

Predicting an Unfavourable Course in Patients Hospitalised for COVID-19: The PREDICT-COVID Study

Michel Ducher

Hospices Civils de Lyon

Christelle Elias

Hospices Civils de Lyon

Nans Florens

Hospices Civils de Lyon

Philippe Vanhems

Hospices Civils de Lyon

jean pierre fauvel (✉ jean-pierre.fauvel@chu-lyon.fr)

Hospices Civils de Lyon

Research

Keywords: COVID-19, SARS-Cov-2, mortality, epidemiology, Bayesian network, disease severity, risk prediction

Posted Date: September 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-68521/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background. Clinical decision tools that have been proposed to predict the clinical course of patients admitted to hospital with COVID-19 are poorly presented and are at high risk of selection bias. The aim of the study was to propose a prediction clinical tool to predict an unfavourable outcome at the admission of a SARS-CoV2 infected patient that was carefully developed using a large learning database and that was developed from models derived from artificial intelligence.

Methods. The PREDICT-COVID study is a *post hoc* analysis of the Noso-Cor study, a multicenter prospective, observational study. All patients infected by SARS-CoV2 hospitalized in one of the 11 Lyon-University hospitals since 8-March-2020 have been included. The PREDICT-COVID database was split in two separate datasets: the learning dataset (80%) was used for the development of the model and the validation dataset (20%) for internal validation. The primary composite outcome was the need for mechanical ventilation or admission into an intensive care unit, or death within 21 days of admission.

Results. Data from 823 patients were analysed: age 70.6 ± 16.9 years; body mass index 26.7 ± 5.4 kg/m² and median number of comorbidities was 2. Out of the 44 recorded variables, 11 that were the most linked to the primary outcome criteria were retained to develop the optimised risk prediction tool. At admission the 5 most informative predictors were, in descending order: C-Reactive Protein, neutrophil-to-lymphocyte ratio, aspartate transaminase, shortness of breath, and prothrombin time. The ten-fold cross validation of the optimised model had an area under the ROC curve of 0.76 ± 0.06 . The performance of the developed Bayesian model to predict the primary outcome of the validation dataset had a mean area under the ROC curve of 0.78, sensitivity of 60%, and specificity of 77%.

Conclusions. The proposed optimised prediction tool that uses 11 routinely determined variables to predict an unfavourable course at admission for COVID-19 had satisfactory performance. For an external validation, the PREDICT-COVID prediction tool is available online at: <https://www.hed.cc/?a=covid&n=NETCRIT21J.neta>

Trial registration: The Noso-Cor study was registered on ClinicalTrials (NCT04290780). The present analysis was registered on ClinicalTrials (NCT04412031) the June 2, 2020.

Background

The COVID-19 pandemic has raised health concerns around the world. In France, as of May 29, cases had reached 145,746, and 18,260 patients had died in hospital [1]. The clinical signs of COVID-19 that lead patients to hospital are well known: fever, cough, fatigue, headache, and shortness of breath [2, 3]. The most frequently reported diagnostic and prognostic predictors of COVID-19 are age, body temperature, lymphocyte count, and lung imaging characteristics. Flu-like symptoms and neutrophil counts are frequently predictive in diagnostic models, while co-morbidities, sex, C-reactive protein and creatinine are frequent prognostic factors [4–10]. With more than 90% of deaths occurring in patients over the age of 60 years, and a predominance of males, French data are consistent with that reported worldwide [1, 4–9]. In

Italy [11], a higher case-fatality rate (7.2%) compared to China (2.3%) was attributed to a higher proportion of elderly people [12], obviously if the methods for classifying and reporting deaths due to COVID-19 disease were the same.

In many countries, the capacity of intensive care units (ICU) was exceeded, and this was also the case in France in the north-east and in the Paris regions [1]. Thus, it is of major interest to forecast as soon as possible the clinical course of COVID-19 for individuals and for healthcare decision makers to anticipate and predict the need for ICU and resuscitation beds. To this end, many prediction models have been developed; so far 3 types have been developed: diagnostic models to detect COVID-19 infection, models to predict hospital admission for COVID-19 pneumonia in the general population, and models to predict progression to a serious or critical condition [10]. As point out by Wynants *et al.* [10], these proposed models are poorly presented and have a high risk of bias, raising concerns that their predictions may be unreliable when applied in daily practice.

Using techniques derived from artificial intelligence, Bayesian methods provide additional insight into the interpretation of data by providing probability estimates of clinical interest and by allowing prior evidence to be taken into account to provide posterior probabilities that reflect natural sequential learning and thus facilitate medical decision making [13]. Using data available from the Noso-Cor study [14], the main objective of the present analysis was to propose a new prediction risk tool, carefully developed using a Bayesian network, and make it available to physicians and healthcare decision makers for predicting an unfavourable course at admission for COVID-19.

Materials And Methods

This is a *post hoc* analysis of the Noso-Cor study [14], that is an international multicenter prospective, observational, hospital-based study in adults and children. The data were restricted to patients hospitalised at one of the 11 Lyon University hospitals. Demographic and clinical data were collected using specifically designed case report forms. An internal monitoring was implemented to detect transcription errors. The Noso-Cor study included all patients with a cough and/or fever greater than 37.8 °C at admission for whom SARS-Cov-2 infection was confirmed by reverse transcriptase polymerase chain reaction (RT-PCR) from a nasopharyngeal swab. The Noso-Cor study was approved by the ethics committee of Ile de France V on March 8, 2020 (*Comité de Protection des Personnes Ile de France 5*). Under French law, patients received a written information form and gave their oral consent to the use of their anonymised data for research purposes. The trial was registered on ClinicalTrials (NCT04290780). The study, which started the March 8, 2020, is still on-going. The present analysis that considered data of patients with a positive SARS-Cov-2 RT-PCR was also registered on ClinicalTrials (NCT04412031) the June 2, 2020.

Database

Anonymised data for patients who were tested positive for Sars-Cov2 and hospitalised in one of the thirteen hospitals of the Hospices Civils de Lyon (Lyon, France) were extracted from the Noso-Cor database REF 16 on May 20, 2020. The data extracted concerned the following. Clinical course of the patient during the 21 days following admission: date of symptoms onset, date of admission to the hospital, date of admission to ICU, as well as vital status at 7, 15, and 21 days. Patient characteristics at admission: age, sex, body mass index (BMI), smoking status, alcoholism, and whether or not they were a healthcare professional. Clinical signs at admission: temperature, cough, fatigue, irritability, abnormal lung auscultation, shortness of breath, pain, myalgia, nausea, runny nose, headache, ageusia, anosmia, diarrhoea, nasopharyngeal discomfort, red eye, confusion, heart failure, sore throat, and coma. Comorbidities: cardiac comorbidities, hypertension, liver comorbidities, hypothyroidism, rheumatological comorbidities, neurological comorbidities, diabetes (type 1 or 2), renal comorbidities, pulmonary comorbidities, cancer, and immunodeficiency; these were condensed into a single variable which is the sum of the total number of comorbidities. Paramedical exams at admission: Lung X-ray, C-Reactive Protein (CRP), haemoglobin, white cell count, lymphocyte count, creatinine, natraemia, serum potassium, prothrombin time, aspartate transaminase (AST), alanine transaminase (ALT), and lactate dehydrogenase (LDH).

For Bayesian modelling, continuous variables were divided into classes of clinical relevance (age, BMI, temperature) or into five classes (best tested discretisation).

The database was separated in two separate datasets at random; the learning dataset that contained 80% of the data was used for the development of the models (Bayesian network and logistic regression) and the validation dataset that contained 20% of the data was used for the internal validation of the model.

Primary outcome

The primary composite outcome included: need for mechanical ventilation, admission into an ICU because of multiple organ dysfunction syndrome or death within 21 days of admission. For all patients included in the present analysis (who were all SARS-CoV-2 –infected), the need for mechanical ventilation, transfer to an intensive care unit, or the occurrence of death was considered, regardless of pre-existing conditions that may have contributed to or caused the need for mechanical ventilation, transfer to an intensive care unit, or death.

Prediction model

A Bayesian model was used to develop the prediction tool. A Bayesian network is a directed acyclic graph that includes nodes and arrows. Each node represents a variable (and its modalities), and each arrow represents a probabilistic dependency between the parent variable and the child variable. The Tree Augmented Naive (TAN) algorithm was used to build the structure of the Bayesian network. The TAN algorithm applies three rules: i) each node is independently linked to the target node (*i.e.* primary criteria); ii) each node is also linked to a unique parent node; iii) amongst all possible structures, the structure that maximises the overall mutual information (mutual information measures the strength of the relationship

between each variable and the target) between nodes is selected. The next step consists in estimating conditional probability tables, using the expectation maximisation algorithm. The expectation maximisation algorithm is an iterative procedure to compute the maximum likelihood estimation in the presence of missing or hidden data. In the maximum likelihood estimation, the aim was to estimate the model parameters for which the observed data were the most likely. This step is the so-called “learning step” that allows to develop the Bayesian network using the learning PREDICT-COVID dataset. The software used to create the model was RapidMiner Studio[®] version 9.7 with the W-BayesNet (TAN) Weka 3[®] Machine Learning Group extension.

Optimisation of the prediction tool

The binding strengths between the variables and the target (unfavourable course) were evaluated by the proportion of variance reduction [15]. This method based on variance decomposition makes it possible to completely explore the space of the inputs to the model and to take into account interactions as well as non-linear responses [15]. The sensitivity analysis makes it possible to classify the variables of the model and to reduce their number, keeping in board the most relevant variables. Thus, to improve the ergonomics of the model and make it usable in clinical practice, the number of variables was reduced while preserving the performance of the prediction network. A regression logistic model was also used to confirm the performances of the Bayesian model. The software used to create the logistic regression model was RapidMiner Studio[®] version 9.7 Weka 3[®] Machine Learning Group extension.

The performance of the optimised developed prediction tool used a 10-fold cross validation to determine mean \pm standard deviation (SD) area under the receiver-operating-characteristics (AUC-ROC) curve and accuracy. Performance of the optimised developed model to predict the clinical course of the patients of the validation dataset was investigated using the AUC-ROC curve, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

The software used was RapidMiner Studio[®] version 9.7 with the W-BayesNet (TAN) Weka 3[®] Machine Learning Group extension.

The prediction tool was published online using Netica[®] version 5.19 (Norsys corporation[®]).

Statistics and validation of the models

Quantitative variables are described in terms of means and SD if normal distribution, or median and interquartile range [IQR], and qualitative variables in percentages. The software used was MedCalc[®] version 11.5.1.0. The validation dataset that contained 20% of the data was used for the internal validation of the model. The performance of the model was assessed by C statistic.

Results

Study population

A total of 1100 patients on May 20, 2020 were selected. Wrongly selected patients (nosocomial infection, n = 266; under 18 years old, n = 3), and patients whose admission date was unknown (n = 8) were excluded from this analysis; thus, data from 823 patients were analysed. The mean \pm SD age of the patients was 70.6 ± 16.9 years (range: 19 to 102 years), BMI was 26.7 ± 5.4 kg/m² (range: 11 to 46 kg/m²), and the median total number of comorbidities was 2 (range: 2 to 8). The most frequent comorbidity was hypertension (45.1%), and 23.4% of the patients were diabetics. The most frequent clinical signs reported at admission were abnormal lung auscultation (68.0%), fatigue (67.0%), and cough (66.7%; Supplementary Table 1). There was no significant difference between patients with an unfavourable and those with a favourable outcome with regards to sex, age, smoking status, temperature, renal comorbidity, diabetes, lung auscultation, fatigue, cough, shortness of breath, diarrhoea, pain, headache, runny nose, myalgia, coma at admission, lung X-Ray, creatininaemia, C-Reactive-Protein, prothrombin time ratio, white blood cells count, neutrophils to lymphocytes ratio and AST were significantly different (Supplementary Table 2).

The mean proportion of missing values was 5.2%, ranging from 0% for age and sex, to 25.0% for smoking status (Supplementary Table 1). Variables with more than 25% missing values (LDH 66.7% and alcohol consumption 43.7%) were excluded from the analysis.

Models to predict the primary composite outcome: need for mechanical ventilation; admission into an ICU because of multiple organ dysfunction syndrome or death within 21 days of admission

Using a Bayesian network, variables were first classified according to the variance reduction, (Table 1). The 11 retained variables that were the most linked to the primary outcome criteria are listed in Table 2. The 5 most informative variables to predict the primary outcome, in descending order, were: CRP, neutrophil to lymphocyte ratio, AST, shortness of breath, and prothrombin time (Table 1).

Table 1

Relative variance reduction between the predictive variables and the composite outcome (need for mechanical ventilation, transfer to an intensive care unit or death within 21 days of hospital admission).

	Variables	Percentage
	Variables included in the optimised Bayesian network	
1	C-Reactive protein (CRP)	12.90
2	Neutrophil to lymphocyte ratio	12.10
3	Aspartate transaminase (AST)	9.74
4	Shortness of breath	5.87
5	Prothrombin time	5.68
6	White blood cell count	5.67
7	Creatininaemia	4.29
8	Temperature	2.80
9	Lung X-ray	2.72
10	Pain	2.40
11	Sex	2.40
	Variables not included in the optimised Bayesian network	
12	Smoking status	2.27
13	Natraemia	1.89
14	Age	1.60
15	Alanine transaminase (ALT)	1.57
16	Headache	1.51
17	Myalgia	1.46
18	Body mass index	1.06
19	Comorbidities	1.04
20	Runny nose	1.02
21	Seizure	0.90
22	Coma	0.77
23	Diarrhea	0.73

Variables		Percentage
24	Fever	0.70
25	Serum potassium	0.67
26	Lung auscultation	0.59
27	Healthcare professional	0.50
28	Weakness	0.50
29	Haemoglobin	0.47
30	Cough	0.47
31	Platelet count	0.44
32	Ageusia	0.42
33	Irritability	0.28
34	Nauseas	0.15
35	Pharynx exudate	0.11
36	Anosmia	0.10
37	Sore throat	0.05
38	Red eye	0.00

Table 2

Characteristics of patients infected with SARS-Cov2 at admission for the 11 variables included in the model and frequency of occurrence of key components of the primary composite criteria within 21 days of admission. The remaining variables are described in the Supplementary Table 1.

Primary criteria within 21 days after admission			N Filled (%)	N	Distribution (%)	
Composite primary criteria			100	301	36.5	
Death				131	15.9	
Need for mechanical ventilation				233	28.2	
Transfer to intensive care units				60	0.7	
Characteristics of the patients		Mean	SD	N Filled (%)	N	Distribution (%)
Sex				100		
M					458	55.8
F					365	44.2
Temperature (Celsius)		37.9	1.0	89.2		
Lower than 37.5					216	29.5
From 37.5 to 39					439	59.7
Greater than 39					79	10.7
Clinical sign			N Filled (%)	N	Distribution (%)	
Shortness of breath			99.9			
Yes					467	56.7
Pain			99.9			
Yes					210	25.5
Paramedical entrance exam		Median	IQR	N Filled (%)	N	Distribution
Lung X-Ray				78.6		
Abnormal					570	88.1
Creatininaemia ($\mu\text{mol/L}$)		82	[66–106]	97.1		
C-Reactive Protein (mg/L)		70	[29–136]	88.5		
Prothrombin time ratio (%)		80	[68–89]	74.2		

Primary criteria within 21 days after admission			N Filled (%)	N	Distribution (%)
White blood cell count (G/L)	6.4	[4.8–8.6]	97.7		
Neutrophil to lymphocyte ratio (%)	5.0	[3.0–8.9]	97.3		
Aspartate transaminase (UI/L)	45	[32–65]	78.0		

The ten-fold cross validation of the optimised Bayesian model developed using the learning dataset (80% of the patients, n = 658) and the 11 analysed variables provided a mean \pm SD AUC-ROC of 0.76 ± 0.06 and a mean \pm SD error rate of $29.2 \pm 6.5\%$.

The ten-fold cross-validation of the logistic regression model developed using the learning dataset (80% of the patients, n = 658) and the 11 analysed variables for the primary composite outcome had a mean \pm SD AUC-ROC of 0.76 ± 0.08 and a mean \pm SD error rate of $27.4 \pm 11.0\%$.

The performance of the developed optimised Bayesian model to predict the primary outcome in the validation dataset (the remaining 20% of the patients, n = 165) had an AUC-ROC of 0.78, an accuracy of 77.6%, a sensitivity of 67.9%, a specificity of 76.2%, a PPV of 59.4%, and NPV of 82.5% (Table 3). The structure of the optimised Bayesian network using 11 variables to assess the primary outcome criteria is presented in Fig. 1.

Table 3

Performance of the developed optimised Bayesian model (11 variables) to predict the primary outcome of the validation dataset (remaining 20% of the patients, n = 165)

Parameters used to evaluate the performance	
AUC-ROC curve	0.78
Accuracy	77.6%
Sensitivity	67.9%
Specificity	76.2%
Positive predictive value	59.4%
Negative predictive value	82.5%
AUC-ROC curve: Area under the Receiver operating characteristic curve.	

Discussion

Using data from a large and recent prospective cohort of COVID-19 patients, a new prediction tool using a Bayesian network was built to predict the need for mechanical ventilation, transfer to an intensive care unit or death within 21 days of hospital admission.

A total of 32 clinical decision tools to predict the clinical course of patients admitted to hospital with COVID-19 have been proposed. Among these, 23 aimed to predict the risk of mortality and 8 were designed to predict progression to a more serious or critical illness. However, as noted by Wynants et al. [10], the proposed clinical prediction tools are poorly presented and are at high risk of selection bias. Sufficient information is not provided to replicate these studies concerning the selection of patients and data, and the methodology used. Few studies report missing data, and, even when this was the case, the authors do not explain how these values were replaced, which is crucial for the construction of clinical prediction tools. Internal validation is not always carried out or clearly presented, and without a careful internal validation an overfitting of the model to the data is to be expected explaining that their reported performance is probably optimistic [10]. Contrary to previously published models, the methodology used to develop the prediction model proposed herein follows the published recommendations concerning information on source data, the presentation of inclusion and exclusion criteria for the prospectively included population, the explanation of the judgment criterion, the management of missing data, the explanation of the model used, the methodology used for internal validation, and the model performance measures and their interpretations [16].

The internal validity tests confirmed that the performance of the PREDICT-COVID prediction tool described herein was satisfactory, and with only 11 variables that are commonly used, the prediction tool is easily usable in clinical practice. The performance of the logistic regression model, which is the most commonly used method in medicine to calculate the risk of an event according to exposure, was similar to that of the Bayesian model. However, Bayesian models have many advantages over logistic regression models; for instance, as they do not use any *a priori* hypothesis, explanatory variables can be co-linear (e.g. white blood cell count and neutrophil to lymphocyte ratio), and as they are based on conditional probabilities, the associations between variables are taken into account even if they are not linear. In addition, Bayesian networks are applicable in case of missing data that are frequent in clinical practice, which could be considered the most important advantage as in absence of a single variable a logistic regression model will not be able to calculate a prediction score for an individual. This explains, in part, why Bayesian models are increasingly used to develop risk prediction tools [17].

A selection bias is highly likely in the proposed clinical prediction tools. For example, Chinese clinical prediction tools were constructed using data from hospitalised populations that are much younger (less than 50 years of age) than those hospitalised in Europe or the USA, and whose mortality is much lower (less than 5%). Other clinical prediction tools were built using learning dataset of less than 100 patients. Furthermore, some clinical prediction tools have been performed on particular populations and use non-routinely performed laboratory or radiological data (e.g. IL-6 [18], coronary calcifications [19]) when a patient is admitted to hospital for COVID-19. It is likely, that prescription of specific laboratory or

radiological exams by a trained physician is more predictive than the results of the specific exams *per se*. A very high mortality rate in these studies confirms this hypothesis.

The clinical characteristics and clinical course of COVID-19 patients hospitalised in Lyon were similar to those reported in California [6], New York City [9], and Italy [5, 7, 8], which supports the generalisability of the proposed prediction tool. As reported in the analysis published by Wynants *et al* [10], the most frequently reported diagnostic and prognostic predictors of covid-19 are age, body temperature, lymphocyte count and lung imaging characteristics. Flu-like symptoms and neutrophil counts are frequently predictive in diagnostic models, while co-morbidities, sex, C-reactive protein and creatinine are frequent prognostic factors. In agreement, 7 of the 11 selected prognostic variables included in the model proposed herein are included among those cited in the other proposed models. Although increasing age is reported as a risk for poor outcomes [8, 9, 20–23], herein patients with an unfavourable outcome were slightly but significantly older, but this was not retained in the prediction model presented herein. Similarly, elevated BMI has also been reported to be a risk factor [22, 23], but this was not the case herein. This suggests that age and BMI may explain a higher rate of hospitalisation in COVID-19 patients rather than an unfavourable outcome. Another interesting point is that among the 11 variables retained herein, the majority were laboratory parameters, and these represented 6 of the 7 most predictive variables. Among these biological parameters, only aspartate transaminase and prothrombin time are not among the parameters most frequently retained by other predictive tools.

The present study has some limitations. The endpoint was independent of the co-morbidities of the patient infected with SARS-Cov-2. This method was chosen because there is no consensus to formally attribute death to SARS-Cov-2 and comorbidities certainly influence the patient's course. Furthermore, as recommended by Piccininni *et al* [8], total mortality captures indirect deaths, such as those related to a healthcare system under crisis, yielding a more complete picture of the pandemic's consequences. The total number of explanatory variables was limited due to the increased mathematical combinatory; and was related to the number of subjects included in the learning dataset. After a classification based on variance reduction, somewhat arbitrarily, only the 11 most relevant variables were selected. Thus, using 11 variables, the prediction model offers a good compromise between performance and ergonomics. As Lindsell *et al*. [24] point out, our model, which can be used as both a prognostic and predictive model, does not guide the health care team on the effectiveness of treatment, which is more informative than a simple prognosis. The main limitation of the present study is that, although carefully developed from especially collected data and controlled by internal monitoring, the results have to be externally validated using similar data sources and prove its effectiveness in a pragmatic trial.

Conclusions

The proposed optimised prediction tool that uses 11 routinely determined variables to predict an unfavourable course at admission for COVID-19 had satisfactory performance. For an external validation, the PREDICT-COVID prediction tool is available online at: <https://www.hed.cc/?a=covid&n=NETCRIT21J.neta>

Declarations

Ethics approval and consent to participate

The Noso-Cor study was approved by the ethics committee of Ile de France V on March 8, 2020 (*Comité de Protection des Personnes Ile de France 5*). Under French law, patients received a written information form and gave their oral consent to the use of their anonymised data for research purposes. The trial was registered on ClinicalTrials (NCT04290780). The study, which started the March 8, 2020, is still on-going. The present analysis that considered data of patients with a positive SARS-Cov-2 RT-PCR was also registered on ClinicalTrials (NCT04412031) the June 2, 2020.

Consent for publication. Not Applicable

Availability of data and materials. Data are available on reasonable request from Pr Vanhems

Competing interests. Authors have no conflicts of interest to disclose.

Funding. This analysis was conducted without any specific funding

Authors' contributions

Ducher, Vanhems and Fauvel contributed to the study conception and design. Material preparation, data collection and analysis were performed by Elias, Florens, Ducher and Fauvel. The first draft of the manuscript was written by Fauvel and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. Philip Robinson for manuscript preparation and critical reading of the manuscript and M Saadatian-Elahi, L. Henaff, and clinical research assistants from the department of epidemiology and infection control department.

References

1. COVID-19 : point épidémiologique du 29 mai 2020. <https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-29-mai-2020>
2. Elizabeth Williamson, Alex J Walker, Krishnan J Bhaskaran et al (2020) OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. Ben Goldacre medRxiv.05.06.20092999; doi: <https://doi.org/10.1101/2020.05.06.20092999>
3. Onder G, Rezza G, Brusaferro S (2020) Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. JAMA published online 23 March. doi:10.1001/jama.2020.4683.

pmid:32203977

4. Deng G, Yin M, Chen X, Zeng F (2020) Clinical determinants for fatality of 44,672 patients with COVID-19. *Crit Care* 24:179
5. Colaneri M, Sacchi P, Zuccaro V, et al (2020) Clinical characteristics of coronavirus disease (COVID-19) early findings from a teaching hospital in Pavia, North Italy, 21 to 28 February 2020. *Euro Surveill* 25: 2000460
6. Lewnard JA, Liu VX, Jackson ML, et al (2020) Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in California and Washington: prospective cohort study [published correction appears in *BMJ*. Jun 4;369:m2205]. *BMJ* 2020;369:m1923
7. Grasselli G, Zangrillo A, Zanella A, et al (2020) Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy [published online ahead of print, 2020 Apr 6]. *JAMA* 323:1574-81
8. Piccininni M, Rohmann JL, Foresti L, Lurani C, Kurth T (2020) Use of all cause mortality to quantify the consequences of covid-19 in Nembro, Lombardy: descriptive study. *BMJ* 369:m1835. Published 2020 May 14. doi:10.1136/bmj.m1835
9. Cummings MJ, Baldwin MR, Abrams D, et al (2020) Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet* 395:1763-70
10. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal [published correction appears in *BMJ*. 2020 Jun 3;369:m2204]. 2020;369:m1328
11. Livingston E, Bucher K. Coronavirus Disease 2019 (COVID-19) in Italy (2020) *JAMA*;323:1335. doi:10.1001/jama.2020.4344
12. Novel Coronavirus Pneumonia Emergency. Response Epidemiology Team. Vital surveillances: the epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China (2020) *China CDC Weekly*. 2020;2(8):113-122. Accessed March 18, 2020. <http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51>
13. Siga MM, Ducher M, Florens N, et al (2020) Prediction of all-cause mortality in haemodialysis patients using a Bayesian network [published online ahead of print, 2020 Feb 10]. *Nephrol Dial Transplant* gfz295. doi:10.1093/ndt/gfz295
14. Saadatian-Elahi M, Picot-Sanchez V, Henaff L, et al (2020) Protocol for a multicentre study of nosocomial SARS-CoV2 transmission The NOSO-COR Project. medRxiv. DOI: 10.1101/2020.04.08.20057471
15. Pappenberger F., Ratto, M.V. Vandenberghe. Review of sensitivity analysis methods. In P.A.Vanrolleghem, éditeur :Modelling aspects of water framework directive implementation, pages 191–265.IWA Publishing, 2010
16. Leisman, D. E., Harhay, M. O., Lederer, D. J., Abramson, M., Adjei, A. A., et al (2020) Development and Reporting of Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and Critical

Care Journals. Critical care medicine 48(5), 623–633.

<https://doi.org/10.1097/CCM.0000000000004246>

17. Maddox TM, Rumsfeld JS, Payne PRO (2019) Questions for artificial intelligence in health care. *JAMA* 321: 31–32
18. Vultaggio A, Vivarelli E, Virgili G, et al (2020) Prompt Predicting of Early Clinical Deterioration of Moderate-to-Severe COVID-19 Patients: Usefulness of a Combined Score Using IL-6 in a Preliminary Study [published online ahead of print, 2020 Jun 19]. *J Allergy Clin Immunol Pract* S2213-2198(20)30611-5
19. Matos J, Paparo F, Mussetto I, et al (2020) Evaluation of novel coronavirus disease (COVID-19) using quantitative lung CT and clinical data: prediction of short-term outcome. *Eur Radiol Exp* 4:39. Published 2020 Jun 26. doi:10.1186/s41747-020-00167-0
20. Liang W, Liang H, Ou L, et al (2020) Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. *JAMA Intern Med*. Published online May 12, 2020 doi:10.1001/jamainternmed.2020.2033
21. Gong J, Ou J, Qiu X, et al (2020) A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China [published online ahead of print, 2020 Apr 16]. *Clin Infect Dis* ciaa443. doi:10.1093/cid/ciaa443
22. Caussy C, Pattou F, Wallet F, et al (2020) Prevalence of obesity among adult inpatients with COVID-19 in France. *Lancet Diabetes Endocrinol* 8:562-564. doi:10.1016/S2213-8587(20)30160-1
23. Gupta S, Hayek SS, Wang W, et al (2020) Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US [published online ahead of print, 2020 Jul 15]. *JAMA Intern Med* e203596. doi:10.1001/jamainternmed.2020.3596
24. Lindsell CJ, Stead WW, Johnson KB (2020) Action-Informed Artificial Intelligence-Matching the Algorithm to the Problem [published online ahead of print, 2020 May 1]. *JAMA* 10.1001/jama.2020.5035. doi:10.1001/jama.2020.5035

Figures

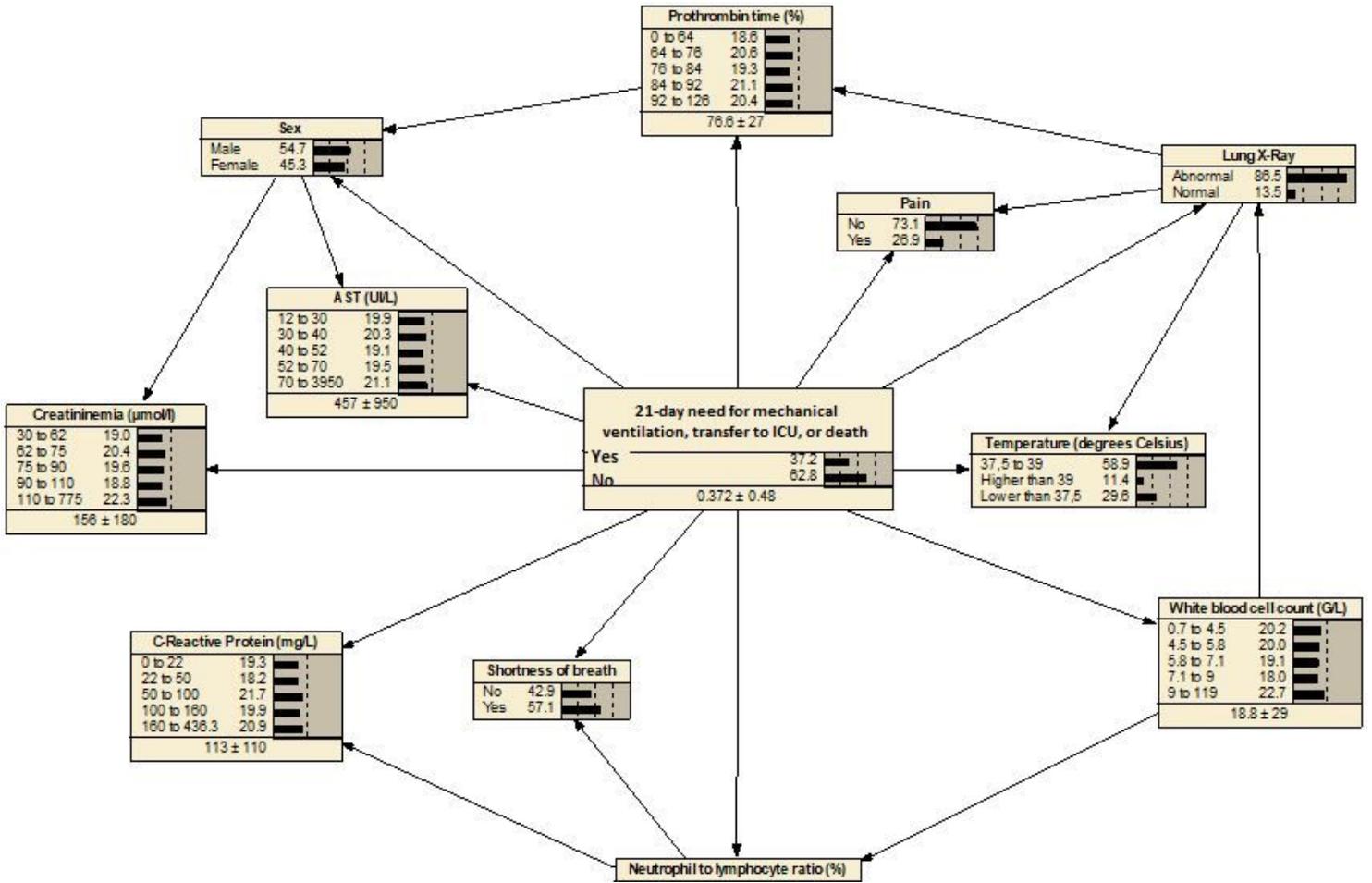


Figure 1

Structure of the optimised Bayesian network to predict an unfavourable course in patients hospitalised for the COVID-19 disease. Values for each modality of each variable represent the distribution expressed as a percentage.