The University Of Sheffield. Clinical Trials Research Unit.

PRIEST

PRIEST Statistical Analysis Plan Version 1

Study Title	**The PRIEST study: Pandemic Respiratory Infection Emergency System Triage**
Short title	**PRIEST**
Funding body:	**NIHR HTA**
Funder's reference	**11/46/07**
UKCRN ref	

Authored by

_____ ___/___/_____

Ellen Lee Date
Statistician, Sheffield CTRU

Approved by

_____ ___/___/_____

Prof. Steve Goodacre Date
PRIEST Chief investigator, The University of Sheffield

_____ ___/___/_____

Dr Paul Baxter Date
Study Steering Committee independent statistician, University of Leeds

_____ ___/___/_____

Prof. Tim Coats Date
Study Steering Committee Chair, University of Leicester

_____ ___/___/_____

Mike Bradburn, Date
Senior Statistician, Sheffield CTRU

Amendments to the SAP since version 1

Version	Date approved	Modifications	Prior to/after unblind review
		-	

Table of contents

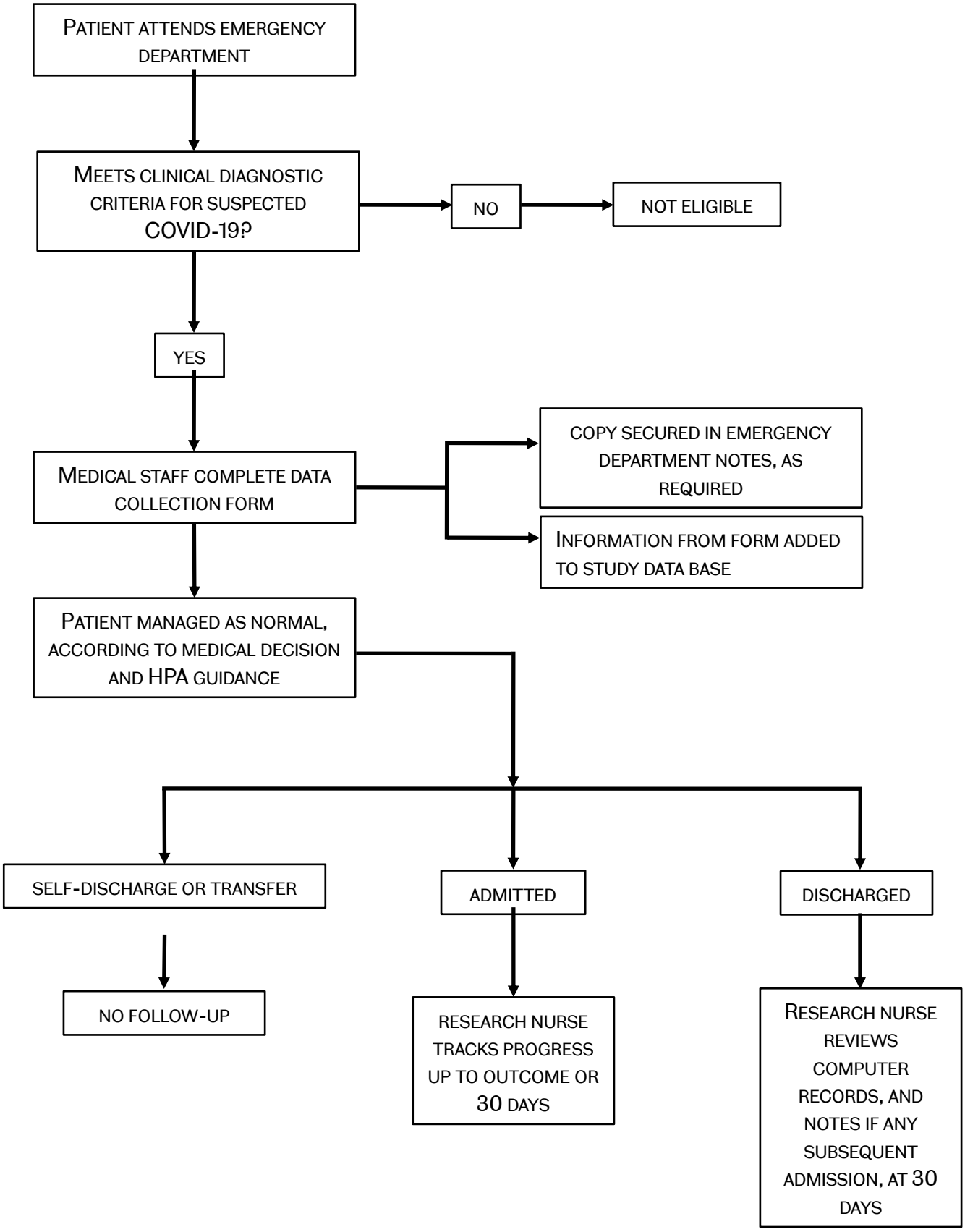
List of abbreviations used

AE	Adverse Event
AUROC	Area Under the Receiver Operating Characteristics
AVPU	Alert, Verbal, Pain, Unresponsive
BP	Blood Pressure
CI	Confidence Interval
CRN	Clinical Research Network
C-Statistic	Concordance statistic
CTRU	Clinical Trials Research Unit
CXR	Chest X-ray
DNR	Do Not Resuscitate
ECG	Electrocardiogram
GCS	Glasgow Coma Score
HPA	Health Protection Agency
HTA	Health Technology Assessment
ICNARC	Intensive Care National Audit & Research Centre
LASSO	Least Absolute Shrinkage and Selection Operator
MI	Multiple Imputation
NHS	National Health Service
NIHR	National Institute for Health Research
ONS	Office for National Statistics
PCA	Principle Components Analysis
PMG	Project Management Group
POPS	Paediatric Observation Priority Score
ROC	Receiver Operating Characteristic
SAP	Statistical Analysis Plan
SOP	Standard operating procedure
SSC	Study Steering Committee
STARD	Standards for Reporting Diagnostic Accuracy
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
WCC	White Cell Count
WHO	World Health Organisation

Summary Page

Study title	The PRIEST study: Pandemic Respiratory Infection Emergency System Triage
Study design	Observational cohort study
Study participants	People attending hospital emergency department with suspected respiratory infections during the COVID-19 pandemic
Sample size	The sample size will ultimately depend upon the size and severity of the pandemic, we anticipate collecting data from 20,000 cases to identify 200 with an adverse outcome.
Follow-up	Follow up will be collected at 30 days post initial attendance.
Statistical analysis	• Assess the performance of existing triage tools (CURB-65, PMEWS, The Swine Flu Hospital Pathway, NEWS2, WHO pneumonia hospitalisation algorithm) for predicting adverse outcome (death or needing respiratory, cardiovascular or renal support) in patients with suspected pandemic respiratory infection • Investigate the predictive value of clinical characteristics and routine tests for adverse outcome • Develop and internally validate new triage methods based on o Presenting clinical characteristics o Presenting clinical characteristics, electrocardiogram, chest X-ray and routine blood test results by developing a multivariable prediction model for adverse outcome • Dependent on the number of cases and events, to externally validate the new triage methods

Figure 1: Study flow diagram



1 Introduction, study design and key study objectives

This statistical analysis plan (SAP) outlines the analysis to be performed for the hospital emergency department part of The PRIEST study. Planned analyses for the prehospital services (NHS 111 and 999) will be detailed in a separate SAP. This SAP will guide the Trial Statistician during the statistical analysis of all quantitative outcomes in order to answer the objectives of the study.

This SAP is written in conjunction with applicable statistical standard operating procedures from the University of Sheffield Clinical Trials Research Unit (CTRU) and study documents (Protocol and Data Validation Specification). It will adhere to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement [1] and the Standards for Reporting Diagnostic Accuracy (STARD) 2015 guidelines [2].

All analyses will be performed in a statistical software package such as STATA version 16 [3].

1.1 Study objectives

The specific objectives outlined in the protocol after each wave of the pandemic, for the hospital (emergency department) are:

1. To determine the discriminant value of emergency department triage methods for predicting severe illness in patients presenting with suspected pandemic respiratory infection
2. To determine the discriminant value of presenting clinical characteristics and routine tests for identifying severe illness
3. To determine the independent predictive value of presenting clinical characteristics and routine tests for severe illness
4. To develop new triage methods based upon presenting clinical characteristics alone or presenting clinical characteristics, electrocardiogram (ECG), chest X-ray and routine blood test results, depending upon the data available and the predictive value of variables evaluated in objective 3

2 Sample Size

The sample size will ultimately depend upon the size and severity of the pandemic. Our pragmatic data collection methods will ensure that we maximise any opportunity to evaluate emergency department triage methods in a pandemic.

Our experience in the 2009 pandemic has shown us that pre-pandemic estimates of case hospitalisation and case fatality rates can be very misleading and that sample size estimates must take into account considerable uncertainty in these estimates. Nevertheless, we have also shown that informative findings can be generated even in a pandemic with a very low rate of adverse outcome.

Given that most cases of suspected pandemic respiratory infection (even in a severe pandemic) do not result in an adverse outcome, the key variable in determining study power is the number of cases with an adverse outcome. A single cohort including at least 150 cases with adverse outcome would allow us to estimate the c-statistic of a triage method, clinical variable or test with a standard error of 0.03 (assuming the true c-statistic was 0.8). The table below shows the standard error resulting from samples with smaller numbers of adverse outcomes.

N with adverse outcome	Standard error (assuming c-statistic was 0.8)
150	0.033
125	0.036
100	0.040
75	0.046
50	0.056

A sample with N=150 adverse outcome would estimate the sensitivity of a dichotomised rule, variable or test with a standard error as outlined in the table below, depending on the sensitivity at the threshold used. Estimates of specificity would obviously be very precise given the anticipated low prevalence of adverse outcome.

Sensitivity	Lower limit of 95% CI
1.00	0.98
0.95	0.90
0.90	0.84
0.85	0.78
0.80	0.73

The same cohort could be used to identify independent predictors of outcome and develop new triage methods (objectives 3 and 4). The number of variables that could be tested as independent predictors of outcome in a multivariable model and for inclusion in a triage method would depend upon the sample size. Based on the rule of thumb of needing at least 10 events for each independent regression variable in a

logistic regression, a cohort with 150 cases with adverse outcome would allow us to test up to 15 parameters [4].

These estimates assume that each triage method and predictor variable will be used and tested on the whole cohort. However, we plan to explore whether different patients require different triage methods, particularly whether a different triage method is required for children and adults. Data from the 2009 H1N1 pandemic suggest that around a quarter to a third of adverse outcomes may occur in children [5] [6]. To increase the probability that we will have at least 50 cases with adverse outcome among children we will aim to recruit a total of 200 cases with adverse outcome rather than 150.

If we assume that the prevalence of adverse outcome is the same as our 2009 cohort (1%) then we would need to collect data from 20,000 cases to identify 200 with an adverse outcome. We have therefore used this estimate in planning, although it is likely to be an overestimate of the total numbers required given the mild nature of the 2009 pandemic. A more severe pandemic would allow more precise estimates to be made with no additional costs or would allow us to reduce the total number of cases required to identify 200 with an adverse outcome.

If we are able to develop a new triage method that appears to have superior discriminant value to existing methods then we would want to validate this method in a new cohort. A sample including 421 cases with adverse outcome would provide 80% power to compare an area under the ROC curve of 0.85 versus 0.90 at 5% significance, assuming a correlation of 0.6 between scores. We have not included validation of a new triage method in our objectives because this would require (a) successful development of a new method and (b) a much larger sample size, with associated costs and assumptions about pandemic severity. However, if the pandemic is severe (i.e. the prevalence of adverse outcome exceeds 3%, so the number with adverse outcome exceeds 450) we will split the cohort into two equal cohorts to allow testing of existing rules and derivation of new rules on one half and validation of new rules, with comparison to existing rules, on the other.

We plan to collect data across 40 hospitals and have based our sample size calculation on the assumption of receiving 500 completed forms, including an average of 5 adverse outcomes, per hospital over the course of the pandemic.

3 Outcome measures

3.1 Primary outcome

Patients who die or require respiratory, cardiovascular or renal support within 30 days of first attendance will be defined as having an adverse outcome. If patients survive to 30 days without requiring respiratory, cardiovascular or renal support they will be defined as having no adverse outcome. A detailed description of how adverse outcome is determined using study data, including the handling of missing data is given in section 8.1.1.

3.2 Secondary outcomes

The following secondary outcomes will be investigated

- Death without respiratory, cardiovascular or renal support: This outcome is less likely to have been preventable, since it either occurred before intervention was possible or intervention was not expected to be beneficial.
- Respiratory, cardiovascular or renal support as a composite outcome (i.e. excluding group 1 above): This outcome is more likely to represent a preventable or prevented death, since life-saving intervention was attempted.
- Subsequent admission, or adverse outcome, in the cohort of adult patients discharged at initial attendance (excluding discharged with DNR decision). This outcome would enable investigation into whether we can predict risky discharge decisions.

The analysis to be performed on these secondary outcomes is given in section 6.8.

4 Weekly descriptive analysis

Whilst study data is being collected, weekly descriptive reports, created by the study statisticians will be produced.

The following will be included in the weekly reports

1. The number and geographical distribution of new cases
2. The proportion with an adverse outcome and details of adverse outcomes
3. Summaries of characteristics and potential prediction factors presented for
    a. The whole cohort
    b. Patients with adverse outcome
    c. Patients who were not admitted but had adverse outcome
    d. Patients who were admitted but did not have adverse outcome
4. Detailed line listings of patients who were not admitted but did have adverse outcome

The report will be made available to the study team, the Project Management Group (PMG) and the Study Steering Committee (SSC). However, the PMG and SSC may only see the report for the week they are meeting.

5 Data Sources and analysis populations

5.1 Data sources

Study data will be extracted from source documents and entered onto the CTRUs in house data management system (PROSPECT). The data management team in the Sheffield CTRU will provide limited post-entry validation (as stipulated in SOP DM005). The trial statistician will conduct any additional validation checks where appropriate before the data lock and sign off (as guided by DM005 and DM012).

Table 1: Details of data collected at each timepoint

Baseline	Pandemic Respiratory Infection Form
Baseline 2	Pandemic Respiratory Infection Form
Baseline 3	Pandemic Respiratory Infection Form
Follow up (30 days after Baseline)	Follow up

5.1.1 Study population and Inclusion/exclusion criteria

Planned inclusion/exclusion criteria are detailed in the protocol. For the adults and children presenting at recruiting hospital emergency departments, inclusion will be determined on the basis of the assessing clinician recording on the patient record that the patient has suspected pandemic infection, which will result in standardised data being collected.

5.2 Analysis populations

The following analysis populations will be studied in the analyses:

Name	**Participants included**
Primary analysis set (adults)	The adult primary analysis set includes all adults (defined as age 16 or older on the date of first attendance) for which presence or absence of adverse outcome could be ascertained (details of the ascertainment of adverse outcome given in section 8.1.1).
Primary analysis set (children)	The child primary analysis set includes all children (defined as age 15 or younger on the date of first attendance) for which presence or absence of adverse outcome could

	be ascertained (details of the ascertainment of adverse outcome given in section 8.1.1).
Secondary analyses sets	The following analysis sets will be used in the subgroup analysis detailed in section 6.7, both will be on adults only due to insufficient numbers in the child cohort: 1. Cases admitted and confirmed as COVID-19 (based on follow up form) 2. Cases admitted and not confirmed as COVID-19 (based on follow up form) 3. Do Not Resuscitate (DNR) decision made on or before day of admission 4. No DNR decision made, or DNR decision made later than day of admission

6 Outline of analyses

6.1 General considerations

Summaries of continuous variables will comprise the number of observations used, mean, median, standard deviation, inter-quartile range, minimum and maximum as appropriate for the distributional form of the data.

Summaries of categorical variables will comprise the number of observations used, and the number and percentage of observations in each category.

For all analyses the patient, rather than the attendance, will be the unit of analysis. If patients attend more than once (the database can collect up to three 'baseline' attendances) and have predictor variable data recorded on more than one occasion, we will only use the predictor variable data from the first attendance in all analyses.

Complete details of data derivations and methods of handling missing data are covered in section8.

6.2 Existing triage tools

We will evaluate the following existing triage methods:

- CURB-65 [7]
- PMEWS [8]
- The Swine Flu Hospital Pathway [9]
- NEWS2 [10]

- The WHO decision-making algorithm for hospitalisation with pneumonia [11]
- POPS [12]
- COAST [13]

The triage tools will be retrospectively applied to the data by the study statisticians. CURB-65 and PMEWS will be evaluated in adults only, NEWS2 will be evaluated in adults excluding pregnant women, POPS will be evaluated in children only. Details of scoring and handling missing data for the triage tools are given in sections 8.2 to 8.5. Existing triage tools will be assessed by plotting the ROC curve and calculating the area under the ROC curve (c-statistic) for discriminating between cases with and without adverse outcome. Sensitivity, specificity, positive predictive value and negative predictive value at key decision making thresholds will be calculated and presented:

Tool	**Threshold**
CURB-65	0-1 versus 2-5
PMEWS	0-2 versus 3+
The Swine Flu Hospital Pathway	0 versus 1+
NEWS2	0-4 versus 5-20, 0-4 versus 50-20 or scoring at least 3 in one category
WHO	0 versus 1
POPS	0-4 versus 5+
COAST	0-2 versus 3+

6.3 Presenting clinical characteristics and routine tests

For all characteristics (candidate predictors) we will present the odds ratio, p-value and 95% confidence interval (CI) from a univariable analysis. For binary characteristics we will also present the distribution of adverse outcomes in each category.. If a continuous characteristic has pre-specified cut-off points the distribution of adverse outcomes will be presented for each category.

The following clinical characteristics and routine tests will be evaluated.

Continuous Characteristic	**Cut offs**	**Adults**	**Children**	**Initial assessment model**	**Full model**
Age	None	✓	✓	✓	✓

Symptom duration		✓	✓	✓	✓
Number of current medications		✓	✓	✓	✓
Urea		✓	✓		✓
Creatinine		✓	✓		✓
Respiratory rate	0= 12-20 or missing | 1= 9-11 | 2= 21-24 | 3= <9 or >24	✓	✓	✓	✓
Pulse rate	0= 51-90 or missing | 1= 41-50 or 91-100 | 2= 111-130 | 3= <41 or >130	✓	✓	✓	✓
Temperature	0= 36.1-38.0 or missing | 1= 35.1-36.0 or 38.1-39.0 | 2= >39.0 | 3= <35.1	✓	✓	✓	✓
Systolic Blood Pressure	0= 111-219 or missing | 1=101-110 | 2= 91-100 | 3= <91 or >219	✓	✓	✓	✓
SaO_2	0= >95 or missing | 1= 94-95 | 2= 92-93 | 3= <93 (add two points for FiO_2>21%)	✓	✓	✓	✓
SaO_2/FiO_2 ratio		✓	✓	✓	✓
GCS Total[a]	1= Severe (<8) | 2= Moderate (9-12) | 3= Mild (13-15)	✓	✓	✓	✓
Na	1= Low (<136) | 2=Normal (136-142) | 3 = High (>142)	✓	✓		✓
K	1= Low (<3.5) | 2=Normal (3.5-5.0) | 3 = High (>5.0)	✓	✓		✓
Hb	1= Low (<130 males, <115 females) | 2=Normal (130-180 males, 115-165 females) | 3 = High (>180 males, 165 females)	✓	✓		✓
Platelets	1= Low (<150) | 2=Normal (150-400) | 3 = High (>400)	✓	✓		✓
WCC	1= Low (<4) | 2=Normal (4-11) | 3 = High (>11)	✓	✓		✓
Neutrophils	1= Low (<2.0) | 2=Normal (2.0-7.5) | 3 = High (>7.5)	✓	✓		✓
Lymphocytes	1= Low (<1.5) | 2=Normal (1.5-4.5) | 3 = High (>4.5)	✓	✓		✓
Lactate		✓	✓		✓

C-Reactive		✓	✓		✓
D-dimer		✓	✓		✓
Troponin		✓	✓		✓

a) Both GCE and AVPU will be analysed in univariable analysis, however in multivariable analysis AVPU will be used and missing AVPU data will be imputed using GCE scores (see section 7.1.2 for the imputation rules).

Binary Characteristic	**Levels**	**Adults**	**Children**	**Initial assessment model**	**Full model**
Sex	Male/Female	✓	✓	✓	✓
Presenting features (Shortness of breath, cough, fever, sore throat, headache, confusion, rash, anosmia, abdominal pain, diarrhoea, vomiting)	Presence/Absence per feature	✓	✓	✓	✓
Previous attendance	Yes/No	✓	✓	✓	✓
Current use of anticoagulants	Presence/Absence	✓	✓	✓	✓
Medical History (Heart disease, renal impairment, steroid therapy, asthma, diabetes, active malignancy, immunosuppression, other chronic lung disease, hypertension)	Yes/No per disease	✓	✓	✓	✓
Paediatric questions (routine vaccination, taking feeds, parental anxiety, premature)	Yes/No per question		✓	✓	✓
known contact with Covid-19 case	Yes/No per lifestyle option	✓	✓	✓	✓

Lifestyle (Pregnant, clinically obese, tobacco/vape user)	Yes/No per lifestyle option	✓		✓	✓
Central capillary refill	Normal/Abnormal	✓	✓	✓	✓
Severe respiratory distress	Yes/No	✓	✓	✓	✓
Respiratory exhaustion	Yes/No	✓	✓	✓	✓
Severe dehydration	Yes/No	✓	✓	✓	✓
CXR	Normal/Abnormal*	✓	✓		✓
ECG	Normal/Abnormal*	✓	✓		✓

* May also be treated as categorical with normal/abnormal/not done depending on frequencies

Categorical Characteristic	**Levels**	**Adults**	**Children**	**Initial assessment model**	**Full model**
Ethnicity*	UK/Irish/other white -South Asian -East Asian -Other Asian -Arab -African -Caribbean -Other black -Other/Mixed	✓	✓	✓	✓
Performance status*	1=Unrestricted normal activity | 2=Limited strenuous activity, can do light | 3=Limited activity, can self care | 4=Limited self care | 5=Bed/chair bound, no self care	✓	✓	✓	✓

*We will consider combining similar categories to reduce the number of predictor variables in the model (each category counts as one predictor), particularly ethnicity where some categories may have low frequencies.

6.4 Development of predictive model and new triage tools

New triage methods will be developed by combining potential predictors of outcome using multivariable logistic regression with Least Absolute Shrinkage and Selection Operator (LASSO) to avoid overfitting [14].

Two new prediction models and subsequent triage scores will be developed: one based on clinical variables measured at initial assessment only (initial assessment model) and the other based on all clinical variables (including blood tests and x-rays) measured in the emergency department (full model). Both models will be built for adult and child populations separately (subject to the number of events in each population, if not enough events are observed in the child population we will not attempt to build prediction models).

Univariable screening of predictors will not be used as a basis for inclusion or exclusion into the predictive model as what matters is the association of predictors with outcome *after* adjustment for other predictors [15]. Using the LASSO begins with a full model of all potentially relevant predictors and simultaneously preforms predictor selection and penalisation during model development to avoid overfitting. This penalisation method essentially shrinks the estimated association between predictor and outcome towards the null, in order to reduce the variability of predictions when the model is used in new datasets.

The potential predictors considered for the predictive model will be those listed in section 6.3; whether the predictor will be considered for the initial assessment model or the full model is given in each table. The sample size calculation specified that at least 20 potential prediction factors could be assessed in a multivariable model but as many as 30 or more could be considered depending on the number of cases and adverse outcome events. Before estimating the relationship between predictors and outcome in a prognostic model, the potential to collapse correlated predictors, or keep one of a set of correlated predictors will be investigated (see section 6.4.1). All continuous predictors will be included as continuous in prediction models in the first instance but the use of cut-offs will be investigated. The potential cut-points to base categorisation of continuous predictors are listed in section 6.3. Non linear prognostic effects for continuous predictors will also be investigated (see section 7.2). Any predictor with more than 50% missing data will not be considered for inclusion in the predictive model. The initial assessment prognostic model will be built first, all prognostic factors that remain in that model after model building will be included with the bloods and tests to be considered for inclusion for the full model.

The stability of derived models will be assessed using bootstrap methods with visual calibration methods (section7.3) [16] [17].

6.4.1 Investigating correlation between prediction factors

The correlation between prediction factors will be investigated informally using scatterplots and investigatory analyses will be performed to see if the dimensions (number of predictor variables) can be reduced using

Principle Components Analysis (PCA), although it is possible that some correlated predictors will be dropped out of the regression model as part of the selection process performed during LASSO regression. For factors that appear highly correlated only one factor will be chosen for inclusion in the prediction model. The availability of data for each factor will be an important consideration in the decision to take forward one of a correlated group of factors, as this represents which factor is more often routinely collected, and also reduces uncertainty in model building.

6.4.1 Converting the predictive model into a triage tool

Integer weights will be assigned to each category of predictor variable according to the coefficient derived from a multivariable model using categorised continuous predictors. This will generate a composite clinical score in which risk of adverse outcome increases with the total score.

Clinical members of the PMG will review the multivariable prediction models and provide input into the derivation of the triage tools. They will input into the following:

- Reviewing the multivariable prediction model for
    - Clinical credibility of predictors included in the model
    - Clinical credibility of the direction and magnitude of effect of predictors in the model
- Deciding whether coefficients can be rounded to integer values
- Deciding if using cut-offs for continuous variables are appropriate and choosing the number and placement of cut-offs (it is expected existing cut-offs will be used as specified in section 6.3 as 'optimal' cut offs found in the models may not translate to other cohorts)
- Reviewing the number of predictors included in the prediction model and deciding if these need to be reduced for ease of use in the emergency department setting (initial discussion in the PMG suggested 8 factors as an upper limit).

A threshold for decision making will be decided upon based on clinical review but will include the following

- A cut-off that ensures sensitivity exceeds 95%

6.5 External validation of prediction models and new triage tools

The population will be split into a development and validation cohort if at least 400 adverse outcomes are recorded (resulting in at least 200 events in each cohort). This will provide sufficient power for the triage tools to be validated [18]. We anticipate there will be insufficient events to split the child cohort and hence this population will not be split. The cohort will be split by date across certain sites. A table of characteristics for the development and validation cohort will be presented to show how the cohorts differ.

The final logistic regression models (both the initial assessment model and full model) will be externally validated by applying the predictive model and triage tools and assessing calibration, discrimination and net benefit in order to assess reproducibility. The validation cohort will also be used to compare the newly developed triage tools against the existing triage tool that performed the best in the development cohort (based on AUROC) by comparing AUROC for these tools.

In order to assess transportability of the new triage tools into different populations or settings, future cohorts must be used to examine heterogeneity in performance across different populations and settings.

6.6 Safety and Harms

In addition to adverse outcome (death and events that required respiratory, cardiovascular or renal support), additional adverse events occurring before 30 day follow up will be recorded if they:

- were life-threatening
- resulted in persistent or significant disability or incapacity
- prolonged hospitalisation

Descriptive statistics of these adverse events will be calculated and reported. The following summaries will be presented:

- Number (%) participants experiencing ≥1 AE
- Number of all AEs including repeat events
- AE by Seriousness (Life threatening, Prolongs hospitalisation, Persistent or significant disability/incapacity)

6.7 Subgroup analysis

All subgroup analysis will be performed for the adult primary analysis set only. The following subgroups will be investigated:

Site

- AUROC will be presented by site for the two newly developed prediction models/triage tools for (if the number of events at each site is sufficient).

COVID-19 confirmed

- The predictive ability of the four existing and two new triage tools, and the predictive models' performance will be assessed separately on admitted patients who had COVID-19 confirmed on the follow up form and those who didn't

DNR decision

- The predictive models' performance and the new triage tools predictive ability will be investigated separately for the subgroup that had DNR decision made on day of first attendance and those that had DNR decision made later or not at all.

6.8 Analysis on secondary outcomes

Predictive multivariable logistic regression models will also be built for the secondary outcomes listed in section 3. The independent relationship between predictor and secondary outcome, and the performance of the existing triage tools will not be investigated on any of the secondary outcomes. The analysis on secondary outcomes will be performed for adults primary analysis set only (section 5.2).

7 Detailed statistical methods and calculations

7.1 Missing data

Cases will be excluded from analysis if we are unable to ascertain if they had adverse outcome or not (i.e. they are missing follow up form). Cases will also be excluded from analysis if more than half of the variables used to develop a prediction model are not available. If particular sites have a large amount of missing data for key variables (for example age, respiratory rate) they will be excluded from analysis. The characteristics of cases with (≥20% missing) and without (<20% missing) missing data will be compared. A threshold of 20% has been chosen for comparison as we anticipate many cases to have some missing data due to the ED setting and the observational nature of this work.

It is likely that a proportion of data for most predictor variables (especially blood results) will be missing. A likely reason is that a measurement would not be made or test performed if it was expected to be normal. If a predictor variable has more than 50% data missing it will not be included in the multivariable predictive models.

Missing data will be handled in the following ways

- In the application of existing triage tools, the recommendations for handling missing data for each tool will be followed (see sections 8.2 to 8.5 for details).
- When investigating the univariate relationship between predictor and outcome (section 6.3) missing data will be treated as missing, or where appropriate treated as normal

- For building the prognostic models missing data will be handled in three ways and the results of these three scenarios will be compared, however the primary method will be multiple imputation
    - Use multiple imputation to impute all missing predictor values
    - Assume missing predictor data is normal (when continuous variable are categorised, or when the outcome has a 'normal' option, for example CXR) or treat as missing if a continuous variable, or if no 'normal' category (see section 10 for details of which predictors will be imputed)
    - Treat all missing predictor data as missing

7.1.1 Multiple imputation

Multiple imputation (MI) will be used to impute missing values for predictor variables in the development of the prognostic model only. At least 20 and ideally 50 multiple imputation data sets will be created using chained equations. Due to the large sample size for this study, it may be computationally intensive to have more than 20 MI data sets. The multiple imputation equation will include outcome and baseline variables with both missing and complete data (for example age, sex, presenting features) to make the missing at random assumption as plausible as possible. A list of which variable will be multiply imputed, together with the factors expected to be related and used in the imputation equations is given in section 10. The imputation model will be consistent with the prognostic model used for analysis, including the functional form of continuous variables.

7.1.2 Missing AVPU

For the multivariable analysis models AVPU will be used and GCS Total will be used to impute AVPU score where AVPU is not available. The following rules will be used:

- GCS 15 = A,
- GCS 9-14 = V,
- GCS 7-9 = P,
- GCS 3-6 = U.

7.2 Modelling non-linear prognostic effects for continuous predictors

Fractional polynomials will be considered for inclusion in the logistic regression model for continuous predictors where there is a known non-linear relationship with outcome. Fractional polynomials have been shown to give similar results to restricted cubic splines; in some situations fractional polynomials better

recover simpler non-linear trends, whereas splines better recover more complex trends [19]. Functional form will be investigated first using univariate analysis and any apparent non-linear relationships will be considered for inclusion in the LASSO model. The powers to be considered are {−2, −1, −0.5, 0, 0.5, 1, 2, 3} as recommended by Royston and Sauerbrei [20].

7.3 Prognostic Model checking

7.3.1 Overall measures of model fit

Model fit for the logistic multivariable regression models will be investigated using the Cox-Snell R^2 [21] and Nagelkerke's R^2 [22] and the likelihood ratio statistic.

7.3.2 Calibration statistics

Calibration (the extent to which predicted and observed risk of adverse outcome agree) of the logistic regression models will be assessed using the calibration slope (where a slope of 1 indicates perfect calibration), and the observed/expected ratio (ideal value 1). These will be presented for the whole cohort on which each model is built (i.e. for adults and for children separately), and will also be derived for the COVID-19 confirmed and DNR subgroups outlined in section 6.7. Calibration will be visualised graphically using calibration plots of observed versus predicted risk.

7.3.3 Discrimination statistics

Discrimination (the extent to which predictions discriminate between patients who did and did not experience adverse outcome) will be measured using the concordance statistic (C-statistic) for each of the logistic regression models, and will be presented with 95% CI.

7.3.4 Internal validation of predictive models

Cross validation of the prediction models will be conducted using 10-sample cross validation and bootstrapping.

10 sample cross validation

Within each of the 10 cycles, the predictive model is developed in all but the one omitted group, and the model performance is estimated in the omitted group, this is repeated for each of the 10 groups, and the performance statistics (discrimination and calibration) are then averaged across the 10 groups.

Bootstrapping

The same model building steps used to develop the original model are applied to each bootstrap sample. The optimism of the model will be calculated as the difference between apparent performance of the bootstrap model in the bootstrap sample and the test performance in the original sample. This procedure will be repeated for at least 100 bootstrap samples and the average of these estimates obtained in order to obtain an optimism corrected estimate of model performance.
We will also examine model stability by investigating which predictors are selected in each of the bootstrap samples.

8 Triage tool scoring and data definitions

8.1 Definitions

8.1.1 Adverse Outcome

Adverse outcome will be calculated according to the following rules:

- Death is defined using mortality status on the follow up form.
- Needing respiratory, cardiovascular or renal support is defined if any of the following have been recorded in the follow up admissions
    - Respiratory: mechanical ventilation, non-invasive ventilation, CPAP, HFNO;
    - Cardiovascular: ECMO, inotropic/vasopressor drugs, central venous pressure measurement, invasive intra-arterial blood pressure measurement;
    - Renal: haemofiltration, haemodialysis, peritoneal dialysis.

  Any support recorded in 'other' (not flagged as the options above) are to be sent to the chief investigator and a clinical member of the PMG to manually review and categorise where appropriate.

8.2 CURB-65

The CURB-65 score uses five parameters, each scoring 1 point when positive and zero if negative, to give a total score between zero and five.
Five parameters:

1. Confusion: GCS-V is less than 4 or GCS total is less than 15 or AVPU is recorded as V, P or U
2. Urea: Raised blood urea nitrogen over 7mmol/litre
3. Respiratory: rate of 30 breaths per minute or more
4. Blood pressure: diastolic BP is 60mmHg or less or systolic BP is 90 mmHg or less
5. Age: 65 years or more

Missing data: The rules above effectively classify missing data as normal. Low rates of missing data are expected for all parameters except urea. The CRB-65 score [7] is recommended for community settings where access to blood testing is more limited. It is the same as CURB-65 but is a 4-point scale that does not include urea with the threshold remaining as <2 for low risk, 2+ for high risk. Therefore an alternative way of handling missing data for urea is to calculate the CRB-65 score. Further sensitivity analysis may be considered if physiological variables have missing rates > 5%. If a patient has fewer than three of the five parameters complete, the score will not be calculated.

8.3 PMEWS

PMEWS uses six physiological parameters and patient parameters to calculate a score from zero to 19. The score is calculated by taking the score in the table below dependent on each of the six physiological parameters then adding points for two patient parameters after if they are positive.

Physiological:

Score	3	2	1	0	1	2	3
Respiratory Rate	≤8			9-18	19-25	26-29	≥30
SaO_2	<89	90-93	94-96	>96			
Pulse Rate	≤40	41-50		51-100	101-110	111-129	≥130
Systolic BP	≤70	71-90	90-100	>100			
Temperature		≤35.0	35.1-36.0	36.1-37.9	38-38.9	≥39	
Neuro				Alert	Confused Agitated*	Voice	Pain Uncon

* confused/agitated will be defined based on GCS-V<4 or GCS total<15, or if confusion is ticked as a presenting feature

Patient:

1. Add 1 point if age>65
2. Add 1 point if either:
    a. Patient lives alone / no fixed abode or
    b. has a co-morbidity (respiratory, cardiac, renal, immunosuppressed, diabetes)
    c. performance status is more than two suggesting limited activity can self-care, limited activity limited self-care, or bed/chair bound no self-care.

Missing data: If data is missing one or two variables then the normal score (zero) should be assumed. If more than three variables are missing the patient should be excluded. These rules effectively classify

missing data as normal. Low rates of missing data are expected for the physiological parameters and age. The remaining parameters only account for one point and therefore would be within the range of using the patient data. Sensitivity analysis will be considered if physiological variables have missing rates > 5%. If AVPU is missing and GCS is recorded, impute the following AVPU scores using GCS: 0 if GCS=15, 1 if GCS=12-14; 2 if GCS=9-11 and 3 if GCS<9.

8.4 Swine Flu Hospital Pathway

The Swine Flu Hospital Pathway consists of seven criteria, if any one of the criteria meet the threshold they should be admitted. The criteria thresholds are different for adults and children, these are in the table below.

Criteria Label	Criteria	Adult Threshold	Child Threshold
A	Severe respiratory distress	Severe breathlessness (severe respiratory distress ticked on form)	Lower chest wall indrawing, sternal recession, grunting or noise breathing when calm (severe respiratory distress ticked on form)
B	Respiratory rate	Over 30 breaths per minute	50+ breaths per minute if under 1, 40+ breaths per minute if 1+ years
C	Oxygen saturation	≤92% on pulse oximetry, breathing air or on oxygen	≤92% on pulse oximetry, breathing air or on oxygen
D	Respiratory exhaustion	New abnormal breathing pattern (respiratory exhaustion ticked on form)	Exhaustion or apnoeic episode (20+ second pause in breathing) (respiratory exhaustion ticked on form)
E	Dehydration or shock[a]	Systolic BP <90 mmHg and /or diastolic BP <60 mmHg. Sternal capillary refill time >2 seconds, reduced skin turgor	Sternal capillary refill time >2 seconds, reduced skin turgor, sunken eyes or fontanelle
F	Altered conscious level[b]	New confusion, striking agitation or seizures	Strikingly agitated or irritable, seizures or floppy infant
G	Other clinical concern[c]	Another clinical concern	Another clinical concern

a) dehydration or shock will be defined in adults if they are labelled as having severe dehydration, or Systolic BP <90 mmHg and /or diastolic BP <60 mmHg, or central capillary refill is categorised as abnormal. In children it will be defined as having severe dehydration or central capillary refill categorised as abnormal. B) Altered conscious will be positive if GCS is less than 15 or AVPU is anything other than A. C) Other clinical concern is not recorded in the data, two scores will be calculated, one where point G is ignored and the Swine Flu hospital pathway is calculated based on the first 6 items, and another where clinical concern is considered positive if NEWS2>4 or any parameter in NEWS2 is given a score of 3.

Missing data: As the measure only requires one score then missing data will be classified as normal. Low rates of missing data are expected for the physiological parameters however it will not be possible to determine whether failure to tick the relevant boxes represented missing data or not. The measure will not be calculated if fewer than three of the parameters are complete.

8.5 NEWS2

The NEWS2 has seven parameters which are scores from zero to three providing an overall score between zero and 20. The scores for each parameter can be found in the table below.

Score	3	2	1	0	1	2	3
Respiratory Rate	≤8		9-11	12-20		21-24	≥25
SaO_2	≤91	92-93	94-95	≥96			
Pulse Rate	≤40		41-50	51-90	91-110	111-130	≥131
Systolic BP	≤90	91-100	101-110	111-219			≥220
Temperature	≤35.0		35.1-36.0	36.1-38.0	38.1-39.0	≥39.1	
Neuro				Alert			Confusion, Voice, Pain, Unresponsive
Air or Oxygen		Oxygen (based on FiO_2>21%, or FiO_2>0 L/min)		Air			

Missing data: Any missing data will be imputed with the value zero, therefore classifying missing as normal. Low rates of missing data are expected for the physiological parameters. Sensitivity analysis will be considered if physiological variables have missing rates > 5%. The score will not be calculated if fewer than three of the parameters are available. The scale for patients with confirmed hypercapnic respiratory failure will not be used. For ROC analysis the raw score outlined above will be used. To calculate sensitivity and specificity at key cut-offs high risk will be defined as scoring at least 5, or scoring 3 points in any one of the above categories.

8.6 The WHO decision-making algorithm for hospitalisation with pneumonia

The WHO decision making algorithm for hospitalisation with pneumonia suggests an adult patient is admitted (score 1) if any of the following are present:

- respiratory rate >30/minute,

- oxygen saturation <90%,
- respiratory distress,
- age >60,
- any of the following comorbidities; hypertension, diabetes, cardiovascular disease, chronic respiratory disease, renal impairment immunosuppression

A patient will score 0 otherwise. Any missing data will be assumed as normal. A score will not be calculated if fewer than three of the above are complete.

8.6.1 WHO algorithm in children

The WHO decision-making algorithm for hospitalisation with pneumonia suggests a child is admitted (score 1) if any of the following are present:

- respiratory rate: >30/minute if over 5 years old; ≥ 40/min if 1-5 years old; or ≥ 50/min if <1year old
- oxygen saturation <90%
- respiratory distress, respiratory exhaustion or severe dehydration ticked
- AVPU is P or U, or GCS<13
- any of the following comorbidities are present; diabetes, cardiovascular disease, chronic respiratory disease, renal impairment, immunosuppression

A patient will score 0 otherwise. Any missing data will be assumed as normal. A score will not be calculated if fewer than three of the above are complete.

8.7 POPS

Paediatric Observation Priority Score (POPS) will be calculated for the child cohort only. The scoring chart is given below:

Paediatric Observation Priority Score (POPS) Chart

This chart is not a substitute for good clinical judgement and any concerns about the condition of a child should be brought to the attention of a senior nurse or doctor

Age	Score	2	1	0	1	2
Any	Sats	<90%	90-94%	>95%	90-94%	<90%
Any	Breathing	Stridor	Audible grunt or wheeze	No distress	Mild or Moderate Recession	Severe Recession
Any	AVPU	Pain	Voice	Alert	Voice	Pain
Any	Gut Feeling	High level concern	Low level concern	Well	Low level concern	Child looks unwell
Any	Other	Oncology Patient	Significant PMH*		Significant PMH*	Congenital heart disease
0-1	Pulse	<90	90 - 109	110 - 160	161 - 180	180+
	RR	<25	25 - 29	30 - 40	41 - 50	50+
	Temp	<35°	35 – 35.9°	36 - 37.5°	37.6 - 39°	39°+
1-2	Pulse	<90	90 - 99	100 - 150	151 - 170	170+
	RR	<20	20 - 24	25 - 35	36 - 50	50+
	Temp	<35°	35 - 35.9°	36 - 38.4°	38.5 - 40°	40°+
2-5	Pulse	<80	80 - 94	95 - 140	141 - 160	160+
	RR	<20	20 - 24	25 - 30	31 - 40	40+
	Temp	<35°	35 - 35.9°	36 - 38.4°	38.5 - 40°	40°+
5-12	Pulse	<70	70 - 79	80 - 120	121 - 150	150+
	RR	<15	15 - 19	20 - 25	26 - 40	40+
	Temp	<35°	35 - 35.9°	36 - 38.4°	38.5 - 40°	40°+
13-16	Pulse	<50	50 - 60	60 - 100	100 - 110	110+
	RR	<12	12 - 14	15 - 20	20 - 25	25+
	Temp	<35°	35 - 35.9°	36 - 38.4°	38.5 - 40°	40°+

Total Score	Priority
0-1	
2-3	
4-7	
8+	Immediate review

Any child scoring above **8** should be considered for transfer to resus

***Significant PMH includes:**

- Ex-premature
- Syndromic conditions
- Cardiac problems
- Asthma
- Diabetes
- Long term steroids
- All other chronic conditions

An adjusted POPS will be scored based on the availability of data for our cohort.

- **Gut feeling** will not be scored.
- **Breathing** will be given a score of 2 if either severe respiratory distress or respiratory exhaustion is ticked, and a score of 0 otherwise
- **Other** will be given a score of 1 if any of premature, heart disease, asthma, diabetes, steroid therapy or other chronic lung disease is ticked, and 2 points if immunosuppression or active malignancy is ticked.

Any missing data will be assumed as normal. A score will not be calculated if fewer than three of the above are complete.

8.8 COAST

COAST will be scored in children only using the 2018 COAST ED charts; there is a separate chart used to score infants (0-1years), preschool (1-4 years), schoolage (5-12 years) and teenage (13-18 years) which are not included in this SAP. COAST is scored by giving one point for each of

- Doctor/nurse/family concern
- abnormal heart rate for age

- abnormal respiratory rate for age
- abnormal SaO2
- mod/severe respiratory distress
- altered consciousness
- pain score

A modified COAST score will be calculated for children, to reflect the data we are collecting

- Excluding pain score
- Only recording severe respiratory distress
- Only recording score for parental concern (parental anxiety)

Hence, the COAST score is out of a possible 6 and will only be scored if at least 3 of the factors are available.
If at least three of the factors are available the missing factors will be assumed normal (score 0).

8.9 Triage tool summary table

Name	No. of items	Score range	Description	Interpretation of score
CURB-65	5	0 - 5		0-1 = low risk, 2 = intermediate risk, 3-5 = high risk
PMEWS	8	0 - 19		0 – 2 = low risk, 3+ = high risk
Swine flu hospital pathway	7	N/A		Any one item is positive then should be considered a risk
NEWS2	7	0 - 20		0 – 4 low risk, any score of 3 on one or 5+ high risk
WHO	5	0,1		0 = do not admit, 1 = admit
POPS	7	0-14		0-4 low risk 5+ high risk
COAST	6	0-6		0-2 low risk, 3+ high risk

9 Implementation of the analysis plan

This SAP will be used as a work description for the statistician(s) involved in the study. All analyses will be performed by statisticians in the study team, under the supervision of the senior statistician, none of the investigators involved in the trial will perform any of the statistical analyses.

Due to the observational nature of the data, statisticians have export rights to the database and are able to query data in an ongoing manner. Any queries will be communicated to the data manager prior to database lock, and any changes to the database during this time will be documented. The database will be locked after an agreement between the statistician, data manager and study manager. It is expected that no data amendments should be required following database lock. However, if an amendment is required, the process is documented in CTRU SOP DM012.

10 Appendix 1 – Missing data imputation details

Variable	**Deterministic imputation – imputing based on 'normal'**	**Multiple imputation** **What features are associated with this & so used to** predict? NB all will be predicted using outcome also
Model 1 – presenting characteristics		
Demographics		
Age	No. Don't include patients that have this missing	No
Sex	No. Don't include patients that have this missing	No
Ethnicity	No. Conduct sensitivity analysis limited to those with available ethnicity data. Otherwise may have to remove from model dependent on level of missing data.	No
Current status		
Symptom duration	No	Yes: Age, presenting features, comorbidities
Number of current medications	Missing = no medications	Yes: Age, Medical history/complications.
Presenting features -Shortness of breath -cough -fever -sore throat -headache -confusion -rash -anosmia -abdominal pain -diarrhoea -vomiting	Absent/not ticked = no Although prevalence checks will be made for presenting features, where the prevalence is much lower than expected we will assume that feature was not recorded across many sites and will explore removing it from the model	N/a

Lifestyle -Pregnant -clinically obese -tobacco/vape user	Absent/not ticked = no Although prevalence sense checks will be conducted as above.	N/a
Paediatric questions -routine vaccination -taking feeds -parental anxiety -premature	Absent/not ticked = no Although prevalence sense checks will be conducted as above.	N/a
Performance status	Assume unrestricted normal activity	Yes: Age, medical history
known contact with Covid-19 case	Not ticked = no	N/a
Severe respiratory distress	Not ticked = no	N/a
Respiratory exhaustion	Not ticked = no	N/a
Severe dehydration	Not ticked = no	N/a
Medical history		
Previous attendance	Not ticked = no	N/a
Current use of antibiotics	Not ticked = no	N/a
Medical History -Heart disease -renal impairment -steroid therapy -asthma -diabetes -active malignancy -immunosuppression - other chronic lung disease -hypertension	Absent/not ticked = no, unless 'none' i.e. no medication is not ticked, if so assume missing.	Yes (for those with all including 'none' missing) age, number of current medications
Physiology		
Central capillary refill		Yes: age, sex and physiology
Respiratory rate	Yes, when categorised see section 6.3 for details	Yes: age, sex and physiology
Pulse rate	"	"
Temperature	"	"
Systolic Blood Pressure	"	"
SaO_2	"	"
SaO_2/FiO_2 ratio	SaO2 assume normal (100), assume missing FiO2 = 21% and compute the from this	Yes: age, comorbidities and other physiology (respiratory rate, chronic lung, pulse, AVPU and BP).

	Could we impute FiO2 from age, comorbidities and other physiology (or just chronic lung disease and resp rate)? Use of supplemental O2 will be determined by respiratory rate and chronic lung disease (and maybe pulse, AVPU and BP).	
AVPU	Using GCS total – see 7.1.2, otherwise assume A	Yes Using GCS deterministically, otherwise using physiology
Model 2 – including further tests		
Bloods		
Na	Yes, when categorised, see section 6.3 for details	Yes: other bloods, age, sex, comorbidities and physiology
K	"	"
Hb	"	"
Platelets	"	"
WCC	"	"
Neutrophils	"	"
Lymphocytes	"	"
Lactate	"	"
C-Reactive	"	"
Urea	"	"
Creatinine	"	"
D-Dimer		
Troponin		
Other investigations		
CXR	Yes assume normal	"
ECG	Yes assume normal	"

11 References

[1] Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. British Journal of Surgery, 102(3), 148-158.

[2] Bossuyt, P. M., Cohen, J. F., Gatsonis, C. A., & Korevaar, D. A. (2016). STARD 2015: updated reporting guidelines for all diagnostic accuracy studies. Annals of translational medicine, 4(4).

[3] StataCorp. (2019). Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.

[4] Peduzzi, P., Concato, J., Kemper, E., Holford, TR., Feinstein, AR. (1996). A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis. Journal of Clinical Epidemiology. 49:1373-9.

[5] Nguyen-Van-Tam, J., Openshaw P. et al. (2010). Risk factors for hospitalisation and poor outcome with pandemic A/H1N1 influenza: United Kingdom first wave (May-September 2009). Thorax. 65: 645-651.

[6] Campbell, A., Rodin, R. et al. (2010). Risk of severe outcomes among patients admitted to hospital with pandemic (H1N1) influenza. CMAJ. 182:349-55.

[7] Lim, W.S., Van der Eerden, M. M., Laing, R. et al. (2003). Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 58: 377–82.

[8] Challen, K., Bright, J., Bentley, A., Walter, D. (2007). Physiological-social score (PMEWS) vs. CURB-65 to triage pandemic influenza: a comparative validation study using community-acquired pneumonia as a proxy. BMC Health Serv Res. 7:33.

[9] Department of Health. (2009). Swine flu clinical package for use when there are exceptional demands on healthcare services. Available from: http://www.aodp.org/Files/DH_Clinical20Package_Swine20Flu.pdf.

[10] Royal College of Physicians. (2017). National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party. London: RCP.

[11] WHO, “Clinical care of severe acute respiratory infections – Tool kit,” 11 April 2020. [Online]. Available: https://www.who.int/publications-detail/clinical-care-of-severe-acute-respiratory-infections-tool-kit. [Accessed 13 May 2020].

[12] D. L. G. F. P. H. C. W. A. &. D. F. Roland, “The paediatric observation priority score: a system to aid detection of serious illness and assist in safe discharge.,” *Open Journal of Emergency Medicine,,* vol. 4, no. 2, pp. 38-44, 2016.

[13] J. M. &. M. I. K. Sandell, “ Paediatric early warning systems (PEWS) in the ED.,” *Emergency Medicine Journal,* vol. 33, pp. 754-755, 2016.

[14] Tibshirani R. (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B. 58(1): 267-88.

[15] Riley, R. D., van der Windt, D., Croft, P., & Moons, K. G. (Eds.). (2019). Prognosis Research in Healthcare: concepts, methods, and impact. Oxford University Press.

[16] Altman, D.G. & Royston, P. (2000). What do we mean by validating a prognostic model? Stat Med. 19(4):453-73.

[17] Austin, P.C., Steyerberg, E. W. (2014). Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. Stat Med. 33(3):517-35.

[18] Collins, G. S., Ogundimu, E. O. & Altman, D. G. (2016). Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Statistics in medicine. 35(2), 214-226.

[19] H. S. W. &. R. P. Binder, “Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response.,” *Statistics in Medicine,* vol. 32, no. 13, pp. 2262-2277, 2013.

[20] P. a. W. S. Royston, “Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables.,” *Chichester, UK: Wiley,* 2008 .

[21] Cox, D. R., & Snell, E. J. (1989). Analysis of binary data (Vol. 32). CRC press.

[22] Nagelkerke, N. J. D. (1991). A note on the general definition of the coefficient of determination. Biometrika. 78:3, 691-692.