion hemorrhage/di or exp brain hematoma/di or exp brain hemorrhage/di or exp brain infarction/di or exp brain ischemia/di or exp carotid artery disease/di or cerebral artery disease/di or exp cerebrovascular accident/di or exp occlusive cerebrovascular disease/di or stroke patient/di

22. 18 and 21

23. neurologic examination/ or diagnosis/ or early diagnosis/ or diagnostic test/

24. 6 and 18 and 23

25. "sensitivity and specificity"/

26. receiver operating characteristic/

27. diagnostic accuracy/

28. exp diagnostic error/

29. observer variation/

30. "limit of detection"/

31. "diagnostic test accuracy study".sh.

32. (sensitivity or specificity).tw.

33. (predictive adj3 value$).tw.

34. ((false adj positive$) or (false adj negative$)).tw.

35. observer variation$.tw.

36. (roc adj curve$).tw.

37. (likelihood adj3 ratio$).tw.

38. or/25-37

39. 6 and 18 and 38

40. 10 or 20 or 22 or 24 or 39

Appendix 4. Data extraction form

Review, reviewer and study Information

Title and ID of review

Reviewer ID

Date of form completion

Study ID (for Revman)

Study characteristics

Title

Authors

Journal Name

Publication date

Study design

Number of centres

Sample size

Funding

Country

Patient characteristics

Age

Sex

Seizure history

The study's inclusion and exclusion criteria

Details of presenting symptoms/reason for query stroke

Index test

Test interventions- the prehospital stroke scale

Test name

Title of test administrator (Paramedic, Emergency physician)

Details of test and stroke specific training regimen for test administrator

Reference standard

Tests on which the final diagnosis is based

Adjudication process

Outcomes

Diagnostic accuracy outcomes

Number of patients in study

Number of patients with a discharge diagnosis of stroke

Number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN)

Sensitivity and specificity with 95% confidence intervals

Additional outcomes

Total mortality

Total morbidity

Modified Rankin Score

Change in Modified Rankin Score at discharge and at 90 days

Door to needle time

Categorization of alternate diagnosis for patients who did not have an ischemic stroke

Withdrawals

Unable to use data – inadequate data has been provided in the publication, specifically, the data set is inadequate to the point where the sensitivity, specificity etc. cannot be calculated from the data set provided

Appendix 5. Quality Assessment Checklist (QUADAS-2)

Patient selection

Description

Describe the methods of patient selection reported in the paper.

Risk of bias

Was a consecutive or random sample of patients enrolled?

- 'Yes' if consecutive or random sampling is explicitly stated.
- 'No' if non-consecutive or convenience sampling is used.
- 'Unclear' if the provided information is insufficient to make a judgment.

Was a case-control design avoided?

- 'Yes' if all patients were recruited from the same population using a single set of inclusion and exclusion criteria.
- 'No' if patients with and without the target condition were recruited from different populations and/or using different sets of inclusion and exclusion criteria.
- ·'Unclear' if the provided information is insufficient to decide .

Did the study avoid inappropriate exclusions?

(Exclusion of patients with a past medical history suggestive of an alternate diagnosis that can mimic a stroke such as a seizure).

(Exclusion of patients with an undifferentiated presentation that may have a stroke according to the intervention test).

(Exclusion of patients based on the absence of any cardinal risk factors for stroke (hypertension, smoker, diabetes, dyslipidemia, prior TIA/stroke).

- 'Yes' if all eligible patients suspected of stroke were included in the study.
- 'No' if eligible patients suspected of stroke were excluded from the studies for reasons that might affect the accuracy of the index test.
- 'Unclear' if not possible to make a judgment based on the information provided in the study.

Prospective study design? (Were participants recruited prospectively for the purpose of the study?)

- 'Yes' if participants recruited prospectively.
- 'No' if study sample selected from a registry (retrospective recruitment).
- 'Unclear' if the provided information is insufficient to make a judgment.

Concerns regarding applicability

- 'Low' if the study included unselected patients suspected of having a stroke at the first point of contact.
- 'High' if otherwise.
- 'Unclear' if the information provided in the paper is insufficient to make a decision.

Index test

Description

Describe the index test and how it was conducted and interpreted.

Risk of bias

Were the index test results interpreted without knowledge of the results of the reference standard?

- 'Yes' or 'No' if stated in the study
- 'Unclear' if not stated

If a threshold was used, was it prespecified?

- 'Yes' if the threshold used to make a referral decision was prespecified.
- 'No' if ROC-optimized or other not prespecified cutoff was used.
- 'Unclear' if the provided information is insufficient to decide.

Concerns regarding applicability

- 'Low' if the index test was performed by prehospital staff in the context of first contact with a patient suspected of stroke.
- 'High' if otherwise.
- 'Unclear' if insufficient information is provided.

Reference standard

Description

Describe the reference standard and how it was conducted and interpreted.

Risk of bias

Is the reference standard likely to correctly classify the target condition?

- 'Yes' if the final diagnosis was based on the results from history, physical examination and non-contrast computed tomography (CT) head scan and/or any other imaging and was adjudicated independently by a neurologist.
- 'No' if the above criteria were not met.
- 'Unclear' if reported data is insufficient to decide.

Were the results from the reference standard interpreted without knowledge of the results from the index test?

- 'Yes' or 'no' if explicitly stated, 'unclear' if not reported.

Concerns regarding applicability

- 'Low' if the target condition as defined by the reference standard was stroke, without discrimination of type or severity.
- 'High' if the target condition was a specific type of stroke (e.g. ischemic or hemorrhagic) or level of stroke severity.

- 'Unclear' if insufficient information is provided.

Flow and timing

Description

Describe any patients who did not receive the index test(s), or reference standard, or both, or who were excluded from the 2x2 table.

Risk of bias

Was there an appropriate interval between index test(s) and reference standard?

- 'Yes' – the blinded neurologist saw the patient and took a history and physical and interpreted the imaging within 14 days of the implementation of the index test.
- 'No' – the study clearly states that patients were not seen by the blinded neurologist within 14 days.
- 'Unclear' – the time of diagnosis by the neurologist is not made available.

Did all patients receive a reference standard?

- 'Yes' if all patients received a reference standard (seen by the blinded neurologist).
- 'No' if a portion of the included patients did not receive a reference standard.
- 'Unclear' if data is insufficient to decide.

Did all patients receive the same reference standard?

- 'Yes' if all patients received the same reference standard.
- 'No' if the final diagnosis of stroke was based on different combination of tests.
- 'Unclear' if there is no sufficient data to make a judgment.

Were all patients included in the analysis?

- 'Yes' if all patients were included.
- 'No' if patients who were enrolled in the study were excluded from the analysis.
- 'Unclear' if insufficient data is available to decide.

Appendix 6. Results from the searches

Initial searches (January 2015):

MEDLINE in Ovid (1946 to January 2015), N = 2156

Embase in Ovid (1980 to January 2015), N = 6182

Cochrane Library (CDSR, DARE, CENTRAL, HTA, NHSEED searched January 2015), N = 143 refs

Total N = 8481

Update (January 2015 to January 2017):

MEDLINE in Ovid (1950 to January 2017), N = 554

Embase in Ovid (1980 to January 2017), N = 2447

Cochrane Central Register of Controlled Trials (CENTRAL; 2017, Issue 1) in the Cochrane Library (searched January 2017), N = 237

Total N = 3238

Update (January 2017 to January 2018):

MEDLINE Ovid (1946 to 30 January 2018), N = 131

Embase Ovid (1974 to 30 January 2018), N = 1035

Cochrane Central Register of Controlled Trials (CENTRAL; 2018, Issue 12) and the Cochrane Database of Systematic Reviews (CENTRAL; 2018, Issue 1) in the Cochrane Library (searched 30 January 2018), N = 54

Nothing new was found in HTA or DARE since previous search

Appendix 7. Summary of the reasons for full-text exclusion of studies

Exclusions of studies at full text screening

143 from the initial searches:

- 71 conference abstracts;
- 39 no new data (e.g. review) or inappropriate study design;
- 25 unsuitable index test;
- 4 unsuitable setting;
- 1 population;
- 2 unsuitable reference standard;
- 1 insufficient data.

29 from the update:

- 4 conference abstracts;
- 12 no new data (e.g. review) or inappropriate study design;
- 1 unsuitable target condition;
- 9 unsuitable index test;
- 2 unsuitable setting;
- 1 unsuitable reference standard.

Appendix 8. Type of stroke missed by ROSIER (false negatives) as reported by the studies included in the meta-analysis

Study	Sample size	Details of the type of stroke missed by the test as stated in the paper
Whiteley 2011	356	"Both the ROSIER and the FAST performed relatively poorly in identifying patients with posterior circulations strokes (18/37, 49% FAST positive and 19/37, 51% ROSIER positive)." p. 1008.
Lee 2015	312	TIA, thalamic infarction, posterior circulation stroke, middle cerebral artery infarction, intracranial hemorrhage, table 5, p. 471.
Jiang 2014	715	Not reported.
Nor 2005	160	"The false-negative group included posterior circulation infarction (n = 5) and lacunar infarction (n = 2). The neurological signs in these cases were gait ataxia (n = 5), sensory deficits (n = 2), and one each of ophthalmoplegia, quadriparesis, and loss of consciousness. Most (six) of these false-negative cases had mild deficits (NIHSS < 3), and would not have been clear candidates for thrombolytic therapy even if they had presented sufficiently early. The remaining patient had an NIHSS score of 24 and presented with drowsiness, gaze palsy, and quadriplegia." p. 371.
Jackson 2008	50	"Two patients with stroke were found to have a ROSIER score of 0, one was admitted unconscious with a large primary intracerebral hemorrhage and was wrongly scored and the second had a cerebellar infarct with no weakness, speech or visual field defect." p. 190.

FAST: Face Arm Speech Time; n: number of participants; NIHSS: National Institutes of Health Stroke Scale; ROSIER: Recognition of Stroke in the Emergency Room; TIA: transient ischemic attack.

Appendix 9. Additional comparative data reported by Purrucker 2015

Stroke scale	Sensitivity (95% CI)	Specificity (95% CI)
CPSS	0.83 (0.76 to 0.88)	0.69 (0.64 to 0.73)
FAST	0.85 (0.78 to 0.90)	0.68 (0.63 to 0.72)
LAPSS 1998	0.44 (0.36 to 0.52)	0.98 (0.96 to 0.99)
LAPSS 2000	0.49 (0.41 to 0.57)	0.97 (0.95 to 0.99)
MASS	0.63 (0.55 to 0.70)	0.94 (0.91 to 0.96)
MedPACS	0.71 (0.64 to 0.78)	0.92 (0.89 to 0.94)
ROSIER	0.80 (0.73 to 0.85)	0.79 (0.75 to 0.83)

The table is based on Table 3 in Purrucker 2015.

CI: confidence interval; FAST: Face Arm Speech Time; LAPSS: Los Angeles Prehospital Stroke Scale; MASS: Melbourne Ambulance Stroke Scale; MedPACS: Medic Prehospital Assessment for Code Stroke; ROSIER: Recognition of Stroke in the Emergency Room.

Appendix 10. Summary of comparative accuracy: comparison between included studies and Purrucker 2015

Scales	Included studies		Purrucker 2015	
	Sensitivity	Specificity	Sensitivity	Specificity
FAST vs ROSIER	=	**=**	?	**<**
CPSS vs MASS	**=**	<	**>**	<
CPSS vs ROSIER	?	<	?	<
CPSS vs LAPSS	>	?	>	<
CPSS vs MedPACS	>	<	trend >	trend <
MASS vs LAPSS	>	**=**	trend >	**trend <**

'Trend >' or 'trend <' means that the confidence intervals overlap but only marginally (see the table in Appendix 9), and there is a clear trend towards one of the scales being more/less sensitive or specific. Discrepant results are given in bold.

CPSS: Cincinnati Prehospital Stroke Scale; FAST: Face Arm Speech Time; LAPSS: Los Angeles Prehospital Stroke Scale; MASS: Melbourne Ambulance Stroke Scale; MedPACS: Medic Prehospital Assessment for Code Stroke; ROSIER: Recognition of Stroke in the Emergency Room.

CONTRIBUTIONS OF AUTHORS

GW and SY formulated the idea for the review, registered the review title and developed the basis of the protocol.

GW wrote the first draft of the protocol with contributions from SW, ZZ and NH.

SY, ZZ, NH and JF reviewed the protocol.

All review authors participated in the selection of studies, data extraction, and methodologic quality assessment.

SY translated from Chinese.

ZZ and NH conducted the statistical analysis and all authors participated in the interpretation of results.

GW and ZZ wrote the first draft of the review, which was then reviewed by SY, NH, JF.

The final presubmission draft of the review was completed by ZZ and approved by GW, SY, NH, JF.

DECLARATIONS OF INTEREST

GW: none.

SY: none.

ZZ: none.

NH: none.

JF: none.

SOURCES OF SUPPORT

Internal sources

- National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula (NIHR CLAHRC South West Peninsula), UK.

  Zhivko Zhelev's contribution to this research was supported by the NIHR Collaboration for Leadership in Applied Health Research and Care South West Peninsula (NIHR CLAHRC South West Peninsula). The views expressed are those of the author(s) and not necessarily those of the National Health Service, the NIHR or the Department of Health and Social Care.

External sources

- No sources of support supplied

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

- We included studies that compared multiple tests provided they met the rest of our inclusion criteria.
- In tailoring QUADAS-2, we added to the patient selection domain a signaling question regarding prospective/retrospective design. The reason for adding this question is explained in the Methods. Upon advice from the Editors, we also added the question "Were the index test results interpreted without knowledge of the results of the reference standard?" This latter question was initially removed from the checklist as all answers would be 'Yes' (the stroke scales are always used before the reference standard tests).
- The Cochrane information Specialists who conducted the electronic searches had no access to MEDION (www.mediondatabase.nl) and, therefore, we did not search this database. Our previous experience suggests that searching this database is unlikely to identify additional studies missed by the main database searches.
- In Appendix 9 and Appendix 10, we reported the results from Purrucker 2015. This study did not meet our inclusion criterion for use of the scales on actual patients and, therefore, was excluded from the main analysis. However, it provides valuable information on the comparative accuracy of multiple scales. Given the paucity of comparative data, we decided to report its results in appendices and include it in a secondary analysis.

INDEX TERMS

Medical Subject Headings (MeSH)

Ischemic Attack, Transient [*diagnosis]; Mass Screening; Randomized Controlled Trials as Topic; Severity of Illness Index; Stroke [*diagnosis]

MeSH check words

Humans