Systematic review and meta-analysis of pre-hospital diagnostic accuracy studies

Caitlin Wilson,[1] Clare Harley,[2] Stephanie Steels[2]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/emermed-2018-207588).

[1]Paramedic Emergency Service, North West Ambulance Service NHS Trust, Bolton, UK
[2]Faculty of Medicine and Health, University of Leeds, Leeds, UK

Correspondence to
Caitlin Wilson, North West Ambulance Service NHS Trust, Bolton, UK; caitlin.wilson@nwas.nhs.uk

999 EMS Research Forum Conference 2018

Received 26 February 2018
Revised 20 June 2018
Accepted 19 August 2018
Published Online First 14 September 2018

ABSTRACT

Introduction Paramedics are involved in examining, treating and diagnosing patients. The accuracy of these diagnoses is evaluated using diagnostic accuracy studies. We undertook a systematic review of published literature to provide an overview of how accurately paramedics diagnose patients compared with hospital doctors. A bivariate meta-analysis was incorporated to examine the range of diagnostic sensitivity and specificity.

Methods We searched MEDLINE, CINAHL, Embase, AMED and the Cochrane Database from 1946 to 7 May 2016 for studies where patients had been given a diagnosis by paramedics and hospital doctors. Keywords focused on study type ('diagnostic accuracy'), outcomes (sensitivity, specificity, likelihood ratio?, predictive value?) and setting (paramedic*, pre-hospital, ambulance, 'emergency service?', 'emergency medical service?', 'emergency technician?').

Results 2941 references were screened by title and/or abstract. Eleven studies encompassing 384 985 patients were included after full-text review. The types of diagnoses in one of the studies encompassed all possible diagnoses and in the other studies focused on sepsis, stroke and myocardial infarction. Sensitivity estimates ranged from 32% to 100% and specificity estimates from 14% to 100%. Eight of the studies were deemed to have a low risk of bias and were incorporated into a meta-analysis which showed a pooled sensitivity of 0.74 (0.62 to 0.82) and a pooled specificity of 0.94 (0.87 to 0.97).

Discussion Current published research suggests that diagnoses made by paramedics have high sensitivity and even higher specificity. However, the paucity and varying quality of studies indicates that further prehospital diagnostic accuracy studies are warranted especially in the field of non-life-threatening conditions.

PROSPERO registration number CRD42016039306.

INTRODUCTION

Paramedics routinely obtain patient histories, take basic observations, perform physical examinations and use diagnostic tools such as ECGs or blood glucose tests. Although there is an ongoing debate in clinical practice and grey literature whether paramedics formally diagnose patients, it is apparent that they use the obtained information to create an impression or field diagnosis which forms the basis for any treatment and transport recommendations.[1]

Accurate diagnosis is important in patients with life-threatening conditions such as myocardial infarction (MI), stroke or sepsis where paramedics may bypass the nearest hospital in favour of a more specialised department, but is also relevant for non-life-threatening conditions, especially in light of ED overcrowding and the increase in see-and-treat, alternative transport and alternative destination pathways.[2,3] Alternative prehospital pathways in the UK, Australia, Canada and USA include paramedics directly referring patients to community services; therefore, correct diagnosis especially of low-acuity clinical presentations is vital to ensure that referrals are appropriate and safe for patients.[4,5]

The accuracy of paramedics' diagnoses can be evaluated through diagnostic accuracy studies. A systematic review of prehospital diagnostic accuracy was deemed necessary as no systematic review in this area has been published to date.

The aim of this systematic review was to evaluate the diagnostic accuracy of paramedics' prehospital diagnosis of patients presenting to an emergency ambulance service compared with the in-hospital diagnosis given by doctors. This study was not disease-specific but looked at prehospital diagnostic accuracy across a range of diagnoses.

METHODS

A systematic review and meta-analysis of the literature was undertaken and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. Prior to commencing this review, a study protocol was developed and registered with PROSPERO.

Information sources

Studies were identified by searching electronic databases, reference checking, citation searching and handsearching the *Journal of Paramedic Practice*, the *Journal of Emergency Primary Health Care* and the *Australasian Journal of Paramedicine*. Searches were performed on MEDLINE (Ovid, 1946–April week 3 2016), CINAHL (EBSCO, 1960–7 May 2016), Embase (Ovid, 1947–7 May 2016), AMED (Ovid, 1985–May 2016) and the Cochrane Database of Systematic Reviews (Cochrane Library, 2005–7 May 2016).

Search strategy

Databases were searched using a wide variety of search terms: 'sensitivity', 'specificity', 'diagnostic accuracy', 'likelihood ratio?' and 'predictive value?' (online supplementary appendix 1). The selected search terms avoided methodological filters, instead using text words and subject headings to describe the review's index test in line with recommendations by de Vet *et al*.[6,7] Search terms describing a target condition were not added due to aiming to retrieve all prehospital diagnostic accuracy studies regardless of target condition.[6] The relevant setting was ensured using the search terms 'paramedic*', 'pre-hospital', 'ambulance', 'emergency service?', 'emergency medical service?' and 'emergency technician?'. The search was limited to languages using the phonological writing system due to a lack of funding for logographical writing.

Study selection

Studies were assessed for eligibility against the following criteria:

- ▶ Population: patients of any age presenting to any emergency ambulance service worldwide.
- ▶ Intervention: prehospital diagnosis given by paramedics or emergency medical technicians (EMTs).
- ▶ Comparator/reference standard: hospital diagnosis given by doctors.
- ▶ Outcomes: sensitivity, specificity, predictive values, likelihood ratios (LRs) (or data to calculate these).
- ▶ Study design: excluded simulated studies, cost–utility analyses, case studies, comments and letters.

Titles and/or abstracts of studies were screened by CW in order to exclude records according to eligibility criteria. Duplicate studies were excluded using the reference management software EndNote V.X7 (Clarivate Analytics). The full text of potentially eligible studies was retrieved and assessed for eligibility by the author.

Data extraction

To extract data from included studies, an amended version of the Joanna Briggs Institute's standardised form was decided on in advance.[8] Information was obtained from published articles and supplementary online information, although authors were contacted for further information if reported methods seemed unclear or results contained apparent errors.

Risk of bias

The risk of bias of included studies was assessed by CW using the validated Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.[9] The tool provided signalling questions for four key areas: patient selection, index test, reference standard and flow/timing which guided CW in assigning high/unclear/low risks of bias scores to each study. Studies were included in the quantitative evidence synthesis if the first and last domain were judged to have a low or unclear risk of bias because index and reference tests were expected to be similar following the study selection process.

Outcome measures

The primary outcome measures were sensitivity and specificity of paramedics' diagnoses. Sensitivity represented the probability that patients diagnosed with a certain condition in hospital were also diagnosed with this condition by paramedics.[10–12] Specificity described the probability that patients for whom a particular diagnosis was excluded in hospital was also excluded by paramedics.[10–12] Estimates of individual studies' sensitivity and specificity values were plotted in a paired forest plot with 95% CIs using RevMan V.5.3 (The Nordic Cochrane Centre).

Secondary outcome measures were positive predictive values (PPVs) and negative predictive values (NPVs), as well as positive and negative LRs (LR+, LR−). PPV was the proportion of patients with a prehospital diagnosis of a particular condition, who were also diagnosed with this condition in hospital; whereby, NPV was the proportion of patients where a certain condition was excluded prehospitally, who also had this condition excluded in hospital.[12 13] LR+ and LR− provided information on how likely patients with a certain hospital diagnosis were to have been given a positive/negative prehospital diagnosis than patients without this condition.[14]

The tertiary outcome measure was inter-rater reliability as measured by kappa coefficients. The kappa statistic was chosen in preference over per cent agreement because it seeks to measure agreement between clinicians beyond that which may be obtained by chance.[15] Interpretation of reported kappa coefficients was conducted using a commonly cited scale by Landis and Koch,[16] as well as a more conservative scale by McHugh[17] which emphasised that kappa values falling short of the 1.0 value symbolising perfect agreement indicated corresponding disagreement.

Evidence synthesis

Pooled estimates for the primary outcome measures were obtained by fitting the bivariate random-effects model using the 'metandi' command in Stata V.13 (StataCorp).[18] The bivariate random-effects model was chosen because it takes into consideration any potential negative correlation between sensitivity and specificity, while accounting for differences between studies, known as heterogeneity.[19 20]

Heterogeneity between studies was investigated visually using a summary receiver operating characteristic plot to evaluate the scatter of points and the prediction ellipse, and statistically in the form of meta-regression.[19] The meta-regression incorporated two planned subgroup analyses to compare whether the primary outcomes of sensitivity and specificity varied by methodological quality or diagnosis type. On a post-hoc basis, further subgroups (ie, study design, country and prehospital clinician qualification) were also investigated. Due to the small number of studies within subgroups, the bivariate random-effects model could not be fitted, so a univariate random-effects model was fitted using the Stata command 'metan'.[21–23] Throughout, $p<0.05$ were deemed to indicate statistical significance.

No formal assessments of publication bias using funnel plots or regression tests were conducted because their effectiveness with diagnostic accuracy studies has not been established.[19 24]

RESULTS

The electronic literature search provided a total of 2936 records, and five further records were identified through additional sources (figure 1). Of these, 2918 were discarded after screening and full-text review which resulted in a total of 23 relevant references as per the original exclusion criteria.

Subsequently, one full-text article using logographical writing was excluded due to a lack of funding for translation and 11 abstracts of conference presentations were excluded due to providing insufficient information to facilitate in-depth quality appraisal which resulted in a total of 11 individual studies.

Study characteristics

Table 1 shows the characteristics of the 11 studies selected for the review which incorporated 14 datasets and were all published in English. Sample sizes varied between studies due to study periods differing from 1[25] to 36 months[26] and study populations consisting of a selection of patients suspected to have a certain condition[25 27–33] or a consecutive cohort of all patients seen.[26 34 35]

In total, the included studies involved 384 985 participants and were either adults or the age range was not stated. The main eligibility criterion for patients was emergency ambulance transport to a defined hospital but additional criteria varied

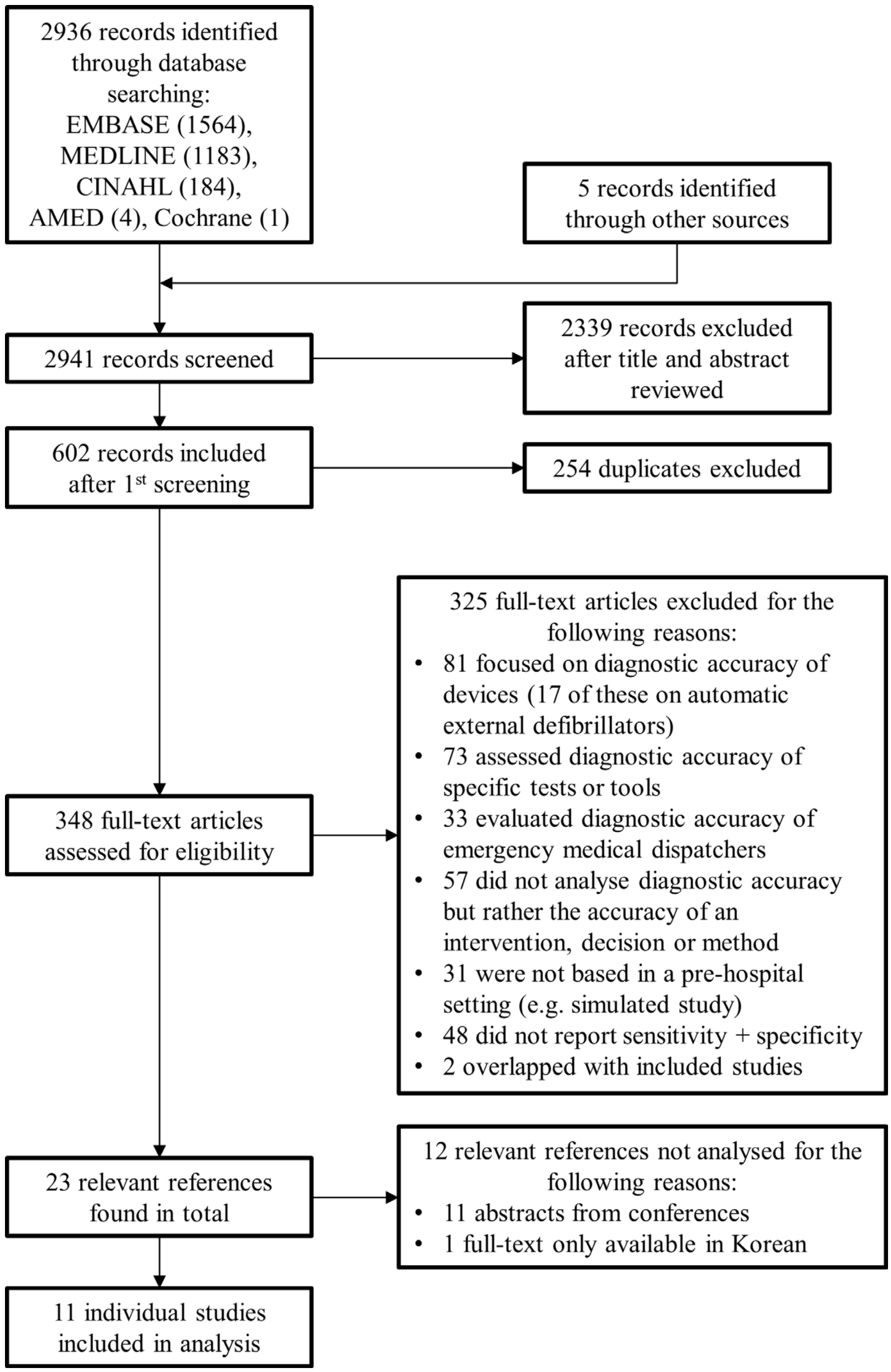

Figure 1 Literature search flow chart.

depending on the study population and type of diagnosis (eg, chest pain,[31 33] suspected sepsis[25] or dispatch code difficulty in breathing).[30] Studies reporting demographic details described a male prevalence between 41.1%[35] and 71%[28] and a mean age between 47.6[34] and 69.8[25] years.

The reference standard in most studies was ED or discharge diagnosis made by an in-hospital physician, although two studies[29 34] used diagnoses recorded on local databases or registries. The index test in all studies was diagnosis by a paramedic or a combination of paramedic/EMT. The types of diagnoses in the study by Ackerman and Waldron[30] encompassed all possible diagnoses and in other studies focused on sepsis, stroke or MI.

All diagnostic accuracy studies are cross-sectional and descriptive in nature; therefore, classifications by Rutjes *et al*[36] were used to further define study designs: single-gate design to describe a consecutive series comparable with a cohort study and two-gate design to describe an approach similar to that of a case–control study.

Quality assessment of included studies

Overall, the included studies were considered to be of high quality following assessment using QUADAS-2 (online supplementary appendix 2). However, three studies[25 28 29] were judged as having a high risk of bias due to employing two-gate designs which are associated with overestimations of sensitivity and specificity.[37] This was illustrated in the study by McClelland and Jones[25] which reported a spurious specificity estimate based on

Table 1 Summary of included prehospital diagnostic accuracy studies

Study	Country	Sample size	Age range	Diagnosis	Index test	Reference standard	Study design
Ackerman and Waldron[30]	USA	244	N/A	Any	Paramedic	Hospital physician (ED)	Retrospective single-gate
Aufderheide *et al*[33]	USA	151	Adult	MI	Paramedic	Hospital physician (discharge)	Prospective single-gate
Brandler *et al*[34]	USA	72 984	≥18 years	Stroke	Paramedic or EMT	Local database (discharge)	Retrospective single-gate
Bray *et al*[29]	Australia	858	N/A	Stroke	Paramedic	Hospital registry (discharge)	Retrospective two-gate
Ducas *et al*[28]	Canada	703	18–85 years	MI	Paramedic	Hospital physician (ED)	Prospective two-gate
Feldman *et al*[32]	USA	151	20–92 years	MI	Paramedic	Hospital physician (ED)	Prospective single-gate
Govindarajan *et al*[26]	USA	308 359	≥18 years	Stroke	Paramedic or EMT	Hospital physician (ED or discharge)	Retrospective single-gate
Green *et al*[27]	Canada	629	≥16 years	Sepsis	Paramedic	Hospital physician (ED)	Prospective single-gate
Le May *et al*[31]	Canada	411	N/A	MI	Paramedic	Hospital physician (ED)	Prospective single-gate
McClelland and Jones[25]	UK	49	>16 years	Sepsis	Paramedic or EMT	Hospital physician (ED or discharge)	Retrospective two-gate
Wojner *et al*[35]	USA	446	N/A	Stroke	Paramedic	Hospital physician (discharge)	Prospective single-gate

NA, not available; ED, emergency department; MI, myocardial infarction; EMT, emergency medical technician.

an unplanned enrolment of a true-negative patient as confirmed through personal email communication with G McClelland on 21 June 2016.

Two further studies were judged to have an unclear risk of bias as it could not be determined whether the high proportion of excluded patients (327/956 in Green *et al*[27]; 556/967 in Le May *et al*[31]) differed systematically from those who were analysed and therefore potentially introduced bias.[9]

Results of individual studies

The paired forest plot in figure 2 illustrates individual studies' sensitivity and specificity. Sensitivity estimates ranged from 32% to 100% and specificity estimates from 14% to 100%. This means that paramedics' ability to correctly identify or exclude a certain diagnosis varied considerably between the studies.

Estimated PPVs ranged from 23% to 84% and NPVs from 4% to 100% (online supplementary appendix 3). This shows that there was considerable variation between the studies regarding the proportion of patients who were given a particular prehospital diagnosis and the corresponding hospital diagnosis. However, there was even greater variation in the proportion of patients who were ruled out from having a certain condition in the prehospital and in-hospital setting.

Estimated LR+ ranged from 0.5 to 168 and LR− from 0.05 to 4. Most studies found that patients with a particular hospital diagnosis were several times more likely than patients with alternative hospital diagnoses to have also been given this particular diagnosis prehospitally and much less likely to have been given an alternative prehospital diagnosis. Although, the magnitude of LR+ varied considerably, only one study[25] reported that patients with a hospital diagnosis of sepsis were less likely than patients without a hospital diagnosis of sepsis to have received a prehospital diagnosis of sepsis (LR+ 0.5, 0.32 to 0.79) and were more likely to have not received a prehospital diagnosis of sepsis (LR− 4, 0.64 to 25).

Diagnosis agreement between paramedics and hospital doctors was presented as per cent agreement in three studies: 78.2%,[27] 78.24%[28] and 94%.[32] Two studies[30 32] reported kappa coefficients ($\kappa_{Ackerman}=0.71$, $\kappa_{Feldman}=0.73$) which were interpreted as indicating substantial agreement.[16] However, following guidance by McHugh,[17] these values were interpreted as indicating moderate agreement based on only 50% and 53% of the studies' data being considered reliable.

Evidence synthesis

The preplanned meta-analysis revealed pooled estimates of sensitivity 0.79 (0.65 to 0.88) and specificity 0.91 (0.80 to 0.96), as illustrated in figure 3. The preplanned exclusion of the three studies assessed to have a high risk of bias, resulted in pooled estimates of sensitivity 0.74 (0.62 to 0.82) and specificity 0.94

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ackerman 2006 cardiac	95	18	15	116	0.86 [0.79, 0.92]	0.87 [0.80, 0.92]
Ackerman 2006 other	64	15	14	151	0.82 [0.72, 0.90]	0.91 [0.86, 0.95]
Ackerman 2006 respiratory	40	12	16	176	0.71 [0.58, 0.83]	0.94 [0.89, 0.97]
Aufderheide 1990 angina	38	33	23	57	0.62 [0.49, 0.74]	0.63 [0.53, 0.73]
Aufderheide 1990 MI	18	37	6	90	0.75 [0.53, 0.90]	0.71 [0.62, 0.79]
Brandler 2015	468	268	282	71699	0.62 [0.59, 0.66]	1.00 [1.00, 1.00]
Bray 2010	185	87	22	564	0.89 [0.84, 0.93]	0.87 [0.84, 0.89]
Ducas 2012	228	152	1	322	1.00 [0.98, 1.00]	0.68 [0.64, 0.72]
Feldman 2005	20	4	5	122	0.80 [0.59, 0.93]	0.97 [0.92, 0.99]
Govindarajan 2015	1223	4015	2564	300557	0.32 [0.31, 0.34]	0.99 [0.99, 0.99]
Green 2016	52	118	19	440	0.73 [0.61, 0.83]	0.79 [0.75, 0.82]
Le May 2006	60	13	3	335	0.95 [0.87, 0.99]	0.96 [0.94, 0.98]
McClelland 2015	18	6	24	1	0.43 [0.28, 0.59]	0.14 [0.00, 0.58]
Wojner 2003	62	209	32	10261	0.66 [0.55, 0.75]	0.98 [0.98, 0.98]

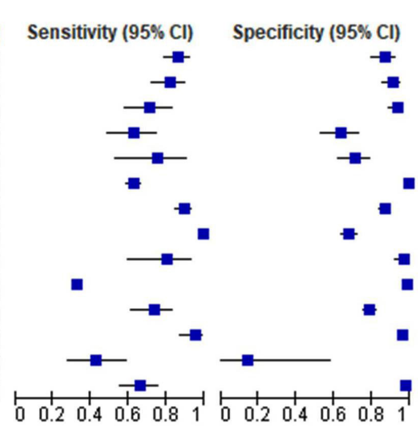

Figure 2 Forest plot displaying sensitivity and specificity results for prehospital diagnostic accuracy.

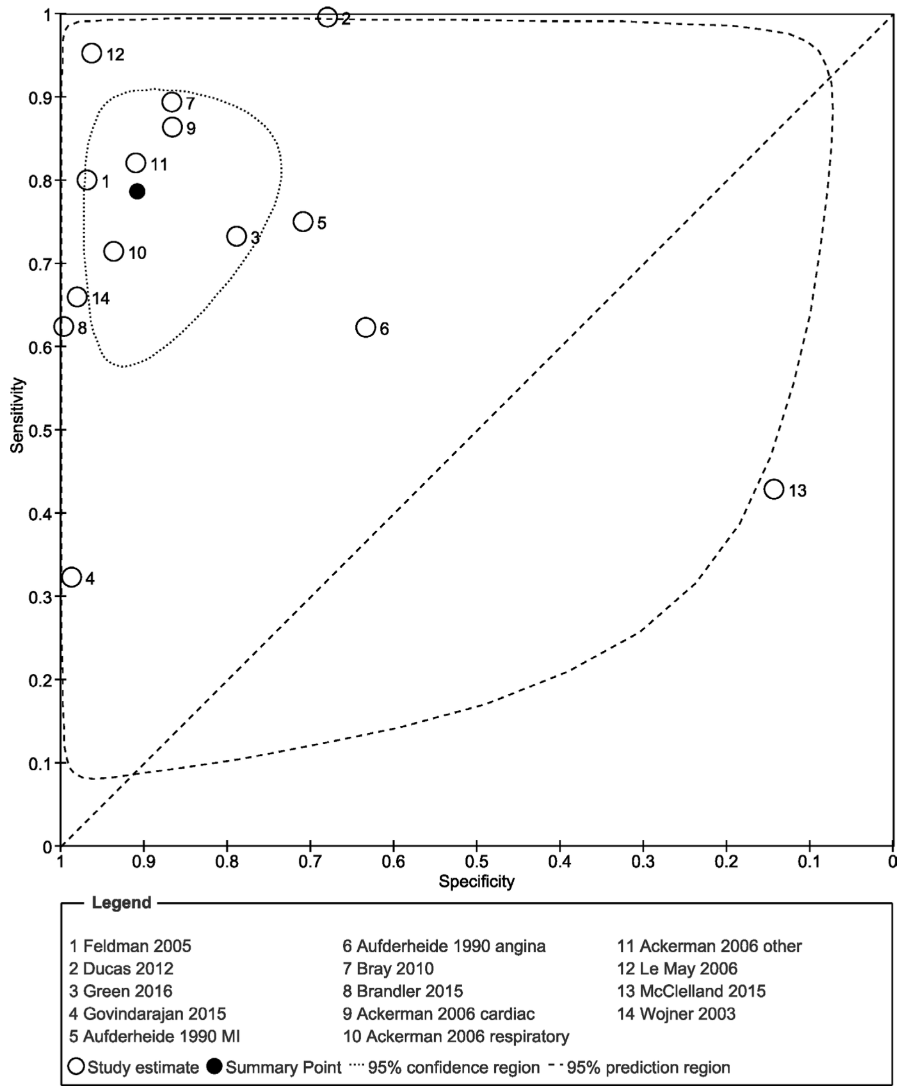

Figure 3 Summary receiver operating characteristic plot of all included studies.

(0.87 to 0.97) (figure 4). The exclusion of low-quality studies emphasised paramedics' high accuracy of identifying a particular diagnosis and slightly lower accuracy of excluding a particular diagnosis.

Statistical heterogeneity between studies was expected due to this being a review of diagnostic accuracy studies, and this was evident in the small degree of overlap between the studies' CIs in the forest plot (figure 2).[19 38] Heterogeneity was further indicated by the 95% prediction region in figure 4 being much larger than the 95% confidence region because the prediction region illustrates variation caused by between-study heterogeneity while the confidence region describes uncertainty associated within each study.[38] The prediction region was even larger when all studies were included (figure 3) than when low-quality ones were excluded (figure 4) which suggests that methodological quality was a source of heterogeneity.

The meta-regression (table 2) indicated statistically significant differences between studies varying in methodological quality, diagnosis type, study design, country and prehospital clinician qualification. However, I^2 statistics ranged from 81.7% to 99.6% for sensitivity and 96.0% to 99.9% for specificity which indicated that the variation between subgroups may be due to heterogeneity between individual studies. An exception was the any diagnosis type subgroup (specificity I^2=49.9%, sensitivity I^2=56.7%) which was to be expected due to all three datasets stemming from the same study.[30] The results of this meta-regression should be treated with caution as findings associated with small study numbers may be coincidental.[38]

An unplanned sensitivity analysis was conducted to investigate an inconsistency found in the flow chart provided by Brandler *et al*.[34] Personal email communication with ES Brandler on 23 June 2016 provided clarification that a simple typing error had resulted in 71 699 true negatives being inputted instead of 71 966. The subsequent sensitivity analysis after adding the 267 missing patients as true negatives showed no variation in pooled sensitivity (0.79, 0.65 to 0.88) but a slight reduction of 0.00002 regarding specificity (0.91, 0.80 to 0.96) which was not deemed clinically significant.

DISCUSSION

This is the first known systematic literature review summarising the evaluation of paramedics' diagnostic accuracy and included 11 international studies encompassing 384 985 patients. Of these, 8 studies representing 11 individual datasets were deemed to be of high methodological quality and were incorporated into a quantitative synthesis

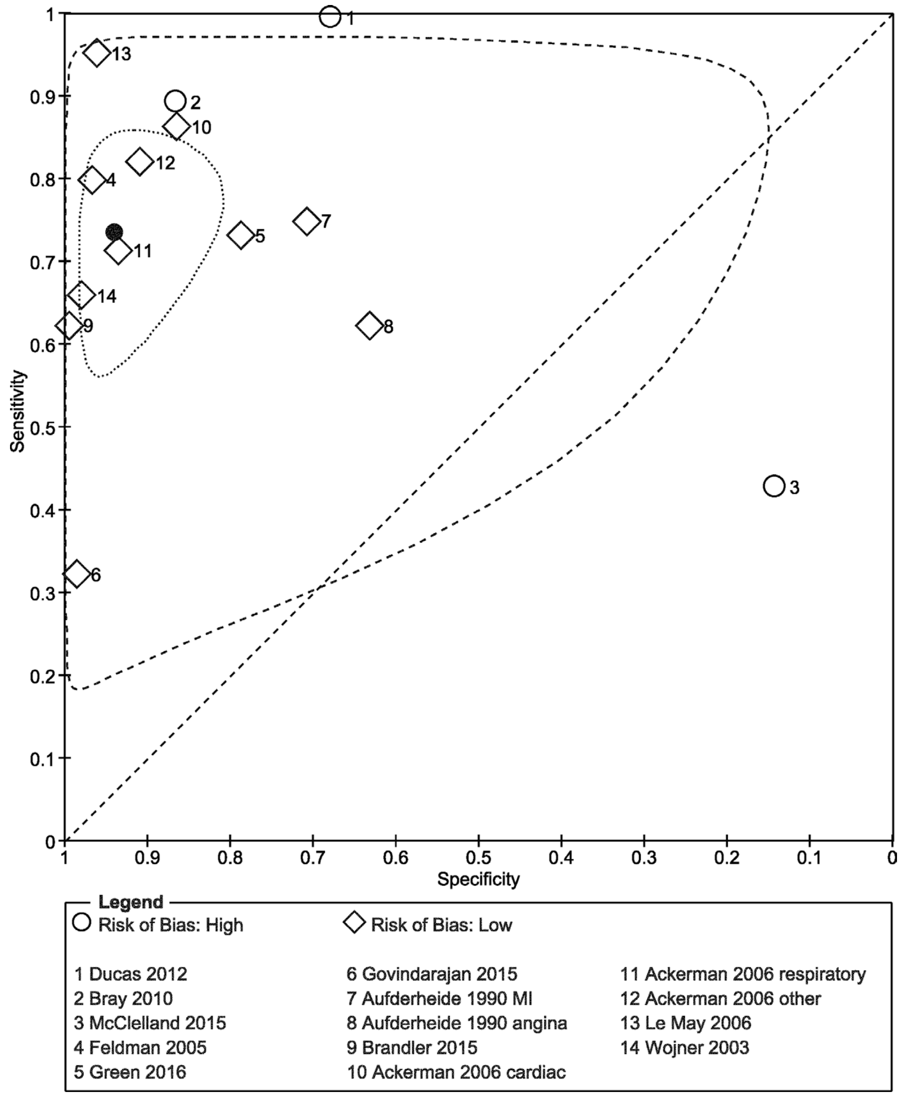

Figure 4 Summary receiver operating characteristic plot of studies with low or unclear risk of bias.

which showed a pooled sensitivity of 0.74 (0.62 to 0.82) and a pooled specificity of 0.94 (0.87 to 0.97). In numbers, this equates to paramedics correctly excluding a particular diagnosis in 94 out of 100 patients and correctly identifying a diagnosis in 74 out of 100 patients. This means that on average paramedics were very good at excluding patients from having a particular condition and sufficiently accurate at correctly identifying patients with a certain condition.

Table 2 Meta-regression results

Variable	Category	Datasets	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	Pooled sensitivity P values	Pooled specificity P values
Methodological quality	Low quality	3	0.80 (0.65 to 0.96)	0.68 (0.50 to 0.86)	<0.001	<0.001
	High quality	11	0.71 (0.55 to 0.88)	0.97 (0.96 to 0.97)		
Diagnosis type	Any	3	0.81 (0.74 to 0.89)	0.91 (0.87 to 0.95)	<0.001	<0.001
	MI	4	0.79 (0.63 to 0.95)	0.83 (0.72 to 0.94)		
	Stroke	3	0.54 (0.29 to 0.80)	0.99 (0.98 to 1.00)		
	Sepsis	1	0.74 (0.64 to 0.84)	0.79 (0.76 to 0.82)		
Study design	Prospective	6	0.76 (0.64 to 0.87)	0.86 (0.79 to 0.92)	<0.001	<0.001
	Retrospective	5	0.67 (0.43 to 0.90)	0.98 (0.97 to 0.99)		
Country	USA	9	0.69 (0.51 to 0.86)	0.98 (0.97 to 0.98)	<0.001	<0.001
	Canada	2	0.84 (0.64 to 1.00)	0.88 (0.70 to 1.00)		
Prehospital clinician qualification	Paramedic	9	0.77 (0.70 to 0.85)	0.87 (0.82 to 0.92)	<0.001	<0.001
	Paramedic/EMT	2	0.47 (0.18 to 0.77)	0.99 (0.98 to 1.00)		

MI, myocardial infarction; EMT, emergency medical technician.

High levels of heterogeneity were present between the studies which may have been due to variations in study quality, diagnosis type and prehospital clinician qualification but could be due to variations between individual studies. The pooled sensitivity value was altered minimally during a sensitivity analysis, although the displayed changes were very slight considering the investigated study encompassed the second largest sample included in the review.

This review indicated slightly higher values and less variability for specificity compared with sensitivity which suggests that paramedics were more reliable at excluding incorrect diagnoses than making correct diagnoses similar to responsibilities assigned to diagnostic tests as supposed to screening tests.[39] This is appropriate given the significant consequences associated with paramedics making a particular diagnosis, for example, thrombolytic treatment when diagnosing MI, bypassing local ED when diagnosing stroke or referrals to community services when diagnosing non-life-threatening conditions.

The findings of this review of paramedics' diagnostic accuracy can be applied to countries that employ the Anglo-American concept of pre-hospital care which involves highly trained paramedics responding to critically ill and injured patients, rather than the Franco-German model which uses emergency physicians.[40] Findings should only be applied to adults as it could not be confirmed that the included studies encompassed paediatric patients.

Limitations

The main limitation of this review was that the study population and type of diagnoses were not the same across the studies. Furthermore, the quality of studies varied with three studies being judged to have poor methodological quality and two studies reporting high levels of missing data which suggests potential bias. Lastly, not all studies reported CIs for their sensitivity and specificity values, although their calculations were possible due to sufficient information being provided to create 2×2 tables.

In terms of the review process, limitations were based on the review's limited scope due to this being a master's thesis as exemplified by studies being selected by a sole researcher and therefore being unable to report on inter-rater variability. Furthermore, abstracts of conference presentations were excluded, despite this possibly providing a more accurate picture due to pre-hospital research still being in its infancy and many studies not being published in full. In addition, the results were biased towards the English language due the review question targeting the Anglo-American prehospital system involving paramedics rather than physicians. Lastly, some eligible studies may have been missed despite the comprehensively conducted search due to poor indexing of diagnostic accuracy studies.[6]

CONCLUSION

The lack of literature that this review can be compared and contrasted with clearly illustrates that further pre-hospital diagnostic accuracy studies are necessary. This is emphasised by paramedics increasingly making significant clinical decisions regarding the treatment and transport of patients with life-threatening conditions such as MI, stroke or sepsis. Other areas within pre-hospital care which could benefit from diagnostic accuracy studies are non-life-threatening conditions because paramedics are increasingly referring patients to community services rather than transporting them to ED. Consequently, it is vital that paramedics correctly diagnose non-life-threatening conditions as this ensures that subsequent referrals to general practitioners or community nurses are appropriate and therefore safe for patients.

Twitter @999_Caitlin

Acknowledgements We would like to thank Darren C Greenwood, Division of Biostatistics, University of Leeds for his assistance with the preparation of the meta-analysis. Thank you also to Graham McClelland and Ethan Brandler for their prompt response to our queries and for giving us permission to cite personal communications.

Contributors CW conceived the study. All authors designed the study and established the search criteria. CW screened the references, performed the data extraction, quality assessment, synthesis and meta-analysis. CW wrote the final manuscript. CH and SS reviewed and approved the final manuscript, as well as providing guidance throughout the project.

Funding Health Education England/National Institute for Health Research Integrated Clinical Academic (ICA) Programme Masters in Clinical Research Studentship Scheme MRES-2015-03-020.

Disclaimer The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Patient consent Not required.

Ethics approval This paper represents a systematic review of published work and separate ethical approval was therefore not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement This paper represents a systematic review of published work; the data included in the paper are available in the published works reviewed.

REFERENCES

1 Heuer JF, Gruschka D, Crozier TA, *et al*. Accuracy of prehospital diagnoses by emergency physicians: comparison with discharge diagnosis. *Eur J Emerg Med* 2012;19:292–6.
2 Cummins NM, Dixon M, Garavan C, *et al*. Can advanced paramedics in the field diagnose patients and predict hospital admission? *Emerg Med J* 2013;30:1043–7.
3 Association of Ambulance Chief Executives. Future national clinical priorities for ambulance services in England April 2014. https://aace.org.uk/wp-content/uploads/2014/05/Future-national-clinical-priorities-for-ambulance-services-in-England-FINAL-2.pdf (accessed 25th February 2018).
4 Ebben RHA, Vloet LCM, Speijers RF, *et al*. A patient-safety and professional perspective on non-conveyance in ambulance care: a systematic review. *Scand J Trauma Resusc Emerg Med* 2017;25:71.
5 Blodgett J, Robertson D, Ratcliffe D, *et al*. Creating a safety net for patients in crisis: paramedic perspectives towards a GP referral scheme. *Journal of Paramedic Practice* 2017;9:11–17.
6 de Vet H, Eisinga A, Riphagen I, *et al*; Chapter 7: searching for studies. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 04 [updated september 2008]: the cochrane collaboration*, 2008.
7 Whiting P, Westwood M, Beynon R, *et al*. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *J Clin Epidemiol* 2011;64:602–7.
8 Joanna Briggs Institute. The Joanna Briggs Institute Reviewers' Manual 2015: The systematic review of studies of diagnostic test accuracy 2015. http://www.joannabriggs.org/assets/docs/sumari/Reviewers-Manual_The-systematic-review-of-studies-of-diagnostic-test-accuracy.pdf (accessed 16th May 2016).
9 Whiting PF, Rutjes AW, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
10 Parikh R, Mathai A, Parikh S, *et al*. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol* 2008;56:45–50.
11 Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatr* 2007;96:338–41.
12 Spitalnic S. Test properties 1: sensitivity, specificity and predictive values. *Hospital Physician* 2004:27–31.
13 Altman DG, Bland JM. Statistics notes: diagnostic tests 2: predictive values. *BMJ* 1994;309:102.
14 Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;329:168–9.
15 Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–3.
16 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
17 McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* 2012;22:276–82.

18 Harbord RM, Whiting P. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *The Stata Journal* 2009;9:211–29.
19 Macaskill P, Gatsonis C, Deeks J, *et al*; Chapter 10: analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 10: the cochrane collaboration*, 2010.
20 Reitsma JB, Glas AS, Rutjes AW, *et al*. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
21 Takwoingi Y, Guo B, Riley RD, *et al*. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res* 2017;26:1896–911.
22 Harbord R. Meta-analysis of diagnostic test accuracy studies: metandi & midas 2010. http://www.stata.com/meeting/uk10/UKSUG10.Harbord.pdf (accessed 12th June 2016).
23 Harris RJ, Bradburn MJ, Deeks JJ, *et al*. metan: fixed- and random-effects meta-analysis. *The Stata Journal* 2008;8:3–28.
24 Tatsioni A, Zarin DA, Aronson N, *et al*. Challenges in systematic reviews of diagnostic technologies. *Ann Intern Med* 2005;142(12 Pt 2):1048–55.
25 McClelland G, Jones J. A pilot study exploring the accuracy of pre-hospital sepsis recognition in the North East Ambulance Service. *Journal of Paramedic Practice* 2015;7:459–65.
26 Govindarajan P, Friedman BT, Delgadillo JQ, *et al*. Race and sex disparities in prehospital recognition of acute stroke. *Acad Emerg Med* 2015;22:264–72.
27 Green RS, Travers AH, Cain E, *et al*. Paramedic recognition of sepsis in the prehospital setting: a prospective observational study. *Emerg Med Int* 2016;2016:1–5.
28 Ducas RA, Wassef AW, Jassal DS, *et al*. To transmit or not to transmit: how good are emergency medical personnel in detecting STEMI in patients with chest pain? *Can J Cardiol* 2012;28:432–7.
29 Bray JE, Coughlan K, Barger B, *et al*. Paramedic diagnosis of stroke: examining long-term use of the Melbourne Ambulance Stroke Screen (MASS) in the field. *Stroke* 2010;41:1363–6.
30 Ackerman R, Waldron RL. Difficulty breathing: agreement of paramedic and emergency physician diagnoses. *Prehosp Emerg Care* 2006;10:77–80.
31 Le May MR, Dionne R, Maloney J, *et al*. Diagnostic performance and potential clinical impact of advanced care paramedic interpretation of ST-segment elevation myocardial infarction in the field. *CJEM* 2006;8:401–7.
32 Feldman JA, Brinsfield K, Bernard S, *et al*. Real-time paramedic compared with blinded physician identification of ST-segment elevation myocardial infarction: results of an observational study. *Am J Emerg Med* 2005;23:443–8.
33 Aufderheide TP, Hendley GE, Thakur RK, *et al*. The diagnostic impact of prehospital 12-lead electrocardiography. *Ann Emerg Med* 1990;19:1280–7.
34 Brandler ES, Sharma M, McCullough F, *et al*. Prehospital stroke identification: factors associated with diagnostic accuracy. *J Stroke Cerebrovasc Dis* 2015;24:2161–6.
35 Wojner AW, Morgenstern L, Alexandrov AV, *et al*. Paramedic and emergency department care of stroke: baseline data from a citywide performance improvement study. *Am J Crit Care* 2003;12:411–7.
36 Rutjes AW, Reitsma JB, Vandenbroucke JP, *et al*. Case-control and two-gate designs in diagnostic accuracy studies. *Clin Chem* 2005;51:1335–41.
37 Whiting P, Rutjes AW, Reitsma JB, *et al*. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;140:189–202.
38 Bossuyt P, Davenport C, Deeks J, *et al*; Chapter 11: interpreting results and drawing conclusions. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 09: the cochrane collaboration*, 2013.
39 Gilbert R, Logan S, Moyer VA, *et al*. Assessing diagnostic and screening tests: Part 1. Concepts. *West J Med* 2001;174:405–9.
40 Dick WF. Anglo-American vs. Franco-German emergency medical services system. *Prehosp Disaster Med* 2003;18:29–37. discussion 35-7.