The COVID-19 Worsening Score (COWS)—a predictive bedside tool for critical illness

ECHOCARDIOGRAPHY
A Journal of Cardiovascular Ultrasound and Allied Techniques

The COVID-19 Worsening Score (COWS) – a predictive bedside tool for critical illness

Journal:	*Echocardiography*
Manuscript ID	ECHO-2020-1025
Wiley - Manuscript type:	Original Investigation
Date Submitted by the Author:	31-Oct-2020
Complete List of Authors:	Boero, Enrico; Ospedale San Giovanni Bosco, Department of Surgery, Anesthesia and Intensive Care Rovida, Serena; Department of Anesthesia and Intensive Care Unit, Saint Bartholomew's Hospital, Barts NHS Trust Schreiber, Annia; University of Toronto, Interdepartmental Division of Critical Care Medicine; St Michael's Hospital Li Ka Shing Knowledge Institute, Keenan Research Centre Berchialla, Paola; University of Torino, Department of Clinical and Biological Sciences Charrier, Lorena; University of Torino, Department of Public Health and Pediatrics Cravino, Marta Maria; Ospedale San Giovanni Bosco, Department of Medicine, Internal Medicine Converso, Marcella; Ospedale San Giovanni Bosco, Department of Medicine, High Dependency Unit Gollini, Paola; Ospedale San Giovanni Bosco, Department of Services, Radiology Puppo, Mattia; University of Torino, Department of Surgical Sciences Gravina, Angela; Ospedale San Giovanni Bosco, Department of Medicine, Internal Medicine Fornelli, Giorgia; Ospedale San Giovanni Bosco, Department of Medicine, Internal medicine Labarile, Giulia; University of Torino, Department of Medical Sciences Sciacca, Santi; Ospedale San Giovanni Bosco, Department of Medicine, Internal Medicine Bove, Tiziana; University of Udine, Department of Anesthesia and Intensive Care Karakitsos, Dimitrios; King Saud Medical City, Critical Care Department; Univ South Carolina, School of Medicine; University of Southern California Keck School of Medicine, Critical Care Department Aprà, Franco; Ospedale San Giovanni Bosco, Department of Medicine Blaivas, Michael; Univ South Carolina, School of Medicine; St. Francis Hospital, Department of Emergency Medicine Vetrugno, Luigi; University of Udine, Department of Anesthesia and Intensive Care
Keywords:	Ultrasound, Real-time imaging technique, Computed Tomography (CT), Diagnostic imaging tools, Mechanical ventilation

The COVID-19 Worsening Score (COWS) – a predictive bedside tool for critical illness

List of authors

Enrico Boero, MD
(Corresponding author)
enrico.boero@hotmail.com
ORCID: 0000-0002-2797-4847
Ospedale San Giovanni Bosco, Department of Surgery, Anesthesia and Intensive Care,
Torino, IT

Serena Rovida, MD
sererovida@gmail.com
Department of Anesthesia and Intensive Care Unit, Saint Bartholomew's Hospital, Barts NHS Trust,
London, UK

Annia Schreiber, MD
anniafleur.schreiber@gmail.com
University of Toronto, Interdepartmental Division of Critical Care Medicine,
Toronto, ON, CAN
St Michael's Hospital Li Ka Shing Knowledge Institute, Keenan Research Centre,
Toronto, ON, CAN

Paola Berchialla, Prof
paola.berchialla@unito.it
University of Torino, Department of Clinical and Biological Sciences,
Torino, Piemonte, IT

Lorena Charrier, Prof
lorena.charrier@unito.it
University of Torino, Department of Public Health and Pediatrics,
Torino, Piemonte, IT

Marta Maria Cravino, MD
mcravino@hotmail.com
Ospedale San Giovanni Bosco, Department of Medicine, Internal Medicine,
Torino, Piemonte, IT

Marcella Converso, MD
lellaconverso@tiscali.it
Ospedale San Giovanni Bosco, Department of Medicine, High Dependency Unit,
Torino, Piemonte, IT

Paola Gollini, MD
paolagol@yahoo.it
Ospedale San Giovanni Bosco, Department of Services, Radiology,
Torino, Piemonte, IT

Mattia Puppo, MD
mattia.puppo@unito.it
University of Torino, Department of Surgical Sciences,
Torino, Piemonte, IT

Angela Gravina, MD angelagravina82@yahoo.it
Ospedale San Giovanni Bosco, Department of Medicine, Internal Medicine,
Torino, Piemonte, IT

Giorgia Fornelli, MD
giorgia.fornelli@gmail.com
Ospedale San Giovanni Bosco, Department of Medicine, Internal medicine,
Torino, Piemonte, IT

Giulia Labarile, MD
labarile.giulia@gmail.com
University of Torino, Department of Medical Sciences,
Torino, Piemonte, IT

Santi Sciacca, MD
santi_sciacca@yahoo.it
Ospedale San Giovanni Bosco, Department of Medicine, Internal Medicine,
Torino, Piemonte, IT

Tiziana Bove, MD, Prof
tiziana.bove@asufc.sanita.fvg.it
University of Udine, Department of Anesthesia and Intensive Care,
Udine, Friuli-Venezia Giulia, IT

Dimitrios Karakitsos, MD, PhD, Prof
karakitsosdimitrios@gmail.com
King Saud Medical City, Critical Care Department,
Riyadh, Riyadh, SA
Univ South Carolina, School of Medicine,
Columbia, SC, USA
University of Southern California Keck School of Medicine, Critical Care Department,
Los Angeles, CA, USA

Franco Aprà, MD
franco.apra@aslcittaditorino.it

Ospedale San Giovanni Bosco, Department of Medicine,
Torino, Piemonte, IT

Michael Blaivas, MD, MBA
mike@blaivas.org
Univ South Carolina, School of Medicine,
Columbia, SC, USA
St. Francis Hospital, Department of Emergency Medicine,
Columbus, GA, USA

Luigi Vetrugno, MD, Prof
luigi.vetrugno@asufc.sanita.fvg.it
University of Udine, Department of Anesthesia and Intensive Care,
Udine, Friuli-Venezia Giulia, IT

Authors contribution

EB and SR equally contributed to concept, drafting article and revision of the manuscript. MB, AS, DK, TB and LV contributed in data interpretation and manuscript revision. PB and LC provided the statistical support for data analysis. MMC, MC, PG, AG, GF, SS, MP, and GL supervised data collection and contributed in manuscript revision. MB and LV share the senior authorship and contributed in critical revision and approval of article.

Conflict of interest

MB has participation in EchoNous Inc, Sonosim Inc, Ethos Medical, 410Medical

LV received travel facilities from Cook medical

Aknowledgments

We would thank Dr. Sergio Livigni for its wise guidance as a supervisor of critical care practice in our institution and for easing the process of data collection and knowledge sharing.

We would thank Dr. Savino Sciascia for its interest in clinical research, helping in text revision and IRB approval.

Introduction

By the beginning of 2020, a novel disease called COVID-19 was recognized, and eventually defined as a pandemic by the WHO[1]. The disease-causing virus, known as SARS-CoV-2, with its high tropism for the lower respiratory tract, can produce an infection with a broad spectrum of symptoms ranging from asymptomatic to severe acute respiratory failure, often requiring intensive care unit (ICU) admission[2].

Since the beginning of the pandemic, many healthcare facilities reorganized entire departments where multidisciplinary teams collaborated to provide care for COVID-19 patients. Massive effort from the worldwide medical community has been put forth to better understand the pathophysiology of this disease, in order to provide appropriate care, optimize hospital resources, and increase efficiency of workflow. In this context, the availability of an easy-to-use standardized scoring system would have been of great help in supporting clinicians with different backgrounds to better identify patients at higher risk of developing a critical illness. Aiming to provide means for a better resource-allocation, several prediction models have been developed over the last few months. Vital parameters, comorbidities and blood test results have been combined to predict disease severity and outcomes for hospitalized COVID-19 patients[3456789101011].

Among them, Liang et al. developed the COVID-GRAM score, which showed success in the early prediction of critical illness development, defined as admission to the ICU, need for invasive mechanical ventilation (IMV), or death[4]. However, the GRAM score requires ten independent variables, including laboratory results, chest X-ray, and requires online calculations to risk stratify patients. Despite its accuracy, its use could be time-consuming as not all required parameters are

readily available in all settings. In fact, during the first pandemic peak, healthcare facilities experienced an unexpected patient influx to the Emergency Department (ED) and medical wards with an average of 60 to 80 COVID-19 patients per hour. Based on this very early Italian experience, such patient influxes made serial radiological imaging unfeasible. For this reason, a less burdensome and rapid prognostic score may be of considerable benefit.

Several of the above-mentioned prognostic scores integrated radiological data (i.e., chest X-ray or CT scan), but no study has yet investigated the performance of lung ultrasound (LUS) as a prognostic tool in COVID-19 patients. LUS is available at the patient's bedside, and its reliability and speed as a tool to evaluate acute respiratory disorders in real-time has been well established [12,13]. Moreover, COVID-19 has a distinctive distribution pattern involving mainly the peripheral and lower regions of the lungs[14], and presumably this is why LUS demonstrated superior sensitivity to CT scan for pleural and subpleural abnormalities[15]. According to the available literature and contingent need, LUS may play a central role in this pandemic where the risk of health care workers' exposure and patients' overflow has been a primary concern.

We hypothesized that a new prognostic score, integrating previously validated variables and LUS findings instead of chest radiography, could work as well as the GRAM score for the early identification of COVID-19 patients developing critical illness. Hence, we firstly tested the GRAM score on our cohort, and then developed and internally validated the new COVID-19 Worsening Score (COWS).

Methods

Study design

We conducted a single-center retrospective cohort validation study of the GRAM score, and subsequently developed and internally validated a new prognostic score.

The study adhere to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines[16].

Study population and Setting

The study was conducted in an Italian tertiary Hospital in Turin (San Giovanni Bosco Hospital). All adult patients with a confirmed diagnosis of SARS-CoV-2 infection admitted to the ED and thereafter to the medical wards during the epidemic peak between February 26 and May 17 were enrolled. Patients with hospital-acquired COVID-19, previous pneumonectomy, or lobar pneumonia on presentation were excluded. SARS-CoV-2 infection disease was confirmed by real-time-polymerase-chain reaction (RT-PCR) performed either on nasal swab or pharyngeal swab. The patients' notes and imaging results were retrieved from electronic medical records, collected in a dedicated COVID-19 database, and retrospectively analyzed. The City of Turin Ethical Committee approved the study on June 3rd, 2020 (protocol #82995). The hospital review board waived patients' consent due to the retrospective nature of the study and anonymous data handling and analysis.

Patients Characteristics and Clinical Outcomes

Patient demographic characteristics, comorbidities, presenting symptoms and date of their onset, clinical signs, laboratory test results, and sonographic and radiological findings (chest X-ray and/or CT) were collected within 48 hours of ED admission. The arterial oxygen partial pressure to fractional-inspired oxygen (P/F) ratio was also recorded.

The adverse outcome referred to as *critical illness* in the results section was defined by the occurrence of at least one of the following three events: admission to ICU, need for invasive mechanical ventilation (IMV), or death [1718] due to COVID-19 within a follow-up of 30 days post admission. Supplementary oxygen support or non-invasive ventilation (NIV) were considered favorable outcomes. Need for IMV and ICU admission were decided based on standard of care criteria[19].

Variables selection

Among the patients' collected data, we selected the ten variables previously identified in the GRAM score. We chose these ten variables due to their ability to predict the severity of respiratory failure and progression to critical illness [20]. Moreover, P/F ratio on admission and number of days from symptoms onset were included in the analysis. Missing data were further searched in available materials such as handover and notes. In patients that underwent a CT scan, we considered the following findings: the number of pulmonary lobes involved the presence of emphysema, and the percentage of well-aerated lung. These radiological features were predictors of ICU admission or death in COVID-19 in a previous study[11].

CT scans (obtained by 64 Slice Discovery HD 750 CT Scanner, General Electric, Boston, MA, USA) and chest X-rays were analyzed by a radiologist with more than ten years of chest imaging experience blinded to patients' outcomes.

The LUS protocol adopted for the study was comprehensive of 6 scanning areas per hemithorax as previously described [21]. Each hemithorax was assessed in one upper and one lower area in the three regions divided by the parasternal, anterior, and posterior axillary lines, respectively. The image focus was placed at the level of the pleural line maintaining the image depth at 8-12 cm[13]. An already validated aeration score was assigned to each area [22], and the final LUS score was calculated as the sum of them.

LUS evaluation was performed by 29 clinicians with more than five years of experience in bedside sonographic imaging, and 7 of them subsequently calculated the lung aeration score on all included patients (EB, MC, GL, MC, AG, GF, SS). When in doubt, a second operator (EB) reviewed both imaging and score.

Low- to high medium-frequency (2-9 MHz) curvilinear probes and three different ultrasound machines were selected for the study (MyLab 5™, MyLab 7™; Esaote, Genoa, Italy, and Sonosite M-Turbo™ Ultrasound System, Fujifilm, Hitchin, UK).

Statistical analysis

Continuous variables were reported as mean and standard deviation (SD) or median with interquartile range (IQR) as appropriate, and categorical variables as numbers and percentages.

Evaluation of the LUS score as a predictor of the adverse evolution of COVID-19 infection was assessed by univariate level. Restricted cubic splines were modeled to assess the non-linear effect, and significance was tested by the Wald χ^2. Significance level was set at 0.05. Finally, the LUS score was dichotomized by the ROC curve analysis. Application of the COVID-GRAM model on our sample was carried out to evaluate its performance in classifying high- and low-risk patients according to the threshold identified by the ROC curve analysis. The evaluation of the COVID-GRAM added with the LUS score was then performed.

Aiming to develop a novel and easy to use prognostic score, a selection strategy based on Bayesian Model Averaging was adopted. The number of comorbidities, LUS score, P/F ratio, dyspnea, duration of symptoms (days) showed a posterior probability of inclusion greater than 30% and were retained in the final logistic regression model labeled as COWS. Thirty percent was chosen as the cutoff through sensitivity analysis to maximize the bootstrapped predictive accuracy of the selected model.

The performance of the model was assessed in terms of Somers concordance index Dxy (the closer to 1, the better), Brier score (scores closer to zero indicate a better prediction), and calibration slope. An internal validation to correct measures of predictive performance for optimism (over-fitting) was performed by bootstrapping 500 samples of the data.

To improve the prediction, a shrinkage bootstrap-based method was applied to re-estimate regression coefficients. The overall optimism across all models was estimated deriving a shrinkage coefficient equal to the average calibration slope from each of the bootstrap samples. The shrinkage coefficient was applied to the original coefficient to account for over-fitting. Finally, the intercept

was re-estimated based on the shrunken coefficients to ensure the overall calibration was maintained, producing the final model. All analyses were carried out using R 4.0.0 [23].

Results

Between February 26th and May 17th, 2020, 274 COVID-19 patients were admitted to the wards from the ED (Figure 1). Baseline clinical characteristics are summarized in Table 1. One hundred and seventy-four patients had a final adverse outcome (critical illness), while 100 patients had a favorable outcome (non-critical illness). Complete data for the study analysis, including LUS findings were available in 143 cases. The mean time between ED admission and outcome was 5.1 days (SD, 5.4; median 3.8; IQR, 1-7).

Performance of GRAM score in this cohort

Necessary data for GRAM score calculation was available in 183 patients. Using the published threshold (40%) for the GRAM score[4] to discriminate between high- and low-risk patients, we identified 51 patients at high risk and 132 at low risk of developing critical illness (Figure 2). When applied to the 143 patients who were integrated in the final analysis, no difference in GRAM score performance was found.

LUS score as a predictor of the main outcome

LUS images were successfully obtained in 211 patients. LUS aeration score ranged from a minimum of 0 to a maximum of 27, with a mean of 12.1 (SD, 7.4; median, 12) in favorable outcome patients and a mean of 16.2 (SD, 7.3; median, 17) in critically ill patients ($p<0.001$). A higher LUS score was

associated with a higher risk ratio (RR) of developing critical illness (RR, 2.05; 95% CI 1.52–2.77) (Figure 3). A value over 15 (out of 0 to 36) on LUS score demonstrated predictive discrimination between favorable and adverse outcomes (area under the curve [AUC], 0.63; 95% CI 0.676–0.634).

Performance of GRAM score powered by LUS

As an intermediate analysis, we investigated if the combination of the dichotomous LUS score with the COVID-GRAM score could increase the performance of GRAM score alone in predicting adverse outcome. We named this combined score GRAM-PLUS (GRAM powered by LUS). This calculation was performed in 143 patients based on the available data. The addition of sonographic findings to the GRAM score slightly reduced the number of patients of the low-risk category as initially established by the GRAM score, and raised the RR from 3.0 to 3.18 (Table 2, Figure 4).

Of note, when GRAM-PLUS was validated in our cohort of patients, unconsciousness and hemoptysis were ignored as these signs were absent. The optimism-adjusted model accuracy index was 0.5193 providing an estimated accuracy of 75.97%.

Performance of COVID-19 Worsening Score (COWS)

By using the Bayesian Averaging Model, we selected five predictive variables with their relative coefficients as follow: LUS score greater than 15, the number of comorbidities, days from the symptom onset, dyspnea at presentation, and P/F ratio (Table 3).

COWS ranged from 0 to 1 and the optimal accuracy was identified at a threshold of 0.183. Using this threshold, the same 143 patients were reclassified in 60 high-risk patients, of whom 35 (58.3%)

developed critical illness, and 83 low-risk patients, of whom 6 (7.2%) developed critical illness (Figure 5). Sensitivity and specificity for critical illness were 0.85 (95% CI, 0.75–0.96) and 0.75 (95% CI, 0.67–0.84), respectively. Positive predictive value (PPV) and negative predictive value (NPV) were 0.58 (95% CI, 0.46–0.71) and 0.93 (95% CI, 0.87–0.98) respectively. The risk ratio increased to 8.07 (95% CI, 4.97–11.1) (Table 2). The equation to calculate COWS is shown (below) in Formulas 1 and 2. Finally, we created a nomogram that can be used to calculate COWS manually (Figure 6).

Risk model formula

(1) $$LP = 1.849 + 0.427 \times C - 0.012 \times PF - 0.069 \times S + 0.946\ (if\ L > 15) - 0.802\ (if\ D\ present)$$

(2) $$\text{COW}\ Score = \frac{e^{LP}}{1 + e^{LP}}$$

Formula 1 and 2. Linear predictor calculation and subsequent COW score calculation. LP = linear predictor; C = number of comorbidities; PF = PF ratio (mmHg); S = days from symptoms onset; L = LUS score; D = dyspnea

Potential role of CT scan in reclassifying false positive and negative patients

After visual inspection of the COWS classification in high- and low-risk patients, we wondered whether second level radiological imaging, such as a thoracic CT scan, might be a viable means to improve the prediction of patients' outcome. Among the 143 patients with complete data, 46 were misclassified by COWS (i.e., they turned out to be false positive or false negative). CT scan data were available only for 59 patients, of whom 55 presented completed data. No statistically significant results were found between the 39 patients incorrectly classified in the high-risk group compared to the 7 incorrectly classified in the low-risk group. The number of involved lobes was greater in

high-risk than in the lower risk group (mean 5 vs. 2.5; P=0.43) as well as the percentage of emphysema (46.7 % in the higher risk vs. 0% in the lower risk group; P=0.485). Percentage of well-aerated lung was also lower in high-risk group (75.0 % vs 87.5 %; P=0.229).

Discussion

In this study, we developed and validated a new prognostic bedside score for early identification of COVID-19 related critical illness and named it COVID-19 Worsening Score (COWS). This new score integrated LUS findings and three selected variables of the previously validated COVID-GRAM score. Since COWS does not require laboratory or radiological results, it enables rapid stratification of patients upon ED arrival. This aspect is critical when considering the large COVID-19 patient influxes seen worldwide, which occasionally necessitated opening of outdoor tent areas and screening of patients in parking lots.

The overall accuracy of COWS is 80%, which is equal to the GRAM score. However, with a negative-predictive value of 93%, COWS better discriminates low-risk patients than the GRAM score, and may thus help in reducing inappropriate ICU admissions and optimizing hospital resources. Moreover, the ability to anticipate clinical worsening could provide benefits to patients, such as shortening the time spent on spontaneous breathing, or on NIV, to prevent patient-self-inflicted lung injury (P-SILI)[24,25].

COVID-GRAM score and COWS are not the only recently proposed scores. Several other prognostic scores were developed based on varying mixes of clinical data, laboratory results and radiological findings. Zhou et al. proposed a predictor of disease severity obtained from combining three

independent variables; the neutrophils to lymphocytes ratio (N/L), C-reactive protein (CRP) and D-dimer values. This product had better predictive performance than single biomarkers as proved by an internal validation study[3]. Several radiological scoring systems were also implemented to assess the severity of the disease and predict patient's outcomes. A chest X-ray (CXR) scoring system on 18-point scale, known as Brixia score, was proposed to quantify and monitor the severity of lung abnormalities[5]. The Brixia score when combined with the patient's age and presence of immunosuppression was shown to predict in-hospital mortality[6]. In a retrospective single center study evaluating 1,198 ED COVID-19 patients the accuracy of both CXR and computerized CT scan for diagnosis of COVID-19 were investigated. Sensitivity and specificity of CXR were 0.56 and 0.60, whereas for CT scan these were 0.85 and 0.50, respectively[7]. Despite its low specificity, CT confirmed the diagnosis of COVID-19 in patients with a false-negative RT-PCR as demonstrated in Chen et al. study. The lower the number of pulmonary consolidations on the CT, the greater the likelihood of a negative RT-PCR, suggesting the central role of CT as screening tool when COVID-19 is strongly suspected [8]. Association between CT findings and patient mortality was also studied[9]. Kunhua et al. investigated the combination of CT findings and clinical features in critical versus non-critical COVID-19 patients. Results indicated that CT could identify patients who needed aggressive treatment and close monitoring[10]. Another retrospective analysis investigated the link between lung aeration at baseline CT with the patient's adverse outcome: the degree of air loss and the presence of 4 or more lung lobes affected by COVID-19 pneumonia were associated with admission to ICU or death[11].

Hence, it may be suggested that CT scanning could be the optimal imaging tool in COVID-19, but it carries the burden of radiation exposure, higher cost and prolonged equipment cleaning time compared to LUS[26]. Moreover, both CT and LUS are not specific for COVID-19 pneumonia[27]. Keeping

a balance between accuracy and availability, LUS is a tool able to identify early signs of pulmonary lesions of COVID-19 pneumonia. Even though LUS cannot determine per se whether patients are infected by SARS-CoV-2, our results showed that in established COVID-19 cases, the higher the LUS score, the greater was the risk of developing critical illness. We identified that a LUS score value higher than 15 helps discriminate between favorable and adverse outcomes in our cohort of patients. This result is consistent with previous findings reported by Soummer et al[28]. Thus, a score based on sonographic (i.e., anatomical), functional and clinical clues may be the most reliable means to provide a quick evaluation of the patient from complementary points of view.

COWS is based on LUS, P/F ratio, dyspnea, number of disease and days from symptoms. For this reason, it acts as both a quick bedside tool and a screening test with a high negative predictive value. These two features suggest its usefulness in the context of the rapid evaluation of multiple patients presenting to the ED to avoid inappropriate resource use on low-risk patients saving costly resources for a minor number of high-risk patients. To this extent, the use of COWS may help increase appropriateness in the deployment of radiological resources, ventilatory equipment, and ICU admissions. Finally, one of the advantages of COWS compared to the GRAM score may also be its quick repeatability over time.

In the likely event of a second-wave massive inflow of patients overwhelming hospital resources, patients may be listed according to the calculated predicted risk, in order to help the decision on resource allocation. In particular, stratifying patients by means of COWS may help set the appropriate monitoring level and aid in the difficult process of applying reverse-triage criteria for ICU access in extreme conditions[29].

In the context of a long-lasting epidemic, where a model of hub-and-spoke COVID-19 hospitals might be used, COWS may speed up the selection of the low-risk patients who may be safely transferred to spokes, keeping high-risk patients in the hub-center.

This study has several limitations. Firstly, it is a retrospective single-center study and the sample size was relatively limited, as complete data were available for 143 patients. Moreover, even if the assessment of internal validity suggests potential usefulness of our newly developed score in clinical practice, however, external validation is needed to enhance the generalizability of our findings. A bigger multicenter, prospective research effort would also be advisable for a greater sample size collection. Secondly, despite COWS' ability to identify low-risk patients, recognition of high-risk patients remains suboptimal, and further adjustment should be applied. Thirdly, we used GRAM' variable selection to build our model, instead of starting from all the possible variables collected in our patients. However, this approach is reasonable as the selected variables were variables with plausible clinical relation with the outcome. Fourthly, the P/F may have been calculated on very different FiO_2 and with different levels of PEEP (from ZEEP to even 10 cmH_2O). Finally, we tried to assess whether a thoracic CT scan might be combined with COWS as a second level exam in selected patients to improve the overall accuracy, but we did not find promising results for this purpose, possibly due to the limited number of observations. Of note, the cross-sectional area of fat tissue at T7-T8 vertebral height, assessed in Colombi et al. [11], was not measured in our study due to CT software limitations.

Conclusion

COVID-19 pandemic has severely challenged hospitals' capacity in providing intensive levels of care. After validating the COVID-GRAM score in our population, we identified a simplified version of the score, by integrating LUS findings, functional, and selected clinical data. The COWS is bedside, quick, and easy to calculate. Its result is able to accurately identify patients who are unlikely to deteriorate or need ICU admission, sparing resources for the minority of COVID-19 patients with a high-risk of developing critical illness.

References

1. World Health Oganization. WHO/Europe | Coronavirus disease (COVID-19) outbreak - WHO announces COVID-19 outbreak a pandemic. Accessed April 17, 2020. http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic

2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497-506. doi:10.1016/S0140-6736(20)30183-5

3. Wu C, Chen X, Cai Y, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med*. 2020;180(7):934. doi:10.1001/jamainternmed.2020.0994

4. Liang W, Liang H, Ou L, et al. 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 2020. doi:10.1001/jamainternmed.2020.2033

5. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. *Radiol Med*. 2020;125(5):509-513. doi:10.1007/s11547-020-01200-3

6. Borghesi A, Zigliani A, Golemi S, et al. Chest X-ray severity index as a predictor of in-hospital mortality in coronavirus disease 2019: A study of 302 patients from Italy. *Int J Infect Dis*. 2020;96:291-293. doi:10.1016/j.ijid.2020.05.021

7. Borakati A, Perera A, Johnson J, Sood T. Chest X-Ray Has Poor Diagnostic Accuracy and Prognostic Significance in COVID-19: A Propensity Matched Database Study. *medRxiv*. Published online July 13, 2020:2020.07.07.20147934. doi:10.1101/2020.07.07.20147934

8. Waller J V., Allen IE, Lin KK, Diaz MJ, Henry TS, Hope MD. The Limited Sensitivity of Chest Computed Tomography Relative to Reverse Transcription Polymerase Chain Reaction for Severe Acute Respiratory Syndrome Coronavirus-2 Infection. *Invest Radiol*. 2020;Publish Ah. doi:10.1097/RLI.0000000000000700

9. Yuan M, Yin W, Tao Z, Tan W, Hu Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. Schildgen O, ed. *PLoS One*. 2020;15(3):e0230548. doi:10.1371/journal.pone.0230548

10. Li K, Wu J, Wu F, et al. The Clinical and Chest CT Features Associated With Severe and Critical COVID-19 Pneumonia. *Invest Radiol*. 2020;55(6):327-331. doi:10.1097/RLI.0000000000000672

11. Colombi D, Bodini FC, Petrini M, et al. Well-aerated Lung on Admitting Chest CT to Predict Adverse Outcome in COVID-19 Pneumonia. *Radiology*. Published online April 17,

2020:201433. doi:10.1148/radiol.2020201433

12. Lichtenstein D, Meziere G, Biderman P, Gepner A, Barre O. The comet-tail artefact: an ultrasound sign of alveolar-interstitial syndrome. *Am J Respir Crit Care Med*. 1997;156:1640-1646. doi:10.1164/ajrccm.156.5.96-07096

13. Volpicelli G, Elbarbary M, Blaivas M, et al. International evidence-based recommendations for point-of-care lung ultrasound. *Intensive Care Med*. 2012;38(4):577-591. doi:10.1007/s00134-012-2513-4

14. Shi H, Han X, Jiang N, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis*. 2020;20(4):425-434. doi:10.1016/S1473-3099(20)30086-4

15. Yang Y, Huang Y, Gao F, Yuan L, Wang Z. Lung ultrasonography versus chest CT in COVID-19 pneumonia: a two-centered retrospective comparison study from China. *Intensive Care Med*. 2020;46(9):1761-1763. doi:10.1007/s00134-020-06096-1

16. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-W73. doi:10.7326/M14-0698

17. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. *Am J Respir Crit Care Med*. 2019;200(7):E45-E67. doi:10.1164/rccm.201908-1581ST

18. Gao H-N, Lu H-Z, Cao B, et al. Clinical Findings in 111 Cases of Influenza A (H7N9) Virus Infection. *N Engl J Med*. 2013;368(24):2277-2285. doi:10.1056/NEJMoa1305584

19. Orsi L, Societa Italiana di Anestesia, Analgesia R e TI. [SIAARTI recommendations for the admission and discharge from Intensive Care and for the limitation of treatments in Intensive Care]. *Minerva Anestesiol*. 2003;69(4):179. http://www.ncbi.nlm.nih.gov/pubmed/12766703

20. Bouadma L, Lescure F-X, Lucet J-C, Yazdanpanah Y, Timsit J-F. Severe SARS-CoV-2 infections: practical considerations and management strategy for intensivists. *Intensive Care Med*. Published online February 2020. doi:10.1007/s00134-020-05967-x

21. Bouhemad B, Mongodi S, Via G, Rouquette I. Ultrasound for "lung monitoring" of ventilated patients. *Anesthesiology*. 2015;122(2):437-447. doi:10.1097/ALN.0000000000000558

22. Bouhemad B, Dransart-Raye O, Mojoli F, Mongodi S. Lung ultrasound for diagnosis and monitoring of ventilator-associated pneumonia. *Ann Transl Med*. 2018;6(21):418. doi:10.21037/atm.2018.10.46

23. R Development Core Team T. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2020.

24. Brochard L, Slutsky A, Pesenti A. Mechanical Ventilation to Minimize Progression of Lung Injury in Acute Respiratory Failure. *Am J Respir Crit Care Med*. 2017;195(4):438-442. doi:10.1164/rccm.201605-1081CP

25. Marini JJ, Gattinoni L. Management of COVID-19 Respiratory Distress. *JAMA - J Am Med Assoc*. Published online 2020. doi:10.1186/cc2392

26. Gaia C, Maria Chiara C, Silvia L, et al. Chest CT for early detection and management of coronavirus disease (COVID-19): a report of 314 patients admitted to Emergency Department with suspected pneumonia. *Radiol Med*. 2020;125(10):931-942. doi:10.1007/s11547-020-01256-1

27. Ai T, Yang Z, Hou H, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*. 2020;296(2):E32-E40. doi:10.1148/radiol.2020200642

28. Soummer A, Perbet S, Brisson H, et al. Ultrasound assessment of lung aeration loss during a successful weaning trial predicts postextubation distress*. *Crit Care Med*. 2012;40(7):2064-2072. doi:10.1097/CCM.0b013e31824e68ae

29. Vergano M, Bertolini G, Giannini A, et al. Clinical Ethics Recommendations for the Allocation of Intensive Care Treatments in exceptional, resource-limited circumstances - Version n. 1 Posted on March, 16. *Crit Care*. 2020;24(165):1-3.

Table I. Demographics and clinical characteristics of the patients who did and didn't develop critical illness

Characteristic	Total (n=274)	No critical illness (n=174)	Critical illness (n=100)	P-value
Age, mean (SD) [range]	67.7 (14.4) [21-96]	64.9 (14.5) [21-96]	72.6 (12.8) [35-89]	<0.000
Gender, male (%)	189 (69.0)	117 (67.2)	72 (72.0)	0.412
Days from symptom onset, mean (SD) [range]	5.8 (4.4) [0-31]	6.4 (4.8) [0-31]	4.8 (3.5) [0-14]	0.009
Number of comorbidities 0 1 2 3 4 5+	(n=268) 71 (24.2) 71 (24.2) 58 (19.8) 38 (13.0) 22 (7.5) 8 (2.7)	(n=171) 56 (32.7) 53 (31.0) 31 (18.1) 20 (11.7) 8 (4.7) 3 (1.8)	(n=97) 15 (15.5) 18 (18.6) 27 (27.8) 18 (18.6) 14 (14.4) 5 (5.1)	<0.000
Malignancy (%)	20 (7.4)	11 (6.4)	9 (9.3)	0.387
Dyspnea (%)	139 (51.7)	75 (43.9)	64 (65.3)	0.001
Hemoptysis (%)	2 (0.74)	1 (0.58)	1 (1.02)	0.597
Unconsciousness (%)	2 (0.74)	-	2 (2.04)	0.132
Abnormal chest radiography/CT (%)	201 (82.4)	124 (77.0)	77 (92.8)	0.002
LUS score at admission, mean (SD) [range]	(n=211) 13.4 (7.6) [0-27]	(n=146) 12.1 (7.4) [0-27]	(n=65) 16.2 (7.3) [0-27]	<0.000
PF ratio at admission, Mean (SD) [range]	(n=245) 263.9 (94.6) [33-647]	(n=164) 297.5 (78.3) [50-647]	(n=81) 196 (88.4) [33-396]	<0.001

Table II. Comparative performance of the three scores (GRAM, GRAM-PLUS and COWS) on 143 patients with available data.

	high risk			low risk			RR (95% CI)	P-value
	Critical illness N (%)	Favorable outcome N (%)	Total	Critical illness N (%)	Favorable outcome N (%)	Total		
GRAM score	22 (56.4)	17 (43.6)	39	19 (18.3)	85 (81.7)	104	3.0 (2.02 - 4.09)	< 0.001
GRAM-PLUS	24 (54.5)	20 (45.5)	44	17 (17.2)	82 (82.8)	99	3.18 (2.07 - 4.25)	< 0.001
COWS	35 (58.3)	25 (41.7)	60	6 (7.2)	77 (92.8)	83	8.07 (4.97-11.1)	< 0.001

Table III. Most predictive variables identified and their effect with 95% confidence intervals

Variable	Effect	95% CI			P-value
Number of comorbidities	1.688	1.216	-	2.344	0.002
LUS score above 15	3.511	1.283	-	9.612	0.015
PF ratio	0.218	0.109	-	0.434	< 0.001
Days from symptom onset	0.595	0.340	-	1.041	0.069
Dyspnea	0.308	0.097	-	0.976	0.045

.

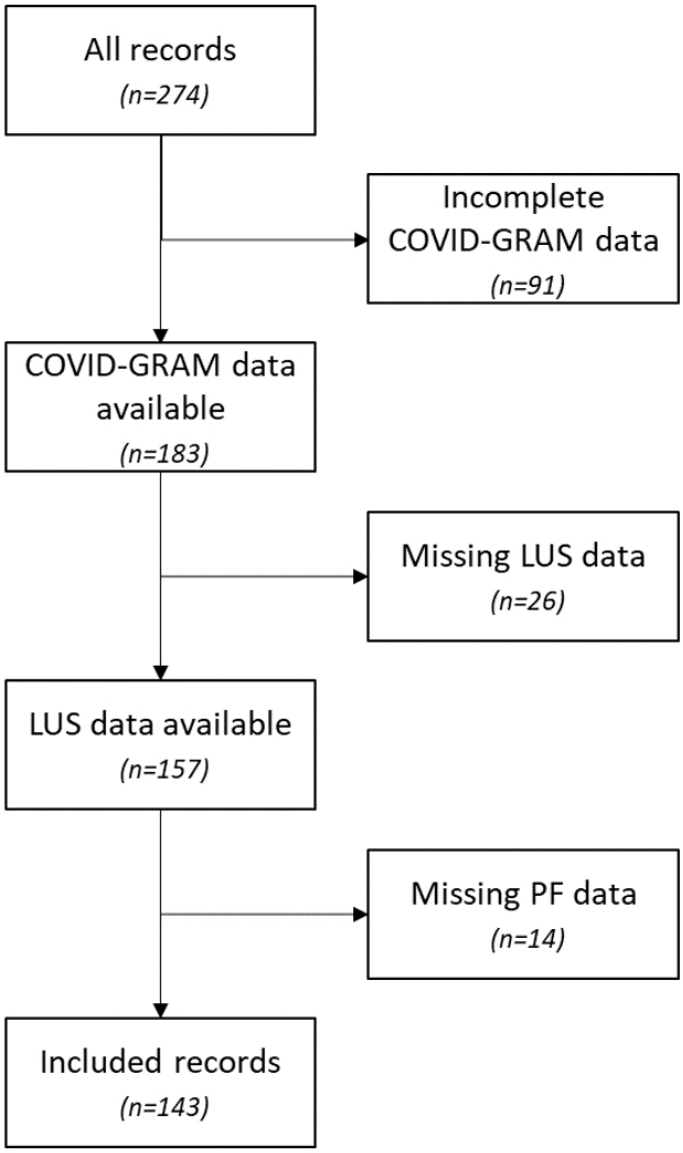
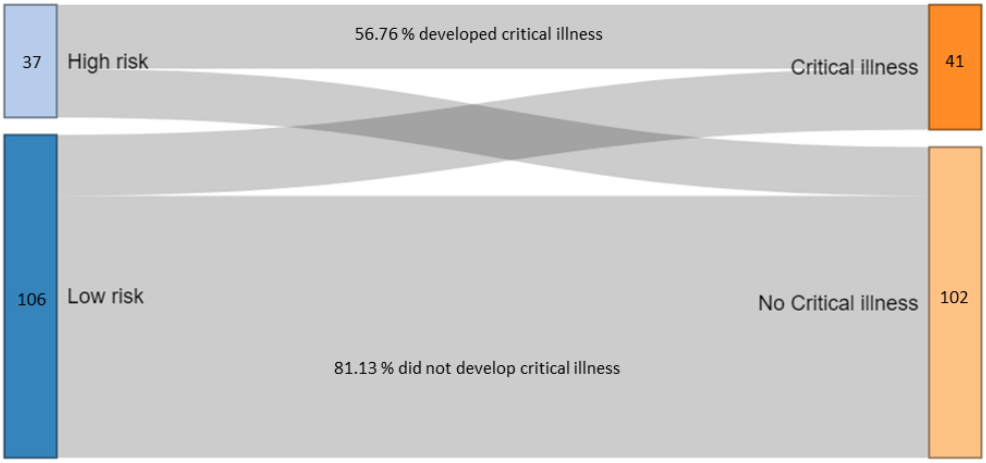
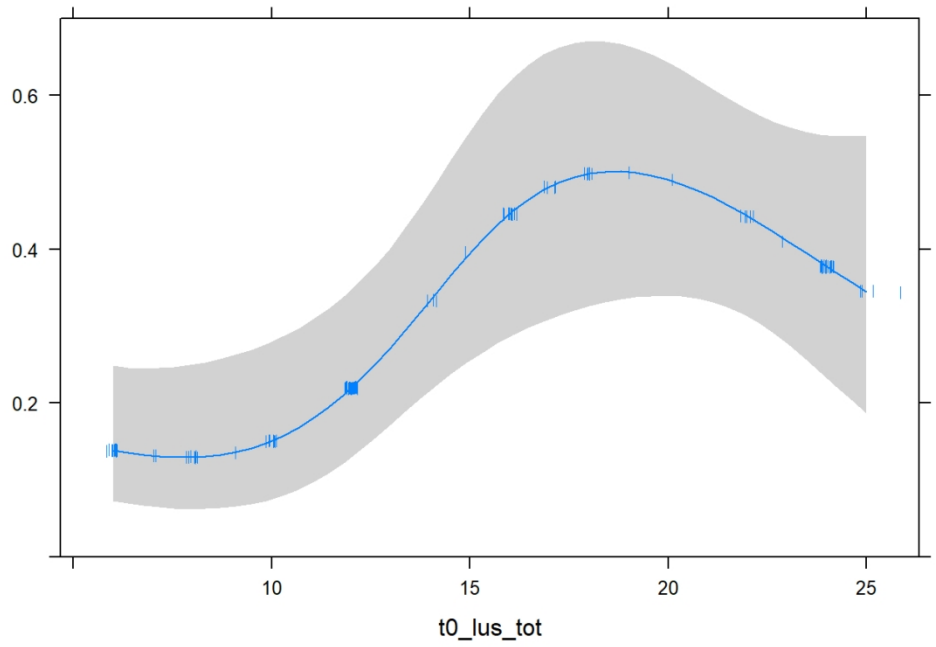


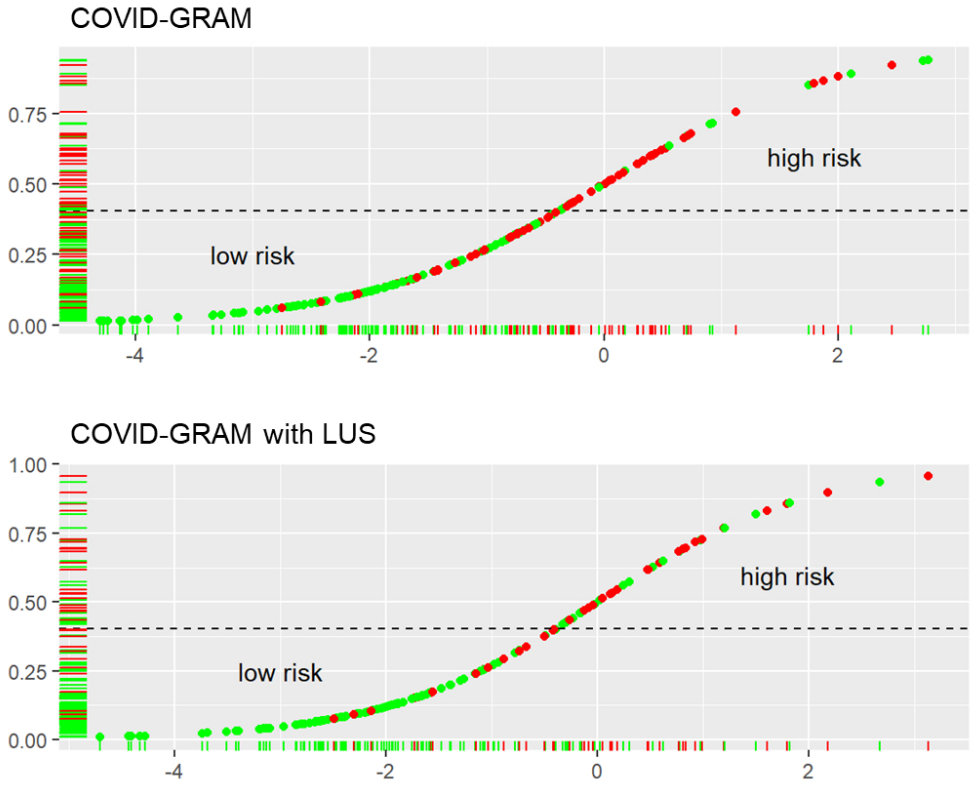
Diagram of included patients.



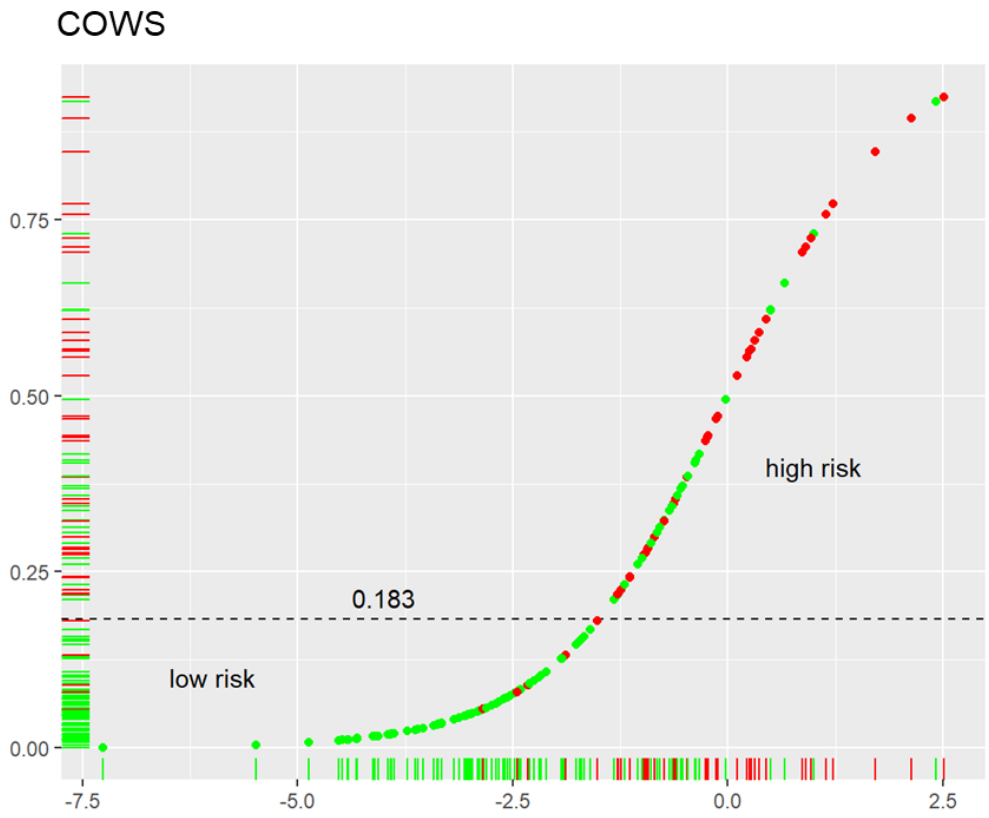
GRAM score derived risk groups (on the left) and outcomes (on the right); grey shadows link classification to outcomes and their width is proportional to the number of patients.



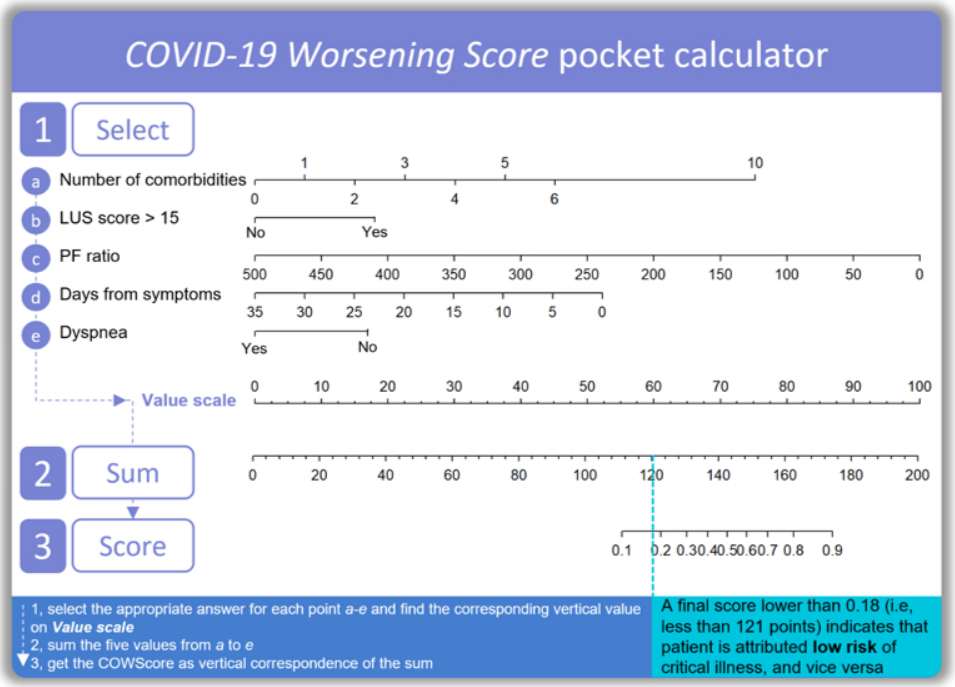
Probability of developing critical illness (Y-axis) according to increasing values of LUS score (X-axis)



Distribution curves of the patients who developed critically illness (red dots) and those who had favorable outcomes (green dots). X- axis: linear predictor; Y-axis: incremental values of GRAM score (panel A) and GRAM-PLUS values (panel B). LUS: lung ultrasound.



Performance of the COWS in classifying high- and low-risk patients. Red dots indicate patients with adverse outcome. Dashed line refers to the COWS threshold.


COVID-19 Worsening Score pocket calculator
1 Select
a Number of comorbidities
1 3 5 10
0 2 4 6
b LUS score > 15
No Yes
c PF ratio
500 450 400 350 300 250 200 150 100 50 0
d Days from symptoms
35 30 25 20 15 10 5 0
e Dyspnea
Yes No
Value scale
0 10 20 30 40 50 60 70 80 90 100
2 Sum
0 20 40 60 80 100 120 140 160 180 200
3 Score
0.1 0.2 0.30.40.50.60.7 0.8 0.9
1, select the appropriate answer for each point a-e and find the corresponding vertical value on Value scale
2, sum the five values from a to e
3, get the COWScore as vertical correspondence of the sum
A final score lower than 0.18 (i.e, less than 121 points) indicates that patient is attributed low risk of critical illness, and vice versa