Accuracy of telephone triage for predicting adverse outcome in suspected COVID-19: An observational cohort study

Carl Marincowitz¹ NIHR Clinical Lecturer Emergency Medicine, PhD, MSc, MRCEM, MB BChir, BA (Hons)

Tony Stone¹ Senior Data Manager, BSc

Peter Bath²

Richard Campbell¹

Janette Turner¹

Madina Hussein²

Richard Pilbery³

Benjamin Thomas⁴

Laura Sutton⁴

Fiona Bell³ Head of Research, PhD, BSc (Hons)

Katie Biggs⁴

Frank Hopfgartner² Senior Lecturer in Data Science, PhD, FHEA

Suvodeep Mazumdar²

Jennifer Petrie⁴

Steve Goodacre¹

1. Centre for Urgent and Emergency Care Research (CURE), Health Services Research School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK, Fax: +44 (0)114 222 0749 Tel: (+44) (0)114 222 4345, Email: C.Marincowitz@Sheffield.ac.uk
2. Information School, University of Sheffield, Regent Court, 211 Portobello St, Sheffield S1 4DP, UK
3. Yorkshire Ambulance Service NHS Trust, Trust HQ, Springhill 2, Brindley Way, Wakefield, United Kingdom, WF2 0XQ
4. Clinical Trials Research Unit (CTRU), Health Services Research School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK, Fax: +44 (0)114 222 0749 Tel: (+44) (0)114 222 4345,

Abstract

Objective: To assess accuracy of telephone triage in identifying patients who need emergency care amongst those with suspected COVID-19 infection and identify factors which affect triage accuracy.

Design: Observational cohort study

Setting: Community telephone triage in the Yorkshire and Humber, Bassetlaw, North Lincolnshire and North East Lincolnshire region.

Participants: 40, 261 adults who contacted NHS 111 telephone triage services provided by Yorkshire Ambulance Service NHS Trust between the 18th March 2020 and 29th June 2020 with symptoms indicating possible COVID-19 infection were linked to Office for National Statistics death registration data, hospital and general practice electronic health care data collected by NHS Digital.

Outcome: Accuracy of triage disposition (self-care/non-urgent clinical assessment versus ambulance dispatch/urgent clinical assessment) was assessed in terms of death or need for organ support at 30, 7 and 3 days from first contact with the telephone triage service.

Results: Callers had a 3% (1, 200/40, 261) risk of adverse outcome. Telephone triage recommended self-care or non-urgent assessment for 60% (24, 335/40, 261), with a 1.3% (310/24, 335) risk of subsequent adverse outcome. Telephone triage had 74.2% sensitivity (95% CI: 71.6 to 76.6%) and 61.5% specificity (61% to 62%) for adverse outcomes at 30 days from first contact. Multivariable analysis suggested some co-morbidities (such as chronic respiratory disease) may be over-estimated as predictors of adverse outcome, while the association of diabetes with adverse outcome may be under-estimated. Repeat contact with the service appears to be an important under recognised predictor of adverse outcomes with both 2 contacts (OR 1.77 95% CI: 1.14 to 2.75) and 3 or more contacts (OR 4.02 95% CI: 1.68 to 9.65) associated with clinical deterioration when not provided with an ambulance or urgent clinical assessment.

Conclusion:

Patients advised to self-care or receive non-urgent clinical assessment had a small but non-negligible risk of serious clinical deterioration. The sensitivity and specificity of telephone triage was comparable to other tools used to triage patient acuity in emergency and urgent care. Repeat contact with telephone services needs recognition as an important predictor of subsequent adverse outcomes.

What is already known on this topic

- Telephone triage has been used to divert patients with suspected COVID-19 to self care or for non-urgent clinical assessments, and thereby help mitigate the risk of health services being overwhelmed by patients who require no specific treatment.
- Concerns have been raised that telephone triage may not be sufficiently accurate in identifying need for emergency care. However, no previous evaluation of accuracy of telephone triage in patients with suspected COVID-19 infection has been completed.

What this study adds

- Patients advised to self care or receive non-urgent clinical assessment had a small but non-negligible risk of deterioration and significant adverse outcomes.
- Telephone triage has comparable performance to methods used to triage patient acuity in other emergency and urgent care settings.
- Accuracy of triage may be improved by better recognition of multiple contact with services as a predictor of adverse outcomes.

Background

During the COVID-19 pandemic, there was a risk that hospitals could be overwhelmed by patients who did not need specific treatment. UK government pandemic planning predicted that, in the advent of an influenza or similar pandemic, there could be around 750,000 excess Emergency Department (ED) attendances in the UK.^{1,2} During the H1N1 outbreak in 2009, ED attendances in the USA increased at some centres by 50%.³ Increased ED attendances were largely for patients who did not require hospitalisation.⁴

To reduce this risk, from the 18th February 2020 onwards, NHS England advised patients with suspected infection to contact the NHS 111 service instead of attending health care providers.⁵ NHS 111 is a national, free to use 24 hour telephone triage service for urgent health problems. Initial triage is carried out by trained, non-clinical call advisors using the NHS Pathways clinical decision support software. The endpoint (disposition) is advice on what to do next, in terms of which service to access and the timeframe that this access should occur. If appropriate, the call can be passed onto a clinician (usually a nurse or paramedic) for further assessment and, depending on local arrangements, callers can speak to other specialist clinicians or appointments can be made with relevant services, including general practitioners.

In the first 6-months of the COVID-19 pandemic, ED attendances in the UK decreased by approximately 25%, probably in part due to displacement of care.⁶ Patients who did attend the ED with suspected COVID-19 infection were high acuity with a mortality rate of 15.5%, with lower acuity patients likely being managed via NHS 111.⁷ Indeed, there were almost 3 million NHS 111 calls made across England in March 2020; a record number and double the number of the previous year.⁸ To cope with the increase in call volume, a specific telephone triage pathway for patients with suspected COVID-19 infection was introduced in early February 2020, which underwent rapid updates as the pandemic progressed. Local NHS 111 services used interim triage methods while awaiting implementation of new telephone triage pathways and, due to excess demand, calls started to be diverted to a national centre on the 4th of March.

Telephone triage services provided by NHS 111 appear to have played an important role in the risk assessment of patients with suspected COVID-19 infection in the UK during the pandemic. However, concerns have been raised that telephone triage may miss patients with serious illness, leading to calls for an inquiry into the effectiveness of NHS 111 to identify those in need of medical intervention.^{9,10} There has been no previous evaluation of the accuracy of the clinical risk-assessment performed by this service or, to our knowledge, other telephone triage services for patients with suspected COVID-19 infection.

Our study aimed to:

- 1) To assess how accurately NHS 111 telephone services identified those who suffered an adverse outcome needing an emergency response.
- 2) To identify any factors that may have affected the accuracy of telephone triage.

Methods

Study Design

The PRIEST study was piloted as the Pandemic Influenza Triage in the Emergency Department (PAINTED) study, part of the UK National Institute for Health Research (NIHR) portfolio of studies to be activated in an influenza pandemic.¹¹ However, it was adapted in February 2020 in response to the COVID-19 pandemic, to include an expanded range of respiratory infections and evaluate pre-hospital urgent and emergency care triage services. This evaluation of NHS 111 telephone services is an observational cohort study that forms part of the PRIEST Study.

Setting

Yorkshire Ambulance Service NHS Trust (YAS) provides 24-hour emergency and health care services for the Yorkshire and Humber, Bassetlaw, North Lincolnshire and North East Lincolnshire region in the north of England; an area of approximately 6,000 square miles and with a population of 5.3 million. In 2018/19, YAS received more than 998,500 emergency (999) and 1,632,514 NHS 111 calls.

Data Sources and linkage

Yorkshire Ambulance Service provided a dataset of NHS 111 calls, triaged using an assessment pathway indicating possible COVID-19 infection, received between the 18th March 2020 and 29th June 2020. All patients within the English National Health Service (NHS) are allocated a unique identification number, the NHS number. Records with no NHS number (less than 2%) were not provided as these records could not be associated with a traceable individual without manual review. The dataset consisted of patient identifiers, demographic data, call details and triage dispositions extracted from routinely collected electronic NHS 111 call records.

Patient identifiers were provided to NHS Digital for them to trace the identities of our cohort (i.e. indicate different sets of identifiers belonging to the same patient) and to supply additional individual level demographic, co-morbidity, and outcome data. NHS Digital manages national health and care data collections from a variety of settings and providers in England.¹² NHS Digital identified records in their collections belonging to patients in our cohort and provided data on patient demographics, limited COVID-related general practice (GP) records, emergency department attendances, hospital inpatient admissions, critical care periods, and death registrations from the Office of National Statistics.

Both YAS and NHS digital removed records belonging to patients who had registered an NHS national data opt-out. The study team excluded patients who had opted out of any part of the PRIEST study and those with inconsistent records (e.g. multiple deaths recorded or death before latest activity). Patient identifiers across all datasets were replaced with a consistent pseudo-identifier to enable the identification of records belonging to individual patients across datasets without revealing patient identifiers.

Inclusion criteria

Our final cohort consisted of all adult (16+) patients at time of first call (index contact) within the YAS NHS 111 calls dataset who were traced by NHS Digital and for whom a final triage disposition, and therefore urgency of recommended triage, was recorded for their index contact.

Patient characteristics

Consistent with methods used to estimate the Charlson comorbidity index from these routine data, comorbidities were included if recorded 12 months before the index contact with NHS 111.^{13 14}

Immunosuppressant drug use only contributes to the immunosuppression co-morbidity if recorded in the 30 days before index contact. Pregnancy status was based on GP records recorded in the previous 9 months. Frailty in patients older than 65 years was derived from the latest recorded (if any) Clinical Frailty Scale score present in the electronic GP records prior to index contact.¹⁵ Patients under the age of 65 were given a score of 1, since the score is not validated in this age group. Smoking status was similarly derived from GP records based on the latest recorded (if any) smoking status prior to the index contact.

Outcome

The primary outcome was death, renal, respiratory, or cardiovascular organ support (identified from death registration and critical care data) at 30 days from index contact.

The secondary outcome was death or organ support at 3 and 7 days from index contact.

Analysis

We conducted a descriptive analysis of patient demographics, co-morbidities and call disposition and used multivariable logistic regression modelling to confirm known patient characteristics associated with the primary adverse outcome in COVID-19 infection. Obesity was excluded from multivariable analysis due to an observed implausible protective association with the primary outcome which we believe to be an artefact of how these data were collected and recorded in the electronic GP data set. Ethnicity was also excluded from multivariable analysis due to the high proportion of missing data (approximately 25%).

To assess how accurately NHS 111 identified patients with adverse outcomes, the call disposition categories of the index contact were divided into a binary classification of either: ambulance dispatched or other urgent clinical assessment required; and self-care or non-urgent assessment (Supplementary Material 1). Advice and call disposition provided by NHS 111 can change over successive calls as a patient's condition changes. Therefore, to assess if deterioration was recognised over multiple calls, a sensitivity analysis was conducted in patients who had an adverse outcome in which the disposition of the call immediately before the adverse outcome was used for binary classification.

We assessed the accuracy of the binary triage classification (ambulance dispatch/urgent clinical assessment versus self-care/non-urgent assessment) in terms of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for the primary outcome with 95% confidence intervals (CI). To assess whether the implementation of different COVID-19 related NHS Pathways affected the accuracy of triage, accuracy was estimated for the whole study period and in two distinct time periods. The first time period (18/03/2020 to 02/06/2020) encompassed the use of Pathways 19.3.3/4/5/7 by YAS and the second period (02/06/2020 (10:30AM) to 29/06/2020) Pathways 19.3.8/9 which incorporated loss of taste or smell as a feature of COVID-19 infection (Supplementary Material 2).

Patient characteristics of false negatives (those advised to self-care/non-urgent assessment who experienced the primary outcome) and true positives (those provided with an ambulance/urgent assessment who experienced the primary outcome) were compared. In patients with the adverse outcome multi-variable logistic regression was used to identify patient characteristics associated with false negative triage. We also compared the characteristics of false positives (those provided with an ambulance/urgent assessment and not conveyed to hospital) and true negatives (those advised to self-care/non-urgent assessment) among those who did not experience the primary composite adverse outcome. We used multivariable logistic regression to identify factors which predicted false positive triage. Frailty was excluded from multivariable modelling due to a high proportion of missing data (approximately 40% of false negatives).

Ethical Approval

The North West—Haydock Research Ethics Committee gave a favourable opinion on the PAINTED study on 25 June 2012 (reference 12/NW/0303) and on the updated PRIEST study on 23rd March 2020, including the analysis presented here. The Confidentiality Advisory Group of the NHS Health Research Authority granted approval to collect data without patient consent in line with Section 251 of the National Health Service Act 2006. Access to data collected by NHS Digital was recommended for approval by its Independent Group Advising on the Release of Data (IGARD) on 11th September 2021 having received additional recommendation for approval for access to GP records from the Profession Advisory Group (PAG) on 19th August 2021.

Patient Public Involvement

The Sheffield Emergency Care Forum (SECF) is a public representative group interested in emergency care research.¹⁶ Members of SECF advised on the development of the PRIEST study and two

members joined the Study Steering Committee. Patients were not involved in the conduct of the study.

Results

All totals presented are rounded to the nearest 5, with small numbers suppressed to comply with NHS Digital data disclosure guidance.

Study population

Figure 1 and Table 1 summarise study cohort derivation and the characteristics of the 40, 261 included individuals. In total, 1, 200 people (3%, 95% CI:2.8% to 3.2%) experienced the primary outcome (death or organ support) within 30 days following first contact with telephone triage services and 670 (56%) of adverse outcomes occurred within 7 days of contact. In our study cohort, 8, 165 patients (20.3%, 95% CI:19.9% to 20.7%) were conveyed or self-presented to the ED and 4, 490 (11.2%, 95% CI:10.9% to 11.5%) were admitted as hospital inpatients within 30 days of index contact.

The median age of the whole cohort was 47 (IQR 32-61), the cohort had a higher proportion of females (56.4%) and had high rates of comorbidity (chronic respiratory disease 25.6%, diabetes 10.5% and hypertension 18.1%). In multivariable modelling (Supplementary Material 3), known predictors of adverse outcomes including increasing age, male sex, diabetes and frailty were associated with an increased risk of the primary composite adverse outcome.

Table 1: Population characteristics

Population Characteristic	Level	Whole Population N=40, 261	Primary Outcome N=1,200
Age (Years)	Median (IQR)* Mean	47 (32-61) 48.4	77 (63-86) 73.3
Sex (N, %)	Male	17,5645 (43.6%)	710 (59%)
Comorbidity (N, %)	Cardiovascular Disease	850 (2.1%)	80 (6.6%)
	Chronic Resp. Disease	10,315 (25.6%)	320 (26.7%)
	Diabetes	4,245 (10.5%)	275 (22.8%)
	Hypertension	7,275 (18.1%)	510 (42.6%)
	Immunosuppression (including steroid)	3, 320 (8.2%)	350 (29.3%)

	use)		
	Active Malignancy	460 (1.2%)	55 (4.6%)
	Obesity	6,205 (15.4%)	105 (8.8%)
	Pregnant	685 (1.7%)	*
	Renal Impairment	415 (1%)	45 (3.7%)
	Smoker	11, 160 (29%)	1, 200 (35.9%)
	Stroke	230 (0.6%)	20 (1.8%)
Number of Drugs Used (N, %)	0	17, 830 (44.3%)	230 (19.2%)
	1-5	17, 300 (43%)	620 (51.5%)
	6-10	4, 550 (11.3%)	315 (26.5%)
	11 or more	585 (1.5%)	40 (3.3%)
Clinical Frailty Scale (N, %)	Unknown	6,610 (16.4%)	605 (50.3%)
	Aged<65	31,860 (79.1%)	330 (27.4)
	1-3	200 (0.5%)	10 (0.9%)
	4-6	670 (1.7%)	60 (5.1%)
	7-9	925 (2.3%)	195 (16.3%)
Ethnicity (N, %)	Unknown	8,950 (22.2%)	410 (34%)
	Asian or Asian British	4,335 (10.8%)	90 (7.3%)
	Black or Black British	850 (2.1%)	15 (1.3%)
	Mixed	540 (1.3%)	10 (0.8%)
	Other Ethnic Groups	615 (1.5%)	10 (1%)
	White	24,975 (62%)	670 (55.7%)
Deprivation Index (N, %)	Unknown	2,360 (5.9%)	135 (11.2%)
	1-2	14,355 (35.7%)	330 (27.6%)
	3-4	7, 365 (18.3%)	205 (17.1%)
	5-6	6,005 (14.9%)	180 (15%)
	7-8	5, 835 (14.5%)	195 (16.1%)
	9-10	4345 (10.8%)	155 (13%)
Index Triage Category (N, %)	Ambulance dispatch/urgent clinical assessment	15, 030 (39.6%)	890 (74.2%)
	Ambulance Response	3, 955 (9.8%)	450 (37.6%)
	Urgent follow up GP Assessment	555 (1.4%)	45 (0.4%)
	Urgent follow up COVID clinical assessment	11, 235 (27.9%)	395 (32.7%)
	Self-care/non-urgent assessment	24, 335 (60.4%)	310 (25.8%)
	Self-Care	12, 925 (32.1%)	85 (7.2%)
	Non-urgent GP assessment	880 (2.2%)	10 (0.1%)

	Further Non-urgent COVID Assessment	10, 510 (26.1%)	215 (17.7%)
Outcome (N, %)	Death	910 (2.3%)	910 (75.6%)
	Deaths due to COVID (including after 30 days)	710 (1.8%)	580 (48.3%)
	Organ support (within 30 days)	425 (1.1%)	425 (35.2%)
Hospitalisation (N, %)	ED Attendance	8,165 (20.3%)	840 (70%)
	Inpatient admission	4,490 (11.2%)	840 (69.9%)
Confirmed Hospital Diagnosis of COVID**	In ED or as inpatient at 30 days	2, 370 (5.9%)	650 (54%)
Number of NHS 111 contacts in study period (N, %)	1	36,480 (90.6%)	1,060 (88.2%)
	2	3,035 (7.5%)	120 (9.7%)
	3 or more	750 (1.9%)	25 (2.1%)
Time to Primary Outcome from index contact- up to and including (N, %)	72 hours	320 (0.8%)	320 (26.5%)
	7 days	670 (1.7%)	670 (56%)

*Interquartile Range (IQR)

**Unrestricted community testing for suspected COVID-19 infection was only available from the 18/05/2020. Confirmed diagnosis is based upon inpatient PCR testing or clinical diagnosis in hospital

Accuracy of NHS 111 triage

Table 2 shows the accuracy of the binary triage disposition (ambulance dispatch/urgent clinical assessment versus self-care/non-urgent assessment) for the composite primary adverse outcome for the whole study period and two different periods of NHS 111 COVID clinical assessment pathway implementation. Supplementary Material 4 shows the accuracy of the binary triage disposition for the adverse outcomes at seven and three days. A triage disposition of ambulance dispatch/urgent clinical assessment achieved a sensitivity of 74.2% (95% CI: 71.6 to 76.6%) to the primary outcome. If advised to self-care/non-urgent clinical assessment, the chance of experiencing an adverse outcome was approximately 1% (NPV: 98.7%, 95% CI: 98.6 to 98.9%). For patients who contacted NHS 111 multiple times, classification of the triage disposition on the basis of the last call before the primary outcome, instead of index contact, did not noticeably affect these estimates (sensitivity 77.3%, 95% CI: 74.8 to 79.6% and (NPV: 98.9%, 95% CI: 98.7 to 99%).

Sensitivity of triage disposition was higher for adverse outcomes at three days from index contact (81.4% 95% CI: 76.6 to 85.5%). Specificity was comparable for adverse outcomes at 30 days (61.5% 95% CI: 61 to 62%) and three days (60.8% 95% CI: 60.2 to 61.3%). In the later period of NHS 111 clinical assessment pathway implementation, sensitivity to adverse outcomes at 30 days increased (85.7% 95% CI: 76.9 to 91.7%) but this was associated with a reduction in specificity (51.5% 95% CI: 50 to 53.1%).

Table 2: Performance of binary NHS 111 triage (ambulance or urgent assessment 4 hours or less) for composite outcome (death or organ support)

Adverse outcome up to 30 days (3%, 2.8-3.2%)			
N=40, 261	Adverse Outcome	No Adverse Outcome	
Ambulance/urgent assessment	890	15, 035	Sensitivity 74.2% (71.6-76.6%) Positive Predictive Value 5.6% (5.2 - 6%)
Self-care/ non-urgent assessment	310	24, 025	Specificity 61.5% (61% - 62%) Negative Predictive Value 98.7% (98.6 - 98.9%)

Time period 1 18/3/20-2/06/2020 (3.1%, 2.9-3.2%)			
N=36, 124	Adverse Outcome	No Adverse Outcome	
Ambulance/urgent assessment	807	13, 079	Sensitivity 73.2% (70.4-75.7%) Positive Predictive Value 5.8% (5.4 – 6.2%)
Self-care/ non-urgent assessment	296	21, 942	Specificity 62.7% (62.1% - 63.2%) Negative Predictive Value 98.7 (98.5 - 98.8%)

Time period 2 2/06/2020-30/06/2020 (2.4%, 1.9-2.9%)			
N= 4,137	Adverse Outcome	No Adverse Outcome	
Ambulance/urgent assessment	84	1, 957	Sensitivity 85.7% (76.9-91.7%) Positive Predictive Value 4.1% (3.3 – 5.1%)
Self-care/ non-urgent assessment	14	2, 082	Specificity 51.5% (50% - 53.1%) Negative Predictive Value

			99.3 (98.9 - 99.6%)
--	--	--	---------------------

Prediction of false negative or false positive triage

Table 3 compares the characteristics of patients who experienced the primary outcome, and either were (true positives) or were not (false negatives) provided with an ambulance or other urgent clinical assessment. In both groups, approximately 50% of people experienced the primary adverse outcome within seven days of first contact, although a higher proportion of true positives experienced the adverse outcome within three days of contact. Multivariable modelling (Supplementary Material 5), showed that younger age (OR 0.99, 95% CI 0.98 to 1.00), multiple contacts (2 contacts, OR 1.77 95% CI: 1.14 to 2.75 and 3 or more contacts, OR 4.02 95% CI: 1.68 to 9.65), and diabetes (OR 1.66, 95% CI 1.13 to 2.45) were associated with increased risk of false negative triage. The effect estimates for multiple NHS 111 contacts were similar if the triage disposition of last call before the primary outcome (2 contacts, OR 1.96 95% CI: 1.11 to 3.48 and 3 or more contacts, OR 7.78 95% CI: 1.02 to 59.43) was used to classify true positives and false negative.

Table 3: Comparison of False Negatives and True positives

Population Characteristic	Level	False Negatives (30 days) N= 310	True Positives (30 days) N=890
Age (Years)	Median (IQR) Mean	71.5 (57-84) 69.9	78 (66-86) 74.5
Sex (N, %)	Male	185 (59%)	525 (58.9%)
Comorbidity (N, %)	Cardiovascular Disease	*	70 (8%)
	Chronic Resp. Disease	70 (22.9%)	250 (28.1%)
	Diabetes	80 (24.8%)	200 (22.1%)
	Hypertension	115 (36.7%)	400 (44.7%)
	Immunosuppression	30 (8.7%)	160 (18.1%)
	Active Malignancy	*	50 (5.5%)
	Obesity	40 (12.3%)	70 (7.6%)
	Pregnant	*	*
	Renal Impairment	*	35

			(4%)
	Smoker	95 (31%)	335 (37.6%)
	Stroke	*	15 1.7% (1-2.8%)
Number of Drugs Used (N, %)	0	75 (23.6%)	160 (17.6%)
	1-5	175 (56.5%)	445 (49.7%)
	6-10	55 (18.1%)	260 (29%)
	11 or more	5 (1.9%)	33 (3.7%)
Clinical Frailty Scale (N, %)	Unknown	130 (42.6%)	470 (53%)
	Aged<65	120 (39.4%)	210 (23.2%)
	1-3	*	10 (1%)
	4-6	*	45 (4.9%)
	7-9	40 (11.9%)	160 (17.9%)
Ethnicity (N, %)	Unknown	100 (31.3%)	310 (34.9%)
	Asian or Asian British	30 (10%)	10 (1%)
	Black or Black British	*	*
	Mixed	*	10 (1%)
	Other Ethnic Groups	*	60 (6.4%)
	White	170 (55.2%)	500 (55.9%)
Deprivation Index (N, %)	Unknown	30 (10%)	105 (11.7%)
	1-2	85 (27%)	250 (28%)
	3-4	55 (18.1%)	150 (16.7%)
	5-6	40 (13.6%)	140 (15.5%)
	7-8	50 (16.8%)	140 (15.8%)
	9-10	45 (14.8%)	110 (12.4%)
Index Triage Category (N, %)	Self-care	85 (27.7%)	NA
	Ambulance Response	NA	450 (50.6%)
	Further COVID Assessment	215 (68.7%)	295 (44.1%)
	Further GP Assessment	10 (3.2%)	45 (5.1%)
Outcome (N, %)	Death	210 (67.1%)	700 (78.6%)
	Deaths due to COVID (including after 30 days)	140 (45.8%)	435 (48.8%)
	Organ support (within 30 days)	140 (45.2%)	225 (45.2%)
Hospitalisation (N, %)	ED Attendance	210 (68.4%)	630 (70.6%)
	Inpatient admission	215 (69%)	625 (70.3%)

Diagnosis of COVID	In ED or as inpatient at 30 days	170 (55.5%)	475 (53.4%)
Number of NHS 111 contacts in study period (N, %)	1	250 (81.3%)	810 (90.6%)
	2	44 (14.2%)	75 (8.2%)
	3 or more	15 (5%)	10 (1.2%)
Time to Primary Outcome (N, %)	72 hours	60 (19%)	260 (29.1%)
	7 days	170 (55.5%)	500 (56.1%)

Table 4 compares the characteristics of patients without the primary outcome who were, either provided with an ambulance/urgent clinical and not conveyed to hospital (False Positives), or not provided with ambulance/urgent clinical response (True Negatives). 24.9% of the cohort were false positives and Supplementary Material 6 presents the results of multivariable modelling to identify factors associated with being a False Positive. Increased risk of being a false positive was most strongly associated with chronic renal impairment (OR1.52 95% CI 1.17 to 1.97), immunosuppression (OR 1.49 95% CI 1.36 to 1.64) and chronic respiratory disease (OR1.31 95% CI 1.22 to 1.40. Weaker predictors included older age, smoking, increased medication use and female sex.

Table 4: Comparison of False Positives and True Negatives

Population Characteristic	Level	False Positives (30 days) N= 10, 000	True Negative (30 days) N=24, 025
Age (Years)	Median (IQR)	49 (33-65)	44 (31-57)
	Mean	49.9	45.2
Sex (N, %)	Male	4,190 (41.9%)	10,460 (43.5%)
Comorbidity (N, %)	Cardiovascular Disease	235 (2.4%)	310 (1.3%)
	Chronic Resp. Disease	3,160 (31.6%)	5,125 (21.3%)
	Diabetes	1,120 (11.2%)	2,100 (8.7%)
	Hypertension	2,010 (20.1%)	3,320 (13.8%)
	Immunosuppression	1070 (10.7%)	1,278 (5.3%)
	Active Malignancy	120 (1.2%)	150 (0.6%)
	Obesity	1,660 (16.6%)	3,490 (14.5%)
	Pregnant	190 (1.9%)	470 (2%)
	Renal Impairment	135 (1.4%)	140 (0.6%)
	Smoker	3,245 (32.4%)	6,200 (25.8%)

	Stroke	70 (0.7%)	85 (0.35%)
Number of Drugs Used (N, %)	0	3, 900 (39%)	12, 230 (50.9%)
	1-5	4, 500 (45%)	9, 830 (40.9%)
	6-10	1, 420 (14.2%)	1, 800 (7.5%)
	11 or more	190 (1.9%)	165 (0.7%)
Clinical Frailty Scale (N, %)	Unknown	1,965 (19.6%)	2,790 (11.6%)
	Aged<65	7,470 (74.7%)	20,725 (86.3%)
	1-3	60 (0.6%)	85 (0.4%)
	4-6	225 (2.2%)	220 (0.9%)
	7-9	290 (2.9%)	210 (0.9%)
Ethnicity (N, %)	Unknown	2,100 (21%)	5,480 (22.8%)
	Asian or Asian British	1,025 (10.3%)	2,790 (11.6%)
	Black or Black British	190 (1.9%)	555 (2.3%)
	Mixed	110 (1%)	355 (1.5%)
	Other Ethnic Groups	130 (1.3%)	420 (1.7%)
	White	6,445 (64%)	14,420 (60%)
Deprivation Index (N, %)	Unknown	650 (6.5%)	1, 180 (4.9%)
	1-2	3, 570 (38.2%)	8, 730 (38.2%)
	3-4	1, 770 (19%)	4, 500 (19.7%)
	5-6	1, 470 (15.7%)	3, 665 (16%)
	7-8	1, 435 (15.4%)	3, 440 (15.1%)
	9-10	1, 105 (11.8%)	2, 510 (11%)
Index Triage Category (N, %)	Self-care	NA	12, 840 (53.4%)
	Ambulance Response	1, 645 (16.5%)	NA
	Further COVID Assessment	7, 935 (79.3%)	10, 300 (42.9%)
	Further GP Assessment	360 (3.6%)	490 (2%)
Number of NHS 111 contacts in study period (N, %)	1	9,230 (92.3%)	21,855 (91%)
	2	645 (6.5%)	1,740 (7.3%)
	3 or more	130 (1.3%)	430 (1.8%)

Discussion

Summary

Our study showed that users of the NHS 111 service who were identified with possible COVID-19 infection had a low (3%) risk of adverse outcome. Telephone triage recommended self-care or non-urgent assessment for the majority (60%), with a very low but non-negligible risk of adverse

outcome (1.3%). The sensitivity (74.2%, 95% CI: 71.6 to 76.6%) and specificity (61.5%, 95% 61% to 62%) of telephone triage to the composite primary outcome is similar to that reported for clinical tools used to triage patient acuity in the ED, at a point on the ROC curve with an equivalent balance of sensitivity and specificity.¹⁷ It is also higher than the reported accuracy of telephone triage for patients with suspected cerebral vascular accidents.¹⁸ Sensitivity of telephone triage was higher for outcomes at three and seven days (Supplementary Material 4), and sensitivity appeared to be increased at the expense of specificity in the later period of clinical assessment pathway implementation (Table 2).

We used multivariable analysis to identify predictors of false negative and false positive triage. The findings need cautious interpretation, given the limited information available during telephone triage, but suggest that some co-morbidities (such as chronic respiratory disease) may be over-estimated as predictors of adverse outcome, while the association of diabetes with adverse outcome may be under-recognised. Perhaps most striking, is that multiple contacts with NHS 111, in which possible COVID-19 infection was identified, was associated with false negative assessment, suggesting that repeat contacts require a more urgent response.

Strengths and limitations

Although telephone triage has been recommended and widely used during the pandemic in the UK to risk assess patients with suspected COVID-19 to limit potential spread of infection, this appears to be the first evaluation of accuracy.¹⁹ We have used a large cohort of patients identified from routinely collected telephone triage records and linked this to nationally collected, patient-level healthcare records to provide robust outcome data. We have assessed performance in a cohort of patients with suspected infection which, in the absence of accurate universally available rapid COVID-19 diagnostic tests, reflects the population which urgent and emergency care services must clinically triage. Unrestricted community testing for those with symptoms suggestive of COVID-19 infection was only available from 18/05/2020 and therefore it is not possible to estimate the proportion of confirmed infections. However, known factors associated with adverse outcomes in COVID-19 infection were found to be predictive of the primary outcome in our cohort including increasing age, male sex, diabetes and frailty.²⁰⁻²²

Due to the use of routinely collected data there were high rates of missing data for some variables, for example, ethnicity and frailty, which prevented inclusion in some analyses. We have also assumed that if co-morbidities were not recorded in the previous 12-months they are not present. The mechanism of how data are collected and recorded in the routine data sets used means that, as identified for obesity, there may be bias in the classification of patients. The estimated prevalence of

obesity in our cohort is 15% (half that reported in the national health survey) and, as weight is not comprehensively and consistently measured by GPs, the observed protective association is likely to reflect unknown characteristics associated with a measurement being taken, rather than obesity itself.²³

We have evaluated the performance of NHS 111 telephone triage as implemented by the Yorkshire Ambulance Service NHS Trust. Although NHS 111 Pathways software algorithms are developed nationally, there may be variability in local implementation which may affect accuracy. During the study period, calls were diverted between regions and to a national centre due to excess demand. The basis on which calls were selected for diversion is not transparent, but it is possible that patients with less complex health care needs were diverted to the national centre, potentially affecting the generalisability of our results. Our study period includes multiple pathway iterations but due to how rapidly assessment pathways were updated it was not possible to assess the accuracy of individual assessment pathways (Supplementary Material 2). A national online assessment tool was implemented from the end of February 2020 and this may have affected the characteristics of the population using telephone triage services for advice.²⁴ However, it was not until June 2020, that the public were advised to use the NHS 111 Online coronavirus service before calling NHS 111.

Implications

Telephone triage performed comparably to triage methods used for patient acuity in the ED and, given the limited information available, including a lack of physiological parameters, this may reflect the best accuracy that could be achievable.^{17,25} In 2019, the estimated population of Yorkshire and the Humber was 5,502,967 (including children).²⁶ On the basis of the number of patients in our cohort and study period, not using telephone triage could have led to around 61 extra ambulances or urgent clinical assessments being provided each day per 1 000, 000 population, without considering diversion to the national centre. Yorkshire Ambulance Service provided a face-to-face response to an estimated 298 incidents per day in March 2020.²⁷ Although there was an inevitable degree of misclassification, NHS 111 telephone triage appears to have effectively helped to mitigate the risk of emergency healthcare services being overwhelmed by lower risk patients. This must be weighed against the small but non-negligible risk that patients who were recommended to self-care or have a non-urgent clinical assessment had of serious adverse outcomes. The acceptable risk of deterioration following such triage is subjective and significant variation in risk tolerance between clinicians and public representatives has been demonstrated.²⁸ As the pandemic developed, and the risk of emergency services being overwhelmed appeared to reduce, the clinical assessment pathways seemed (arguably rightly) to prioritise sensitivity over specificity. The use of similar

telephone triage methods may need to be tailored to the resource constraints and risk tolerance of different health care settings.

Research and information regarding risk factors for adverse outcomes in COVID-19 accumulated rapidly during the initial phase of the pandemic. Early clinical guidelines for the risk stratification of patients with suspected COVID infection, on the basis of previous influenza epidemics, emphasised the importance of respiratory co-morbidities and may have underestimated the risk associated with sex and diabetes.²⁹ The results of our multivariable modelling reflects this, with the importance of smoking and chronic respiratory disease appearing to be overestimated and diabetes underestimated. Later NHS clinical guidelines incorporated this evolving research base and emphasised the risk associated with diabetes.³⁰ However, the association we found with multiple NHS 111 COVID-19 related contacts and risk of under triage does not appear to have been previously identified and may reflect that patients with repeat contacts represent an unrecognised high risk group. Patients with early representation after discharge from the ED are considered clinically high risk for adverse outcomes and misdiagnosis and this is likely to be reflected in patients who contact NHS 111.³¹

Future research could use our methods for a national evaluation of NHS 111 performance, and to assess regional variation. As there are differences in how providers organise NHS 111 services and how NHS 111 clinical triage pathways were implemented, accuracy may vary. Different models for telephone triage in urgent and emergency care exist internationally.^{18 32 33} NHS 111's use of trained, non-clinical call advisors for initial assessment contrasts with other national triage services where assessments are performed by nurses and other clinicians.³³ As has been previously recommended, research is needed to determine the optimal configuration of such services in terms of accuracy and cost-effectiveness.³²

Conclusions

We have conducted the first evaluation of accuracy of telephone triage for need for emergency treatment in patients with suspected COVID-19 infection. Telephone triage appears to have had an important role in managing lower-risk patients and potentially preventing many patients who required no specific treatment from attending hospitals or other care providers. This must be weighed against the small but non-negligible risk of serious adverse outcomes in patients advised to self-care or have a non-urgent clinical assessment. Repeat contact with triage services may need more recognition as an important predictor of subsequent deterioration. Future research is needed

to determine acceptable risk of deterioration in patients advised to self-care and the optimal configuration of telephone triage services.

Author Disclosure Statement

No competing financial interests exist.

Funding

CM is a National Institute for Health Research (NIHR) Clinical Lecturer in Emergency Medicine (Grant Number Not Applicable/NA). This publication presents independent research funded by the National Institute for Health Research and University of Sheffield. The views expressed are those of the author(s) and not necessarily those of the University of Sheffield, the NHS, the NIHR or the Department of Health and Social Care

Authors' contributions

The idea for the study was conceived by SG, JT, TS, FB, PB and CM. Data processing and linkage was completed by TS and RC. The analyses were completed by CM and MH with specialist statistical advice from SG, JT, LS and PB. All authors contributed to interpretation of results, read and approved the final manuscript.

Data sharing:

The data used for this study are subject to data sharing agreements with NHS digital and Yorkshire Ambulance Service which prohibits further sharing of individual level data. The data sets used are obtainable from these organisations.

Figures:

Figure 1: STROBE flow diagram of selection of study population

1. Department of Health. Pandemic Flu: A National Framework for Responding to an Influenza Pandemic. London: Department of Health; 2007.
2. Goodacre S, Challen K, Wilson R, et al. Evaluation of triage methods used to select patients with suspected pandemic influenza for hospital admission: cohort study. *Health Technol Assess* 2010;14(46):173-236. doi: 10.3310/hta14460-03 [published Online First: 2010/10/07]
3. Scarfone RJ, Coffin S, Fieldston ES, et al. Hospital-Based Pandemic Influenza Preparedness and Response: Strategies to Increase Surge Capacity. *Pediatric Emergency Care* 2011;27(6):565-72. doi: 10.1097/PEC.0b013e31821dc9d1
4. Nelissen S, Beullens K, Sabbe M, et al. The swine flu emergency department: the relationship between media attention for the swine flu and registrations in an emergency medicine unit.

- Prehosp Disaster Med* 2014;29(2):141-5. doi: 10.1017/s1049023x14000247 [published Online First: 2014/03/20]
5. <https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/03/coronavirus-primary-care-briefing.pdf>.
6. Mulholland RH, Wood R, Stagg HR, et al. Impact of COVID-19 on accident and emergency attendances and emergency and planned hospital admissions in Scotland: an interrupted time-series analysis. *J R Soc Med* 2020;113(11):444-53. doi: 10.1177/0141076820962447 [published Online First: 2020/10/06]
7. Goodacre S, Thomas B, Lee E, et al. Characterisation of 22445 patients attending UK emergency departments with suspected COVID-19 infection: Observational cohort study. *PLOS ONE* 2020;15(11):e0240206. doi: 10.1371/journal.pone.0240206
8. <https://www.health.org.uk/news-and-comment/charts-and-infographics/how-has-nhs-111-shaped-public-demand-for-the-nhs-in-england>.
9. Kmietowicz Z. Covid-19: Bereaved families, unions, and charities demand immediate public inquiry to save lives. *BMJ* 2020;371:m4729. doi: 10.1136/bmj.m4729
10. <https://www.hsib.org.uk/investigations-cases/response-nhs-111-covid-19-pandemic/>.
11. Goodacre S, Irving A, Wilson R, et al. The PAndemic INfluenza Triage in the Emergency Department (PAINTED) pilot cohort study. *Health Technology Assessment* 2015;19(3)
12. <https://digital.nhs.uk/data-and-information>.
13. Campbell MJ, Jacques RM, Fotheringham J, et al. Developing a summary hospital mortality index: retrospective analysis in English hospitals over five years. *BMJ* 2012;344:e1001. doi: 10.1136/bmj.e1001
14. Street A, Maynou L, Gilbert T, et al. The use of linked routine data to optimise calculation of the Hospital Frailty Risk Score on the basis of previous hospital admissions: a retrospective observational cohort study. *The Lancet Healthy Longevity* 2021;2(3):e154-e62. doi: 10.1016/S2666-7568(21)00004-0
15. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. *CMAJ* 2005;173(5):489-95. doi: 10.1503/cmaj.050051
16. Hirst E, Irving A, Goodacre S. Patient and public involvement in emergency care research. *Emergency Medicine Journal* 2016;33(9):665-70.
17. Zachariasse JM, van der Hagen V, Seiger N, et al. Performance of triage systems in emergency care: a systematic review and meta-analysis. *BMJ Open* 2019;9(5):e026471. doi: 10.1136/bmjopen-2018-026471
18. Erkelens DC, Rutten FH, Wouters LT, et al. Accuracy of telephone triage in patients suspected of transient ischaemic attack or stroke: a cross-sectional study. *BMC Family Practice* 2020;21(1):256. doi: 10.1186/s12875-020-01334-3
19. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. *BMJ* 2020;368:m1182. doi: 10.1136/bmj.m1182
20. Goodacre S, Thomas B, Sutton L, et al. Derivation and validation of a clinical severity score for acutely ill adults with suspected COVID-19: The PRIEST observational cohort study. *PLoS One* 2021;16(1):e0245840. doi: 10.1371/journal.pone.0245840 [published Online First: 2021/01/23]
21. Izcovich A, Ragusa MA, Tortosa F, et al. Prognostic factors for severity and mortality in patients infected with COVID-19: A systematic review. *PLoS One* 2020;15(11):e0241955. doi: 10.1371/journal.pone.0241955 [published Online First: 2020/11/18]
22. Knight SR, Ho A, Pius R, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. *BMJ* 2020;370:m3339. doi: 10.1136/bmj.m3339
23. <https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2019>.
24. <https://www.england.nhs.uk/2020/03/nhs-111-online-to-help-people-with-coronavirus/>.

25. Thomas B, Goodacre S, Lee E, et al. Prognostic accuracy of emergency department triage tools for adults with suspected COVID-19: the PRIEST observational cohort study. *Emergency Medicine Journal* 2021;emermed-2020-210783. doi: 10.1136/emermed-2020-210783
26. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2019#population-age-structures-of-uk-countries-and-english-regions>.
27. <https://www.england.nhs.uk/statistics/statistical-work-areas/ambulance-quality-indicators/ambulance-quality-indicators-data-2019-20/>.
28. O'Keeffe ST. A cross-sectional study of doctors', managers' and public representatives' views regarding acceptable level of risk in discharges from the emergency department. *Qjm* 2015;108(7):533-8. doi: 10.1093/qjmed/hcu246 [published Online First: 2014/12/19]
29. Piroth L, Cottenet J, Mariet A-S, et al. Comparison of the characteristics, morbidity, and mortality of COVID-19 and seasonal influenza: a nationwide, population-based retrospective cohort study. *The Lancet Respiratory Medicine* 2021;9(3):251-59. doi: 10.1016/S2213-2600(20)30527-0
30. <https://www.england.nhs.uk/2020/05/nhs-expands-offer-of-help-to-people-with-diabetes-during-coronavirus-outbreak/>.
31. Hiti EA, Tamim H, Makki M, et al. Characteristics and determinants of high-risk unscheduled return visits to the emergency department. *Emergency Medicine Journal* 2020;37(2):79. doi: 10.1136/emermed-2018-208343
32. Turner J, Coster J, Chambers D, et al. What evidence is there on the effectiveness of different models of delivering urgent care? A rapid review. 2015
33. van Ierland Y, van Veen M, Huibers L, et al. Validity of telephone and physical triage in emergency care: The Netherlands Triage System. *Family Practice* 2011;28(3):334-41. doi: 10.1093/fampra/cmq097

Figure 1: STROBE flow diagram of study population selection