



This is a repository copy of *The diagnostic accuracy of pre-hospital assessment of acute respiratory failure*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/168942/>

Version: Accepted Version

---

**Article:**

Fuller, G.W. [orcid.org/0000-0001-8532-3500](https://orcid.org/0000-0001-8532-3500), Goodacre, S., Keating, S. et al. (9 more authors) (2020) The diagnostic accuracy of pre-hospital assessment of acute respiratory failure. *British Paramedic Journal*, 5 (3). pp. 15-22. ISSN 1478-4726

10.29045/14784726.2020.12.5.3.15

---

© 2020 The College of Paramedics. This is an author-produced version of a paper subsequently published in *British Paramedic Journal*. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

## **The Diagnostic Accuracy of Prehospital Assessment of Acute Respiratory Failure**

Dr Gordon W Fuller\*, Professor Steve Goodacre, Dr Samuel Keating, Ms Esther Herbert, Professor Gavin Perkins, Mr Matthew Ward, Mr Andy Rosser, Ms Imogen Gunson, Mr Joshua Miller, Mr Mike Bradburn, Professor Tim Harris, Professor Cindy Cooper

**Corresponding author\*:** Dr Gordon Ward Fuller

**Work address:** Centre for Urgent and Emergency Care Research, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

**Telephone:** +44 (0) 7968 280641

**E-mail:** [g.fuller@sheffield.ac.uk](mailto:g.fuller@sheffield.ac.uk)

**ORCHID:** 0000-0001-8532-3500

Professor Steve Goodacre, Centre for Urgent and Emergency Care Research, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Email: [s.goodacre@sheffield.ac.uk](mailto:s.goodacre@sheffield.ac.uk)

Dr Sam Keating, Clinical Trials and Research Unit, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Email: [s.keating@sheffield.ac.uk](mailto:s.keating@sheffield.ac.uk)

Ms Esther Herbert, Clinical Trials and Research Unit, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Email: [e.herbert@sheffield.ac.uk](mailto:e.herbert@sheffield.ac.uk)

Professor Gavin Perkins, Warwick Clinical Trials Unit, University of Warwick, Coventry, CV4 7AL

Email: [G.D.Perkins@warwick.ac.uk](mailto:G.D.Perkins@warwick.ac.uk)

Mr Matthew Ward, West Midlands Ambulance Service, Trust Headquarters, Millennium Point, Waterfront Business Park, Waterfront Way, Brierley Hill, West Midlands, DY5 1LX

Email: matthew.ward@wmas.nhs.uk

Mr Andy Rosser, West Midlands Ambulance Service, Trust Headquarters, Millennium Point, Waterfront Business Park, Waterfront Way, Brierley Hill, West Midlands, DY5 1LX

Email: andy.rosser@wmas.nhs.uk

Ms Imogen Gunson, West Midlands Ambulance Service, Trust Headquarters, Millennium Point, Waterfront Business Park, Waterfront Way, Brierley Hill, West Midlands, DY5 1LX

Email: imogen.gunson@wmas.nhs.uk

Mr Joshua Miller, West Midlands Ambulance Service, Trust Headquarters, Millennium Point, Waterfront Business Park, Waterfront Way, Brierley Hill, West Midlands, DY5 1LX

Email: joshua.miller@wmas.nhs.uk

Mr Mike Bradburn, Clinical Trials and Research Unit, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Email: m.bradburn@sheffield.ac.uk

Professor Tim Harris, Blizard Institute, Barts and The London School of Medicine and Dentistry  
4 Newark Street, London, E1 2AT

Email: tim.harris@bartshealth.nhs.uk

Professor Cindy Cooper, Clinical Trials and Research Unit, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Email: C.L.Cooper@sheffield.ac.uk

**Word count:** 2,934

**Tables:** 3

**Figures:** 1

**Key words:** diagnostic accuracy, acute respiratory failure, emergency medical services,  
sensitivity, specificity

## **ABSTRACT**

### **Introduction**

Acute respiratory failure (ARF) is a common medical emergency. Prehospital management includes controlled oxygen therapy, supplemented by specific management options directed at the underlying disease. The aim of the current study was to characterise the accuracy of paramedic diagnostic assessment in acute respiratory failure.

### **Methods**

A nested diagnostic accuracy and agreement study comparing prehospital clinical impression to the final hospital discharge diagnosis, was conducted as part of the ACUTE (Ambulance CPAP: Use, Treatment effect and economics) trial. Adults with suspected ARF were recruited from the UK West Midlands Ambulance Service. The prehospital clinical impression of the recruiting ambulance service clinician was prospectively recorded and compared to the final hospital diagnosis at 30 days. Agreement between prehospital and hospital diagnostic assessments was evaluated using raw agreement and Gwets AC1 coefficient.

### **Results**

77 participants were included. Chronic obstructive pulmonary disease (32.9%) and lower respiratory tract infection (32.9%) were the most frequently suspected primary prehospital diagnoses for ARF, with secondary contributory conditions recorded in 36 patients (46.8%). There was moderate agreement between the primary prehospital and hospital diagnoses, with raw agreement of 58.5% and a Gwets AC1 coefficient of 0.56 (95% CI 0.43 to 0.69). In 5 cases a non-respiratory final diagnosis was present, including; myocardial infarction, ruptured abdominal aortic aneurysm, liver failure, and sepsis.

### **Conclusions**

Prehospital assessment of ARF is challenging, with limited accuracy compared to the final hospital diagnosis. A syndromic approach, providing general supportive care, rather than a

specifically disease orientated treatment strategy is likely to be most appropriate for the prehospital environment.

## Introduction

Acute respiratory failure (ARF) is a common medical emergency which occurs when heart or lung disease result in inadequate blood oxygen levels and/or increased blood carbon dioxide levels (Greene & Peters, 1994). It is caused by a number of common cardiac or respiratory diseases, including heart failure, pneumonia, and exacerbations of chronic obstructive pulmonary disease (COPD) and asthma (Chapman, 1984). There are approximately 9,000 ARF cases in England per year, with a high 14% risk of death within 30 days (Pandor et al., 2015). ARF has substantial health services costs, with patients often requiring prolonged hospital stays, ventilatory support and critical care admissions (Ray et al., 2006). ARF was responsible for over three million National Health Service (NHS) bed days in England in 2014 (Department of Health, 2014). Accurate diagnosis and optimised clinical management of ARF therefore has the potential to improve both health outcomes and cost-effectiveness.

Current United Kingdom (UK) prehospital clinical practice guidelines recommend a standard management approach of oxygen therapy for the treatment of acute respiratory failure; supplemented by specific management options directed at the underlying disease (NICE, 2010; AACE, 2019; Ponikowski et al., 2016). Prehospital administration of continuous positive airways pressure (CPAP) has been promoted as an additional potentially beneficial treatment strategy in some cases of ARF (Goodacre et al., 2014). An accurate prehospital diagnosis may help paramedics tailor therapy to the underlying cause of ARF and improve outcomes. Misdiagnosis could lead to inappropriate treatment, and even harm, for example instigating CPAP in patients with a pneumothorax (Davidson et al., 2016; BTS, 2012).

However, clinical assessment in the prehospital environment is often challenging. Details of previous medical history are often unavailable, dyspnoeic patients may not be able to provide a history, the uncontrolled environment can hamper examination, resuscitation of unstable patients may need to be prioritised, and limited diagnostic tools are available. Furthermore, patients with ARF frequently suffer from multiple cardiorespiratory co-morbidities, or could have concurrent disease processes. There is limited data available investigating prehospital diagnosis of the dyspnoeic patient. The aim of the current study was to characterise the

performance of paramedic clinical assessment in ARF. Specific objectives were to calculate diagnostic accuracy and agreement between prehospital and hospital diagnoses.



## Methods

A nested, pre-planned, diagnostic accuracy and agreement study, comparing prehospital clinical impression to the final hospital discharge diagnosis, was conducted as part of the ACUTE (Ambulance CPAP: Use, Treatment effect and economics) pilot trial. Study conduct and reporting was performed in accordance with STARD and **GRRAS** guidelines for diagnostic accuracy and reliability studies (Bossuyt et al., 2015; J. Kottner et al., 2011).

### *Study population*

The ACUTE trial was an individual patient randomised controlled external pilot trial to determine whether a definitive pragmatic randomised controlled trial (RCT) comparing prehospital **CPAP** to standard oxygen therapy for acute respiratory failure was feasible, acceptable and cost-effective. The trial was pre-registered (ISRCTN12048261) and the protocol has been reported in detail previously (Fuller et al., 2018). Briefly, patients with suspected acute respiratory failure (ARF) were recruited from **four** ambulance hubs in the United Kingdom West Midlands Ambulance Service (WMAS) between August 2017 and July 2018. ARF was defined as respiratory distress with peripheral oxygen saturation below British Thoracic Society (BTS) target levels (88% for patients with COPD, or 94% for other conditions), despite supplemental oxygen (titrated low flow oxygen for COPD, or titrated high flow oxygen in other conditions). (O'Driscoll, Howard, Earis, & Mak, 2017) Eligibility criteria are presented in Table **One**. Patients were allocated to either prehospital CPAP (O<sub>2</sub> system) with supplemental oxygen or standard oxygen therapy using identical equipment boxes. ("O<sub>2</sub> CPAP unit," 2018) Feasibility outcomes included: incidence of recruited eligible patients (target 120); proportion recruited in error; adherence to the allocation schedule and treatment; and retention at 30 days. Effectiveness outcomes comprised survival at 30 days; **proportion undergoing endotracheal intubation**; admission to critical care; and length of hospital stay.

**Table 1. Eligibility criteria for the ACUTE trial**

<b>Inclusion criteria</b>	<b>Exclusion Criteria</b>
<p>Respiratory distress with peripheral oxygen saturations below:</p> <ul style="list-style-type: none"> <li>• 88% for patients with COPD*</li> <li>•94% for other conditions*</li> </ul> <p>*Despite supplemental oxygen (titrated low flow oxygen for COPD, or titrated high flow oxygen in other conditions)</p>	<ul style="list-style-type: none"> <li>•Hospital CPAP treatment available within 15 minutes of eligibility</li> <li>• &lt;18 years</li> <li>•Terminal illness</li> <li>•Pre-existing lack of capacity</li> <li>•Documented not for resuscitation status</li> <li>•Acutely incapacitated patients with known valid advanced directive declining non-invasive ventilation or participation in research</li> <li>•Oxygen alert card</li> <li>•Anticipated inability to apply CPAP (e.g. facial deformity)</li> <li>•Respiratory failure due to chest trauma</li> <li>•Contraindication to CPAP (suspected pneumothorax, respiratory arrest, epistaxis, vomiting, hypotension)</li> <li>•Previous enrolment in the ACUTE trial</li> <li>•Pregnancy</li> <li>•Patients unable to communicate with ambulance service clinicians</li> </ul>

*Data collection*

A patient recruitment data collection form, contained within each equipment box, was completed by recruiting ambulance service clinicians every time a patient was enrolled in the trial. This recorded trial specific information including the prehospital clinical impression. At 30 days, research paramedics reviewed hospital records (including case notes, information systems, and discharge letters), with patient consent, to collect details of clinical progress including the final medical diagnosis.

### *Index test and reference standard*

The index test under consideration was the trial paramedic's clinical impression recorded at the scene of incident. Both the most likely clinical diagnosis, and the presence of any contributing conditions, were recorded **prospectively by attending paramedics** as a **pre-specified six** category nominal variable, comprising: 'heart failure', 'asthma', 'lower respiratory tract infection'; 'chronic obstructive pulmonary disease'; 'pulmonary embolism', or 'other'. These categories were chosen based on the most common causes of ARF, and conditions benefiting from specific treatment strategies. Conditions specified in the free text 'other' option were coded *post hoc* by **two** ACUTE co-investigators with any disagreements resolved by discussion to achieve a consensus decision. **The reference standard was the final hospital diagnosis accounting for the presenting respiratory distress provided by the hospital clinical team. Similarly to the index test, both the primary diagnosis, and any contributory conditions were collected. These were determined retrospectively from the hospital case notes or discharge summary and recorded using the same nominal categories by two ACUTE co-investigators. Hospital clinicians had access to routine prehospital patient records, but not the trial case report form containing the index test classification.**

### *Statistical analysis*

The statistical analysis proceeded in **three** stages. Firstly, sample characteristics were described using summary statistics, cross tabulation and a mosaic plot. Secondly, agreement between prehospital and hospital diagnostic assessments was evaluated (K. Gwet, 2001; Jan Kottner & Streiner, 2011). Raw agreement was initially calculated as the proportion of cases with an identical prehospital and hospital diagnosis (K. L. Gwet, 2008; J. Kottner et al., 2011). To account for the possibility that some agreement might be expected due to chance the Gwets AC1 coefficient was also determined (K. L. Gwet, 2008). This statistic was chosen in preference to Cohen's Kappa statistic as it does not depend upon an assumption of independence between different ratings, is robust to marginal probabilities, and less affected by rating prevalence (N. Wongpakaran, T. Wongpakaran, D. Wedding, & K. L. Gwet, 2013). **Landis and Koch's benchmark values were chosen as the most established thresholds to interpret the magnitude of**

**agreement coefficients with:** 0–0.20 indicating slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1 almost perfect agreement (Landis & Koch, 1977). Agreement was calculated for the primary diagnoses alone; and for combined primary and secondary diagnoses, ignoring the precedence placed on each condition and counting any match. Thirdly, the prehospital primary clinical impressions (index tests) were compared to the final hospital diagnosis (reference standard) with sensitivity and specificity calculated for the most common diagnostic categories. All results were calculated with their 95% confidence intervals. Complete case analyses were conducted with missing or non-interpretable data highlighted where relevant. **As a pre-specified trial sub-study a power calculation was not performed, and confidence interval width indicates the precision of results.** Statistical analyses were carried out in R (R Foundation for Statistical Computing, Vienna, Austria) and AgreeStat 2011.3 (advanced Analytics, Gaithersburg, MD, USA).

#### *Ethics and funding*

If possible, verbal consent was obtained for enrolment in the ACUTE trial at the scene of incident, with subsequent written informed consent confirmed for further participation. Patients lacking capacity were enrolled according to a hierarchical consent process complying with English Mental Capacity Act 2005 (Johnston & Liddle, 2007). Ethical approval was confirmed with NHS Leeds East Research Ethics Committee (**31st October 2016, reference 16/YH/0406**). The University of Sheffield provided sponsorship and monitoring oversight of the project. Funding was provided by the National Institute for Health Research's HTA Programme (HTA Project: 15/08/40).

## Results

### *Study sample*

Over the trial recruitment period 77 participants were enrolled from 364 potentially eligible patients **by 41 individual ambulance service clinicians**. Slightly more participants were allocated to the CPAP intervention arm (42 cases), than to the standard oxygen control arm (35 cases). Included patients were predominantly older (median 71 years), male (62%) and with marked respiratory distress (median oxygen saturations 78.5%, respiratory rate 34 breaths/minute, and breathlessness score of 9/10). Patient characteristics of enrolled patients are summarised in Table **Two**. A valid prehospital primary diagnosis was available for 76/77 patients. In one case the primary clinical impression was recorded as 'other', but lacked interpretable information to assign an underlying aetiology for ARF. A final hospital primary diagnosis was available for 65 patients **who were included in the complete case agreement and diagnostic accuracy analyses (Table Two)**. Consent was declined for data collection in **nine** cases, clinical records were unavailable in **two** cases, and in **one** case there was no clear underlying diagnosis apparent in the notes.

**Table 2. Baseline characteristics of ACUTE trial participants**

Baseline variable	Descriptive statistic	All	Included	Excluded
		N = 77	N = 65	N = 12
Age	n	77	65	12
	Median (IQR)	71.00 (62.00, 77.00)	70.00 (62.00, 77.00)	73.50 (66.75, 86.00)
Sex	Male	48 (62.3%)	42 (64.6%)	6 (50.0%)
	Female	29 (37.7%)	23 (35.4%)	6 (50.0%)
Ancillary disease specific prehospital treatments delivered	Yes	61 (79.2%)	53 (81.5%)	8 (66.7%)
	No	16 (20.8%)	12 (18.5%)	4 (33.3%)
Clinician's assessment of patient's breathlessness at enrolment (VAS 0-10)	n	76	64	12
	Median (IQR)	9.00 (8.00, 10.00)	9.00 (8.00, 10.00)	8.00 (7.00, 10.00)
First systolic blood pressure (mmHg)	n	70	59	11
	Median (IQR)	134.50 (112.25, 152.00)	136.00 (109.50, 152.00)	127.00 (117.50, 135.50)
First Glasgow Coma Score	n	77	65	12
	Median (IQR)	15.00 (14.00, 15.00)	15.00 (14.00, 15.00)	15.00 (15.00, 15.00)
First oxygen saturations (%)**	n	76	64	12
	Median (IQR)	78.50 (74.75, 86.00)	78.50 (74.00, 87.00)	79.50 (75.00, 82.25)
First pulse rate (bpm)	n	75	63	12
	Median (IQR)	115.00 (100.00, 124.00)	115.00 (95.50, 124.50)	114.00 (102.25, 124.00)
First Respiratory rate (breaths/min)	n	77	65	12
	Median (IQR)	34.00 (28.00, 40.00)	34.00 (28.00, 40.00)	36.00 (27.00, 37.00)

IQR: Interquartile range; bpm: beats per minute; COPD: Chronic Obstructive Pulmonary Disease; LRTI: Lower Respiratory Tract Infection.

### *Prehospital diagnosis*

COPD (n=25/76, 32.9%) and LRTI (n=25/76, 32.9%) were the most commonly suspected primary prehospital diagnoses. In **six** cases (n=76, 7.9%) a non-respiratory primary diagnosis was recorded, comprising: ruptured abdominal aortic aneurysm (n=1), liver failure (n=1), sepsis (not specified further, n=2), and urinary tract infection (n=2). A secondary diagnosis was recorded for 36 patients (n=77, 46.8%); with a single contributory condition suspected in 29 patients (n=77, 37.7%), and **two** supplementary diagnoses made for **seven** patients (n=77, 9.1%). LRTI (n=9/77, 11.7%) and heart failure (n=10/77, 13.0%) were the most common concomitantly diagnosed conditions.

### *Hospital diagnosis*

The most common final diagnoses were COPD (21/65, 32.3%) and LRTI (n=28/65, 43.1%). In four cases a non-respiratory final diagnosis was given including: myocardial infarction, ruptured abdominal, liver failure and sepsis (not specified further). Secondary conditions accounting for ARF were diagnosed in 27 patients (n=65, 41.5%); with 1 additional condition recorded for 23 cases; **two** contributory diseases given for **three** cases; and **three** further supporting diagnoses for **one** case. The commonest secondary diagnoses were COPD (n=7/65, 10.8%) and heart failure (n=8/65, 12.3%). Notably **two** patients were diagnosed with a pneumothorax in hospital (one primary diagnosis, one secondary diagnosis, both requiring intercostal drains). Prehospital and final hospital diagnoses are summarised in Table **Three**.

**Table 3. Prehospital and hospital ARF diagnosis.**

		Total	Notes
Primary prehospital ARF diagnosis (n=76)*	COPD	25 (32.9%)	
	LRTI	25 (32.9%)	
	Heart failure	14 (18.4%)	
	Asthma	4 (5.3%)	
	Pulmonary fibrosis	2 (2.6%)	
	Other	6 (2.6%)	Sepsis (2); AAA (1); Liver failure (1); UTI (2)
Secondary contributory ARF prehospital diagnoses (n=77)**	Present	36 (59.7%)	
	COPD	7 (9.1%)	
	LRTI	9 (11.7%)	
	Heart failure	10 (13.0%)	
	Asthma	5 (6.5%)	
	Pulmonary fibrosis	1 (1.3)	
	Other	6 (7.8%)	PE (1); Sepsis (1); Myocardial infarction (1); Pericarditis (1); Guillain-barre syndrome (1); Overdose (1)
Primary final hospital ARF diagnosis (n=65)	COPD	21 (32.3%)	
	LRTI	28 (43.1%)	
	Heart failure	6 (9.2%)	
	Asthma	2 (3.1%)	
	Pulmonary fibrosis	1 (1.5%)	
	Other	7 (10.8%)	Sepsis (1); PE (1); AAA (1); Liver failure (1); Lung cancer (1); Myocardial Infarction (1); Pneumothorax (1)
Secondary contributory ARF final hospital diagnoses (n=65)**	Present	27 (41.5%)	
	COPD	7 (10.8%)	
	LRTI	4 (6.2%)	
	Heart failure	8 (12.3%)	
	Asthma	3 (4.6%)	
	Pulmonary fibrosis	1 (1.5%)	
Other	9 (3.8%)	Sepsis (2); Lung Cancer (2); Bronchiectasis (1); Pneumothorax (1); Morbid obesity (2); Anaemia (1)	

\*\*More than one secondary contributory diagnosis possible. COPD: Chronic Obstructive Pulmonary Diseases; PE: Pulmonary Embolism; LRTI: Lower Respiratory Tract Infection; AAA: Abdominal Aortic Aneurysm; UTI: Urinary Tract Infection.



### *Agreement*

There was limited reproducibility between the primary prehospital and hospital diagnoses, with raw agreement of 58.5% (n=38/65). However, if both primary and secondary diagnoses were considered together, counting any match and ignoring the precedence placed on each condition, there was higher raw agreement of 76.9% on at least one causative disease for ARF (n=50/65). Chance corrected agreement between prehospital and hospital primary diagnosis was moderate as demonstrated by a Gwets AC1 coefficient of 0.56 (95% CI 0.43 to 0.69). When both primary and secondary diagnoses were assessed together, there was substantial chance corrected agreement on at least one condition with a Gwets AC1 coefficient of 0.75 (95% CI 0.64-0.87). Agreement between prehospital and hospital diagnoses is presented in a mosaic plot in Figure [One](#) and is tabulated in the web appendix.

### *Diagnostic accuracy*

The performance of ambulance service clinicians' assessment was then investigated by calculating diagnostic accuracy metrics for the most prevalent conditions (COPD, LRTI, and heart failure). Other conditions were not evaluated due to low sample size with consequent imprecision and intractability. Whilst each condition was identified more correctly than not, all three were commonly missed as the primary diagnosis: the sensitivities for COPD, LRTI and heart failure were 71% (95% CI 48% to 89%), 54% (34% to 73%) and 67% (22% to 96%) respectively. The specificity was higher (COPD 84.1% (69.9% to 93.4%), LRTI 86.5% (71.2% to 95.5%), and heart failure 86.4% (75.0% to 94.0%)). When both primary and secondary diagnoses were assessed together diagnostic accuracy was improved. Considering the index test and reference standard to be positive if the condition was recorded in either the primary or secondary diagnosis gave sensitivities of: COPD 95.2%; LRTI 69.2%; Heart failure 85.7%, meaning all three conditions were typically identified even if not as the primary diagnosis. Specificities in this contingency were: COPD 84.1%; LRTI 92.3%; Heart failure 96.6%.

## Discussion

COPD (n=25/76, 32.9%) and LRTI (n=25/76, 32.9%) were the most frequently suspected primary prehospital diagnoses for ARF, with secondary contributory conditions recorded in 36 patients (n=77, 46.8%). There was moderate agreement between the primary prehospital and hospital diagnoses, with raw agreement of 58.5% (n=38/65) and a Gwets AC1 coefficient of 0.56 (95% CI 0.43 to 0.69). In **seven** cases a final diagnosis was present where CPAP would not be expected to be effective, or could be harmful, including; myocardial infarction, ruptured abdominal aortic aneurysm, liver failure, sepsis, and pneumothorax (n=7/65, 10.8%).

Respiratory distress with low oxygen saturations is common to many conditions, with symptoms and clinical signs shared between differential diagnoses, often making assessment challenging (Chapman, 1984; Delerme & Ray, 2008). It is therefore unsurprising that accuracy of the prehospital clinical impression was limited, that non-specific working diagnoses such as 'sepsis' were used, some non-cardiorespiratory conditions were diagnosed, and concurrent disease processes were suspected in the majority of cases. COPD and a LRTI were the most commonly diagnosed conditions, and distinction between these two entities is known to be difficult, **even in hospital with the benefit of time, access to testing and specialist review** (Finney et al., 2019).

Given that the most important treatment for ARF is provision of oxygen, and other treatment modalities currently available to **UK** paramedics (e.g. nebulizers) have few side effects, it could be argued that an exact prehospital diagnosis is unnecessary prior to definitive hospital care (Chapman, 1984). However, if CPAP or non-invasive ventilation is available, then it is important to recognise conditions representing relative or absolute contraindications (Hess, 2013).

Although low numbers of patients were studied, it is reassuring that all cases with a final diagnosis of asthma were recognised by paramedics, but potentially concerning that there were 2 patients with undetected pneumothorax.

This is the first study to investigate the diagnostic assessment of patients with ARF presenting to EMS. Previous literature has focused on either less unwell dyspnoeic patients, or examined specific diseases including COPD, asthma, or heart failure (Christie, Costa-Scorse, Nicholls,

Jones, & Howie, 2016; Williams, Finn, Celenza, Teng, & Jacobs, 2013; Williams et al., 2015). Although limited by retrospective chart review designs, this body of research has demonstrated similar findings to the current study. Christie et al. reported only moderate agreement between paramedic and hospital diagnosis in a New Zealand cohort, with many cases having no clearly documented working diagnosis (Christie et al., 2016). The sensitivity for prehospital heart failure, asthma and COPD diagnoses was only 29%, 66% and 39% respectively, in Australian EMS studies by Williams and colleagues (Williams et al., 2013; Williams et al., 2015). **More widely, a recent systematic review reported a pooled sensitivity of 0.74 (0.62 to 0.82) and a pooled specificity of 0.94 (0.87 to 0.97) for paramedic diagnosis of myocardial infarction, sepsis, stroke, or all diagnoses. (Wilson, Harley, & Steels, 2018).**

Participating WMAS ambulance hubs serve a mixed rural, semi-rural and urban population, and the sample of recruited ARF patients should ensure good external validity to similar EMS settings. However, clinical trial populations may not be fully representative of undifferentiated prehospital ARF patients after application of eligibility criteria and consent procedures. The ACUTE trial specifically excluded patients with pre-existing lack of capacity, or unable to communicate with trial paramedics; groups where prehospital diagnosis is likely to be even more challenging. Generalizability to areas where disease prevalence differs to the UK, or with alternative EMS models (e.g. physician rather than paramedic assessment), is less certain.

The prospective, preplanned, data collection, using a defined nominal categorisation for ARF, are strengths of this study. However, there are a number of limitations that could adversely affect the internal validity of the results. Firstly, there is the potential for reference standard misclassification as the final diagnosis was recorded from the hospital record or discharge letter, rather than determination through formal expert case review. Secondly, although comparing favourably to other published reproducibility studies, the sample size is relatively low, resulting in imprecise results consistent with either poor or moderate agreement. The sample size constraint also meant we did not attempt to model any clustering for differential effects of paramedics. These findings should therefore be considered as exploratory, requiring confirmation in larger studies. Thirdly, mainly due lack of consent, some reference standard

data was missing. Although this represented a relatively small number (<10%) of patients, with similar characteristics to included cases, selection bias is possible if excluded patients differed systematically from the study population. Finally, we pre-specified the relatively liberal Landis and Koch scale for benchmarking agreement coefficients. Although well established and widely used, this may overstate agreement compared to other benchmarks e.g. Fleiss' or McHugh's proposed scales (Fleiss, Levin, & Paik, 2003; McHugh, 2012).

In conclusion, prehospital assessment of ARF is challenging, with limited diagnostic accuracy compared to the final hospital diagnosis. **A syndromic approach, providing general supportive care, rather than a specifically disease orientated treatment strategy is likely to be most appropriate for ARF in the prehospital environment.**

## References

- AACE (2019). *UK ambulance services clinical practice guidelines: 2019*. Class Professional Publishing. London: Association of Ambulance Chief Executives.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., . . . Cohen, J. F. (2015). STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*, *351*, h5527.
- BTS (2012). National Respiratory Audit Programme Annual Report 2011/2012. London: British Thoracic Society**
- Chapman, G. W. (1984). Pathophysiology and Treatment of Acute Respiratory Failure. *Journal of the National Medical Association*, *76*(2), 201-206.
- Christie, A., Costa-Scorse, B., Nicholls, M., Jones, P., & Howie, G. (2016). Accuracy of working diagnosis by paramedics for patients presenting with dyspnoea. *Emergency Medicine Australasia*, *28*(5), 525-530.
- Davidson, C., Banham, S., Elliott, M., Kennedy, D., Gelder, C., Glossop, A., . . . Thomas, L. (2016). British Thoracic Society/Intensive Care Society Guideline for the ventilatory management of acute hypercapnic respiratory failure in adults. *BMJ open respiratory research*, *3*(1), e000133-e000133.
- Delorme, S., & Ray, P. (2008). Acute respiratory failure in the elderly: diagnosis and prognosis. *Age and Ageing*, *37*(3), 251-257.
- Department of Health. (2014). Hospital Episode Statistics. Retrieved 20th August, 2020, from <http://www.hesonline.nhs.uk>**
- Finney, L. J., Padmanaban, V., Todd, S., Ahmed, N., Elkin, Sarah L., & Mallia, P. (2019). Validity of the diagnosis of pneumonia in hospitalised patients with COPD. *ERJ Open Research*, *5*(2), 00031-02019.
- Fleiss, J. L., Levin, B. A., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed. / Joseph L. Fleiss, Bruce Levin, Myunghee Cho Paik. ed.). Hoboken, N.J. ; [Chichester]: Wiley-Interscience.
- Fuller, G. W., Goodacre, S., Keating, S., Perkins, G., Ward, M., Rosser, A., . . . Cooper, C. (2018). The ACUTE (Ambulance CPAP: Use, Treatment effect and economics) feasibility study: a pilot randomised controlled trial of prehospital CPAP for acute respiratory failure. *Pilot and feasibility studies*, *4*, 86-86.
- Goodacre, S., Stevens, J. W., Pandor, A., Poku, E., Ren, S., Cantrell, A., . . . Plaisance, P. (2014). Prehospital noninvasive ventilation for acute respiratory failure: systematic review, network meta-analysis, and individual patient data meta-analysis. *Academic Emergency Medicine*, *21*(9), 960-970.

- Greene, K. E., & Peters, J. I. (1994). Pathophysiology of acute respiratory failure. *Clinics in Chest Medicine*, 15(1), 1-12.
- Gwet, K. (2001). *Handbook of Inter-Rater Reliability: How to Estimate the Level of Agreement Between Two or Multiple Raters*. Gaithersburg, MD: STATAXIS Publishing Company.
- Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 61(Pt 1), 29-48.
- Hess, D. R. (2013). Noninvasive ventilation for acute respiratory failure. *Respiratory Care*, 58(6), 950-972.
- Johnston, C., & Liddle, J. (2007). The Mental Capacity Act 2005: a new framework for healthcare decision making. *Journal of Medical Ethics*, 33(2), 94-97.
- Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B. J., Hrobjartsson, A., Streiner, D. L. (2011). Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*, 64(1), 96-106.
- Kottner, J., & Streiner, D. L. (2011). The difference between reliability and agreement. *Journal of Clinical Epidemiology*, 64(6), 701-702.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica: Casopis Hrvatskoga Društva Medicinskih Biokemicara*, 22(3), 276-282.
- NICE (2010). **Chronic obstructive pulmonary disease in over 16s: diagnosis and management Clinical guideline [CG101]** London: National Institute for Health and Care Excellence.
- O'Driscoll, B. R., Howard, L. S., Earis, J., & Mak, V. (2017). BTS guideline for oxygen use in adults in healthcare and emergency settings. *Thorax*, 72(Suppl 1), ii1-ii90.
- O\_two CPAP unit. (2018). Retrieved 10th January 2020, from [http://otwo.com/emergency-cpap/o\\_two-single-use-cpap/](http://otwo.com/emergency-cpap/o_two-single-use-cpap/)
- Pandor, A., Thokala, P., Goodacre, S., Poku, E., Stevens, J. W., Ren, S., . . . Penn-Ashman, J. (2015). Prehospital non-invasive ventilation for acute respiratory failure: a systematic review and cost-effectiveness evaluation. *Health Technology Assessment*, 19(42), v-vi, 1-102.
- Ponikowski, P., Voors, A. A., Anker, S. D., Bueno, H., Cleland, J. G., Coats, A. J., . . . van der Meer, P. (2016). 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *European Journal of Heart Failure*, 18(8), 891-975.

- Ray, P., Birolleau, S., Lefort, Y., Becquemin, M. H., Beigelman, C., Isnard, R., . . . Boddaert, J. (2006). Acute respiratory failure in the elderly: etiology, emergency diagnosis and prognosis. *Critical Care (London, England)*, 10(3), R82.
- Williams, T. A., Finn, J., Celenza, A., Teng, T. H., & Jacobs, I. G. (2013). Paramedic identification of acute pulmonary edema in a metropolitan ambulance service. *Prehospital Emergency Care*, 17(3), 339-347.
- Williams, T. A., Finn, J., Fatovich, D., Perkins, G. D., Summers, Q., & Jacobs, I. (2015). Paramedic Differentiation of Asthma and COPD in the Prehospital Setting Is Difficult. *Prehospital Emergency Care*, 19(4), 535-543.
- Wilson, C., Harley, C., & Steels, S. (2018) Systematic review and meta-analysis of pre-hospital diagnostic accuracy studies. *Emergency Medicine Journal*, 35(12), 757-764.
- Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K.L. (2013) et al. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol*, 13, 61.

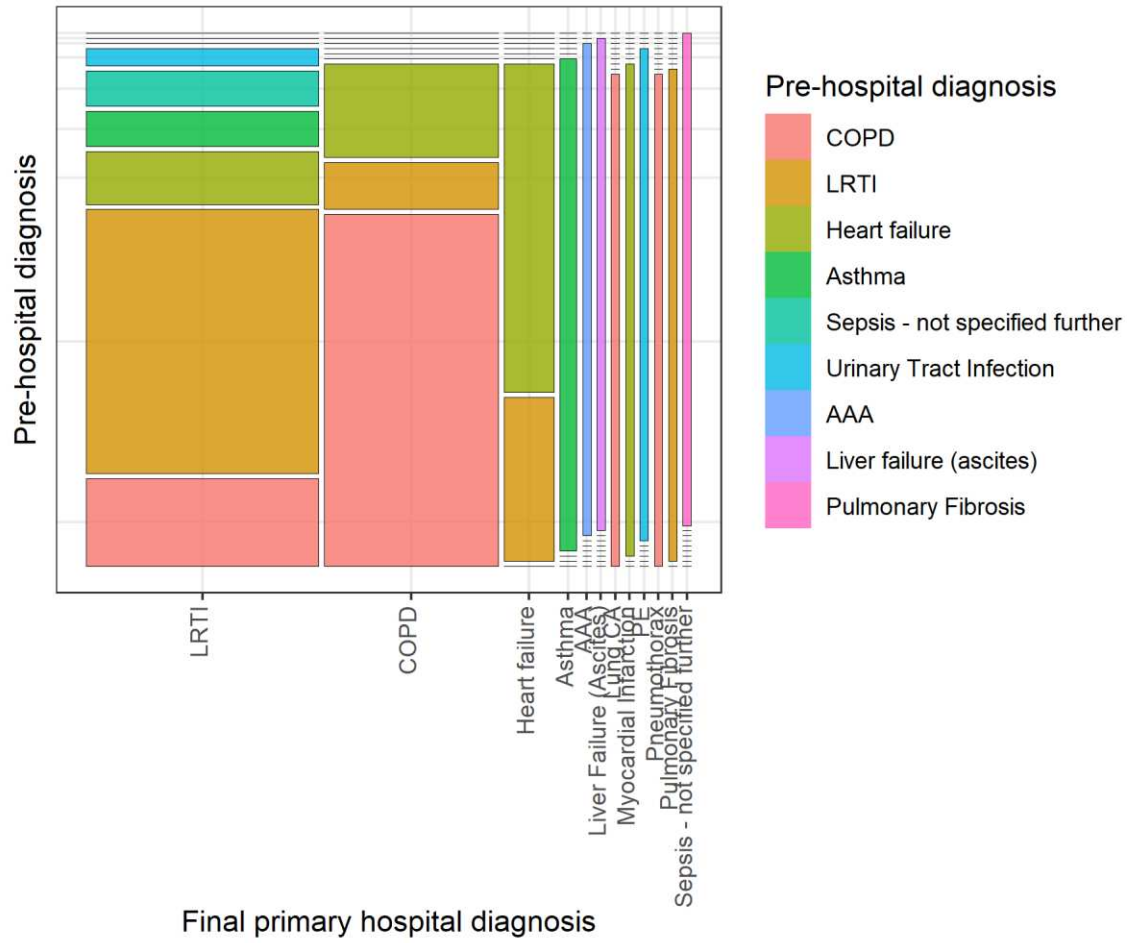


Figure 1. Mosaic plot demonstrating agreement between prehospital and hospital diagnoses



**Author contributions:** GF and SG conceived and designed the study. AR, MW, IG and JM collected data. All authors were involved in the analysis and the interpretation of data. GF drafted the report. All authors revised the work critically for important intellectual content and were involved in the final approval of the version to be published. All authors agree to be accountable for all aspects of the work.

**Conflicts of interest:** Professor Steve Goodacre is Deputy Director of the NIHR HTA Programme, Chair of the NIHR HTA Commissioning Board and member of the NIHR HTA Funding Strategy Group. Professor Gavin Perkins is a NIHR Senior Investigator and member of the Programme Grants for Applied Research Board. Professor Cindy Cooper is a member of the NIHR Clinical Trials Unit Standing Advisory Committee and of the UK Clinical Research Collaboration Registered Clinical Trials Unit Network Executive Group.

**Ethics:** Ethical approval was confirmed with NHS Leeds East Research Ethics Committee. The University of Sheffield provided sponsorship and monitoring oversight of the project.

**Funding:** Funding was provided by the National Institute for Health Research's HTA Programme (HTA Project: 15/08/40).