Conventional and Bayesian workflows for clinical prediction modelling of severe Covid-19 outcomes based on clinical biomarker test results: LabMarCS: Laboratory Markers of COVID-19 Severity - Bristol Cohort

Brian Sullivan^a, Edward Barker^a, Philip Williams^c, Louis MacGregor^a, Ranjeet Bhamber^a, Matt Thomas^b, Stefan Gurney^c, Catherine Hyams^{a,b}, Alastair Whiteway^b, Jennifer A Cooper^a, Chris McWilliams^a, Katy Turner^a, Andrew W. Dowsey^a, Mahableshwar Albur^b

^aUniversity Of Bristol ^bSouthmead Hospital. North Bristol NHS Trust ^cUniversity Hospitals Bristol and Weston NHS Trust

Abstract

We describe several regression models to predict severe outcomes in COVID-19 and challenges present in complex observational medical data. We demonstrate best practices for data curation, cross-validated statistical modelling, and variable selection emphasizing recent Bayesian methods. The study follows a retrospective observational cohort design using multicentre records across National Health Service (NHS) trusts in southwest England, UK. Participants included hospitalised adult patients positive for SARS-CoV 2 during March to October 2020, totalling 843 patients (mean age 71, 45% female, 32% died or needed ICU stay), split into training (n=590) and validation groups (n=253). Models were fit to predict severe outcomes (ICU admission or death within 28-days of admission to hospital for COVID-19, or a positive PCR result if already admitted) using demographic data and initial results from 30 biomarker tests collected within 3 days of admission or testing positive if already admitted. Cross-validation results showed standard logistic regression had an internal validation median AUC of 0.74 (95% Interval [0.62,0.83]), and external validation AUC of 0.68 [0.61, 0.71]; a Bayesian logistic regression (with horseshoe prior) internal AUC of 0.79 [0.71, 0.87], and external AUC of 0.70 [0.68, 0.71]. Variable selection performed using Bayesian predictive projection determined a four variable model using Age, Urea, Prothrombin time and Neutrophil-Lymphocyte ratio, with a median internal AUC of 0.79 [0.78, 0.80], and external AUC of 0.67 [0.65, 0.69]. We illustrate best-practices protocol for conventional and Bayesian prediction modelling on complex clinical data and reiterate the predictive value of previously identified biomarkers for COVID-19 severity assessment.

Introduction

Globally as of 26 April 2023, there have been 764 million 2010 2 confirmed cases of COVID-19, including 6.9 million deaths, 21 3 with 24.6 million cases in the UK, resulting in over 207,000 22 4 deaths (WHO Coronavirus (COVID-19) Dashboard, https: 23 5 //covid19.who.int/). COVID-19 has a wide spectrum of 24 6 clinical features ranging from asymptomatic to severe systemic 25 7 illness with a significant attributable mortality, while clinical $_{_{26}}$ 8 manifestations are variable especially in the most vulnerable 27 9 groups and immunocompromised people [1]. COVID-19 is a 28 10 multi-system disease resulting in the derangements of home-29 11 ostasis affecting pulmonary, cardiovascular, coagulation, haema-30 12 tological, oxygenation, hepatic, renal and fluid balance [2, 3, 4, 31 13 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. Although the major-14 ity of people with COVID-19 will have mild or no symptoms, 33 15 a small but significant proportion will suffer from a severe in- $_{_{34}}$ 16 fection needing hospitalisation for supportive care, oxygen, or $_{35}$ 17

36

37

38

admission to intensive care units (ICU) for respiratory support.

Early identification of hospitalised COVID-19 patients who are likely to deteriorate, i.e. transfer to ICU or who may die, is vital for clinical decision making. Healthcare systems across the world including highly developed countries continue to face challenges in terms of capacity and resources to manage this pandemic, as lock down measures have been relaxed, including opening of schools and businesses.

To date, published prediction models have evaluated caselevel factors that might predict poor outcomes (critical illness or death). A recent living systematic review [17] identified 265 prognostic models for mortality and 84 for progression to severe or critical state. The majority of the studies looked at vital signs, age, comorbidities, and radiological features. Models were unlikely to include a broad range of variables concerning co-infection, biochemical factors (outside of C-reactive protein), and other haematological factors on an individual patient level. Most of the prognostic models did not describe the target population or care setting adequately, did not fully describe the regression equation, showed high or unclear risk of bias and/or were inadequately evaluated for performance.

^{*}Dowsey and Albur are joint senior authors

^{**}Corresponding authors

Email address:

brian.sullivan@bristol.ac.uk,mahableshwar.albur@nbt.nhs.uk (MahableshWarFalthis preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

39 Goals

The present study analyzes a range of laboratory blood marker 40 values across metabolic pathways affected by COVID-19 infec-41 tion (i.e. a core set of biomarkers feasible for clinical collec-42 tion) and evaluates predictive models of severe outcomes. The 43 main objectives of the study are: (1) Examine statistical associ-44 ations of routinely measured physiological and blood biomark-45 ers, and age and gender, to predict severe COVID-19 outcomes. 46 (2) Develop cross-validated logistic regression prediction mod-47 els using the candidate biomarkers, highlighting biomarkers wor-48 thy of future research. (3) Use variable selection techniques in-49 cluding least absolute shrinkage and selection operator (LASSO) 50 regularisation [18] and Bayesian Projective Prediction [19] to 51 illustrate the process of creating a reduced model that maintains 52 reasonable performance and is more feasible to use clinically. 53 (4) In each of these steps, demonstrate best analytic practices 54 for explaining clinical data curation and statistical modelling 55 decisions, with an emphasis on showcasing the capabilities of 56 recent Bayesian methods. 57

58 Methods

59 Study Cohort and Demographics

Pseudonymised data was obtained from Laboratory Infor-60 mation Management Systems (LIMS) linking patient data for 61 laboratory markers to key clinical outcomes. Three hospitals in 62 the Southwest region of England, UK, participated in the study, 63 two of which were tertiary teaching hospitals and the third was 64 a district general hospital (DGH). A system-wide data search 65 was conducted on the LIMS for all patients who tested positive 66 for SARS-CoV-2 by polymerase chain reaction (PCR) at these 67 three hospitals during the first wave of COVID-19 pandemic 68 (01/03/2020 to 31/10/2020). The serial pathology data col-69 lected as a part of standard of care of patients admitted with/for 70 COVID-19 were included- bacteriology, virology, mycology, 71 haematology, and biochemistry. All patients testing negative 72 for SARS CoV-2 by PCR were excluded. All laboratory mark-73 ers including clinical outcomes from LIMS were extracted and 74 the final dataset was anonymized with no patient identifying 75 data to link back. 76

77 Inclusion and exclusion criteria

We included all adult patients admitted to study hospitals
and tested positive for SARS CoV-2 by PCR. Pediatric patients
(<18 years old) and staff/healthcare workers and their house-
hold contacts were excluded. Figure 1 depicts the decision
flow for inclusion and exclusion of patient data.

83 Data Covariates

The LabMarCS dataset includes a variety of host, clinical severity indices, microbiological, immunological, haematological and biochemistry parameters used as predictive variables in the regression models. A full list of recorded data items is shown in Figure 2



Figure 1: Flowchart of patient exclusion and inclusion criteria. The initial set of 1159 candidate patients was narrowed to a training set (n=590) and a validation set (n=253).

Biomarker	Abbreviation	Place Recorded / Reason	No. of	% of	Reference Range/ Criteria	Clinical Categories	Criteria Description
Blood Clottina Tests			Readings	Patients			
Activated partial	APTT	Admisson	422	50%	Normal between 21-33 seconds	Normal, Abnormal	Normal: <33; Mild: 33-49.5; Moderate:
Prothrombin Time	РТ	Admisson	435	52%	Normal between 9.5-13	Normal, Abnormal	Abnormal: >=13
Blood Gas Tests				1			1
Carbon Dioxide	CO2	Arterial/ Point of care	154	18%	Normal: 4.6-6.4 seconds	Normal, Abnormal	Abnormal if outside range
Lactate	poctLAC	Arterial/ Point of care	154	18%	0.5-2.2 mmol/L	Normal, Abnormal	Abnormal if <0.5 or >2.2
Oxygen	02	Arterial/ Point of care	154	18%	11.0-14.4 seconds	Normal, Abnormal	Abnormal if <11 or >14.4
Bicarbonate Excess	BE	Arterial or Venous / Point of care	418	50%	22-29	Normal, Abnormal	Abnormal if outside range
pH acid/base scale	рН	Arterial or Venous / Point of care	417	49%	7.35-7.45	Normal, Abnormal	Abnormal if outside these bounds
Coinfection Battery							
Blood Culture	bc_coinfection	Admisson	843	100%	34 bacterial strains tested	Positive, Negative	Postive if one or more postive
Respiratory	resp_coinfection	Admisson	843	100%	34 bacterial strains tested	Positive, Negative	Postive if one or more postive
Urine	urine_coinfection	Admisson	843	100%	34 bacterial strains tested	Positive, Negative	Postive if one or more postive
Viral	viral_cointection	Admisson	843	100%	10 viral infections tested	Positive, Negative	Postive if one or more postive
Diabeles		Deint of Core / Decord		1	1		1
Glucose	Glucose	Often Not Digitized	222	26%	Non-fasting: 3.0-7.8 mmol/L	Normal, Abnormal	Abnormal if outside range
Full Blood Count Tests		Often Not Digitized		I			
Hemoglobin	НВ	Admisson	772	92%	Male 130-170 g/L, Female 120- 150 g/L	Normal, Mild, Moderate, Severe	Normal: >gender specific criteria; Mild: 100 to gender specific criteria; Moderate: 80-100; Severe: <80
Platelet Count	PLT	Admisson	770	91%	150-450 10^9/L	Normal, Mild, Moderate, Severe	Normal: >150; Mild: 100-150; Moderate: 50-100; Severe: <50
Lymphocytes	Lymphocyte	Admisson	772	92%	1.5-4.5 10^9/L	Normal, Mild, Moderate, Severe	Normal 1.5-4.5; Mild 1-1.5; Moderate 0.5- 1; Severe: <0.5 or >4.5
Neutrophils	Neutrophil	Admisson	772	92%	2.0-7.5 10^9/L	Normal, Mild,	Normal 2-7.5; Mild 1-2; Moderate: 0.5-1;
Neutrophil - Lymphocyte						Moderate, Severe	Severe: <0.5 or > 7.5
Ratio	NLR	Admisson	772	92%	0.78 and 3.53	Moderate, Severe	Severe: >18
White Cell Count	wcc	Admisson	772	92%	4.0-11.0 10^9/L	Normal, Mild, Moderate Severe	Normal: 4-11; Mild: 1-4; Moderate: 0.5-1; Severe: <0.5 and >11
Urea & Electrolytes Tests		.		•	•		
C-Reactive Protein	CRP	Admisson	759	90%	< 6 mg/L	Normal, Abnormal	Abnormal if greater than criteria
Estimated Glomerular	eGFR	Admisson	707	84%	>90	Normal, Abnormal	Abnormal if greater than criteria
Urea	urea	Admisson	754	89%	2.5-7 10^9/L	Normal, Abnormal	Abnormal if outside these bounds
Investigatory Tests							
Brain / B-type natriuretic peptide	BNP	Cardiac Function	47	6%	Men under 70: <100pg/ml, Women under 70: <150 pg/ml, All 70yr and over: <300 pg/ml	Normal, Abnormal	Abnormal if greater than age/gender specific criteria
D-Dimer	DDM		111	13%	Age (Years) D-dimer (ng/ml) <60	Normal, Abnormal	Abnormal if greater than age-specific criteria
Ferritin	FER		115	14%	Male: 33-490, Female(0-44): 15- 445, Female(45+yrs): 30-470	Normal, Mild, Moderate, Severe	Normal: <age appropriate="" criteria;<br="" gender="">Mild: >criteria-735; Moderate: 735-2450; Severe: >2450</age>
Fibrinogen	fib		104	12%	1.8-4.0 g/L	Normal, Mild, Severe.	Normal: >1.8; Mild: 1-1.8; Severe: <1
Glycated haemoglobin	HBA1c	Diabetes	17	2%	>=48 mmol/mol	Normal, Abnormal	Abnormal if greater than criteria
Lactate dehydrogenase	LDH	Investigatory	66	8%	240-480 IU/L	Normal, Mild, Moderate, Severe	Normal: <=480; Mild: >480-720; Moderate: >720-1440; Severe: >1440
Procalcitonin	PCT	ITU / Bacterial Infection	39	5%	Normal range: <0.2ng/mL	Normal, Abnormal	Abnormal: >=0.2
Triglycerides	trig	Investigatory	19	2%	0.5-1.7 mmol/L	Normal, Abnormal	Abnormal if outside these bounds
Troponin-T	trop	Cardiac Function	177	21%	Normal: <14ng/L	Normal, Abnormal	Abnormal if greater than criteria
Covid CT	Covid CT		843	100%	Threshold unique to type of test. Lab reports categorical 'positive' variable alongside CT value	Positive, Negative	Only positives included in current study
Other Data							
Age	Age		843	100%		Continuous	All ages >=18
Gender	Gender		843	100%		Male, Female	
Covid Positive on Admission	OnAdmission		843	100%		True, False	Tested only in univariate evaluation
Outcome	Outcome		843	100%		Discharge, ICU, Death	

Figure 2: Variables recorded in the LabMarCS dataset, including plain text description, abbreviation, place of record, frequency in the dataset, and criteria used for converting continuous readings into categorical values.



Figure 3: Example a single patient's time series laboratory biomarker data. See Figure 2 for biomarker abbreviations. Biomarker results are normalised to span 0 to 1 via offsetting by the absolute value of the minimum value and dividing by the maximum value.

89 Outcomes

For all sites, the primary prediction outcome was death or transfer to the ICU within 28 days of the critical date. This critical date was either the point of admission to hospital, or the date of the first positive COVID-19 PCR test result if the₁₂₇ patient was already admitted. This generally corresponds to WHO-COVID-19 Outcomes Scale Score 6–10 (severe) versus¹²⁸ 0–5 (mild/moderate) [20].

97 Patient Timelines

The collected laboratory biomarkers are continuous mea-132 98 sures and provide a time-series representation of the course of a¹³³ 99 patient's admission. Figure 3 shows an example of a single pa-134 100 tient's readings over the course of 18 days between testing pos-101 itive for COVID-19 and being released from hospital care. This¹³⁵ 102 provides a representative example of the heterogeneity seen in_{136} 103 our dataset, i.e. not all tests are taken and others are taken regu-137 104 larly or intermittently (further examples in Supplementary Fig-138 105 ures S2 - S6). 106 139

107 Transformation of Biomarker Data

Prediction modelling of irregularly sampled time-series data₁₄₂ 108 is a challenging open research question [21]. In this study we_{143} 109 focused on established and available tools for conventional and₁₄₄ 110 Bayesian prediction. To balance inclusion of test data not avail-145 111 able on the day of admission and the need for clinical decisions₁₄₆ 112 to be guided soon after admission, we chose to consider the first₁₄₇ 113 value recorded for each biomarker within three days of their₁₄₈ 114 'critical date'. We additionally considered the worst or best₁₄₉ 115 readings within 1, 5 or 7 days, but found the first reading within₁₅₀ 116 3 days to be the most realistic compromise. In addition, we₁₅₁ 117 transformed continuous biomarkers into categorical variables152 118 via reference ranges for clinical use in the typical healthy popu-153 119 lation ranges, see Figure 2. As an example, Figure 4 shows the₁₅₄ 120 histogram of readings for all values recorded for Neutrophils, 155 121 including clinical thresholds to transform into categorical data.156 122 No missing data imputation was performed, instead missing-157 123 ness was coded as as an additional category 'Test not taken'. 158 124 For further elaboration these modelling choices and the chal-125

lenges, please see Discussion section.



Figure 4: Example distribution of biomarker readings for Neutrophil Training and Validation Data. Vertical lines indicate clinical thresholds for bounds on Normal, Mild, Moderate, and Severe categorization.

Statistical Analysis

130

131

140

141

Analytics were carried out using the R statistical language (v4.13) and R Studio (Prairie Trillium release). We used the following packages: Standard logistic regression analyses used the R Stats GLM package (v3.6.2); LASSO analyses, GLMnet (v4.1-4); and for Bayesian analyses, BRMS (v2.17) and ProjPred (v2.1.2). Source code for this analysis pipeline can be found at https://github.com/biospi/LABMARCS.

Analysis of Individual Biomarkers

Before running full regression models, we examined the independent contribution of individual biomarkers in predicting ICU entry or death via standard logistic regressions and Bayesian logistic regressions with either a flat (aka uniform) or horseshoe prior. This allowed calculation of p-values and odds ratios for each biomarker. A 5-fold cross-validation repeated 20 times was run for each biomarker to estimate median AUC and 95% interquartile intervals. Stratified cross-validation data shuffling was precomputed so all models used the same starting data. Stratified cross-validation separates patients by outcome (two groups of patients with severe outcomes and those without), and shuffles both into 5 groups (yielding an 80/20 training/test for each fold). These groups are combined ensuring all training and test datasets reflect the actual portion of patients with severe outcomes for that particular biomarker and not a random sample of that portion, which helps guarantee model convergence for biomarkers with high data missingness.

Only complete cases of training data available for each biomarker were considered, i.e. we did not include data for variables marked 'Test not taken', to focus on the predictive power of test results. Each individual biomarker model included age and gender (except univariate age and gender models) and were compared against a standard model including only age and gender. Regressions were fit using all associated dummy variables for a

216

217

234

235

given biomarker (e.g. 'Mild', 'Moderate', 'Severe') using 'Nor-214
 mal' as the reference. 215

162 Analysis Using All Valid Biomarker Data

After individual biomarker evaluation, logistic regression²¹⁸ 163 models considering all valid biomarkers (Prediction Using In-219 164 dividual Variables section) and demographic variables were fit²²⁰ 165 to the data. Their predictions were tested via internal and ex-221 166 ternal validation using the stratified cross-validation procedures²²² 167 detailed above, expect models were fit using all available train-223 168 ing data using 'Test Not Taken' for absent data. The models²²⁴ 169 include a standard logistic regression, a logistic regression reg-225 170 ularised with LASSO, and two Bayesian models using a flat and²²⁶ 171 a horseshoe prior [22]. LASSO encourages models to converge²²⁷ 172 on sparse solutions with most coefficients set to zero to achieve²²⁸ 173 variable reduction as discussed in the Reduced Variable Models²²⁹ 174 section. Bayesian horseshoe prior models similarly encourage230 175 sparse solutions but without making hard decisions on variable²³¹ 176 inclusion - this can subsequently be performed using Projective232 177 Prediction. 233 178

179 Analysis Using Reduced Variable Models

While a model using all biomarker data may have strong²³⁶ predictive power, it is clinically desirable to have a strong prediction with the least amount of biomarkers possible to save²³⁸ on resources devoted to biomarker collection. We used two²³⁹ methodologies to choose reduced variable models to predict²⁴⁰ COVID-19 severe outcomes, LASSO and Bayesian Projective²⁴¹ Prediction.²⁴²

243 LASSO is an optimization constraint that shrinks parame-187 ters according to their variance, reduces over-fitting, and en-244 188 ables variable selection [18]. The optimal degree of regulari-²⁴⁵ 189 sation is determined by tuning parameter λ within each cross-²⁴⁶ 190 validation fold through a nested cross-validation step. LASSO²⁴⁷ 191 has a drawback of having biased coefficient and log-odds esti-248 192 mates, as such after evaluating LASSO models we run a stan-²⁴⁹ 193 dard GLM model on the reduced biomarker panel selected with $^{^{250}}$ 194 the LASSO. 195

To evaluate LASSO coefficient estimates, we performed re-251 197 peated nested stratified cross-validation (5-folds the for the inner LASSO loop; 5-folds for the outer loop, and 20 repeats).252 198 For a particular dataset fit, LASSO optimises for a sparse rep-253 199 resentation with many coefficients set to zero. Across cross-254 200 validated trials these variables will vary. LASSO fits are statis-255 201 tically biased and are better suited as a guide for variable selec-256 202 tion in a reduced variable standard GLM. As recommended in257 203 Heinze et al [23], we consider the frequency of how often a par-258 204 ticular biomarker has a coefficient greater than zero and count259 205 across cross-validation trials. 260 206

For determining unbiased effect sizes for the reduced vari-261 able set with a standard GLM, it was chosen that if at least one262 categorical level for a particular biomarker (e.g. 'Severe') was263 selected by the LASSO, all levels for that biomarker were in-264 cluded in the model. This resulted in a final set of 'LASSO in-265 spired' variables that were then fit with standard logistic GLM.266 Note this approach, and more generally fitting multiple models to the same dataset, is subject to the problem of selective inference (aka multiple comparison error), see [24, 25] and the related R package [26]. Given our focus is not on reporting p-values, but instead cross-validation and generalisation from training data to validation data, these concerns are minimized.

The second variable selection method explored was Bayesian Projective Prediction [19], a technique for assessing reduced variable models against a complete 'reference' model, which in our case is a Bayesian logistic regression with a horseshoe prior [22]. Priors such as the horseshoe can be applied to provide adaptive shrinkage to covariates in Bayesian models directly so that full posterior distributions of odds estimates can be generated in an unbiased way. Unlike the LASSO, this does not shrink coefficients to zero exactly as the inherent uncertainty is not ignored. To perform hard variable selection, the recent approach of Projective Prediction can be used to compare the fit of submodels of the reference model through projections and approximate leave-one-out (LOO) cross-validation. Under the hood, Projective Prediction uses LASSO (or forward search) to select submodels for comparison, but retains the Bayesian inference for coefficient ranking and odds-ratio estimates. This approach allows one to evaluate the trade-off between AUC performance and the number of variables included in the model and use a reduced model projection at a desired AUC cutoff. Further projective prediction allows the flexibility to train one model on all valid available data, perform variable selection, and then use any projected sub-model with reduced variables to predict outcomes for novel data. Projective prediction models were evaluated using cross-validation procedures described in prior sections. Note, the analysis of the projective prediction model using all training data uses LOO for variable selection, which is computationally intensive. To speed variable selection computation during our cross-validation analysis, we used 'naive' variable selection, which only considers the training data from current fold as is, and does not perform any further internal cross-validation (the projective prediction package allows naive, k-fold, and LOO).

Results

Cohort Description

The initial cohort included 1159 patients which was narrowed down to 843 patients who met all inclusion criteria described above, see Figure 1. 57% of patients were hospitalised for COVID-19 and the remainder had nosocomial infection. For our statistical models, the training cohort (n=590) was defined as all adults admitted to hospital and testing positive for SARS-Cov-2 by PCR, or testing positive while already admitted between March and October 2020. For external validation, we held the DGH cohort (n=253) out of training. Figure 5 depicts the distribution of ages and genders in the training and validation data sets. Patients in the training set had a mean age of 70, were 44% female, and 29% had severe outcomes. The validation set had a mean age of 75, were 47% female, and 38% had a severe outcome.



Figure 5: Distribution of age and gender for hospitalized patients with coronavirus disease 2019 (COVID-19) for (Top) training data (n=590) and (Bottom)³¹⁹ hold out validation data (n=253) cohorts.³²⁰

267 Prediction Using Individual Variables

Figure 6 shows descriptive statistics on individual biomarker, 268 readings and their odds ratio contributions in a 5-fold 20-repeat₃₂₄ 269 stratified cross-validated logistic regression including the par-325 270 ticular biomarker and age and gender. Figure 7 details perfor-326 271 mance using the area under the receiver operating characteristic₃₂₇ 272 curve (AUC) metric, comparing biomarker models (a particular₃₂₈ 273 biomarker plus age and gender) to a model using only age and₃₂₉ 274 gender. Due to the categorical representation of the biomarkers,330 275 individual levels may be significant while another is not (e.g. 276 'Severe' is a predictor, but 'Mild' is not). Statistically signifi-331 277 cant predictors (i.e. odds ratios deviating from one with p-value 278 332 at 0.05 or lower) associated with increasing risk of a severe out-279 333 come (as shown in Figure 6) include age, and the biomark-280 ers: Activated Partial Thromboplastin Time (Mild), Prothrom-281 335 bin time (Abnormal), blood pH (Abnormal), Haemoglobin (Se-282 336 vere), Platelet count (Moderate), Lymphocytes (Moderate, Se-283 vere), Neutrophils (Severe), Neutrophil-Lymphocyte Ratio (Mild, 284 Moderate, Severe), C-Reactive Protein (Abnormal), Urea (Ab-285 normal), and Troponin-T (Abnormal). Nosocomial transmis-286 340 sion was included due to the high number of cases in our cohort 287 but was not a significant predictor and excluded from further $\frac{342}{342}$ 288 analyses. Due to small numbers preventing cross validation, $\frac{343}{343}$ 289 Triglycerides, Glycated Haemoglobin, and Procalcitonin (also) 290 invalid due to being recorded only in ICU) were excluded from $_{345}^{345}$ 291 further analysis and require future research. 292 346

Regression Models Using All Valid Biomarker Data

Each model was evaluated via 5-fold stratified cross-validation with 20 repeats (100 models total). As such, each model is trained with a randomised sample of 80% of the training data set (n=472). Internal validation evaluates model predictions on the 20% (n=118) held out. External validation uses the same model, but is instead tested on the never trained on validation data set (n=253). Missing data for each biomarker is coded as 'Test Not Taken' and is included as a predictor variable. Figure 8 shows the performance of these models (AUC, Sensitivity, Specificity). For comparison, Figure 9 shows the performance of each model using all valid training data (n=590) and testing on the same data (internal validation) and testing on the held out external validation data (n=253).

Models trained on the full data have improved AUC scores, but do not provide a direct uncertainty estimate. For a single model this could be done via bootstrapping, but would not include uncertainty in model fit. Instead we compute interquantile ranges using 5-fold 20-repeat cross-validation of models. Cross-validation results provide 95% inter-quantile ranges that clearly illustrate that in general, all models perform similarly, with a median AUC in the mid 0.70's in internal validation, and near the high 0.60's in external validation. The Bayes model with horseshoe prior slightly outperforms all others, as shown in the AUC difference column showing the distribution of pair-wise differences across folds and repeats. The calibration of the models is reasonably good on the full data, all training data, but has poor calibration on the validation set, see Supplementary Figure S9.

Reduced Variable Models

318

322

The models detailed above are moderately good predictors of severe COVID-19 outcomes, but for clinicians with limited time and resources, reduced models can balance predictive performance with ease of clinical use by using only the most informative biomarkers. To address this, we use two variable selection approaches, LASSO and projective prediction, that allow the creation of reduced models with fewer biomarkers but similar performance to the larger models.

LASSO Models

After performing 5-fold 20 repeat cross-validation we examined the frequency of how often a particular biomarker has a coefficient greater than zero and count across cross-validation trials. Figure S10 shows the frequency of variables having a coefficient great than zero in the cross-validated LASSO analysis. If we select variables that appear at least 50% of the time, our reduced model would include: Age, CRP (abnormal), FER (mild), FIB (mild), HB (severe), PLT (mild, moderate, severe), Lymphocytes (Severe), Neutrophils (Mild, Severe), NLR (Severe), APTT (mild, moderate), PT (abnormal), blood pH (abnormal), Urea (abnormal), and positive viral and blood culture co-infections.

For the reduced variable standard GLM, this resulted in a model using the 15 biomarkers above for all categorical levels, and was evaluated via both cross-validation and as fit to all

					Standard Logistic GLM		Bayesian Logistic (Flat Prior)	Bayesian Logistic (Horse Shoe Prior)
Biomarker	Binary Categorical Variable	% of Patients with Biomarker Recording	# TRUE (% of TRUE Observations with Severe Outcome)	# FALSE (% of FALSE Observations with Severe Outcome)	P Value	Odds Ratio [2.5%, 97.5%]	Odds Ratio [2.5%, 97.5%]	Odds Ratio [2.5%, 97.5%]
Demographics / Other					-		-	
Age	-	100%	-	-	3.22E-05	1.02 [1.01, 1.04]	1.02 [1.01, 1.04]	1.02 [1.01, 1.04]
Gender	Female	100%	257 (26%)	333 (32%)	0.08	0.72 [0.50, 1.03]	0.72 [0.50, 1.03]	0.79 [0.54, 1.07]
Age & Gender	- Female	100%	- 257 (26%)	- 333 (32%)	2.83E-05 0.06	1.02 [1.01, 1.04] 0.70 [0.49, 1.02]	1.02 [1.01, 1.04] 0.70 [0.48, 1.02]	1.02 [1.01, 1.04] 0.83 [0.55, 1.07]
Nosocomial Transmission	TRUE	100%	240 (30%)	350 (29%)	0.65	0.92 [0.63, 1.33]	0.91 [0.63, 1.32]	0.94 [0.70, 1.21]
Blood Clotting Tests				•				
Activated partial	Mild	54%	30 (63%)	291 (320%)	2.44E-03	3.44 [1.57, 7.88]	3.6 [1.52, 8.30]	2.79 [1.12, 6.19]
thromboplastin time	Moderate	54%	4 (100%)	317 (34%)	0.98	9.91E+06 [0.00, NA]	5.3E+106 [7.9E+05, Inf]	6.84 [0.92, 218.16]
Prothrombin Time	Abnormal	56%	45 (58%)	288 (31%)	2.96E-03	2.73 [1.41, 5.36]	2.18 [1.01, 4.67]	2.26 [1.01, 4.63]
Blood Gas Tests	I.u. 1	249/	60 (500()	57 (540()		4 44 [0 70 0 00]		
Carbon Dioxide	Abnormal	21%	68 (59%)	57 (51%)	0.33	1.44 [0.70, 2.99]	1.46 [0.69, 3.01]	1.08 [0.82, 1.92]
Lactate	Abnormal	21%	13 (54%)	112 (55%)	0.96	1.03 [0.32, 3.44]	1.03 [0.32, 3.50]	1.00 [0.58, 1.64]
Divigen Ricarbonato Excoss	Abnormal	64%	103 (33%)	20 (55%)	0.98	1 30 [0 82 2 05]	1.01 [0.37, 2.09]	1.00 [0.05, 1.54]
nH acid/base scale	Abnormal	63%	136 (46%)	232 (31%)	1.05E-04	2 45 [1 56 3 87]	2 48 [1 59 3 96]	2 23 [1 33 3 57]
Coinfection	, ano in an	03/0	150 (40/6)	230 (2070)	1.052 04	2.45 [1.50, 5.67]	2.40 [1.55, 5.50]	2.23 [1.55, 5.57]
Blood Culture	TRUE	100%	5 (0%)	585 (30%)	0.98	3.20E-07 [NA. 2.94E+22]	2.17E-140 [0. 2E-05]	0.45 [0.01, 1.26]
Respiratory	TRUE	100%	6 (50%)	584 (29%)	0.20	2.95 [0.52, 16.62]	2.94 [0.48, 18.21]	1.30 [0.69, 5.62]
Urine	TRUE	100%	12 (25%)	579 (30%)	0.63	0.72 [0.15, 2.53]	0.64 [0.14, 2.46]	0.93 [0.36, 1.56]
Viral	TRUE	100%	7 (71%)	583 (29%)	0.06	4.95 [1.04, 35.13]	5.86 [1.06, 44.82]	1.92 [0.82, 12.53]
Diabetes							-	
Glucose	Abnormal	30%	49 (45%)	126 (32%)	0.11	1.77 [0.88, 3.54]	1.80 [0.90, 3.63]	1.31 [0.85, 2.60]
Full Blood Count Tests	Mild		176 (26%)	269 (27%)	0.12	1 20 [0 01 2 00]	1 27 [0 01 2 09]	1 14 [0 00 1 74]
Hemoglohin	Moderate	92%	48 (33%)	495 (30%)	0.13	1.38 [0.91, 2.08]	1.37 [0.91, 2.08]	1.14 [0.90, 1.74]
inemoglobin	Severe	52/0	11 (55%)	532 (30%)	0.02	4.08 [1.16, 15.06]	4.22 [1.20, 15,15]	1.63 [0.83, 7.16]
	Mild		67 (39%)	474 (29%)	0.07	1.65 [0.95, 2.83]	1.63 [0.93, 2.86]	1.33 [0.90, 2.31]
Platelet Count	Moderate	92%	17 (65%)	524 (29%)	0.01	4.21 [1.54, 12.65]	4.40 [1.53, 13.35]	2.58 [0.97, 8.33]
	Severe		4 (75%)	537 (30%)	0.12	6.16 [0.76, 126.83]	9.34 [0.84, 305.58]	1.86 [0.69, 17.37]
	Mild		151 (27%)	392 (31%)	0.12	1.69 [0.89, 3.34]	1.71 [0.89, 3.26]	1.06 [0.73, 1.86]
Lymphocytes	Moderate	92%	217 (31%)	326 (30%)	0.03	1.96 [1.07, 3.75]	1.99 [1.08, 3.70]	1.21 [0.88, 2.18]
	Severe		84 (48%)	459 (27%)	4.99E-04	3.48 [1.75, 7.17]	3.53 [1.74, 7.11]	2.00 [1.00, 4.35]
Neutrenhile	Mild	0.2%	23 (13%)	520 (31%)	0.23	0.47 [0.11, 1.43]	0.41 [0.09, 1.34]	0.71 [0.22, 1.29]
Neutrophils	Sovoro	92%	3 (33%)	540 (30%)	1 995 02	1./1 [0.08, 19.15]	1.29 [0.04, 21.86]	1.07 [0.30, 3.80]
	Mild		237 (28%)	306 (32%)	3.69E-03	2 50 [1 38 4 79]	2 60 [1.28, 2.92]	1.73 [1.07, 2.70]
Neutrophil - Lymphocyte	Moderate	92%	137 (39%)	406 (27%)	3.18E-05	3.97 [2.12, 7.81]	4.13 [2.19, 8.23]	2.90 [1.42, 5.80]
Ratio	Severe		54 (54%)	489 (28%)	2.61E-06	6.38 [2.99, 14.14]	6.72 [3.06, 15.22]	4.48 [1.91, 10.15]
	Mild		57 (23%)	486 (31%)	0.34	0.72 [0.36, 1.38]	0.70 [0.35, 1.32]	0.84 [0.47, 1.23]
White Cell Count	Moderate	92%	2 (50%)	541 (30%)	0.45	3.03 [0.11, 83.24]	3.11 [0.08, 117.32]	1.15 [0.43, 5.00]
	Severe		85 (42%)	458 (28%)	0.02	1.84 [1.12, 3.00]	1.83 [1.10, 3.00]	1.48 [0.97, 2.68]
Urea & Electrolytes Tests	Abnormal	019/	490 (229/)	47 (49/)	1 405 02	10 22 [2 09 62 44]	12 12 [2 49 77 20]	7 91 [2 44 21 55]
Estimated Glomerular	Abnormal	82%	350 (38%)	131 (18%)	0.06	1 76 [0 98 3 23]	1 79 [0 98 3 28]	1 42 [0 93 2 83]
Filtration Rate Urea	Abnormal	89%	262 (47%)	264 (15%)	4.23E-11	4.27 [2.79, 6.63]	4.33 [2.85, 6.79]	4.13 [2.69, 6.33]
Investigatory Tests				- ()		[[[]]]]		
Brain / B-type natriuretic	Abnormal	70/	20 (52%)	14 (20%)	0.12	2 01 [0 72 27 00]	4 79 [0 76 24 21]	1 52 [0 74 0 02]
peptide	Abhormai	7 76	50 (55%)	14 (29%)	0.15	5.91 [0.75, 27.00]	4.78 [0.76, 54.51]	1.52 [0.74, 9.05]
D-Dimer	Abnormal	12%	52 (42%)	18 (33%)	0.67	1.29 [0.40, 4.43]	1.34 [0.40, 4.88]	1.11 [0.61, 2.62]
	Mild	14%	11 (64%)	72 (39%)	0.09	3.61 [0.84, 17.70]	4.00 [0.85, 19.54]	1.30 [0.78, 5.08]
Ferritin	Noderate	14%	28 (46%)	55 (40%)	0.27	1.79 [0.64, 5.15]	1.85 [0.66, 5.30]	1.09 [0.74, 2.36]
	Severe	14%	b (33%)	77 (43%)	0.94	0.93 [0.11, 5.90]	0.82 [0.09, 5.71]	0.95 [0.35, 1.89]
Fibrinogen	Severe	5%	3 (67%)	20 (46%)	0.10	3.41 [0.23, 105.85]	3.41 [0.23, 105.85]	1.15 [0.46, 5.30]
Glycated haemoglobin*	Abnormal	3%	11 (9%)	4 (0%)	1.00	2.98E+08 [0, NA]	8.30E+11 [0.22, 4E+46]	1.28 [0.32, 15.50]
Lactate debuderance	Mild	6%	12 (67%)	25 (56%)	0.49	2.61 [0.19, 71.00]	3.67 [0.18, 166.51]	1.14 [0.58, 3.98]
Lactate denydrogenase	Moderate	6%	16 (63%)	21 (57%)	0.78	1.81 [0.09, 77.89]	1.81 [0.09, 77.89]	1.01 [0.39, 2.45]
	Severe	6%	5 (60%)	32 (59%)	0.34	4.63 [0.22, 178.20]	7.51 [0.19, 465.95]	1.06 [0.43, 3.67]
Procalcitonin*	Abnormal	4%	21 (86%)	4 (100%)	1.00	1.15E-07 [NA, 1.6E+184]	2.3E-07 [1.05E-31, 5.16]	0.80 [0.08, 2.72]
Triglycerides*	Abnormal	3%	10 (90%)	5 (100%)	1.00	1.68E-09 [NA, Inf]	4.3E-07 [1.98E-30, 1.31]	0.74 [0.04, 3.05]
Troponin-T	Abnormal	24%	91 (44%)	51 (22%)	0.03	2.96 [1.17, 7.96]	3.07 [1.25, 7.91]	1.71 [0.91, 5.40]

* Biomarkers not included in subsequent models due to small sample size, and recorded only in ICU (PCT)

Figure 6: Individual biomarker evaluation including descriptive statistics and logistic regression model outcomes (Standard, Bayesian with flat prior, and Bayes with horseshoe prior), including age and gender (except univariate age and gender models). Regressions were fit using all associated dummy variables for a given biomarker (e.g. normal, mild, moderate, severe) and using only complete cases of training data, i.e. not using a variable for 'Test not taken.' Categorical variables use a reading of 'Normal' as a reference in the fitted model, except 'Male' used as the reference category for gender.

	Standard I	ogistic GLM	Bayesian Log	istic (Flat Prior)	Bavesian Logistic (Horse Shoe Prior)			
	Cross-Validat	ted 80/20 Split	Cross-Validat	ted 80/20 Split	Cross-Validated 80/20 Split			
Demographic / Biomarker	Median AUC [2.5%,97.5%]	Median AUC Difference to Age & Gender Standard [2.5%,97.5%]	Median AUC [2.5%,97.5%]	Median AUC Difference to Age & Gender Standard [2.5%,97.5%]	Median AUC [2.5%,97.5%]	Median AUC Difference to Age & Gender Standard [2.5%,97.5%]		
Demographics / Other								
Age	0.62 [0.52, 0.74]	0.00 [-0.10, 0.03]	0.61 [0.45, 0.74]	0.01 [-0.10, 0.06]	0.61 [0.51, 0.73]	0.00 [-0.13, 0.03]		
Gender	0.54 [0.45, 0.62]	0.07 [-0.06, 0.17]	0.54 [0.46, 0.64]	0.08 [-0.08, 0.20]	0.54 [0.47, 0.62]	0.07 [-0.11, 0.17]		
Age & Gender	0.62 [0.51, 0.71]	0.00, [0.00, 0.00]	0.62 [0.51, 0.73]	0.00, [-0.02, 0.01]	0.61 [0.52, 0.72]	0.00, [-0.02, 0.02]		
Nosocomial Transmission	0.61 [0.47, 0.69]	0.00 [-0.05, 0.04]	0.61 [0.48, 0.69]	0.00 [-0.02, 0.04]	0.61 [0.43, 0.70]	0.00 [-0.11, 0.02]		
Blood Clotting Tests								
Activated partial	0.65 [0.46, 0.79]	0.04[0.22.0.04]	0.65 [0.47, 0.78]	0.04[0.24.0.06]	0.64 [0.45, 0.77]	0.04 [0.15, 0.06]		
thromboplastin time	0.05 [0.40, 0.78]	-0.04 [-0.23, 0.04]	0.65 [0.47, 0.78]	-0.04 [-0.24, 0.06]	0.04 [0.45, 0.77]	-0.04 [-0.15, 0.06]		
Prothrombin Time	0.64 [0.46, 0.77]	-0.03 [-0.22, 0.06]	0.64 [0.47, 0.76]	-0.03 [-0.16, 0.05]	0.64 [0.50, 0.75]	-0.03 [-0.13, 0.04]		
Blood Gas Tests								
Carbon Dioxide	0.54 [0.42, 0.69]	0.01 [-0.16, 0.24]	0.55 [0.42, 0.68]	0.00 [-0.15, 0.21]	0.56 [0.43, 0.69]	0.00 [-0.17, 0.13]		
Lactate	0.57 [0.42, 0.72]	0.00 [-0.20, 0.10]	0.56 [0.40, 0.72]	-0.01 [-0.16, 0.19]	0.57 [0.42, 0.76]	0.00 [-0.23, 0.18]		
Oxygen	0.57 [0.43, 0.74]	0.00 [-0.21, 0.14]	0.56 [0.42, 0.71]	0.00 [-0.16, 0.12]	0.54 [0.42, 0.70]	0.03 [-0.15, 0.18]		
Bicarbonate Excess	0.57 [0.47, 0.70]	0.00 [-0.09, 0.11]	0.57 [0.45, 0.70]	0.00 [-0.08, 0.13]	0.58 [0.43, 0.72]	0.00 [-0.05, 0.13]		
pH acid/base scale	0.65 [0.48, 0.75]	-0.07 [-0.18, 0.08]	0.65 [0.48, 0.74]	-0.07 [-0.26, 0.07]	0.64 [0.47, 0.75]	-0.06 [-0.21, 0.09]		
Coinfection								
Blood Culture	0.62 [0.52, 0.71]	-0.01 [-0.02, 0.00]	0.62 [0.52, 0.72]	-0.01 [-0.03, 0.01]	0.62 [0.51, 0.72]	-0.00 [-0.13, 0.02]		
Respiratory	0.61 [0.51, 0.73]	-0.00 [-0.02, 0.01]	0.62 [0.50, 0.73]	-0.00 [-0.05, 0.02]	0.62 [0.50, 0.73]	-0.00 [-0.02, 0.02]		
Urine	0.61 [0.47, 0.70]	0.00 [-0.00, 0.04]	0.61 [0.48, 0.70]	0.00 [-0.03, 0.04]	0.62 [0.50, 0.71]	0.00 [-0.05, 0.02]		
Viral	0.61 [0.48, 0.76]	-0.00 [-0.02, 0.01]	0.61 [0.47, 0.76]	-0.00 [-0.03, 0.02]	0.61 [0.46, 0.76]	-0.00 [-0.04, 0.03]		
Diabetes								
Glucose	0.60 [0.44, 0.77]	-0.00 [-0.09, 0.15]	0.60 [0.45, 0.76]	-0.01 [-0.11, 0.17]	0.58 [0.46, 0.74]	-0.01 [-0.10, 0.19]		
Full Blood Count Tests								
Hemoglobin	0.61 [0.52, 0.72]	-0.00 [-0.06, 0.07]	0.61 [0.53, 0.72]	-0.00 [-0.06, 0.08]	0.61 [0.51, 0.71]	-0.00 [-0.03, 0.05]		
Platelet Count	0.63 [0.51, 0.74]	-0.01 [-0.07, 0.04]	0.63 [0.51, 0.73]	-0.01 [-0.07, 0.03]	0.63 [0.51, 0.73]	-0.01 [-0.09, 0.04]		
Lymphocytes	0.65 [0.56, 0.76]	-0.03 [-0.10, 0.03]	0.65 [0.56, 0.76]	-0.04 [-0.17, 0.04]	0.63 [0.53, 0.73]	-0.02 [-0.08, 0.03]		
Neutrophils	0.63 [0.53, 0.74]	-0.02 [-0.09, 0.06]	0.63 [0.53, 0.75]	-0.02 [-0.07, 0.06]	0.64 [0.54, 0.74]	-0.02 [-0.07, 0.04]		
Neutrophil - Lymphocyte Ratio	0.67 [0.59, 0.75]	-0.05 [-0.15, 0.04]	0.67 [0.59, 0.75]	-0.05 [-0.14, 0.05]	0.67 [0.58, 0.74]	-0.05 [-0.16, 0.05]		
White Cell Count	0.62 [0.50, 0.70]	-0.01 [-0.08, 0.07]	0.62 [0.49, 0.71]	-0.01 [-0.06, 0.08]	0.63 [0.49, 0.71]	-0.00 [-0.10, 0.06]		
Urea & Electrolytes Tests								
C-Reactive Protein	0.66 [0.54, 0.74]	-0.04 [-0.11, 0.01]	0.66 [0.55, 0.74]	-0.03 [-0.10, 0.02]	0.66 [0.55, 0.74]	-0.04 [-0.12, 0.01]		
Estimated Glomerular Filtration Rate	0.62 [0.48, 0.74]	-0.01 [-0.03, 0.04]	0.63 [0.47, 0.74]	-0.00 [-0.03, 0.05]	0.62 [0.44, 0.74]	0.00 [-0.03, 0.03]		
Urea	0.71 [0.60, 0.81]	-0.09 [-0.19, -0.01]	0.71 [0.60, 0.80]	-0.09 [-0.20, -0.01]	0.70 [0.60, 0.81]	-0.09 [-0.19, -0.00]		
Investigatory Tests								
Brain / B-type natriuretic	0.65 [0.28, 0.05]		0.65 [0.40, 1.00]	0.02 [0.45 0.25]		0.00[0.40_0.25]		
peptide	0.05 [0.38, 0.95]	-0.05 [-0.38, 0.25]	0.65 [0.40, 1.00]	-0.03 [-0.45, 0.25]	0.05 [0.40, 0.90]	0.00 [-0.40, 0.25]		
D-Dimer	0.65 [0.44, 0.86]	0.01 [-0.11, 0.18]	0.65 [0.42, 0.90]	0.00 [-0.13, 0.18]	0.65 [0.40, 0.85]	0.00 [-0.17, 0.14]		
Ferritin	0.60 [0.41, 0.81]	-0.01 [-0.24, 0.23]	0.60 [0.44, 0.83]	-0.01 [-0.30, 0.19]	0.59 [0.41, 0.79]	0.00 [-0.25, 0.21]		
Fibrinogen	0.67 [0.44, 1.00]	0.00 [-0.44, 0.33]	0.67 [0.44, 1.00]	0.00 [-0.44, 0.33]	0.67 [0.44, 1.00]	0.00 [-0.44, 0.33]		
Glycated haemoglobin*	NA	NA	NA	NA	NA	NA		
Lactate dehydrogenase	0.67 [0.42, 1.00]	0.00 [-0.25, 0.33]	0.67 [0.42, 1.00]	0.00 [-0.27, 0.33]	0.67 [0.40, 1.00]	0.00 [-0.33, 0.27]		
Procalcitonin*	NA	NA	NA	NA	NA	NA		
Triglycerides*	NA	NA	NA	NA	NA	NA		
Troponin-T	0.54 [0.39, 0.72]	0.01 [-0.23, 0.21]	0.56 [0.44, 0.72]	-0.00 [-0.23, 0.19]	0.57 [0.44, 0.74]	0.01 [-0.24, 0.16]		

* Biomarkers not included in subsequent models due to small sample size. and recorded only in ICU (PCT)

Figure 7: Predictive performance of the individual biomarker models in Figure 6 as described by the median area under the curve (AUC) in receiver operating curve (ROC) analysis and median difference between an Age and Gender reference model and the same model (negative values indicate the reference has worse performance) with the particular biomarker included (except univariate age and gender models). Regressions were fit using all associated dummy variables for a given biomarker (e.g. mild, moderate, severe) and using only complete cases of training data (n=590), i.e. not using a variable for 'Test not taken.' 95% interquantile ranges calculated via 5-fold cross-validation with 20 repeats (100 models total). Categorical variables use a reading of 'Normal' as a reference in the fitted model, except 'Male' used as the reference category for gender.

		Internal Val	idation	External Validation					
Model	AUC [2.5%, 97.5%]	Specificity at 90% Sensitivity [2.5%, 97.5%]	Specificity at 95% Sensitivity [2.5%, 97.5%]	AUC Difference	AUC [2.5, 97.5]	Specificity at 90% Sensitivity [2.5%, 97.5%]	Specificity at 95% Sensitivity [2.5%, 97.5%]	AUC Difference	
Standard Logistic GLM	0.74 [0.62, 0.83]	0.40 [0.02, 0.61]	0.14 [0.00, 0.48]	-0.04 [-0.14, 0.03]	0.68 [0.61, 0.71]	0.27 [0.18, 0.37]	0.13 [0.02, 0.23]	-0.02 [-0.07, 0.01]	
GLM with LASSO regularisation	0.78 [0.70, 0.85]	0.47 [0.27, 0.63]	0.34 [0.04, 0.54]	-0.01 [-0.04, 0.02]	0.69 [0.66, 0.71]	0.32 [0.24, 0.38]	0.19 [0.14, 0.26]	-0.01 [-0.03, 0.00]	
Bayesian GLM (Flat Prior)	0.74 [0.62, 0.83]	0.42 [0.01, 0.63]	0.23 [0.00, 0.52]	-0.04 [-0.14, 0.02]	0.67 [0.60, 0.70]	0.27 [0.16, 0.36]	0.13 [0.01, 0.22]	-0.03 [-0.08, 0.00]	
Bayesian GLM (Horse Shoe Prior)	0.79 [0.71, 0.87]	0.49 [0.35, 0.68]	0.37 [0.10, 0.58]	Reference	0.70 [0.68, 0.71]	0.34 [0.29, 0.40]	0.23 [0.17, 0.28]	Reference	
LASSO inspired GLM (15 biomarkers)	0.80 [0.73, 0.87]	0.51 [0.34, 0.66]	0.39 [0.01, 0.60]	0.01 [-0.04, 0.05]	0.67 [0.65, 0.69]	0.28 [0.22, 0.34]	0.16 [0.10, 0.20]	-0.03 [-0.04, -0.01]	
Projective Prediction (28 Biomarkers)	0.78 [0.71, 0.87]	0.50 [0.29, 0.65]	0.36 [0.07, 0.56]	-0.01 [-0.02, 0.04]	0.70 [0.68, 0.71]	0.34 [0.28, 0.40]	0.23 [0.16, 0.28]	0 [-0.01, 0.01]	
Projective Prediction (3 Biomarkers)	0.79 [0.78, 0.80]	0.50 [0.46, 0.53]	0.41 [0.34, 0.46]	0.00 [-0.07, 0.08]	0.67 [0.65, 0.69]	0.27 [0.23, 0.31]	0.19 [0.15,0.24]	-0.03 [-0.05, -0.01]	

Figure 8: Cross-validated performance of models trained using valid biomarker data. 95% inter-quantile ranges are presented for each estimate. Specificity is obtained by evaluating at a set sensitivity of either 90% or 95%. All reduced variable models include age, and a stated number of biomarkers. The reduced variable standard GLM uses 15 biomarkers that had non-zero coefficients on >=50% LASSO Cross-validation trials. If at least one categorical level for a particular biomarker (e.g. severe) met this requirement, all levels for that biomarker were included in the model. The 3 biomarker projective prediction model uses all categorical levels for Urea, PT, and NLR. Pairwise AUC difference is presented in comparison to the Bayesian (Horse shoe prior) model.

	Internal Validation					External Validation				
Model	Accuracy	Accuracy AUC Brier Sensitivity Specificity				Accuracy	AUC	Brier	Sensitivity	Specificity
Standard Logistic GLM	0.82	0.87	0.13	0.93	0.56	0.66	0.69	0.13	0.82	0.40
Standard GLM with LASSO regularisation	0.77	0.83	0.23	0.94	0.39	0.62	0.69	0.38	0.93	0.13
LASSO inspired GLM (15 biomarkers)	0.79	0.84	0.14	0.91	0.50	0.67	0.69	0.14	0.88	0.34
Bayesian GLM (Flat Prior)	0.82	0.86	0.18	0.92	0.58	0.64	0.68	0.36	0.79	0.40
Bayesian GLM (Horse Shoe Prior)	0.79	0.84	0.21	0.94	0.45	0.63	0.71	0.37	0.89	0.22
Projective Prediction (28 Biomarkers)	0.79	0.83	0.21	0.94	0.44	0.64	0.71	0.36	0.90	0.24

Figure 9: Performance of models using all valid biomarker data trained on all training data available (n=590). Internal validation is trained on all of the training data and tested on the same. External validation uses the same model and is tested on held out validation data set (n=253). Missing data for each biomarker is coded as 'Test Not Taken'. Specificity and sensitivity evaluated using a probability threshold of 0.5 (i.e. assumes a well-calibrated model). All reduced variable models include age, and a stated number of biomarkers. The reduced variable standard GLM uses 15 biomarkers that had non-zero coefficients on >=50% LASSO Cross-validation trials. If at least one categorical level for a particular biomarker (e.g. severe) met this requirement, all levels for that biomarker were included in the model. The 3 biomarker projective prediction model uses uses all categorical levels for Urea, PT, and NLR.

384

385

available training data. This model had performance very simi-373
lar to the models using all valid biomarker data, with a median374
external validation AUC of 0.68 [0.63, 0.72], see Figures 8 and375
9. 376

Note, 'Test Not Taken' is a significant predictor for LDH₃₇₇ and Lactate on over 50% of cross-validation trials. The poten-₃₇₈ tial significance of missing data is complex and is addressed in₃₇₉ the Discussion section. Due to this confounding, biomarkers₃₈₀ whose top predictive contribution was from 'Test Not Taken'₃₈₁ were excluded from both LASSO reduced variable models and₃₈₂ projective prediction models described below. 383

358 Projective Prediction Models

When all biomarkers were considered, projective prediction₃₈₆ 359 identifies the following predictors in the top 20, in order of con-387 360 tribution to AUC: Urea (abnormal), Age, PT (abnormal), NLR₃₈₈ 361 (Severe), pH (abnormal), Lymphocytes (severe), APPT(mild), 389 362 eGFR (abnormal), Neutrophils (Severe), APPT(moderate), CRP₃₉₀ 363 (abnormal), DDM (abnormal), Hemoglobin (severe). Thus age₃₉₁ 364 and 12 biomarkers are candidates for a reduced model. Note,392 365 several predictors of 'Test Not Taken' were also selected in-366 cluding Lactate, O2, CO2, LDH, Ferritin and Fibrinogen. As 367 mentioned above, these biomarkers are set aside due to this³⁹³ 368 confounding. Supplementary Figures S11 and S12 display 369 the output from projective prediction ranking the contribution 370 of each variable to the model. A model using a projection in-395 371 corporating all biomarker and demographic data is equivalent to396 372

the standard Bayesian GLM we evaluated in the prior section, see Figures 8 and 9.

Reduced variable projections were evaluated by manual inspection of AUC performance among groups of models using the top biomarkers. Guided by the projective prediction ranking, we ran a model using only the top biomarker, using only the top two, the top three, and so on. As described above we omit biomarkers with significant contributions from 'Test Not Taken' and include all categorical levels for a given biomarker as long as one level is highly ranked. Ultimately, we found a 3 biomarker projective prediction model using age and including urea, prothrombin time, neutrophil-lymphocyte ratios had similar performance to larger models with a median internal validation AUC of 0.79 [0.78, 0.80], and external validation AUC of 0.67 [0.65, 0.69], as shown in Figures 8 and 9. Odds ratios for the full Bayesian model and the reduced 3-biomarker model can be found in Supplementary Table S13. The calibration of the model is reasonably good on the training data but has poor calibration on the validation set, see Supplementary Figure S14.

Discussion

Challenges of Complex Medical Data

Curating the LabMarCS data is challenging as the data are heterogeneous in multiple ways. Biomarkers are recorded for

460

461

472

473

different reasons, e.g. routine upon admission, investigatory451
tests, or tests primarily or exclusively taken in ICU. Further,452
some biomarkers are typically recorded together (but not al-453
ways) as part of a test suite, including: Urea and electrolytes,

full blood count, COVID-19 and co-infection swab test, blood₄₅₄ clotting, and blood gas tests (arterial or venous). The schedule₄₅₅ when some these markers are recorded vary by patient and clin-₄₅₆ ical decision, leading to records being present in highly varying₄₅₇ amounts, e.g. only 3% up to 100% of patients depending on the₄₅₈ particular biomarker, see Supplementary Figure S1.

407 Modelling Choices

When constructing and evaluating models, there are many₄₆₂ 408 choice points that should be explicitly highlighted with justifi-463 409 cation, be it based on convenience, computational complexity,464 410 clinical advice, or a heuristic. The space of potential models465 411 is vast and most studies will constrain the model search space,466 412 delineating why these choices are made will facilitate under-467 413 standing and reproduction by other researchers. These include468 414 key choices relating to: patient inclusion/exclusion criteria, data469 415 missingness protocols, data transformations, training and vali-470 416 dation data selection, and performance evaluation. 471 417

418 Missing Data

Missingness, in the context of this study and in healthcare474 419 data more generally, can sometimes be informative and miss-475 420 ing not at random (MNAR), with the presence or absence of a476 421 test correlated with the its measurement or the study outcome. 422 Imputation of missing data relies on key statistical assumptions⁴⁷⁷ 423 that imputed variables are missing at random (MAR) or missing₄₇₈ 424 completely at random (MCAR). Conversations with our clini-479 425 cal co-authors established some routinely collected biomark-480 426 ers might be inferred to be MAR. However, the routines iden-481 427 tified were specific to a small a subset of our cohort and not482 428 likely to extrapolate. We ultimately erred to be conservative and₄₈₃ 429 avoid all imputation, and instead include the presence/absence484 430 of missing values as a covariate itself [27, 28]. As such, in the485 431 current study we chose to use placeholders for 'Test not taken'486 432 if there was no recorded value for a particular biomarker within487 433 434 the evaluated 3-day window. 488

This approach allows the possibility that a 'Test Not Taken'489 435 may be a significant predictor. This has many potential mean-490 436 ings, as it may convey that when a patient is doing well and₄₉₁ 437 unlikely to experience a severe outcome, clinicians are unlikely492 438 to request some biomarker tests. Alternatively, if a patient is in493 439 palliative care and has a poor prognosis, a clinician may con-494 440 sider further testing unnecessary. As such, the likelihood of a495 441 test being administered may follow an inverted-U function as496 442 patients to healthy or too ill may not have tests administered.497 443 Furthermore, as our data was collected early in the pandemic,498 444 445 there may be other underlying clinical decisions or resource499 limitations that drove why some tests were taken but not others.500 446 Lastly, because we only consider results from the first 3 days501 447 from a patients critical date, it may be that some tests were sim-502 448 ply taken later in a patient's stay due to operational constraints,503 449 and hence may be more predictive as they were taken closer to504 450

the outcome. When these instances occurred, we were conservative and excluded biomarkers with 'Test Not Taken' as the most informative category from our reduced variable models.

Data Transforms - Time Windows

Ideally clinicians make decisions based on readings on the day of admission. However, not all tests are administered on admission. To balance inclusion of test data not available on the day of admission and the need for clinical decisions to be guided soon after admission, we chose to consider the first value recorded for each biomarkers within three days of their 'critical date', i.e. date of admission if already COVID-19 positive, or if already in hospital, the date of testing COVID-19 positive. However, given the richness of the time series data collected, further research into models that leverage this extra information is needed.

Focusing on early detection reflects the intent for the model to improve early stage clinical decision making when potential treatments or changes in care may be introduced. This focus on the first reading in a 3-day interval loses information, but greatly simplifies the modelling approach. Note, this choice is not without risk of reducing statistical power, increasing the risk of false positives, and underestimation of the extent of variation in biomarker readings and outcomes between groups [29]. It is likely that representing biomarker data as time-series (assuming regular measures across patients) would add considerable information for continuous monitoring.

Data Transforms - Continuous vs. Categorical

A key modelling decision must be made on whether to use continuous data or transformed categorical data. Clinicians often use biomarker thresholds to provide semantic categories (e.g. normal, mild, moderate, severe) which sometimes use non-linear or discontinuous mappings that require special care if using continuous data. While clinical thresholds are likely established with evidence, it may be the case that thresholds for one use may not apply to a novel use. This led [30, 31] to use machine learning approaches to build categorisation models on continuous biomarker data dependent on the training data at hand. However, using machine learning to establish categorisation thresholds on our biomarker data is difficult with a small training data set and the heterogeneity of biomarker recordings. If missing data imputation is performed, it raises another decision point on whether to impute the continuous or the transformed categorical data.

Another important factor to recognise is that some biomarkers lack a linear relationship between a reading and a semantic category. Biomarkers can have a lower and upper bound for what is considered normal, and both below and above this range reflects clinically meaningful yet sometimes separate abnormalities. The modelling needs to factor in non-linearity when persevering continuous data or trying to map to a categorical space. In our position, categorical transformation had an advantage, as they allowed us to collaborate with ICU consultants while using pre-established clinically acceptable ranges to define our categorisation, see Figure 2. 555

567

568

578

579

586

587

Training and Validation Data Selection 505

There are multiple ways that our data set could be split be-506 tween training and validation sets, e.g. randomly sampling 1/3507 of the data to hold out as a validation set. Random selection 508 of training data should in principle generate data more repre-509 sentative of the validation set left out. However, hospitals may⁵⁵⁹ 510 have differing practices and non-stratified randomization may 511 inflate performance at the cost of real world generalisation. We 512 chose to separate our training and validation datasets by hospi-513 tal to provide a stronger test of generalisation that should mimic⁵⁶³ 514 generalisation to novel hospitals completely outside the original 515 training data . 516 566

Model Performance Evaluation and Dissemination 517

There are a variety of ways statistical model performance₅₆₉ 518 can be evaluated. Here we have chose here to emphasize cross-570 519 validated estimates of AUC, sensitivity, and specificity. Inter-571 520 quartile intervals over these measures reveal that the variety of 572 521 models perform in similar ways. With a larger data set, trade-573 522 offs may become more apparent. Model calibration on the val-574 523 idation set is a clear weak point. While the models have $a_{_{575}}$ 524 reasonable calibration for training data, generalisation perfor-576 525 mance is weak and suggestive of the lack of sufficient data. 526 577

Comparison to Contemporary Models 527

We found several biomarkers previously highlighted by other 528 groups to have significant predictive power, including: Urea,581 529 Neutrophil-Lymphocyte Ratio, Lymphoctyes, APTT, eGFR, and 530 CRP. Our highly reduced 3-biomarker model (plus age) uses 531 Urea (highlighted by all prior models), Neutrophil-Lymphocyte 532 ratio (highlighted by [32, 11, 31]), and APPT (highlighted by₅₈₃ 533 [31]). With a larger dataset, further vetting of these and other₅₈₄ 534 biomarkers would be possible, but it gives reassurance that de-585 535 spite limitations, we find similar predictive biomarkers.

Advantages of Bayesian Modelling 537

536

While the predictive performance across models presented 538 here is generally quite similar, the Bayesian horse shoe model⁵ 539 had slightly better cross-validated predictive performance and 540 there are several reasons for researchers to favor Bayesian ap-590 541 proaches. Coefficients estimated via Bayes should on average₅₉₁ 542 deliver better predictive performance than standard GLM. Ad-592 543 ditionally, if a sparse model is needed, a horseshoe prior can_{593} 544 provide advantages similar to LASSO without biased coeffi-545 cient estimates. Computationally, Bayesian techniques can be 546 slow due Markov Chain Monte Carlo used to sample the coeffi-594 547 cient space. If one is interested in variable selection, projective 548 prediction offers the ability to take a single Bayesian model fit, 549 run a variable selection algorithm to rank variable contributions, 550 and then arbitrarily create sub-model projections with any num-596 551 ber of original variables. While the initial model fit and variable 552 selection are computationally intensive, sub-model projections⁵⁹⁷ 553 598 are fast to create and performance test. 554

Summary & Conclusions

Limitations: This is a retrospective cohort study in Southwest England where case numbers have varied widely, and were below national incidence rates during the first wave. This results in less precise parameter estimates for prediction models (less power/smaller sample size) and likely reduced generalizability of the model to other settings. The timing of biomarker collection was highly varied both within and between patients, with many types of readings missing.

Strengths: The primary strength of our study is the granularity of serial laboratory data available linked to clinical outcomes. This study was performed during the first wave where there was the original Wuhan strain circulating amongst the unvaccinated naïve population without any specific immunomodulating therapies such as steroids or antiviral agents, reflecting the "true" homeostasis derangements at a population level.

In particular, this study describes the variety of challenges present in complex medical data sets and how statistical bestpractices can be applied to such data, highlighting the benefits of recent Bayesian methodology. Our study reiterates the predictive value of previously identified biomarkers for COVID-19 severity assessment (e.g. age, urea, prothrombin time, and neutrophil-lymphocyte ratio). Both the full and reduced variable models have moderately good training performance, but improved external validation is needed for all models to be clinically viable. The methods presented here should generalise well to a larger dataset.

Ethics approval

The study [IRAS project ID: 283439] underwent a rigorous ethical and regulatory approval process, and a favourable opinion was gained from Research Ethics Service, Wales REC 7, c/o Public Health Wales, Building 1, Jobswell Road, St David's Park, SA31 3HB on 11/09/2020.

Funding

This work is funded by Health Data Research UK via the Better Care Partnership Southwest (HDR CF0129), Medical Research Council Research Grant MR/T005408/1, the Elizabeth Blackwell Institute for Health Research, University of Bristol, and the Wellcome Trust Institutional Strategic Support Fund.

Declaration of competing interest

The authors have no competing interests.

Acknowledgements

This research was supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration West (NIHR ARC West). The views expressed in this article are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

599

600

601

673

602 **References**

627

628

629

630

631

632

633

634

644

645

646

647

- 674 [1] A. B. Docherty, E. M. Harrison, C. A. Green, H. Hardwick, R. Pius, 675 603 L. Norman, K. A. Holden, J. M. Read, F. Dondelinger, G. Carson, L. Mer-676 604 son, J. Lee, D. Plotkin, L. Sigfrid, S. Halpin, C. Jackson, C. Gam-677 605 ble, P. W. Horby, J. S. Nguyen-Van-Tam, I. Investigators, J. Dunning, 678 606 P. J. M. Openshaw, J. K. Baillie, M. G. Semple, Features of 16,749 hos-679 607 pitalised UK patients with COVID-19 using the ISARIC WHO Clinical 608 Characterisation Protocol, medRxiv (2020) 2020.04.23.20076042doi: 609 10.1101/2020.04.23.20076042. 610 682
- [2] C. Wu, X. Chen, Y. Cai, J. Xia, X. Zhou, S. Xu, H. Huang, L. Zhang, X. Zhou, C. Du, Y. Zhang, J. Song, S. Wang, Y. Chao, Z. Yang, J. Xu, X. Zhou, D. Chen, W. Xiong, L. Xu, F. Zhou, J. Jiang, C. Bai, J. Zheng, X. Zhou, D. Chen, W. Xiong, L. Xu, F. Zhou, J. Jiang, C. Bai, J. Zheng, Siang, Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumoria in Wuhan, China, JAMA internal medicine 180 (7) (2020) 934–943.
 doi:10.1001/jamainternmed.2020.0994.
- [3] L. Bowles, S. Platton, N. Yartey, M. Dave, K. Lee, D. P. Hart, V. Macbinsing Donald, L. Green, S. Sivapalaratnam, K. J. Pasi, P. MacCallum, Lupus
 Anticoagulant and Abnormal Coagulation Tests in Patients with Covid-19, New England Journal of Medicine 383 (3) (2020) 288–290. doi:
 10.1056/NEJMc2013656.
- [4] N. Tang, D. Li, X. Wang, Z. Sun, Abnormal coagulation parameters are⁶⁹⁵ associated with poor prognosis in patients with novel coronavirus pneu-⁶⁹⁶ monia, Journal of thrombosis and haemostasis: JTH 18 (4) (2020) 844-⁶⁹⁷ 847. doi:10.1111/jth.14768.
 - [5] H. Han, L. Yang, R. Liu, F. Liu, K.-L. Wu, J. Li, X.-H. Liu, C.-L. Zhu, 699
 Prominent changes in blood coagulation of patients with SARS-CoV-700
 2 infection, Clinical Chemistry and Laboratory Medicine 58 (7) (2020)701
 1116–1120. doi:10.1515/cclm-2020-0188.
 - [6] X. Bi, Z. Su, H. Yan, J. Du, J. Wang, L. Chen, M. Peng, S. Chen, B. Shen, J. Li, Prediction of severe illness due to COVID-19 based on an analysis₇₀₄ of initial Fibrinogen to Albumin Ratio and Platelet count, Platelets 31 (5)₇₀₅ (2020) 674–679. doi:10.1080/09537104.2020.1760230.
- [7] F. Liu, L. Li, M. Xu, J. Wu, D. Luo, Y. Zhu, B. Li, X. Song, X. Zhou, Prognostic value of interleukin-6, C-reactive protein, and procalcitonin₇₀₈ in patients with COVID-19, Journal of Clinical Virology: The Official₇₀₉ Publication of the Pan American Society for Clinical Virology 127 (2020)₇₁₀ 104370. doi:10.1016/j.jcv.2020.104370.
- [8] G. Vaseghi, M. Mansourian, R. Karimi, K. Heshmat-Ghahdarijani,712
 P. Rouhi, M. Shariati, S. H. Javanmard, Inflammatory markers in Covid-713
 Patients: A systematic review and meta-analysis, medRxiv (2020)714
 2020.04.29.20084863doi:10.1101/2020.04.29.20084863.
 - [9] Q. Ruan, K. Yang, W. Wang, L. Jiang, J. Song, Clinical predictors of 716 mortality due to COVID-19 based on an analysis of data of 150 patients 717 from Wuhan, China, Intensive Care Medicine 46 (5) (2020) 846–848.718 doi:10.1007/s00134-020-05991-x.
- [10] B. E. Young, S. W. X. Ong, S. Kalimuddin, J. G. Low, S. Y. Tan, J. Loh, 720 648 O.-T. Ng, K. Marimuthu, L. W. Ang, T. M. Mak, S. K. Lau, D. E. An-721 649 derson, K. S. Chan, T. Y. Tan, T. Y. Ng, L. Cui, Z. Said, L. Kurupatham,722 650 M. I.-C. Chen, M. Chan, S. Vasoo, L.-F. Wang, B. H. Tan, R. T. P. Lin,723 651 V. J. M. Lee, Y.-S. Leo, D. C. Lye, Singapore 2019 Novel Coronavirus724 652 Outbreak Research Team, Epidemiologic Features and Clinical Course₇₂₅ 653 of Patients Infected With SARS-CoV-2 in Singapore, JAMA 323 (15)726 654 (2020) 1488-1494. doi:10.1001/jama.2020.3204. 655
- [11] J. Liu, Y. Liu, P. Xiang, L. Pu, H. Xiong, C. Li, M. Zhang, J. Tan, Y. Xu, 728
 R. Song, M. Song, L. Wang, W. Zhang, B. Han, L. Yang, X. Wang, 729
 G. Zhou, T. Zhang, B. Li, Y. Wang, Z. Chen, X. Wang, Neutrophil-to-730
 Lymphocyte Ratio Predicts Severe Illness Patients with 2019 Novel Coro-731
 navirus in the Early Stage, medRxiv (2020) 2020.02.10.20021584doi: 732
 10.1101/2020.02.10.20021584.
- L. E. Gralinski, A. Bankhead, S. Jeng, V. D. Menachery, S. Proll, S. E. 734
 Belisle, M. Matzke, B.-J. M. Webb-Robertson, M. L. Luna, A. K. Shukla, 735
 M. T. Ferris, M. Bolles, J. Chang, L. Aicher, K. M. Waters, R. D. Smith, 736
 T. O. Metz, G. L. Law, M. G. Katze, S. McWeeney, R. S. Baric, Mech-737
 anisms of severe acute respiratory syndrome coronavirus-induced acute 738
 lung injury, mBio 4 (4) (Aug. 2013). doi:10.1128/mBio.00271-13. 739
- [13] Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang, S. Liu, P. Zhao,⁷⁴⁰
 H. Liu, L. Zhu, Y. Tai, C. Bai, T. Gao, J. Song, P. Xia, J. Dong, J. Zhao,⁷⁴¹
 F.-S. Wang, Pathological findings of COVID-19 associated with acute res-⁷⁴²
 piratory distress syndrome, The Lancet Respiratory Medicine 8 (4) (2020)⁷⁴³
 420–422. doi:10.1016/S2213-2600(20)30076-X.

- [14] M. Arentz, E. Yim, L. Klaff, S. Lokhandwala, F. X. Riedo, M. Chong, M. Lee, Characteristics and Outcomes of 21 Critically III Patients With COVID-19 in Washington State, JAMA 323 (16) (2020) 1612–1614. doi:10.1001/jama.2020.4326.
- [15] N. Wilson, A. Kvalsvig, L. T. Barnard, M. G. Baker, Case-Fatality Risk Estimates for COVID-19 Calculated by Using a Lag Time for Fatality - Volume 26, Number 6—June 2020 - Emerging Infectious Diseases journal - CDC, Emerging Infectious Diseases (2020). doi:10.3201/ eid2606.200320.
- [16] H. Barrasa, J. Rello, S. Tejada, A. Martín, G. Balziskueta, C. Vinuesa, B. Fernández-Miret, A. Villagra, A. Vallejo, A. San Sebastián, S. Cabañes, S. Iribarren, F. Fonseca, J. Maynar, Alava COVID-19 Study Investigators, SARS-CoV-2 in Spanish Intensive Care Units: Early experience with 15-day survival in Vitoria, Anaesthesia, Critical Care & Pain Medicine 39 (5) (2020) 553–561. doi:10.1016/j.accpm.2020.04. 001.
- [17] L. Wynants, B. V. Calster, G. S. Collins, R. D. Riley, G. Heinze, E. Schuit, M. M. J. Bonten, D. L. Dahly, J. A. Damen, T. P. A. Debray, V. M. T. de Jong, M. D. Vos, P. Dhiman, M. C. Haller, M. O. Harhay, L. Henckaerts, P. Heus, M. Kammer, N. Kreuzberger, A. Lohmann, K. Luijken, J. Ma, G. P. Martin, D. J. McLernon, C. L. A. Navarro, J. B. Reitsma, J. C. Sergeant, C. Shi, N. Skoetz, L. J. M. Smits, K. I. E. Snell, M. Sperrin, R. Spijker, E. W. Steyerberg, T. Takada, I. Tzoulaki, S. M. J. van Kuijk, B. C. T. van Bussel, I. C. C. van der Horst, F. S. van Royen, J. Y. Verbakel, C. Wallisch, J. Wilkinson, R. Wolff, L. Hooft, K. G. M. Moons, M. van Smeden, Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal, BMJ 369 (2020) m1328. doi:10.1136/bmj.m1328.
- [18] R. Tibshirani, Regression Shrinkage and Selection Via the Lasso, Journal of the Royal Statistical Society: Series B (Methodological) 58 (1) (1996) 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.
- [19] J. Piironen, M. Paasiniemi, A. Vehtari, Projective inference in highdimensional problems: Prediction and feature selection, Electronic Journal of Statistics 14 (1) (2020) 2155–2197. doi:10.1214/20-EJS1711.
- [20] J. C. Marshall, S. Murthy, J. Diaz, N. Adhikari, D. C. Angus, Y. M. Arabi, K. Baillie, M. Bauer, S. Berry, B. Blackwood, et al., A minimal common outcome measure set for covid-19 clinical research, The Lancet Infectious Diseases 20 (8) (2020) e192–e197.
- [21] M. van der Schaar Lab, Time series in healthcare: challenges and solutions, https://www.vanderschaar-lab.com/time-series-in-healthcare/ (2022).
- [22] C. M. Carvalho, N. G. Polson, J. G. Scott, Handling sparsity via the horseshoe, in: Artificial Intelligence and Statistics, PMLR, 2009, pp. 73–80.
- [23] Variable selection A review and recommendations for the practicing statistician - Heinze - 2018 - Biometrical Journal - Wiley Online Library, https://onlinelibrary.wiley.com/doi/full/10.1002/binj.201700067 (2018).
- [24] Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society. Series B (Methodological) 57 (1) (1995) 289– 300.
- [25] J. Taylor, R. J. Tibshirani, Statistical learning and selective inference, Proceedings of the National Academy of Sciences 112 (25) (2015) 7629–7634. doi:10.1073/pnas.1507583112.
- [26] R. Tibshirani, R. Tibshirani, J. Taylor, J. Loftus, S. Reid (Sep 2019). [link].

URL https://github.com/selective-inference/R-software

- [27] R. H. Groenwold, Informative missingness in electronic health record systems: the curse of knowing, Diagnostic and prognostic research 4 (1) (2020) 1–6.
- [28] S. Van Buuren, Flexible imputation of missing data, CRC press, 2018.
- [29] D. G. Altman, P. Royston, The cost of dichotomising continuous variables, BMJ : British Medical Journal 332 (7549) (2006) 1080.
- [30] S. R. Knight, A. Ho, R. Pius, I. Buchan, G. Carson, T. M. Drake, J. Dunning, C. J. Fairfield, C. Gamble, C. A. Green, 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 370 (2020).
- [31] J. Zhou, S. Lee, X. Wang, Y. Li, W. K. K. Wu, T. Liu, Z. Cao, D. D. Zeng, K. S. K. Leung, A. K. C. Wai, et al., Development of a multivariable prediction model for severe covid-19 disease: a population-based study from hong kong, NPJ digital medicine 4 (1) (2021) 1–9.



Figure S1: Heat map displaying missing values across recorded biomarkers. Light blue indicates a value is missing and dark blue indicate it is present



Figure S2: Example biomarker time series for a patient admitted to hospital COVID-19 positive and who subsequently died almost two weeks later.

[32] E. Carr, R. Bendayan, D. Bean, M. Stammers, W. Wang, H. Zhang, 744 T. Searle, Z. Kraljevic, A. Shek, H. T. T. Phan, W. Muruet, R. K. Gupta, 745 A. J. Shinton, M. Wyatt, T. Shi, X. Zhang, A. Pickles, D. Stahl, R. Zak-746 eri, M. Noursadeghi, K. O'Gallagher, M. Rogers, A. Folarin, A. Karwath, 747 K. E. Wickstrøm, A. Köhn-Luque, L. Slater, V. R. Cardoso, C. Bour-748 deaux, A. R. Holten, S. Ball, C. McWilliams, L. Roguski, F. Borca, 749 J. Batchelor, E. K. Amundsen, X. Wu, G. V. Gkoutos, J. Sun, A. Pinto, 750 B. Guthrie, C. Breen, A. Douiri, H. Wu, V. Curcin, J. T. Teo, A. M. 751 Shah, R. J. B. Dobson, Evaluation and improvement of the National Early 752 Warning Score (NEWS2) for COVID-19: A multi-hospital study, BMC 753 Medicine 19 (1) (2021) 23. doi:10.1186/s12916-020-01893-3. 754

755 Supplementary materials



Figure S3: Example biomarker time series for a patient admitted to hospital with subsequent nosocomial transmission and discharge a week later.







Figure S5: Example biomarker time series for a patient admitted to hospital and ICU, with subsequent nosocomial transmission and discharge about one week later.



Figure S6: Example biomarker time series for a patient with two hospital admissions and testing COVID-19 positive on the first, with discharge almost two weeks after second admission.



Figure S7: Distribution of D-Dimer readings with clinical classification requiring age and gender bands

Biomarker	Binary Categorical Variable	P-Value	Odds Ratio CI [2.5%, 97.5%]	
Demographics / Other				
Age Gondor	- Eomalo	6.30E-05	1.04 [1.02, 1.06]	
Blood Clotting Tests	Territale	0.45	1.21 [0.75, 1.57]	
Activated partial	Mild	0.07	2.79 [0.93, 8.35]	
thromboplastin time	Moderate	0.99	3.23E+07 [0, Inf]	
	Not Taken	0.50	0.45 [0.05, 4.51]	
Prothrombin Time	Not Taken	0.46	2.38 [0.24, 23.86]	
Blood Gas Tests				
Carbon Dioxide	Abnormal	0.80	1.13 [0.44, 2.95]	
	Not Taken	NA 0.62	NA	
Lactate	Not Taken	0.03	0.16 [0.04, 0.72]	
Oxygen	Abnormal	0.95	1.05 [0.26, 4.15]	
Oxygen	Not Taken	NA	NA	
Bicarbonate Excess	Abnormal Not Taken	0.24	1.48 [0.76, 2.87]	
	Abnormal	0.13	1.59 [0.87, 2.91]	
pH acid/base scale	Not Taken	0.99	9.4E06 [0, Inf]	
Coinfection				
Blood Culture	TRUE	0.99	0 [0, Inf]	
Urine	TRUE	0.36	0.45 [0.06, 3.24]	
Viral	TRUE	0.02	16.64 [1.7, 162.66]	
Diabetes				
Glucose	Abnormal	0.36	1.55 [0.61, 3.92]	
Full Blood Count Tests	Not Taken	0.60	0.84 [0.44, 1.61]	
Full blood count rests	Mild	0.05	1.76 [1.01. 3.09]	
Hemoglobin	Moderate	0.99	1.01 [0.40, 2.51]	
Hemoglobin	Severe	0.15	4.62 [0.58, 37.05]	
	Not Taken	0.99	1.3E9 [0, Inf]	
	Moderate	0.01	5.81 [1.21, 28.03]	
Platelet Count	Severe	0.10	20.44 [0.57, 734.84]	
	Not Taken	0.99	0 [0, Inf]	
	Mild	0.12	1.98 [0.83, 4.73]	
Lymphocytes	Severe	0.23	2.61 [0.74, 9.22]	
	Not Taken	NA	NA	
	Mild	0.02	0.05 [4E-3, 0.59]	
Neutrophils	Severe	0.73	0.22 [3.1E-5, 1.4E3]	
	Not Taken	NA	NA	
	Mild	0.82	1.1 [0.48, 2.52]	
Neutrophil - Lymphocyte	Moderate	0.80	1.15 [0.39, 3.44]	
Ratio	Severe Not Taken	0.41 NA	1.89 [0.41, 8.69] NA	
	Mild	0.72	0.83 [0.29, 2.38]	
White Cell Count	Moderate	0.73	0.21 [2.5E-5, 1.6E3]	
White cell count	Severe	0.83	1.11 [0.43, 2.83]	
l Irea & Electrolytes Tests	Not Taken	NA	NA	
orcu a Liccitolytes rests	Abnormal	0.07	4.46 [0.91, 21,93]	
C-Reactive Protein	Not Taken	0.94	0.90 [0.07, 12.34]	
Estimated Glomerular	Abnormal	0.41	0.72 [0.33, 1.58]	
Filtration Rate	Not Taken	0.04	0.25 [0.06, 0.95]	
Urea	Abnormal	3.71E-04	2.74 [1.57, 4.77]	
Investigatory Tests	Not Taken	0.95	0.94 [0.11, 7.76]	
Brain / B-type natriuretic	Abnormal	0.72	1.47 [0.18, 11.76]	
peptide	Not Taken	0.81	1.24 [0.21, 7.40]	
D-Dimer	Abnormal	0.31	0.42 [0.08, 2.24]	
	Not Taken	0.23	0.37 [0.07, 1.89]	
	Moderate	0.10	1.52 [0.30, 7.58]	
Ferritin	Severe	0.84	1.30 [0.10, 17.54]	
	Not Taken	0.63	1.30 [0.45, 3.72]	
Elbaire and	Mild	0.21	11.93 [0.26, 552.59]	
Fibrinogen	Severe Not Taken	0.42	1.07 [0.27 4.22]	
	Mild	0.92	14.88 [0.46, 477.41]	
	Moderate	0.31	5.41 [0.20, 145.69]	
Lactate dehydrogenase	Severe	0.48	4.20 [0.08, 217.62]	
	Not Taken	0.61	2.13 [0.12, 39.40]	
Troponin-T	Abnormal	0.68	1.31 [0.37, 4.65]	
Intercept	Not Taken	0.58	1.40 [0.43, 4.53]	
Intercept	-	0.004	0.001 [0.0, 0.11]	

14 Figure S8: Standard logistic regression odds ratio and confidence intervals per biomarker using all valid biomarker training data available (n=590). Note most biomarkers include a 'Test Not Taken' stand in variable.



Figure S9: Model calibration depicting a standard GLM model trained on: (Top) all data and tested on all data (Middle); training data (n=590) and tested on the same; (Bottom) training data and tested on validation data (n=293). A well calibrated model should evenly distribute outcome probabilities, i.e. be close to unity.



Figure S10: Frequency of LASSO logistic regression variables having a coefficient greater or less than 0. Red and black lines indicate thresholds for 20% and 50% frequency.



Figure S11: Heatmap representation of the LOO variable selection output from Bayesian projective prediction ranking predictive power as a function in change of AUC. The color of an individual cell shows the proportion of times in the LOO process a variable was chosen at that particular rank of predictive strength. Note this demonstrates a reduced 15-biomarker model (51 variables total), where biomarkers that had 'Test Not Taken' ranked as their most important predictive element were removed from the model.

		Difference	ELPD	Standard	Difference	Standard
Solution Terms	AUC	Difference	LOO	Error	Difference	Error
<na></na>	0	-0.8	-358.5	9.7	-61.6	10.7
UreaAbnormal	0.5	-0.3	-326.3	11.4	-29.4	8.3
poctLACNA	0.6	-0.2	-310.6	12.2	-13.6	6.4
O2NA	0.6	-0.2	-310.6	12.2	-13.6	6.4
CO2NA	0.6	-0.2	-310.5	12.1	-13.6	6.3
Age	0.8	0	-302.2	11.9	-5.3	5
PTAbnormal	0.8	0	-299.2	12	-2.2	4.3
NLRSevere	0.8	0	-307.6	12.3	-10.7	4
LDHNA	0.8	0	-304.7	12.5	-7.8	3.7
poctpHAbnormal	0.8	0	-302.3	12.4	-5.4	3.4
LymphocytesSevere	0.8	0	-302.9	12.4	-5.9	3.4
APTTMild	0.8	0	-301.4	12.4	-4.5	3.4
eGFRAbnormal	0.8	0	-299	12.4	-2	3.3
NeutrophilsSevere	0.8	0	-301.8	12.6	-4.8	3.1
APTTModerate	0.8	0	-302.5	12.8	-5.6	3
FERNA	0.8	0	-304.7	12.8	-7.8	2.9
fibNA	0.8	0	-302.4	12.8	-5.4	2.8
CRPAbnormal	0.8	0	-303.1	12.7	-6.2	2.8
CO2Abnormal	0.8	0	-301.2	12.8	-4.3	2.7
DDMAbnormal	0.8	0	-302.4	12.7	-5.5	2.6
HBSevere	0.8	0	-302.9	12.8	-6	2.6

Figure S12: Summary statistics of Bayesian projective prediction ranking the contribution of each variable by change in AUC and expected log-predictive density (ELPD)

Biomarker	Binary Categorical Variable	Bayesian Horshoe Odds Ratios CI [2.5%, 97.5%]	Projective Prediction 3-Biomarker Model Odds Ratios CI [2.5%, 97.5%]		
Demographics / Other					
Age	-	1.03 [1.02, 1.05]	1.02 [1, 1.03]		
Gender	Female	0.99 [0.74, 1.3]			
Blood Clotting Tests					
Activated partial	Mild	1.57 [0.87, 4.63]			
thromboplastin time	Moderate	3.24 [0.81, 112.13]			
	Not Taken	0.93 [0.55, 1.32]			
Prothrombin Time	Abnormal	1.66 [0.91, 4.18]	2.32 [1.15, 4.61]		
	Not Taken	0.99 [0.64, 1.55]	0.63 [0.41, 0.9]		
Blood Gas Tests			1		
Carbon Dioxide	Abnormal	1.05 [0.69, 1.81]			
	Not Taken	0.64 [0.13, 1.6]			
Lactate	Abnormal	0.89 [0.36, 1.53]			
	Not Taken	0.63 [0.13, 1.58]			
Oxygen	Abnormal	1.1 [0.66, 2.52]			
	Absormal	0.63 [0.13, 1.56]			
Bicarbonate Excess	Abrioritian Not Takan	1.15 [0.64, 1.91]			
	Absormal	1.1 [0.68, 2.17]			
pH acid/base scale	Abrioritian Not Takan	1.3 [0.91, 2.31]			
Coinfaction	NULTAKEIT	1.17 [0.76, 2.54]			
Blood Culture	TDUE	0.6[0.04 1.4]	1		
Biolog Culture	TRUE	1 1 [0 56 2 10]			
Liripo	TRUE	0.94 [0.20, 3.19]			
Viral	TRUE	1 81 [0 77 14 70]			
Diabetes	INUE	1.01 [0.77, 14.79]			
5.300103	Abnormal	1 1 [0 76 2 01]	1		
Glucose	Not Takar	1.1 [0.70, 2.01]			
Full Blood Count Tests	NULTAKEIT	0.90 [0.00, 1.5]			
run biobu count rests	Mild	1 11 [0 95 1 7]	1		
	Moderate	1.11 [0.85, 1.7]			
Hemoglobin	Severe	1 57 [0 77 8 07]			
	Not Taken	1 3 [0 55 8 27]			
	Mild	1.35 [0.89, 2.83]			
	Moderate	1.58 [0.83, 6.12]			
Platelet Count	Severe	1.41 [0.64, 11.32]			
	Not Taken	0.97 [0.34, 2.29]			
	Mild	1.05 [0.76, 1.6]			
lum also and a	Moderate	1.02 [0.75, 1.48]			
Lymphocytes	Severe	1.2 [0.84, 2.35]			
	Not Taken	1.31 [0.56, 8.86]			
	Mild	0.66 [0.13, 1.27]			
Neutrophils	Moderate	1 [0.33, 2.96]			
	Severe	1.14 [0.82, 1.95]			
	Not Taken	1.31 [0.57, 8.9]			
	Mild	1.04 [0.77, 1.55]	1.32 [0.92, 2.29]		
Neutrophil - Lymphocyte	Moderate	1.08 [0.79, 1.78]	1.98 [1.34, 3.38]		
Ratio	Severe	1.38 [0.86, 3.41]	2.55 [1.5, 5.34]		
	Not Taken	1.31 [0.57, 8.84]	3.22 [0.94, 14.32]		
	IVIIId	0.91 [0.48, 1.35]			
White Cell Count	Noderate	1.04 [0.37, 3.61]			
	Not Takor	1.12 [0.78, 2]			
Lirea & Electrolytes Tests	Not laken	1.25 [.035, 6.15]			
or ca de Lieur oryles rests	Absorget	2.04 [0.02 7.04]			
C-Reactive Protein	Autiormal	2.04 [0.93, 7.61]			
	NOL Taken	1.11 [0.55, 3.54]			
Estimated Glomerular	Abnormal	1.03 [0.71, 1.6]			
Filtration Rate	Not Takon	0 72 [0 26 1 16]			
	Abnormal	2 83 [1 72 / 65]	3 28 [2 2 5 2]		
Urea	Not Taken	1 07 [0 56 2 52]	0.84 [0.25, 1.81]		
Investigatory Tests	Not lakel	1.07 [0.30, 2.33]	0.04 [0.20, 1.01]		
			1		
Brain / B-type natriuretic	Abnormal	1.05 [0.64, 1.99]			
peptide	Not Taken	0.95 [0.53 1 49]			
	Abnormal	1.07 [0.66 2.08]			
D-Dimer	Not Taken	0.85 [0.4, 1.3]			
	Mild	1.43 [0.75, 6.87]			
F 1 1 1	Moderate	1.07 [0.65, 2.14]	1		
Ferritin	Severe	1.01 [0.44, 2.35]			
	Not Taken	0.93 [0.56, 1.34]			
	Mild	1.32 [0.61, 8.49]			
Fibrinogen	Severe	0.92 [0.26, 2.21]			
	Not Taken	0.87 [0.39, 1.37]			
	Mild	1.29 [0.7, 5.43]			
	Moderate	1.07 [0.56, 2.51]	1		
Lactate dehydrogenase	Severe	1.05 [0.47, 2.96]			
	Not Tal	0 71 [0 34 4 33]			
	NOL Taken	0.71 [0.24, 1.22]			
	Abnormal	1 01 [0 68 1 55]			
Troponin-T	Ashorman	1.01 [0.06, 1.55]			
	Not Taken	1.02 [0.71 ,1.57]			
Intercept					
Intercept		0.04 [0.0, 0.28]	0.05 [0.02, 0.11]		

Figure S13: Odds ratios for full Bayesian model and geduced 3-biomarker model via projective prediction



Figure S14: Model calibration depicting Projective Prediction 3-biomarker model tested on: (Top) training data (n=590); (Bottom) validation data (n=293). Note the models do not have points for each of the 10 probability bins because some ranges, e.g. 0.9-1.0 had no patients in this band as judged by the model output. A well calibrated model should evenly distribute outcome probabilities, i.e. be close to unity.