

Original Paper

Cite this article: Ishikawa G, Argenti G, Fadel CB (2021). Non-specific blood tests as proxies for COVID-19 hospitalisation: are there plausible associations after excluding noisy predictors? *Epidemiology and Infection* **149**, e23, 1–14. <https://doi.org/10.1017/S0950268821000078>

Received: 12 October 2020

Revised: 3 January 2021

Accepted: 7 January 2021

Key words:

COVID-19; C-reactive protein; creatinine; eosinophils; hospitalisation; laboratory tests; monocytes; neutrophils; prediction

Author for correspondence:

G. Ishikawa,

E-mail: gersonishikawa@utfpr.edu.br

Non-specific blood tests as proxies for COVID-19 hospitalisation: are there plausible associations after excluding noisy predictors?

G. Ishikawa¹ , G. Argenti² and C. B. Fadel³

¹Professor and researcher, Universidade Tecnológica Federal do Parana (UTFPR), Ponta Grossa, Brazil;

²Researcher, Postgraduate Programme in Health Sciences, Universidade Estadual de Ponta Grossa (UEPG), Ponta Grossa, Brazil and ³Professor and researcher, Universidade Estadual de Ponta Grossa (UEPG), Ponta Grossa, Brazil

Abstract

This study applied causal criteria in directed acyclic graphs for handling covariates in associations for prognosis of severe coronavirus disease 2019 (COVID-19) cases. To identify non-specific blood tests and risk factors as predictors of hospitalisation due to COVID-19, one has to exclude noisy predictors by comparing the concordance statistics (area under the curve – AUC) for positive and negative cases of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2). Predictors with significant AUC at negative stratum should be either controlled for their confounders or eliminated (when confounders are unavailable). Models were classified according to the difference of AUC between strata. The framework was applied to an open database with 5644 patients from Hospital Israelita Albert Einstein in Brazil with SARS-CoV-2 reverse transcription – polymerase chain reaction (RT-PCR) exam. C-reactive protein (CRP) was a noisy predictor: hospitalisation could have happened due to causes other than COVID-19 even when SARS-CoV-2 RT-PCR is positive and CRP is reactive, as most cases are asymptomatic to mild. Candidates of characteristic response from moderate-to-severe inflammation of COVID-19 were: combinations of eosinophils, monocytes and neutrophils, with age as risk factor; and creatinine, as risk factor, sharpens the odds ratio of the model with monocytes, neutrophils and age.

Introduction

Coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) stands out for its high rate of hospitalisation and long hospital stay and in intensive care units (ICUs). COVID-19 disease severity can be mild, moderate, severe and critical [1]. While 81% of those infected with COVID-19 have mild or moderate symptoms, World Health Organization estimates that 14% of those infected with COVID-19 are severe and require hospitalisation and oxygen support, and 5% are critical and admitted to ICUs [1]. Reported median hospital length of stay (LoS) was from 4 to 21 days (outside China) and ICU LoS was from 4 to 19 days [2].

The severity of COVID-19 states is associated with many risk factors. Early reports suggest advanced age, morbidities, multi-morbidities and immunosuppression [3, 4]. The enlarging list includes cardiac, chronic lung, cerebrovascular, chronic kidney and liver diseases, cancer, diabetes, obesity, hypertension, dyspnoea, fatigue and anorexia [1, 5, 6].

Early identification of severe cases allows for optimising emergency care support [1] and improving patient outcomes [7]. However, patients who do not yet meet supportive care criteria may fail to receive the necessary care, when there is rapid deterioration or inability to promptly go to a hospital. In the transition from moderate-to-severe cases there can be avoidable delays in life support interventions with non-optimised treatments.

Together with high hospitalisation rates [1] and lengthy stay [2], the superposition of COVID-19 waves and sustained transmission [8] are causing prolonged depletions of health care resources in many countries. Prognosis tools may play a role in planning and in improving the access to supportive treatments by allowing timely allocation of scarce resources to better cope with COVID-19. Indeed, there is widespread interest in predictive models of COVID-19 outcomes [7, 9], but a review of 50 prognostic models concluded that they are at high risk of bias [9]. As they focus on statistical findings, our concern is with lack of minimum causal criteria to identify associations that are effectively related to COVID-19.

In this context, a path to optimised supportive treatments is more reliable assessments of the transition from moderate-to-severe cases of COVID-19 inflammation. We choose non-specific blood tests as they are widely available, and hospitalisation decision as a proxy to characterise the transition from moderate-to-severe cases (when not constrained by inpatients availability). After formalising an analytical framework with causal reasoning, the goal is to

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

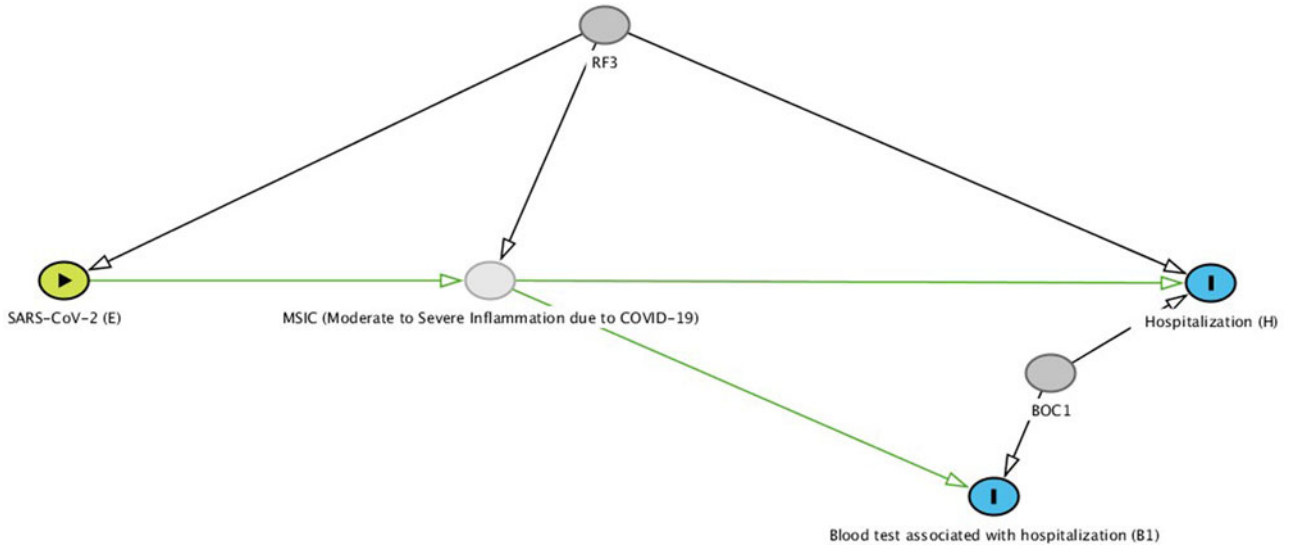


Fig. 1. Initial hypothetical directed acyclic diagram with the main causal path of a moderate-to-severe COVID-19 inflammation (MSIC), one risk factor (RF3) and one confounder (BOC1) of the focal outcomes (H and B1). Legend: MSIC is a latent variable (unmeasured); outcomes are H: hospitalisation (H = {regular ward, semi-intensive care, ICU}); and B: blood test (B = {B1}).

identify candidate sets of blood tests associated with hospitalisation (with risk factors), excluding noisy predictors that are not related to COVID-19 inflammation.

Methods

Whereas causal effects are clearly predictive, prediction studies usually refer to non-causal analysis that uses observational data to make predictions beyond the observed ones and confounding bias is generally considered a non-issue [10]. However when one needs more reliable predictions, confounding bias and causality should be accounted for in associations. This study applies analytical tools from the causal effect estimation of directed acyclic graph (DAG) theory [11] to investigate associations considering covariates.

The strength of the association depends on the specificity and sensitivity of the inflammation pattern, as a kind of distinctive signature of the disease. A low association can also occur and means that the pattern with that set of variables allows weak inferences. If a substantial association due to COVID-19 is identified and it is also stable and representative of the target population, then these blood tests may be useful as proxies in surveillance protocols and screening interventions.

Theoretical framework

The theory of DAG provides graphical notation and a non-parametric probabilistic terminology to describe and evaluate causal relationships [11]. The use of DAGs in epidemiology is emergent [12] and it is especially helpful with multiple potential confounders [12, 13] that may introduce systematic bias [10, 14]. In DAGs, confounding associations between two variables may come from unblocked backdoor paths [13] that can be graphically identified because they share parent nodes. With a formal definition of backdoor path, for instance, DAG provides a general explanation of the Simpson's paradox [15], where a phenomenon appears to reverse the sign of the estimated association in disaggregated subsets in comparison to the whole population. As a

framework, DAG supplies analytical tools to evaluate which adjustment is mandatory (to predict a non-causal sign reverse) and which covariate should be omitted (to estimate the causal effect), thereby enforcing the elicitation of qualitative causal assumptions [11, 12, 14].

A hypothetical DAG model with latent variable was conceived to evaluate the influence of various types of covariates on the focal association. Initially, we drew the main causal path from exposure to outcome. The DAG in Figure 1 starts from the infection by SARS-CoV-2 (exposure E) that, in some cases, leads to 'Moderate-to-severe inflammation due to COVID-19' (MSIC, hypothetical latent variable ($E \rightarrow MSIC$)), and that inflammation causes two outcomes (mutual dependent relationship ($H \leftarrow MSIC \rightarrow B$)): (H) hospitalisation decision; and ($B = \{B_1, \dots, B_k\}$) blood tests measured at hospital admission. The blood tests are selected according to their strength with hospitalisation. The focal outcomes under investigation are hospitalisation (H) and blood tests (B).

Considering the initial DAG plausible, we hypothesised candidate covariates that are parents of the variables and may open back-door paths, Figure 1 shows one risk factor (RF3) and one confounder (BOC1). Figure 2 is an enhancement of the initial DAG with potential risk factors, confounders of the focal association and other covariates. Risk factors contribute directly to the development of COVID-19 inflammation ($RF = \{RF_1, \dots, RF_L\}$, mutual causation relationships ($RF_i \rightarrow MSIC \leftarrow RF_j$)) and they can also affect other variables. Figure 2 also distinguishes the covariates in terms of their confounding potential on the association between H and B. Covariates that affect both focal outcomes are identified as Both-Outcomes-Confounders ($BOC = \{BOC_1, \dots, BOC_m\}$), as they are correlated to the focal outcomes but not to COVID-19, and when affect one outcome as Single-Outcome-Covariate ($SOC = \{SOC_1, \dots, SOC_n\}$). These covariates are not exhaustive but to generate causal graph criteria for handling confounding factors.

Causal relationships in DAGs are defined with the $do(\cdot)$ operator that performs a theoretical intervention by holding constant the value of a chosen variable [11, 16]. The association caused by COVID-19 inflammation can be understood as a comparison of

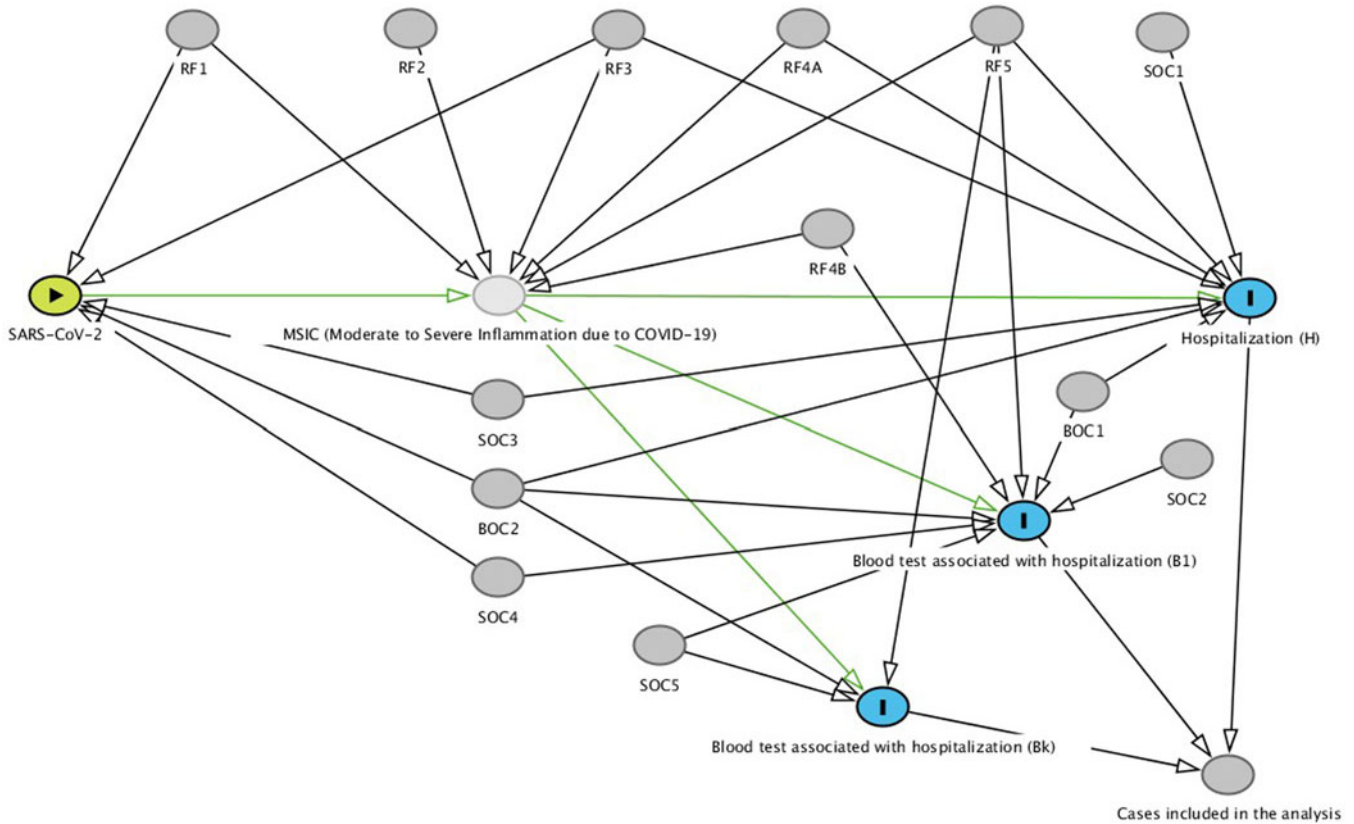


Fig. 2. Hypothetical directed acyclic diagram of a COVID-19 inflammation causal path with risk factors, confounders and other covariates. Legend: Exposure = SARS-CoV-2 (E) (acute respiratory syndrome coronavirus 2); outcomes are H: hospitalisation (H = {regular ward, semi-intensive care, ICU}), and B: blood tests (B = {B1, ..., BK}); Covariates are RF: risk factor (RF = {RF1, ..., RF4A, RF4B, RF5}), SOC: single outcome covariate (SOC = {SOC1, ..., SOC5}) and BOC: both outcomes confounder (BOC = {BOC1, BOC2}).

the conditional probabilities of hospitalisation (H) given a set of blood tests (B) under intervention to SARS-CoV-2 infection ($do(SARS-CoV-2) = 1$) and intervention without infection ($do(SARS-CoV-2) = 0$):

$$P[H|B = b, do(SARS-CoV-2 = 1)] \quad (1)$$

$$P[H|B = b', do(SARS-CoV-2 = 0)] \quad (2)$$

where $P(H|B = b, do(SARS-CoV-2 = 1))$ represents the population distribution of H (hospitalisation) given a set of blood tests equal to b , if everyone in the population had been infected with SARS-CoV-2. And $P(H|B = b', do(SARS-CoV-2 = 0))$ if everyone in the population had not been infected. Of interest is the comparison of these distributional probabilities for each intervention.

The interventions with $do(\cdot)$ generate two modified DAGs:

- The $do(SARS-CoV-2 = 0)$ eliminates all arrows directed towards SARS-CoV-2 and to MSIC (Fig. 3). Ignoring the floating covariates, there are single arrow covariates pointing to hospitalisation (RF3, RF4A, SOC1, SOC3) and to blood tests (RF4B, SOC2, SOC4) and fork covariates pointing to both outcomes (BOC1, BOC2, RF5).
- Similarly, the modified graph of $do(SARS-CoV-2 = 1)$ is equal to the former by adding single arrows from RF1 and RF2 to MSIC; and converting RF3, RF4A, RF4B and RF5 to fork types with arrows directed to MSIC.

As most covariates are either unmeasured or unknown, the effect of their absence can be evaluated following the d-separation concept [11]. This concept attempts to separate (make independent) two focal sets of variables by blocking the causal ancestors (or back-door paths) and by avoiding statistical control for mutual causal descendants [11]. Differently, to preserve the association between descendants of MSIC (Fig. 2), the focal outcomes (H and B) must remain d-connected (dependent on each other only through MSIC) and their relations with other covariates (that may introduce systematic bias) should be d-separated (conditionally independent). Figure 3, at the negative stratum, shows the confounders that may introduce systematic bias into both outcomes: BOC1, BOC2, RF5. The influence of these confounders on the focal association can be estimated with the modified model at the negative strata. A strong association of the outcomes without infection can be due to these confounders and suggest efforts to measure and control for them (as they have to be d-separated). Another pragmatic possibility is to exclude the noisy exams affected by these confounders. The other covariates are single arrows or they affect only one outcome (H or B) – their absence should not be critical because they are likely to be discarded due to poor discriminative performance.

Model assessment with naïve estimation

A naïve estimation of equations (1) and (2) is to assume that they are equal to their conditional probabilities available in a given dataset at each stratum. The cost of this simplification is that the analysis is no longer causal (in a counterfactual sense, because

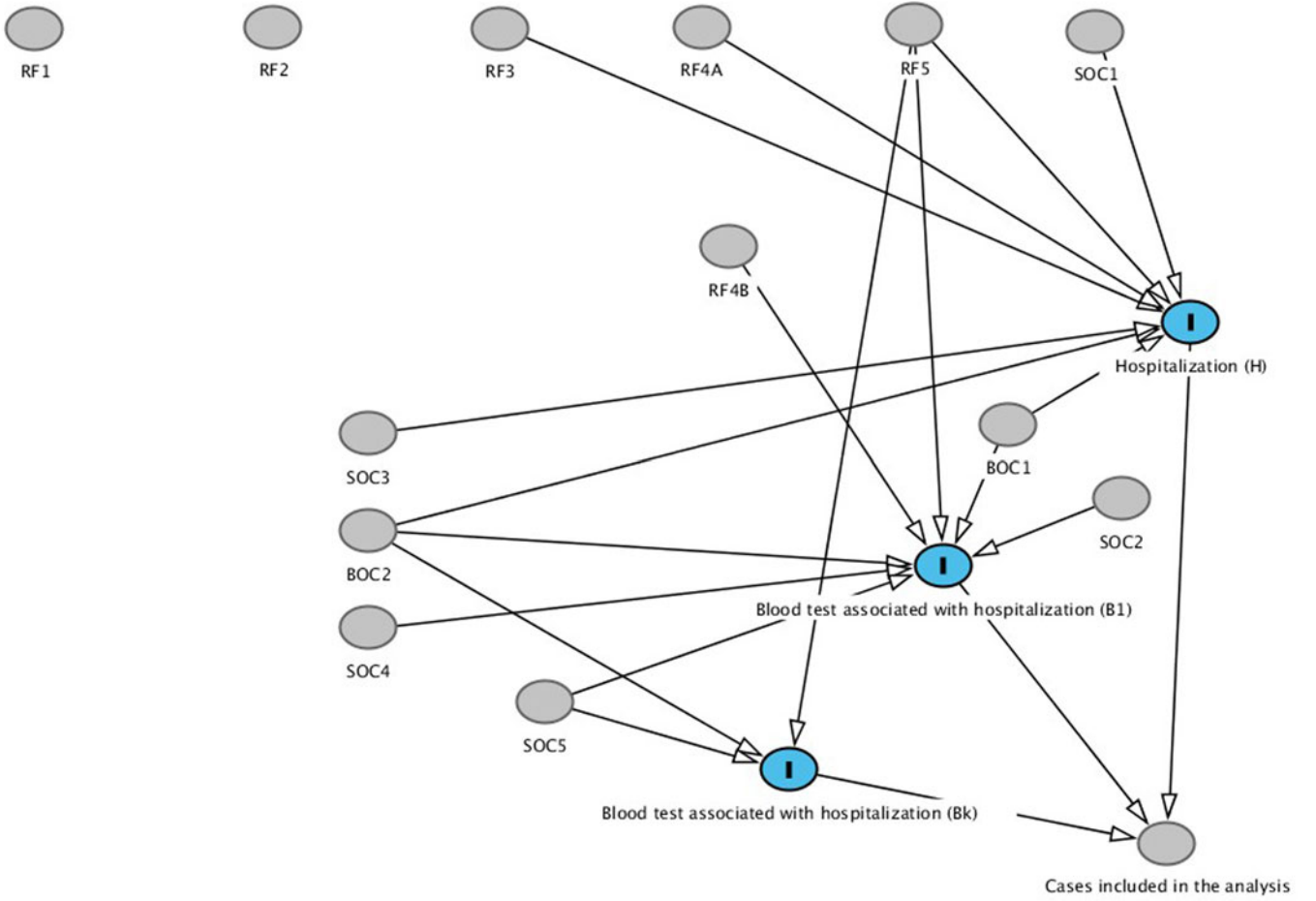


Fig. 3. Modified directed acyclic diagram with intervention at no exposure ($do(SARS-CoV-2 = 0)$) to evaluate the influence of covariates on the focal outcomes (H and B). Legend: Exposure = SARS-CoV-2 (E) (acute respiratory syndrome coronavirus 2); Outcomes are H: hospitalisation ($H = \{\text{regular ward, semi-intensive care, ICU}\}$), and B: blood tests ($B = \{B1, \dots, Bk\}$); Covariates are RF: risk factor ($RF = \{RF1, \dots, RF4A, RF4B, RF5\}$), SOC: single outcome covariate ($SOC = \{SOC1, \dots, SOC5\}$) and BOC: both outcomes confounder ($BOC = \{BOC1, BOC2\}$).

we are not contrasting the whole population infected and the whole population not infected [10, 11, 16]) and the estimation becomes an association between two disjoint sets that each represents separate parts of the target population.

$$\begin{aligned} P[H|B = b, do(SARS-CoV-2 = 1)] \\ = P[H|B = b, SARS-CoV-2 = 1] \end{aligned} \quad (3)$$

$$\begin{aligned} P[H|B = b', do(SARS-CoV-2 = 0)] \\ = P[H|B = b', SARS-CoV-2 = 0] \end{aligned} \quad (4)$$

As hospitalisation is a dichotomous variable, this conditional probability, $P(H|B = b, SARS-CoV-2 = 1)$, can be computed through a logistic regression of hospitalisation (dependent variable) given a set of blood tests at $SARS-CoV-2 = 1$. From the modified graph with intervention, $P(H|B = b', SARS-CoV-2 = 0)$ is calculated with the same model parameters but applied to cases at the negative stratum. It is implicit that there is the conditioning by a proper set of covariates for each model.

The concordance statistic of a logistic regression model is a measure of its predictive accuracy and is calculated as the area under curve (AUC) of the receiver operating characteristic

(ROC) [10, 17]. A way to compare the discriminative ability of (3) and (4) is to subtract the AUC values at each stratum. A difference of 0.0 means no specific association with COVID-19 (i.e. equivalent responses for both strata) and 0.5 means perfect focal association of the outcomes and perfect differentiation among strata (i.e. perfect response at the positive stratum and random response at the negative).

$$\begin{aligned} \Delta_{\text{Discriminative Ability Naive}} \\ = AUC(P[H|B = b, SARS-CoV-2 = 1]) \\ - AUC(P[H|B = b', SARS-CoV-2 = 0]) \end{aligned} \quad (5)$$

The comparison of the models with AUC values at the negative stratum of SARS-CoV-2 is a necessary improvement in the assessment of prognostic models. This is similar to the null values concept in measures of associations of two groups with two outcomes [10], but generalised for continuous multivariable prognostic models.

Model selection criteria

The above framework guided our approach to identify sets of blood tests associated with the hospitalisation due to COVID-19 together with

Table 1. Univariate logistic regression models with blood tests for predicting hospitalisation

	At SARS-CoV-2 = 1						At SARS-CoV-2 = 0					
	N	B	P	OR	OR 95% CI		N	B	P	OR	OR 95% CI	
					Lower	Upper					Lower	Upper
zBasophils	83	-0.374	0.229	0.688			519	-0.375	0.010	0.687		
zHaematocrit	83	-0.123	0.658	0.884			520	-0.976	0.000	0.377		
zHaemoglobin	83	-0.073	0.785	0.930			520	-1.009	0.000	0.365		
zLeucocytes	83	0.617	0.167	1.854			519	0.658	0.000	1.931		
zMCH	83	-0.253	0.280	0.776			519	-0.289	0.011	0.749		
zMCHC	83	0.118	0.629	1.126			519	-0.259	0.023	0.772		
zMCV	83	-0.331	0.176	0.718			519	-0.196	0.094	0.822		
zMPV	81	-0.465	0.079	0.628			518	-0.229	0.062	0.795		
zPlatelets	83	-0.272	0.433	0.762			519	0.101	0.363	10.107		
zPotassium	58	-0.482	0.145	0.618			313	0.161	0.210	1.174		
zRed_blood_cells	83	0.087	0.707	1.091			519	-0.791	0.000	0.453		
zRDW	83	0.140	0.560	1.150			519	0.648	0.000	1.912		
zSerum_glucose	33	-0.172	0.734	0.842			175	0.713	0.001	2.041		
zSodium	58	-0.530	0.097	0.589			312	-0.232	0.077	0.793		
zUrea	59	0.468	0.275	1.597			338	0.403	0.004	1.496		
Age_quantile ^a	558	0.199	0.000	1.220	1.137	1.310	5086	-0.03	0.044	0.968	0.938	0.999
zCreatinine ^b	62	1.002	0.019	2.723	1.177	6.301	362	-0.116	0.367	0.891	0.693	1.145
zCRP ^b	70	1.857	0.004	6.406	1.805	22.73	436	1.012	0.000	2.751	2.015	3.756
zEosinophils ^b	83	-2.768	0.001	0.063	0.012	0.332	519	-0.312	0.036	0.732	0.547	0.980
zLymphocytes ^b	83	-0.794	0.006	0.452	0.256	0.796	519	-0.537	0.000	0.584	0.451	0.758
zMonocytes ^b	83	-0.629	0.006	0.533	0.339	0.838	518	-0.321	0.021	0.726	0.552	0.953
zNeutrophils ^b	75	1.412	0.000	4.104	1.957	8.605	438	0.509	0.001	1.663	1.244	2.224

SARS-CoV-2 (acute respiratory syndrome coronavirus 2): result of the exam for SARS-CoV-2 RT-PCR (0 = negative; 1 = positive) (reverse transcription – polymerase chain reaction).

N, Cases included in the analysis; B, coefficient of the univariate logistic regression; P, coefficient significance; OR, odds ratio (exp(B)); CI, confidence interval. MCH, mean corpuscular haemoglobin; MCHC, mean corpuscular haemoglobin concentration; MCV, mean corpuscular volume; MPV, mean platelet volume; RDW, red blood cell distribution width; zName, means that the variable was converted and made available in a standardised format (mean = 0; standard deviation = 1).

^aAge was converted in quantiles in the range of 0–19, mean value is 9.32.

^bBlood tests selected for screening as potential predictors of COVID-19 inflammation.

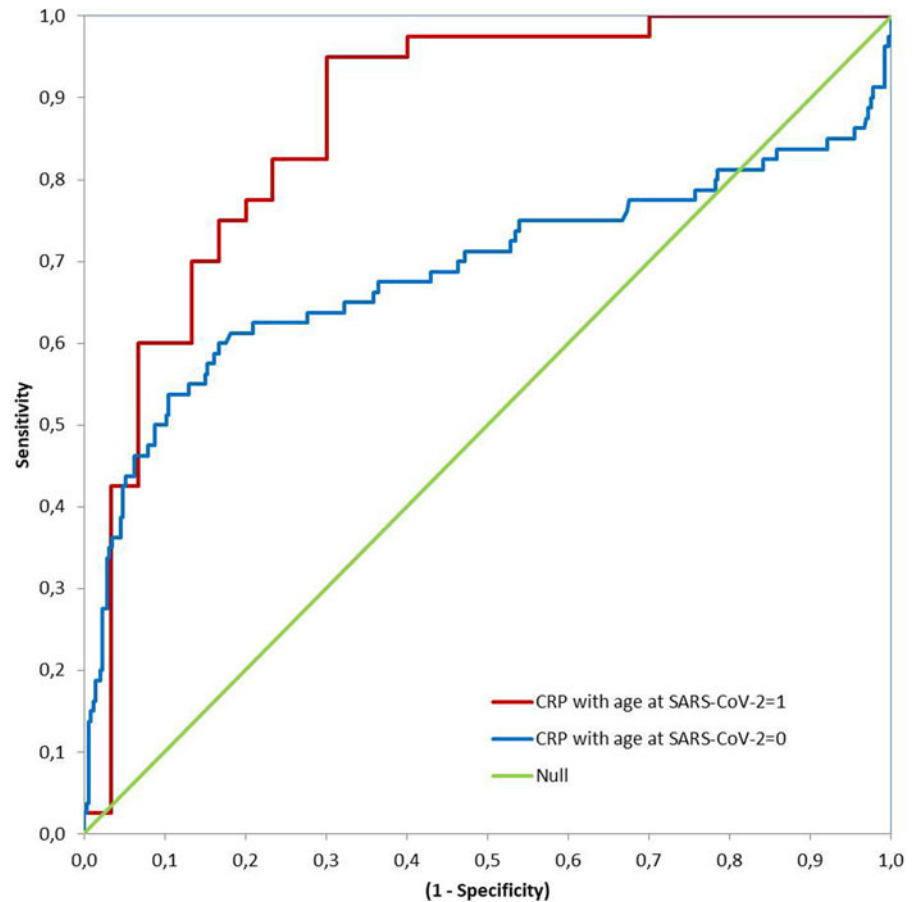


Fig. 4. ROC curves of the logistic regression model for hospitalisation prediction with CRP controlled for age quantile at both strata (with and without exposure to SARS-CoV-2). Legend: Null – area of the null hypothesis model is 0.5.

- Acceptable overall statistical properties of each model at the positive stratum of SARS-CoV-2, without and with bootstrap procedure.
- Consistency of the blood test coefficients across models with one variable and with multiple variables: considering causal effects, coefficients should not change signal when properly conditioned across models [15].
- Elimination of models with high AUC at the negative stratum of SARS-CoV-2 and classification of the sets of blood tests by the difference of AUC between strata.

Source dataset

We identified one public observational database in which, at least partially, we could apply the framework and generate candidate prognostic models. Hospital Israelita Albert Einstein (HIAE), Sao Paulo/Brazil, made public a database (HIAE_dataset) [18] in the *kaggle* platform of 5644 patients screened with SARS-CoV-2 RT-PCR (reverse transcription–polymerase chain reaction) exam and a few collected additional laboratory tests during a visit to this hospital from February to March 2020. All blood tests were standardised to have mean of zero and unitary standard deviation. As this research is based on public and anonymised dataset, it was not revised by any institutional board. The logistic regression models were evaluated with IBM SPSS version 22.0 and the causal map with DAGitty.net version 3.0.

Results

Of the 5644 patients, 558 presented positive results for SARS-CoV-2 RT-PCR. Of the 170 patients hospitalised (in

regular ward, semi-intensive unit or ICU), 52 were positive (9.3% rate of hospitalisation due to COVID-19). Patient age quantile, from 0 to 19, with sample mean of 9.32, was the only demographic variable available. Age was not conditionally independent with SARS-CoV-2 RT-PCR exam. Only 0.9% were positive in the age quantile 0, 1 and 2 (8 positive cases in 883 exams) while the incidence (not weighted) in the age quantile from 3 to 19 was $11.7\% \pm 2.6\%$.

In the first round, 15 blood tests were discarded because of poor performance of the univariate model when SARS-CoV-2 = 1 (Table 1). The remaining blood tests were creatinine, C-reactive protein (CRP), eosinophils, lymphocytes, monocytes and neutrophils (Table 1). Only creatinine was not related with the immune system directly and was evaluated as a risk factor. Of the 5644 patients, 602 patients presented values of eosinophils, 602 lymphocytes, 601 monocytes, 513 neutrophils, 506 CRP and 424 creatinine. Regarding missing cases, all observations with the required data were included (available-case analysis).

CRP is a biomarker of various types of inflammation [19, 20]. At SARS-CoV-2 = 1, the model with CRP and age had good discriminative ability with AUC of 0.872. But at SARS-CoV-2 = 0, AUC = 0.680 was also substantial and the difference of the discriminative ability $\Delta = 0.192$ was moderate (candidate models should present higher differences); the corresponding ROC curve in Figure 4 shows overlapping curves up to sensitivity of 0.5–0.6. Models with CRP demonstrated sensitivity to resampling within the dataset [17], the coefficient significance moved from 0.005 to 0.144. Similar effects were found in models that include CRP with other blood tests and sensitivity to bootstrapping was reduced by dichotomising CRP (reactive/not-reactive). Models

Table 2. Potential candidate logistic regression models for predicting hospitalisation with blood tests and age quantile (different models for each stratum)

		At SARS-CoV-2 = 1					At SARS-CoV-2 = 0				
		B	P	OR	OR 95% CI		B	P	OR	OR 95% CI	
					Lower	Upper				Lower	Upper
Model 1	Age_quantile	0.223	0.001	1.250	1.091	1.432	0.002	0.906	1.002	0.963	1.043
	zEosinophils	-2.506	0.004	0.082	0.015	0.441	-0.314	0.036	0.731	0.545	0.980
	Constant	-4.233	0.000	0.015			-1.650	0.000	0.192		
Model 2	Age_quantile	0.249	0.000	1.282	1.120	1.468	0.000	0.995	1.000	0.961	1.041
	zMonocytes	-0.693	0.008	0.500	0.300	0.834	-0.321	0.021	0.726	0.552	0.954
	Constant	-2.931	0.002	0.053			-1.668	0.000	0.189		
Model 3	Age_quantile	0.303	0.001	1.354	1.137	1.612	0.055	0.050	1.057	1.000	1.117
	zNeutrophils	1.299	0.002	3.665	1.617	8.308	0.493	0.001	1.637	1.223	2,192
	Constant	-3.940	0.002	0.019			-2.687	0.000	0.068		
Model 4	Age_quantile	0.240	0.001	1.271	1.103	1.466	0.003	0.885	1.003	0.963	1.044
	zEosinophils	-2.109	0.012	0.121	0.023	0.630	-0.290	0.050	0.748	0.560	1.000
	zMonocytes	-0.506	0.057	0.603	0.358	1.015	-0.292	0.032	0.746	0.572	0.975
	Constant	-4.005	0.000	0.018			-1.701	0.000	0.183		
Model 5	Age_quantile	0.299	0.002	1.349	1.119	1.626	0.053	0.058	1.055	0.998	1.115
	zEosinophils	-2.004	0.025	0.135	0.023	0.780	0.191	0.181	1.211	0.915	1.603
	zNeutrophils	1.175	0.010	3.240	1.319	7.954	0.586	0.001	1.797	1.292	2.500
	Constant	-4.927	0.001	0.007			-2.712	0.000	0.066		
Model 6	Age_quantile	0.362	0.001	1.436	1.166	1.770	0.056	0.050	1.057	1.000	1.118
	zMonocytes	-1.010	0.014	0.364	0.163	0.816	-0.018	0.919	0.982	0.697	1.384
	zNeutrophils	0.968	0.033	2.632	1.080	6.413	0.487	0.002	1.628	1.191	2.224
	Constant	-4.089	0.005	0.017			-2.687	0.000	0.068		
Model 7	Age_quantile	0.363	0.001	1.437	1.149	1.797	0.053	0.059	1.055	0.998	1.115
	zEosinophils	-1.951	0.036	0.142	0.023	0.884	0.194	0.183	1.214	0.913	1.615
	zMonocytes	-0.925	0.023	0.397	0.178	0.882	0.018	0.920	1.018	0.716	1.448
	zNeutrophils	0.897	0.069	2.453	0.933	6.447	0.593	0.001	1.810	1.264	2.592
	Constant	-5.174	0.003	0.006			-2.712	0.000	0.066		
Model 8	Age_quantile	0.470	0.006	1.600	1.148	2.230	0.071	0.023	1.074	1.010	1.142
	zCreatinine	2.121	0.020	8.338	1.400	49.648	-0.267	0.166	0.766	0.525	1.117
	zMonocytes	-1.540	0.013	0.214	0.064	0.724	-0.076	0.690	0.927	0.639	1.344
	zNeutrophils	1.981	0.018	7.251	1.401	37.528	0.560	0.001	1.751	1.249	2.454
	Constant	-4.542	0.031	0.011			-2.512	0.000	0.081		

SARS-CoV-2 (acute respiratory syndrome coronavirus 2): result of the exam for SARS-CoV-2 RT-PCR (0 = negative; 1 = positive) (reverse transcription - polymerase chain reaction).
 B, coefficient of the variable; P, value of the statistical significance of the coefficient; OR, odds ratio of B (exp(B)); C.I., confidence interval.

Table 3. Discriminative ability of potential candidate models for predicting hospitalisation from non-specific blood tests

		Model								
		1	2	3	4	5	6	7	8	
Variables included in the model:	zEosinophils	●			●	●		●		
	zMonocytes		●		●		●	●	●	
	zNeutrophils			●		●	●	●	●	
	Age quantile (0–19)	●	●	●	●	●	●	●	●	
	Creatinine								●	
Model for positive stratum (SARS-CoV-2 = 1)	AUC (area under ROC curve)	0.839	0.810	0.862	0.856	0.899	0.897	0.910	0.940	
	Standard error	0.046	0.049	0.044	0.043	0.036	0.036	0.034	0.029	
	Asymptotic significance	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	AUC 95% CI asymptotic	Lower bound	0.748	0.715	0.775	0.772	0.828	0.826	0.844	0.883
		Upper bound	0.929	0.906	0.948	0.941	0.970	0.967	0.976	0.997
	Classification table (cut value = 0.5)	Percentage correct H = 0	70.0	70.0	75.0	75.0	72.2	72.2	75.0	81.0
		Percentage correct H = 1	79.1	79.1	84.6	83.7	87.2	82.1	89.7	82.9
		Overall percentage	74.7	74.7	80.0	79.5	80.0	77.3	82.7	82.1
	Cases included in the analysis	H = 0	40	40	36	40	36	36	36	21
		H = 1	43	43	39	43	39	39	39	35
		Total	83	83	75	83	75	75	75	56
Same model of positive stratum applied to SARS-CoV-2 = 0 cases	AUC	0.562	0.542	0.665	0.564	0.603	0.645	0.600	0.627	
	Standard error	0.036	0.043	0.045	0.037	0.044	0.047	0.044	0.049	
	Asymptotic significance	0.069	0.214	0.000	0.061	0.012	0.000	0.016	0.003	
	AUC 95% CI Asymptotic	Lower bound	0.492	0.459	0.576	0.492	0.518	0.553	0.513	0.532
		Upper bound	0.632	0.626	0.754	0.636	0.689	0.737	0.686	0.723
	Cases included in the analysis	H = 0	433	432	382	432	382	382	382	244
		H = 1	86	86	56	86	56	56	56	55
Total		519	518	438	518	438	438	438	299	
Difference of the discriminative ability (naïve)	0.277	0.268	0.197	0.292	0.295	0.252	0.310	0.313		
Overall discriminative performance order	5	6	8	4	3	7	2	1		

AUC, area under the ROC curve; CI, confidence interval;

H, hospitalisation (0 = false; 1 = regular ward, semi-intensive care, or ICU).

SARS-CoV-2 (acute respiratory syndrome coronavirus 2): result of the exam for SARS-CoV-2 RT-PCR (0 = negative; 1 = positive) (reverse transcription – polymerase chain reaction).

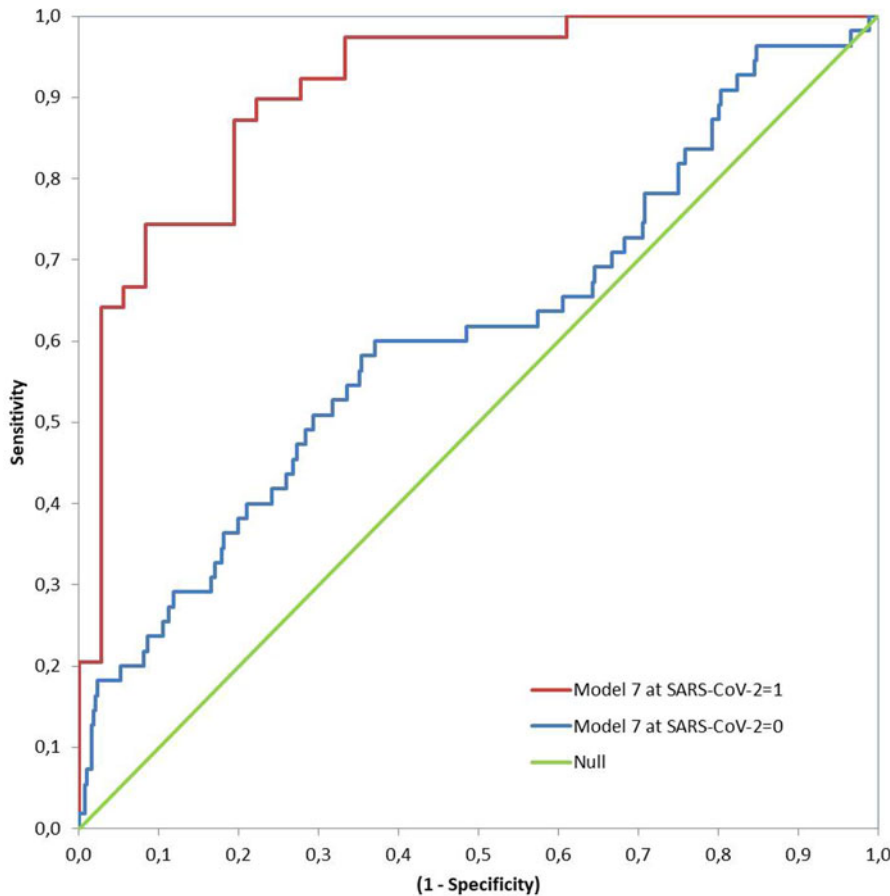


Fig. 5. ROC curves of model 7 to predict hospitalisation at both strata (with and without exposure to SARS-CoV-2). Legend: Null – area of the null hypothesis model is 0.5; model 7 – logistic regression with eosinophils, monocytes and neutrophils controlled for age quantile.

with CRP_reactive, neutrophils and age-generated AUC of 0.901 and 0.730 in the positive and negative strata ($\Delta = 0.171$), and CRP_reactive, monocytes, neutrophils and age-generated AUC of 0.921 and 0.706, respectively ($\Delta = 0.215$). CRP is a predictor of hospitalisation in general, but high levels of AUC at the negative stratum mean that CRP is a response with significant bias due to other causes than COVID-19. Differently from other prognostic studies [21–26], CRP was excluded as candidate.

The neutrophils to lymphocytes ratio (NLR) is considered a possible indicator of severity [21, 24, 27, 28] of COVID-19, but NLR could not be evaluated as all variables were standardised (division by zero). Lymphocytes presented inconsistent behaviour across models. Single exam models indicated lymphopenia at SARS-CoV-2 = 1, as expected [29, 30]. But lymphocytes reversed the sign in the model with neutrophils and age (SARS-CoV-2 = 1), possibly, due to collinearity between them (Pearson's correlation of -0.925 and -0.937 at positive and negative strata, both significant at 0.01 (two-tail)). As there are indications of collinearity issues at both strata, lymphocyte and neutrophils should not be in the same model as independent variables, and this is an indication that NLR may be a noisy association with hospitalisation. As models with combinations of neutrophils were slightly better than with lymphocyte, lymphocyte was dropped from analysis.

In the second round, combinations of eosinophils, monocytes and neutrophils with age were tested systematically. Table 2 presents parameters of models combining eosinophils, monocytes and neutrophils (with age) and the best model with creatinine (as risk factor). Table 3 presents AUCs for each model with the difference of discriminative ability between strata.

Considered individually, eosinophils, monocytes and neutrophils generated models with good performance to estimate the probability of hospitalisation (models 1, 2, 3 with $AUC > 0.810$ at positive stratum). The combinations of these blood tests generated models (4, 5, 6, 7) with better discriminative ability ($AUC > 0.856$ at SARS-CoV-2 = 1). The AUC at SARS-CoV-2 = 0 is a simplified measure of the systematic bias in both outcomes: models 1, 2 and 4 presented low values (with $AUC < 0.564$) and the others presented relevant noisy associations (AUC from 0.600 up to 0.665), but with better difference in discriminative ability $\Delta > 0.252$ in models with two or more exams.

Two patterns of associations were more salient: (1) age as a risk factor with combinations of eosinophils, monocytes and neutrophils as predictors; (2) age and creatinine as risk factors with monocytes and neutrophils as predictors. The interpretation of the conditional probabilities will focus on models 7 and 8, but models with at least two blood tests (4–8) are potential candidate associations. Considering creatinine as a marker of the renal function, model 8 is the overall best model with significant coefficients at $P < 0.05$ and has the highest difference of discriminative ability between strata ($\Delta = 0.313$). Comparative ROC curves for models 7 and 8 are shown in Figures 5 and 6, where there is a substantial discriminative difference between both strata of SARS-CoV-2; confidence intervals at 95% of AUC values are in Table 3.

When the coefficients of model 7 (Table 2) are converted to conditional probabilities we find that at average age quantile (9.32) and average monocyte and neutrophil levels, there is a hospitalisation probability of 51.1% with eosinophils at -1 standard deviation (s.d.); and 90.2% when age quantile is 15. Model 8 with creatinine has different responses: age quantile coefficient

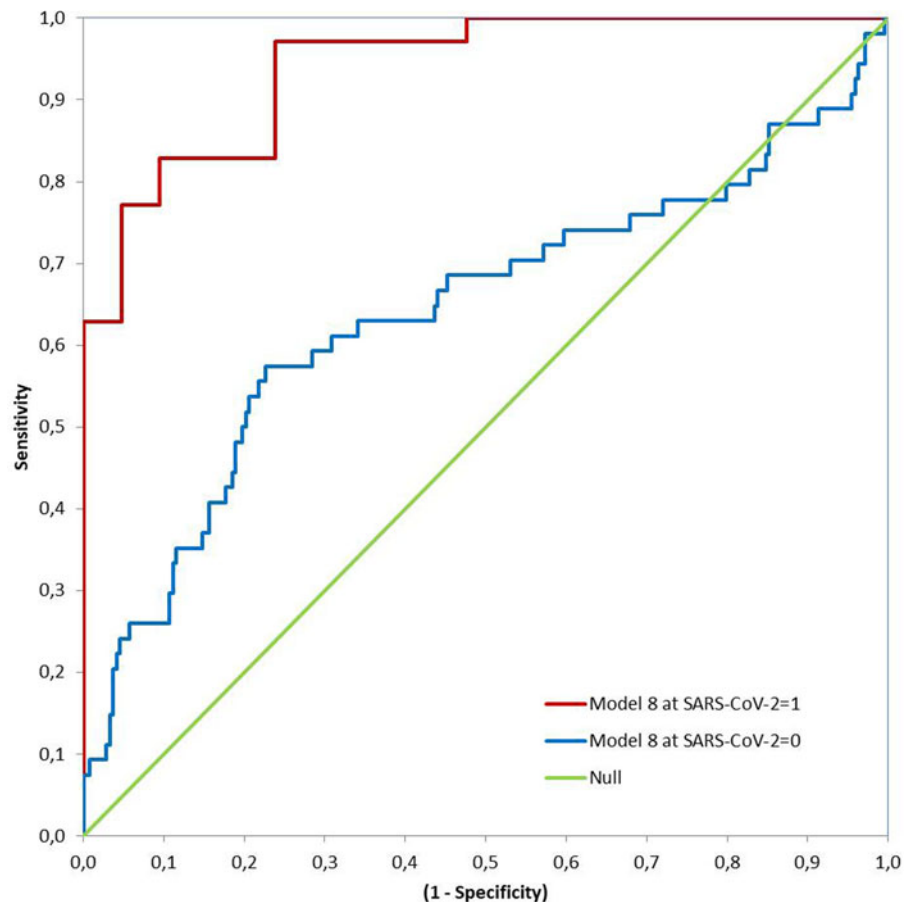


Fig. 6. ROC curves of model 8 to predict hospitalisation at both strata (with and without exposure to SARS-CoV-2). Legend: Null – area of the null hypothesis model is 0.5; model 8 – logistic regression with monocytes and neutrophils controlled for creatinine and age quantile.

is more pronounced and the odds ratio of creatinine is steep (8.338), so average levels of creatinine result in a probability of hospitalisation $>50\%$ for age quantile >9 (with monocytes and neutrophils at average). When creatinine is $+1$ s.d. at age quantile 9, hospitalisation probability is 85.9% (monocytes and neutrophils at average). Only below average levels of creatinine lower hospitalisation probabilities. Monocytes and neutrophils are also steeper than model 7. At age quantile 9, $+1/2$ s.d. of creatinine, $-1/2$ s.d. of monocytes and $+1/2$ s.d. of neutrophils result in a hospitalisation probability of 92.5%.

Model biases may be due to missing cases selection. Most likely, missing data are not at random (MNAR). We performed the bootstrapping procedure to identify potential sensitivity to resampling and, indirectly, to selection bias. The selected models maintained the magnitude and statistical significance of the coefficients. Apparently, as no significant deviation was detected, the missing cases bias may not be an issue.

ROC and AUC calculations used the same data for model fitting. Because of limited sample size, it was not suitable to apply the approach of splitting the database for training and then prediction. After dividing the sample in two groups, most coefficients were not significant at $P > 0.10$ (Table 4) at least in one group. Notwithstanding, classification tables were coherent between subsets and we found no clear indication of model misspecification.

Discussion

We focused on models with discriminative ability to identify peculiar responses in the transition from moderate-to-severe

inflammation only due to COVID-19. The AUC evaluation at the negative SARS-CoV-2 stratum to estimate the influence of unwanted confounders into the focal association together with equivalent criteria of severity state at both strata is, to the best of our knowledge, a needed improvement in prognosis studies of COVID-19.

In comparison to other prediction studies, we identified a few focused on the transition from moderate-to-severe cases of COVID-19 [21–28]. None of them considered data from the negative stratum of SARS-CoV-2, therefore, these models are biased by not excluding noisy predictors.

We eliminated variables with ‘high’ AUC at SARS-CoV-2 = 0, so that variables with more peculiar responses to COVID-19 were included. Reactive levels of CRP together with SARS-CoV-2 RT-PCR exam may be a predictor of hospitalisation, but this can happen due to causes other than COVID-19 (most cases of COVID-19 are asymptomatic to mild). To include it in a model, one should control for all other causes of CRP reactive.

We evaluated age and creatinine as risk factors. Controlling for age improved the AUC of all models at the positive stratum of SARS-CoV-2. The difference between risk factor and outcome among blood tests is subtle. The emergent literature is cautious about whether eosinopaenia may be a risk factor [31] and whether creatinine (and other renal markers) may be associated with COVID-19 renal inflammatory response [32]. As an acute inflammatory kidney response to COVID-19, the interpretation changes and further refinement of the framework is necessary. If eosinopaenia is a risk factor, the prevalence of this condition should be considered and must be properly diagnosed at admission, and the models should be reviewed with new data.

Table 4. Tentative parameters for models 4–8 with split dataset at the positive stratum of SARS-CoV-2: sample size is unsuitable for training and then prediction

		Cases 1 to 5030					Cases 5031 to 5644				
		B	P	OR	OR 95% CI		B	P	OR	OR 95% CI	
					Lower	Upper				Lower	Upper
Model 4	Age_quantile	0.441	0.002	1.555	1.178	2.052	0.022	0.849	1.022	0.817	1.278
	zEosinophils	-1.730	0.083	0.177	0.025	1.250	-6.905	0.026	0.001	0.000	0.431
	zMonocytes	-0.586	0.144	0.556	0.253	1.221	0.175	0.717	1.192	0.461	3.079
	Constant	-6.673	0.002	0.001			-4.040	0.117	0.018		
	Cases (N)	47					36				
	% correct (H = 0)	86.2					54.5				
	% correct (H = 1)	72.2					92.0				
	Overall correct (%)	80.9					80.6				
Model 5	Age_quantile	0.419	0.004	1.520	1.140	2.027	0.049	0.792	1.050	0.731	1.509
	zEosinophils	-1.259	0.196	0.284	0.042	1.916	-7.798	0.026	0.000	0.000	0.389
	zNeutrophils	0.612	0.284	1.844	0.603	5.640	2.301	0.038	9.987	1.131	88.22
	Constant	-6.403	0.004	0.002			-4.037	0.222	0.018		
	Cases (N)	45					30				
	% correct (H = 0)	77.8					88.9				
	% correct (H = 1)	77.8					100.0				
	Overall correct (%)	77.8					96.7				
Model 6	Age_quantile	0.467	0.005	1.596	1.153	2.209	0.228	0.211	1.256	0.879	1.795
	zMonocytes	-0.916	0.068	0.400	0.149	1.071	-0.999	0.235	0.368	0.071	1.915
	zNeutrophils	0.548	0.327	1.729	0.579	5.165	1.617	0.058	5.036	0.949	26.73
	Constant	-5.790	0.008	0.003			-1.731	0.500	0.177		
	Cases (N)	45					30				
	% correct (H = 0)	88.9					77.8				
	% correct (H = 1)	72.2					1000				
	Overall correct (%)	82.2					93.3				
Model 7	Age_quantile	0.504	0.005	1.655	1.162	2.359	0.060	0.774	1.062	0.703	1.605
	zEosinophils	-1.475	0.201	0.229	0.024	2.200	-8.005	0.042	0.000	0.000	0.747
	zMonocytes	-0.898	0.069	0.407	0.154	1.074	-1.024	0.402	0.359	0.033	3.933
	zNeutrophils	0.342	0.564	1.407	0.441	4.492	2.076	0.065	7.974	0.882	72.13
	Constant	-6.988	0.008	0.001			-3.731	0.264	0.024		
	Cases (N)	45					30				
	% correct (H = 0)	88.9					88.9				

(Continued)

Table 4. (Continued.)

	Cases 1 to 5030					Cases 5031 to 5644					
	B	P	OR	OR 95% CI		B	P	OR	OR 95% CI		
				Lower	Upper				Lower	Upper	
	% correct (H = 1)	72.2				1000					
	Overall correct (%)				82.2					96.7	
Model 8	Age_quantile	2.321	0.155	10.184	0.415	249.9	0.245	0.279	1.278	0.819	1.994
	zMonocytes	-4.518	0.183	0.011	0.000	8.467	-1.689	0.168	0.185	0.017	2.039
	zNeutrophils	3.774	0.280	43.55	0.047	40 739	2.758	0.035	15 769	1.211	205.3
	zCreatinine	4.304	0.221	73.96	0.075	72 999	2.420	0.205	11 240	0.266	475.4
	Constant	-25.16	0.155	0.000			-0.919	0.772	0.399		
	Cases (N)	29					27				
	% correct (H = 0)	92.9					85.7				
	% correct (H = 1)	93.3					95.0				
	Overall correct (%)	93.1					92.6				

Note: The cut off at 5030 cases was selected to generate valid parameters with similar quantities of available cases at SARS-CoV-2 = 1 because lower/higher thresholds generated invalid parameters for model 8 due to perfect discrimination. SARS-CoV-2, acute respiratory syndrome coronavirus 2; RT-PCR, reverse transcription - polymerase chain reaction; B, coefficient of the variable; P, value of the statistical significance of the coefficient; OR, odds ratio of B (it is equal to exp(B)); CI, confidence interval.

Results of classification table cut-off value of 0.5 with percentage of correct non-hospitalisation (H = 0) and correct hospitalisation (H = 1).

As we drop noisy predictors, we are effectively dealing with hypothesis about the physiopathology of COVID-19 inflammation. Although not as frequent as the mentions of neutrophils, there are studies on the complex role of eosinophils [31, 33] and monocytes [34, 35] in COVID-19 inflammation indicating eosinopaenia in severe cases and monocytopenia in some phase of the cytokine storm and other COVID-19 pathologies [36].

We selected two patterns of blood tests that are associated with hospitalisation due to COVID-19 inflammation: age with combinations of eosinophils, monocytes and neutrophils; and age and creatinine with monocytes and neutrophils. The model findings are aligned with the known physiopathology of COVID-19 but in a more integrative framework of analysis (not as individual predictors, but as a set that is related to risk factors). The selected blood tests are broadly available even in regions with scarce health care resources. It is unlikely that we will have just one or two overall best models; given different sets of risk factors, we should expect a few representative patterns of the COVID-19 inflammation from moderate to severe.

Limitations and future directions

The models are candidates only and the results cannot be representative beyond the patient health profiles of this reference hospital in Sao Paulo/Brazil that attends a high social-economic segment [37]. The sample refers to the initial phase of the pandemics in Brazil and the patterns may change with medicine prescriptions and other adaptations of SARS-CoV-2. The reduced quantity of available cases did not allow the dataset split for training and prediction. Further efforts are needed to increase internal and external validity across populations, as the prognostic ability is also a function of the variability of the development of COVID-19 inflammation.

As there is no unambiguous way to characterise 'moderate-to-severe COVID-19 inflammation', the inclusion of an unmeasured variable reduces the predicted conditional independences from the DAG. But still this framework can help in the identification and estimation of risk factors. This cross-sectional data (single point time) cannot inform if creatinine (or eosinophil) is risk factor or effect of COVID-19 inflammation. In future data collection efforts, participants should be followed over time, from diagnosis to hospitalisation; ideally from exposure throughout the lifecycle and also with the follow-up of negative cases.

Causal studies are intrinsically predictive [10], therefore, we need to advance prognosis research within causal frameworks. As most studies will be observational, data collection with ample selection of variables for matching estimators (e.g. stratification) [16] will be required to reduce systematic bias.

All candidate models can be reproduced from the dataset [18]. We believe most hospitals can apply this framework to generate similar models appropriate to the target population in which they are inserted by making efforts to collect blood tests and potential risk factors at admission, and other clinical data. By making these databases public (anonymised and with standardised data), they will allow future external validation in larger target populations.

Finally, in the wider context of COVID-19 epidemiology, the collapse of health systems due to opportunistic pathogens is a symptom of threats that requires system-level measures during and after the pandemics [38]. This research is concerned with

hospital care. As a bottleneck, even small gains may have multiplicative effects on health systems. In countries with porous containment efforts, hospital occupancy is a critical metric [39] to alternate between 'soft lockdown' and economic activity with 'constrained mobility'. As some regions with sustained transmission are hesitant and being pushed towards these states, they are poorly capturing the benefits of the switching strategy (Parrondo's paradox applied to epidemics [40]) – because they are struggling in trial and error mode to establish thresholds of when to restrain (and open) and at what pace. Due to the fast saturation of hospital infra-structures with overshooting in these regions, the tendency of excessive losses in each transition is hard to manage. In this context, we believe that the application of prognosis tools can improve the timely access to supportive care in countries with sustained COVID-19 transmission.

Acknowledgements. We are grateful to Antonio Magno Lima Espesquit and Sonia Mara de Andrade who contributed with suggestions to this research. We are also indebted to Hospital Israelita Albert Einstein for making the dataset available, and the referees for their detailed comments.

Author contributions.

G. Ishikawa: Conceptualisation, methodology and formal analysis. G. Argenti: Conceptualisation, formal analysis and clinical and epidemiological validation. C. B. Fadel: Clinical and epidemiological validation and critical review. All authors: Writing, editing, visualisation, review and final approval of the paper

Conflict of interest. The authors declare no conflicts of interest.

This paper has not been published previously in whole or part.

The data that support the results of this study are openly available in reference number [18].

Although this research received no specific grant from any funding agency, commercial or not-for-profit sectors, as institutionally required we inform that 'this study was financed in part by the Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior – Brasil (CAPES) – Finance Code 001'.

References

1. **World Health Organization** (2020) Clinical management of COVID-19: Interim guidance. WHO publications [Internet]. [cited 27 May 2020]; Available at <https://www.who.int/publications/i/item/clinical-management-of-covid-19>.
2. **Rees EM et al.** (2020) COVID-19 length of hospital stay: a systematic review and data synthesis. *BMC Medicine* **18**, 270.
3. **Guan W et al.** (2020) Clinical characteristics of coronavirus disease 2019 in China. *The New England Journal of Medicine [Internet]* **382**, 1708–1720. doi: <https://doi.org/10.1056/NEJMoa2002032>.
4. **Italy: SARS-CoV-2 Surveillance Group** (2020) Characteristics of COVID-19 patients dying in Italy. Epidemiology for public health: Istituto Superiore di Sanità [Internet]. [cited 24 April 2020]; Available at <https://www.epicentro.iss.it/en/coronavirus/sars-cov-2-analysis-of-deaths>.
5. **CDC** (2020) Human infection with 2019 novel coronavirus person under investigation (PUI) and case report form. Atlanta, GA: US Department of Health and Human Services, CDC. Available at <https://www.cdc.gov/coronavirus/2019-ncov/downloads/pui-form.pdf>.
6. **Liu X et al.** (2020) Risk factors associated with disease severity and length of hospital stay in COVID-19 patients. *Journal of Infection* **81**, e95–e97.
7. **Marin BG et al.** (2020) Predictors of COVID-19 severity: a literature review. *Reviews in Medical Virology* **n/a**, e2146. doi: <https://doi.org/10.1002/rmv.2146>.
8. **Lai JW and Cheong KH** (2020) Superposition of COVID-19 waves, anticipating a sustained wave, and lessons for the future. *BioEssays* **42**, e2000178.
9. **Wynants L et al.** (2020) Prediction models for diagnosis and prognosis of Covid-19: systematic review and critical appraisal. *BMJ* **369**, m1328.
10. **Westreich D** (2020) *Epidemiology by Design: A Causal Approach to the Health Sciences*, 1st Edn. New York: Oxford University Press, pp. 41–104.

11. **Pearl J** (2009) *Causality: Models, Reasoning, and Inference*, 2nd Edn. Cambridge: Cambridge University Press, pp. 65–106.
12. **Foraita R, Spallek J and Zeeb H** (2014) Directed acyclic graph. In Ahrens W and Pigeot I (eds), *Handbook of Epidemiology*, 2nd Edn. New York: Springer, pp. 1481–1518.
13. **Greenland S, Pearl J and Robins JM** (1999) Causal diagrams for epidemiologic research. *Epidemiology (Cambridge, Mass.)* **10**, 37–48. PMID: 9888278.
14. **Glymour MM and Greenland S** (2008) Causal diagrams. In Rothman KJ, Greenland S and Lash TL (eds), *Modern Epidemiology*, 3rd Edn. Philadelphia: Lippincott Williams & Wilkins, pp. 183–212.
15. **Pearl J** (2014) Comment: understanding Simpson's paradox. *The American Statistician* **68**, 8–13.
16. **Morgan SL and Winship C** (2015) *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, 2nd Edn. New York: Cambridge University Press, pp. 77–139.
17. **Hosmer DW, Lemeshow S and Sturdivant RX** (2013) *Applied Logistic Regression*, 3rd Edn. Hoboken: John Wiley & Sons, pp.173–181.
18. **Hospital Israelita Albert Einstein** (2020) Diagnosis of COVID-19 and its clinical spectrum: AI and Data Science supporting clinical decisions (from 28th Mar to 3rd Apr). Kaggle [Internet]. [cited 8 Apr 2020]; Available at <https://www.kaggle.com/dataset/e626783d4672f182e7870b1bbe75fae66bdfb232289da0a61f08c2ceb01cab01>.
19. **Lelubre C et al.** (2013) Interpretation of C-reactive protein concentrations in critically ill patients. *BioMed Research International* **2013**, Article ID 124021, 11.
20. **Black S, Kushner I and Samols D** (2004) C-reactive protein. *Journal of Biological Chemistry* **279**, 48487–48490.
21. **Cheng B et al.** (2020) Predictors of progression from moderate to severe coronavirus disease 2019: a retrospective cohort. *Clinical Microbiology Infection* **26**, 1400–1405.
22. **Tan L et al.** (2020) Validation of predictors of disease severity and outcomes in COVID-19 patients: A descriptive and retrospective study [published online ahead of print, 19 May 2020]. *Med (NY)*. doi: <https://doi.org/10.1016/j.medj.2020.05.002>.
23. **Zhu Z et al.** (2020) Clinical value of immune-inflammatory parameters to assess the severity of coronavirus disease 2019. *International Journal of Infectious Disease* **95**, 332–339.
24. **Shang W et al.** (2020) The value of clinical parameters in predicting the severity of COVID-19 [published online ahead of print, 21 May 2020]. *Journal of Medical Virology* **92**, 2188–2192. doi: <https://doi.org/10.1002/jmv.26031>.
25. **Zhou C et al.** (2020) Predictive factors of severe coronavirus disease 2019 in previously healthy young adults: a single-center, retrospective study. *Respiratory Research* **21**, 157.
26. **Bhargava A et al.** (2020) Predictors for severe COVID-19 infection. *Clinical Infectious Diseases* **71**, 1962–1968. doi: <https://doi.org/10.1093/cid/ciaa674>.
27. **Yang AP et al.** (2020) The diagnostic and predictive role of NLR, d-NLR and PLR in COVID-19 patients. *International Immunopharmacology* **84**, 106504.
28. **Wang C et al.** (2020) Preliminary study to identify severe from moderate cases of COVID-19 using combined hematology parameters. *Annals of Translational Medicine* **8**, 593.
29. **Zhao Q et al.** (2020) Lymphopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a systemic review and meta-analysis. *International Journal of Infectious Diseases* **96**, 131–135, ISSN 1201-9712.
30. **Huang I and Pranata R** (2020) Lymphopenia in severe coronavirus disease-2019 (COVID-19): systematic review and meta-analysis. *Journal of Intensive Care* **8**, 36.
31. **Lindsley AW, Schwartz JT and Rothenberg ME** (2020) Eosinophil responses during COVID-19 infections and coronavirus vaccination. *The Journal of Allergy and Clinical Immunology* **146**, 1–7.
32. **Qian JY, Wang B and Liu BC** (2020) Acute kidney injury in the 2019 novel coronavirus disease. *Kidney Diseases* **6**, 318–323.
33. **Xie G et al.** (2020) The role of peripheral blood eosinophil counts in COVID-19 patients. *Allergy* **00**, 1–12.
34. **Alzaid F et al.** (2020) Monocytopenia, monocyte morphological anomalies and hyperinflammation characterise severe COVID-19 in type 2 diabetes [published online ahead of print, 2020 Aug 20]. *EMBO Molecular Medicine* **12**, e13038. doi: <https://doi.org/10.15252/emmm.202013038>.
35. **Martinez F et al.** (2020) Monocyte activation in systemic Covid-19 infection: assay and rationale. *EBioMedicine* **59**, 102964.
36. **Pence B** (2020) Severe COVID-19 and aging: are monocytes the key? *GeroScience* **42**, 1051–1061. doi: <https://doi.org/10.1007/s11357-020-00213-0>.
37. **Teich VD et al.** (2020) Epidemiologic and clinical features of patients with COVID-19 in Brazil. *einstein (Sao Paulo)* **18**, eAO6022.
38. **Cheong KH and Jones MC** (2020) Introducing the 21st century's new four horsemen of the coronapocalypse. *BioEssays* **42**, 2000063.
39. **Martins CM et al.** (2020) Predictive model for COVID-19 incidence in a medium-sized municipality in Brazil (Ponta Grossa, Parana). *Texto & Contexto - Enfermagem* **29**, e20200154.
40. **Cheong KH, Wen T and Lai JW** (2020) Relieving cost of epidemic by Parrondo's paradox: a COVID-19 case study. *Advanced Science*, 2002324. doi: <https://doi.org/10.1002/adv.202002324>.