





Article

External Validation of COVID-19 Risk Scores during Three Waves of Pandemic in a German Cohort—A Retrospective Study

Lukas Häger¹, Philipp Wendland² , Stephanie Biergans³, Simone Lederer³, Marius de Arruda Botelho Herr³, Christian Erhardt³, Kristina Schmauder⁴, Maik Kschischo², Nisar Peter Malek¹, Stefanie Bunk¹, Michael Bitzer¹ , Beryl Primrose Gladstone^{1,5}  and Siri Göpel^{1,5,*} 

¹ Department of Internal Medicine 1, University Hospital Tübingen, Otfried-Müllerstrasse 10, 72076 Tübingen, Germany

² Department of Mathematics and Technology, University of Applied Sciences Koblenz, Joseph-Rovan-Allee 2, 53424 Remagen, Germany

³ Medical Data Integration Center (meDIC), University Hospital Tübingen, Schaffhausenstraße 77, 72072 Tübingen, Germany

⁴ Institute of Medical Microbiology and Hospital Hygiene, University Hospital Tübingen, Elfriede-Aulhorn-Str. 6, 72076 Tübingen, Germany

⁵ Clinical Research Unit for Health Care Associated Infections, German Center for Infection Research (DZIF), Otfried-Müllerstrasse 10, 72076 Tübingen, Germany

* Correspondence: siri.goepel@med.uni-tuebingen.de



Citation: Häger, L.; Wendland, P.; Biergans, S.; Lederer, S.; de Arruda Botelho Herr, M.; Erhardt, C.; Schmauder, K.; Kschischo, M.; Malek, N.P.; Bunk, S.; et al. External Validation of COVID-19 Risk Scores during Three Waves of Pandemic in a German Cohort—A Retrospective Study. *J. Pers. Med.* **2022**, *12*, 1775. <https://doi.org/10.3390/jpm12111775>

Academic Editors: Weikuan Gu and Lotfi Aleya

Received: 21 September 2022

Accepted: 24 October 2022

Published: 28 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Several risk scores were developed during the COVID-19 pandemic to identify patients at risk for critical illness as a basic step to personalizing medicine even in pandemic circumstances. However, the generalizability of these scores with regard to different populations, clinical settings, healthcare systems, and new epidemiological circumstances is unknown. The aim of our study was to compare the predictive validity of qSOFA, CRB65, NEWS, COVID-GRAM, and 4C-Mortality score. In a monocentric retrospective cohort, consecutively hospitalized adults with COVID-19 from February 2020 to June 2021 were included; risk scores at admission were calculated. The area under the receiver operating characteristic curve and the area under the precision–recall curve were compared using DeLong's method and a bootstrapping approach. A total of 347 patients were included; 23.6% were admitted to the ICU, and 9.2% died in a hospital. NEWS and 4C-Score performed best for the outcomes ICU admission and in-hospital mortality. The easy-to-use bedside score NEWS has proven to identify patients at risk for critical illness, whereas the more complex COVID-19-specific scores 4C and COVID-GRAM were not superior. Decreasing mortality and ICU-admission rates affected the discriminatory ability of all scores. A further evaluation of risk assessment is needed in view of new and rapidly changing epidemiological evolution.

Keywords: COVID-19 risk score; NEWS; COVID-GRAM; 4C-Mortality score; pandemic; ICU

1. Introduction

COVID-19 spread around the world in alarming speed, burdening healthcare systems and hospitals. A large number of COVID-19 patients in life-threatening conditions had to be treated in hospitals that were provisionally set up, often by less experienced staff, indicating the need for an easy and objective clinical scoring model to identify high-risk patients [1]. Furthermore, risk scores are the first step in personalizing medicine to stratify individual patients to medical treatment. Several COVID-19-specific risk scores have been developed and validated during the first wave involving different populations and clinical settings: The COVID-GRAM risk score (GRAM: Guangzhou Institute of Respiratory Health Calculator at Admission) was developed in a large cohort of 1590 patients from 575 hospitals in China [2] with the aim of predicting critical illness in COVID-19 patients defined as admission to ICU (intensive care unit), mechanical ventilation, or death. The

risk score summarizes 10 variables based on a logistic model and requires—due to the complexity—an online tool for usage ([2], see Appendix A). In 2020, the 4C-Mortality Index (Coronavirus Clinical Characterisation Consortium) was developed by including 260 hospitals and 35,463 patients in England, Scotland, and Wales to predict in-hospital mortality [3]. In contrast to the younger patient cohort (mean age of 49 years) with a lower mortality rate of 3.2%, which helped develop the COVID-GRAM risk score, the 4C-Mortality Index was developed in an older cohort (mean age was 73 years) with a higher in-hospital mortality rate of 32.2% [3]. Such differences question the generalizability of these scores and warrant further validation of the scores in different populations and various clinical settings. Moreover, several preexisting risk scores developed for other infectious diseases and intensive care medicine such as NEWS (National Early Warning Score), CRB-65 (confusion, respiratory rate, blood pressure—age 65), and qSOFA (quick sequential organ failure assessment) are also in use, though their predictivity among COVID-19 patients remains unclear. For example, during the study period, NEWS was routinely being used as a risk assessment tool to initiate a timely clinical response such as nurse–doctor–contact or notification of ICU for COVID-19 patients at the University Hospital Tübingen. Though several authors examined the predictive performance of some of the scores in COVID-19 risk stratification [4–10], the performance of multiple scores has not been compared in the same patient population. Over the course of the pandemic, epidemiological circumstances and therapeutic options rapidly changed: remdesivir received conditional approval in the treatment of COVID-19 patients in July 2020, and the benefit of dexamethasone was proven by the RECOVERY Collaborative Group [11]. In December 2020, new SARS-CoV-2 variants with high transmissibility and a potential immune escape were reported in the United Kingdom and South Africa, which were declared by the WHO as variants of concern [12]. In 2021, the European Union licensed several COVID-19 vaccinations, which had shown high efficacy and safety in clinical trials. The progress of immunization and vaccination, different COVID-19 variants, various healthcare systems, and the increasing efficacy of treatment have had a great impact on COVID-19 mortality risk. Nevertheless, there is a residual risk reported for critical illness even among the vaccinated population [4]. Until today, there is no risk score recommended in the German COVID-19 guidelines [13].

Amidst this background, a retrospective cohort study was conducted with the primary objective of evaluating the performance of common clinical scores (NEWS, qSOFA, and CRB-65) and COVID-19-specific clinical scoring models (COVID-GRAM, 4C-Mortality score) for the prediction of ICU admission and in-hospital mortality and comparing their predictive performance in a cohort of hospitalized COVID-19 patients in Germany. To the best of our knowledge, this is the first study to compare all of the abovementioned scores in a defined population, comparing their discriminative abilities in different waves during the course of the pandemic.

2. Materials and Methods

2.1. Study Design and Setting

We conducted a monocentric retrospective cohort study at the University Hospital Tübingen (tertiary care hospital) located in Tübingen, Germany, during the first three waves of the COVID-19 pandemic.

2.2. Study Population

The study recruited patients ≥ 18 years admitted to the university hospital due to COVID-19 between 1 March 2020 and 30 May 2021. SARS-CoV-2 infection was confirmed by positive real-time polymerase chain reaction (RT-PCR). Exclusion criteria were as follows:

- Patients not hospitalized for COVID-19 disease;
- Patients with a patient decree determining a DNR/DNI (do not resuscitate/do not intubate) situation;
- Patients transferred to our ICU from other hospitals, for example, due to the need for extracorporeal membrane oxygenation (ECMO).

2.3. Definition of Cohorts

The study period was divided into three cohort periods according to the classification of the waves by the Robert-Koch Institute [14] and based on key epidemiological factors: the number of COVID-19 cases, prevalence of different SARS-CoV-2 variants in Germany and at the University Hospital Tübingen, and availability of standardized specific therapy and vaccination. The prevalence of virus variants was assessed according to own data when genotyping was available and general evidence [12,15]. COVID-19 vaccination rates in Germany were assessed according to the results of the COVIMO study group [16]. The study population was thus divided into cohort 1 from 1 March to 30 June 2020, cohort 2 from 1 July 2020 to 7 March 2021, and cohort 3 from 8 March to 30 May 2021 (see Table 1). We decided to slightly deviate from the Robert-Koch Institute's classification of the COVID-19 waves due to the decreasing incidence of hospital admissions in summer (only 7 patients fulfilled the inclusion criteria between July and September 2020).

Table 1. Definition of cohorts.

	Cohort 1	Cohort 2	Cohort 3
Admission period	1 March 2020–30 June 2020	1 July 2020–7 March 2021	8 March 2021–30 May 2021
Variants	High prevalence of wild type	High prevalence of B.1.177	High prevalence of alpha variant
Specific therapy	None	Remdesivir, steroids	Remdesivir, steroids
Vaccination	None	Completely vaccinated: 0.9–3.3% * (vaccination available since 26 December 2020)	Completely vaccinated: 6.9–27.4% *

* Vaccination rates according to [16].

2.4. Data Collection and Score Validation

We retrospectively collected clinical, demographical, and outcome data for each cohort by using the clinical information and documentation systems. Comorbidities included were chronic respiratory disease, cardiovascular disease, chronic liver disease, chronic kidney disease, HIV (human immunodeficiency virus) infection or AIDS (acquired immunodeficiency syndrome), organ transplantation, diabetes mellitus, malignancy, and chronic neurological conditions. Demographical and epidemiological data collected were age, sex, body mass index, COVID-19 vaccination, and DNR/DNI status. The scores studied were the Quick Sequential Organ Failure Assessment score (qSOFA), National Early Warning Score (NEWS), CRB-65 score, COVID-GRAM risk score, and 4C-Mortality Score (see Table 2). Each of the scores was separately calculated for each patient using the admission data. In case of missing values at admission, we decided to collect the earliest parameter on the day of admission or day 1 after admission to improve the power of the study. This included the laboratory parameter serum urea concentration and radiological data. In case of missing direct bilirubin, we used total bilirubin. Glasgow Coma Scale was retrospectively calculated (for the workflow, see supplementary Figure S1).

Table 2. Parameter of risk scores.

qSOFA	CRB-65	NEWS	COVID-GRAM	4C-Score
Respiratory rate	Confusion	Respiratory rate	X-ray abnormality	Age
Consciousness	Respiratory rate	Oxygen saturation	Age	Sex at birth
Blood pressure	Blood pressure	Suppl. oxygen	Hemoptysis	Number of comorbidities
	Age	Temperature	Dyspnea	Respiratory rate
		Blood pressure	Unconsciousness	Oxygen saturation
		Heart rate	Number of comorbidities	Glasgow Coma Scale
		Consciousness	Cancer history	Urea
			Neutrophil/lymphocytes	CRP level
			Lactate dehydrogenase	
			Direct bilirubin	

2.5. Statistical Analysis

After the inclusion of patients by the above-defined inclusion and exclusion criteria, in a second step, patients with missing values were excluded. In the case of at least one missing value, we did not calculate the specific score (Supplementary Table S1, for characteristics of excluded patients, see Supplementary Table S2). The primary outcomes were endpoints of critical illness defined by ICU admission and in-hospital mortality. Discriminative indices of the selected scores including sensitivity, specificity, and positive and negative predictive values were calculated. Confidence intervals were assessed via a bootstrapping method. The discriminative abilities of the scores were assessed and compared by using the area under the receiver operating characteristic curve (AU-ROC) and the area under the precision–recall curve (AU-PRC). Equality of the AU-ROCs was tested using DeLong’s method and a bootstrapping method as well [17,18]. We decided to show the bootstrap findings in the results. To correct for multiple testing, we used Bonferroni–Holm adjusted *p*-values. An adjusted *p*-value of 0.05 or less was regarded as significant. Data were analyzed by R Version 4.1.2 using the packages readxl, pROC, precrec, boot, and ggplot2 [19–22]. We reported continuous variables as mean with the first and third quartiles and categorical variables as a number with the percentage of the cohort. Continuous variables were compared using the Mann–Whitney U-test, and categorical variables were compared by the use of Fisher’s exact test.

3. Results

3.1. Study Population

Overall, 756 patients were admitted to the hospital with SARS-CoV-2 infection from February 2020 to May 2021, of which 347 (45.9%) patients were included (see Table 3). Of these, 134 (38.6%), 187 (53.9%), and 26 (7.5%) formed the first, second, and third cohorts, respectively. A total of 153 patients (44.1%) were women, and the average age at the time of admission was 65.4 years (IQR 57.0 to 78.0). A total of 173 (49.9%) showed COVID-19-specific findings in X-ray, 184 (53%) had dyspnea, 179 (51.6%) had cough, and 189 (52.2%) had fever. The most common comorbidities were cardiovascular diseases, hypertension, and diabetes. A total of 146 patients (42.1%) were treated with steroids, and 131 (37.8%) required specific medication with remdesivir. One (0.3%) participant was vaccinated at least once against COVID-19; the type of vaccine is unknown. The overall in-hospital mortality rate was 9.2% (*n* = 32), and 23.6% (*n* = 82) required ICU treatment. The characteristics of the study cohort are shown in Table 3.

Table 3. Characteristics of study population.

Characteristics	Overall No (%) or Mean (Q1–Q3)	First Cohort No (%) or Mean (Q1–Q3)	Second Cohort No (%) or Mean (Q1–Q3)	Third Cohort No (%) or Mean (Q1–Q3)	<i>p</i> -Value *
Study population	347 (100)	134 (38.6)	187 (53.9)	26 (7.5)	-
In-hospital mortality	32 (9.2)	23 (17.2)	9 (4.8)	0 (0)	0.044
ICU admission	82 (23.6)	50 (37.3)	30 (16.0)	2 (7.7)	0.002
Average age	65.4 (57.0–78.0)	67.0 (58.5–80.0)	65.1 (57.0–76.0)	59.6 (51.8–68.3)	1
Women	153 (44.1)	57 (42.5)	84 (44.9)	12 (46.2)	1
COVID-19 vaccination	1 (0.3)	0 (0)	0 (0)	1 (3.8)	-
Respiratory disease	72 (20.7)	22 (16.4)	44 (23.5)	6 (23.1)	1
Cardiovascular disease	246 (70.9)	99 (73.9)	131 (70.1)	16 (61.5)	1
Diabetes	82 (23.6)	32 (23.9)	46 (24.6)	4 (15.3)	1
Hypertension	204 (58.8)	86 (64.2)	106 (56.7)	12 (46.2)	1
Liver disease	27 (7.8)	10 (7.5)	12 (6.4)	5 (19.2)	1
Chronic kidney disease	41 (11.8)	12 (9.7)	25 (13.4)	3 (11.5)	1
HIV or AIDS	0 (0)	0 (0)	0 (0)	0 (0)	-
Organ transplantation	11 (3.2)	4 (3.0)	6 (3.2)	1 (3.8)	1
Malignancy (active)	45 (13.0)	9 (6.7)	34 (18.2)	2 (7.7)	0.236
Malignancy (history)	29 (8.4)	20 (14.9)	7 (3.7)	2 (7.7)	0.070
Neurological conditions	79 (22.8)	29 (21.6)	44 (23.5)	6 (23.1)	1
COVID-19 findings	173 (49.9)	82 (61.2)	79 (42.2)	12 (46.2)	0.272
Oxygen saturation (%)	93.4 (92.0–97.0)	91.9 (90.0–96.0)	94.4 (93.0–97.0)	94.2 (93.0–97.0)	0.034
Heart rate (/min)	86.5 (74.0–96.0)	88.9 (76.0–100.0)	85.1 (72.0–94.0)	83.4 (69.0–91.0)	1
Respiratory rate (/min)	21.1 (16.0–24.0)	22.4 (17.0–26.0)	20.6 (16.3–24.0)	17.8 (15.0–20.0)	1
Syst. Blood pressure (mmHg)	132.0 (118.0–145.0)	129.3 (110.0–140.0)	133.8 (120.0–148.0)	127.1 (116.3–138.3)	1
Temperature (°C)	37.1 (36.4–37.7)	37.3 (36.5–38.1)	36.8 (36.2–37.4)	37.5 (36.8–38.3)	0.036
GCS	14.8 (15.0–15.0)	14.8 (15.0–15.0)	14.9 (15.0–15.0)	14.8 (15.0–15.0)	1
Dyspnea	184 (53.0)	71 (53.0)	103 (55.1)	10 (38.5)	1
Cough	179 (51.6)	84 (62.7)	88 (47.1)	7 (26.9)	0.751
Fever	181 (52.2)	86 (64.2)	80 (42.8)	15 (57.7)	0.026
Leukocyte count (/μL)	6828.2 (4355.0–8355.0)	6586.8 (4385.0–7970.0)	7064.0 (4305.0–8575.0)	6389.1 (4543.0–7665.0)	1
Neutrophil count (10 ³ /μL)	5.2 (3.0–6.4)	4.9 (3.1–6.3)	5.4 (2.9–6.8)	5.3 (3.4–6.4)	1
Lymphocyte count (10 ³ /μL)	1.1 (0.6–1.2)	1.0 (0.6–1.1)	1.3 (0.6–1.3)	1.0 (1.0–1.2)	1
Urea (mmol/L)	49.5 (27.0–62.0)	47.1 (24.8–62.0)	52.2 (28.0–65.0)	43.5 (29.0–56.8)	1
CRP (mg/dL)	8.0 (1.0–11.9)	8.9 (2.6–13.3)	7.7 (1.7–11.2)	6.3 (1.9–11.5)	1
Bilirubin (mg/dL)	0.7 (0.4–0.8)	0.7 (0.5–0.9)	0.6 (0.4–0.8)	0.6 (0.4–0.8)	0.507
LDH (U/L)	346.6 (229.8–378.5)	328.6 (229.5–389.0)	361.4 (225.0–365.5)	344.0 (240.0–411.0)	1
Steroids	146 (42.1)	24 (17.9)	103 (59.9)	19 (73.1)	-
Remdesivir	131 (37.8)	0 (0)	112 (55.1)	19 (73.1)	-

* Bonferroni–Holm adjusted *p*-values for the comparison of the first and second cohorts.

3.2. Comparison of Cohorts

The ICU-admission and in-hospital-mortality rates significantly decreased from the first to the second wave (p -values of 0.002 and 0.044, respectively). The mean oxygen saturation at admission was significantly higher in the second cohort (p -value of 0.034), whereas documented fever decreased (p -value of 0.036). We did not find significant differences between the first and the second wave considering age, sex, and comorbidities. Due to the low sample size of the third-wave cohort, statistical tests were less significant. Only the mean respiratory rate was significantly lower in the third cohort compared with the first cohort (p -value of 0.04). Nevertheless, we could observe a decreasing trend in the ICU-admission rate, in-hospital mortality rate, and average age also between the second and third cohorts.

3.3. Predictive Performance of the Scores

We included the whole study population in the comparison of the predictive performance of the scores. Due to the low sample size of the cohort, three statistical tests for the comparison between cohort 3 and the other two cohorts had a lower power (Figure 1). Therefore, the third cohort was excluded from the comparison between the cohorts (discriminatory indices, Supplementary Table S3). Overall, the NEWS model performed the best regarding ICU admission, which was confirmed by ROC-AUC (0.83; CI 0.76–0.88), whereas the 4C-Score showed a higher PR-AUC (0.64; CI 0.50–0.78) than NEWS. We found significant differences in the ROC-AUC compared with qSOFA (0.70; CI 0.64–0.77); however, differences to the other models were not statistically significant. The 4C-Score had the highest ROC-AUC (0.81; CI 0.69–0.90) with regard to in-hospital mortality. Again, differences to the other scores were not statistically significant. qSOFA, CRB-65 score, and COVID-GRAM performed lower (Figure 2).

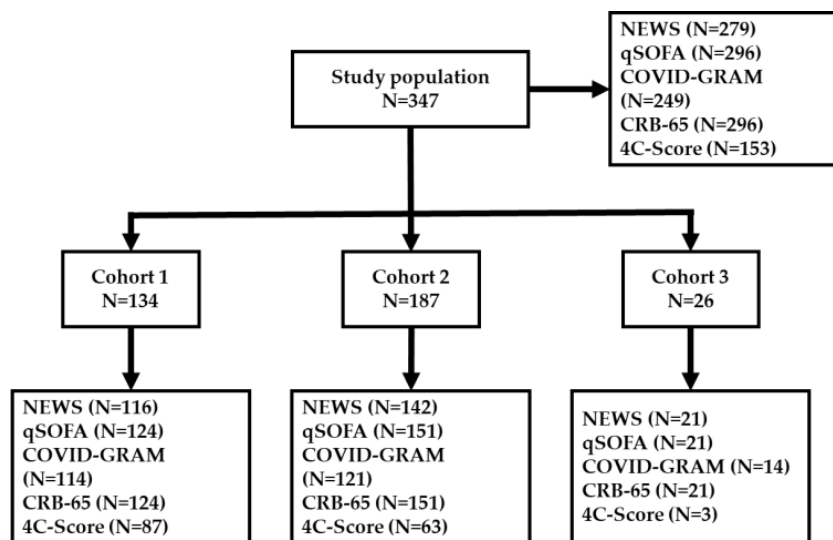


Figure 1. Overview of score calculation. Abbreviations: NEWS, National Early Warning Score; qSOFA, quick sequential organ failure assessment; COVID, coronavirus disease; GRAM, Guangzhou Institute of Respiratory Health Calculator at Admission; CRB, confusion, respiratory rate, blood pressure.

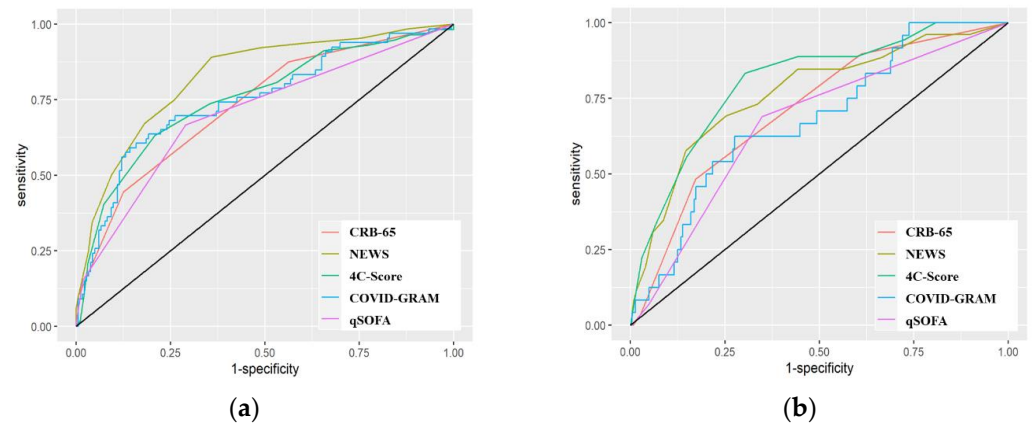


Figure 2. Plots of area under the receiver operating characteristics of the scores. (a) ICU admission (b) and in-hospital mortality. Abbreviations: CRB, confusion, respiratory rate, blood pressure; NEWS, National Early Warning Score; COVID, coronavirus disease; GRAM, Guangzhou Institute of Respiratory Health Calculator at Admission; qSOFA, quick sequential organ failure assessment.

3.4. Comparison between the Cohorts

The performance of NEWS and 4C-Score was better in cohort 1 than in cohort 2. The NEWS had a ROC-AUC of 0.88 (CI 0.80–0.94) in the first cohort and 0.71 (CI 0.60–0.81) in the second cohort with regard to ICU admission, whereas the ROC-AUC was 0.75 (CI 0.62–0.86) in the first cohort and 0.73 (CI 0.53–0.89) in the second cohort with regard to in-hospital mortality. Differences were statistically significant between the cohorts concerning ICU admission (p -value of 0.011). However, differences were not statistically significant regarding in-hospital mortality. The 4C-Score showed a ROC-AUC of 0.84 (CI 0.73–0.93) in the first cohort and 0.58 (CI 0.42–0.72) in the second cohort concerning ICU admission. The ROC-AUC was 0.87 (CI 0.78–0.94) in the first cohort and 0.59 (CI 0.33–0.84) in the second cohort concerning in-hospital mortality. ICU admission differences between the cohorts were statistically significant (p -value of 0.002), as well as in-hospital mortality differences (p -value of 0.045). We did not find any significant difference within the cohorts regarding qSOFA, CRB-65, and COVID-GRAM (details in supplementary Figure S2).

4. Discussion

We evaluated the performance of various COVID-19-specific as well as commonly used risk scores in a retrospective cohort of COVID-19 patients over the course of three waves of the pandemic. In our study, COVID-specific scores including COVID-GRAM and 4C-Score failed to show significant superiority compared with the NEWS model. This was shown especially for the prediction of ICU admission. Liang et al. reported an AUC of 0.88 (CI 0.85–0.91) in the derivation cohort of COVID-GRAM for the composite endpoint of ICU admission, need for invasive ventilation, and death [2]. This was not well-reflected in our findings with AUC 0.75 for ICU admission and 0.65 for in-hospital mortality. Knight et al. reported an AUC of 0.79 (CI 0.78–0.79) in the derivation cohort of the 4C-Score for the endpoint in-hospital mortality [3]. In our first cohort—which might reflect the original derivation cohort—we found an AUC 0.84 for ICU admission and 0.87 for in-hospital mortality, thus confirming their results. The usage of COVID-GRAM and 4C-Mortality score in clinical assessment is more complex compared with NEWS using more variables and requiring radiological assessment in addition to the laboratory parameters as well as information on relevant comorbidities; COVID-GRAM needs to be calculated by a calculator available online. NEWS includes more vital signs than both the COVID-specific scores; the assessment can be easily made on the bedside. In every-day clinical life, this is a clear advantage. The different baseline characteristics (especially the average age, ICU admission rates, and mortality rates) of the derivation cohorts and differences in epidemiological

conditions (e.g., healthcare systems, need for triage) might be the reason for differences of performance in our cohort. It has been demonstrated before that the impact of changing vital sign categories on prognosis in terms of mortality is larger in older patients [23]. This can be especially discussed for the derivation cohort of COVID-GRAM with a mean age of 48.9 years in comparison with our cohort with a mean age of 65 years. 4C performed better in our cohort than COVID-GRAM but did not show significant superiority to NEWS. Interestingly, we could not reproduce the AUC findings of the COVID-GRAM. This implies the importance of evaluating risk scores in different settings. Our findings are in line with a previous study conducted in Italy, which also described the differences between COVID-GRAM (AUC 0.785, CI 0.723–0.838), 4C-Score (AUC 0.799, CI 0.738–0.851), and NEWS (AUC 0.764, CI 0.700–0.819) as not statistically significant with regard to all-causes in-hospital death [7]. Another investigation figured out NEWS2 (an advanced version of the NEWS examined in this study, including hypercapnic respiratory failure instead of oxygen saturation) as prognosticating critical illness in COVID-19 better than COVID-GRAM (AUC of 0.87, CI 0.80–0.93 for NEWS2 and of 0.77, CI 0.68–0.85 for COVID-GRAM) [5]. CRB-65 and qSOFA as commonly used risk scores for pneumonia and sepsis were both clearly outperformed by NEWS—thus highlighting the COVID-specific importance of monitoring of oxygen saturation in addition to the respiratory rate possibly influenced by the silent hypoxemia that is characteristic for severe COVID-19 [24].

The in-hospital-mortality and ICU-admission rates significantly decreased during the study period comparing the three defined cohorts. Especially the ICU-admission rate in our cohort is comparable with national data—according to the data derived from a German federal hospital payment institute, the national ICU-admission rate of hospitalized patients was 30% in the first wave and 14% in the second wave [25]—in our cohort, 37% and 16%, respectively. The in-hospital mortality in our cohort was at 17% in the first and 5% and 0% in the second and third waves, respectively. In a study aggregating health insurance data of 158,490 German patients, the in-hospital mortality of the three waves was at 22.2%, 21.7% and 14.8% [26]. Of note, significant differences between hospitals were seen in this study.

One reason for the difference that can be discussed is a significantly lower mean age in our cohort than in the above-mentioned cohort (67 and 65 years versus 72 and 74 years in the first and second wave, respectively). Our hospital has a specific expertise in ARDS treatment, which might have modified the treatment outcome. For the third wave, patient numbers in our cohort were too low to provide statistically significant numbers.

A decrease in the in-hospital-mortality and ICU-admission rates has also been described in other countries, mostly interpreted as the composite effect of better treatment and the protective effect of immunization and vaccination within the population [27]. By the definition of the cohorts that we choose, this can also be assumed, investigating three different “waves” of patients with different treatment options and availability of vaccination. The effect of the vaccination rollout during cohort 3 is presumably reflected in the reduced average age of mostly unvaccinated hospitalized patients, since vaccination rollout was prioritized in the beginning for patients of higher age and with severe comorbidities. As the in-hospital-mortality and ICU-admission rates decreased, the discriminatory ability of the scores also decreased. Due to a small sample size, we did not calculate the AUC and PR-AUC of the third cohort, but it can be assumed that with further decreasing in-hospital-mortality and low-ICU-admission rates, the trend continues.

A further evaluation of risk assessment according to rapidly changing epidemiological circumstances is needed, especially considering new variants, treatment, and prevention options. According to our findings, a risk stratification of patients at admission could be recommended in Germany using the simple bedside score NEWS.

5. Limitations

Due to the retrospective nature of the study, we had to exclude patients with missing variables, thus possibly creating a bias toward more severely ill patients in which more parameters are documented. Furthermore, the sample size in the third cohort was respectably

smaller than those in the other cohorts. The study was conducted at a university hospital (tertiary care hospital and center for extracorporeal membrane oxygenation), creating a further possible bias. Since the analysis was carried out until June 2021, the succeeding variants of concern Delta and Omicron are not reflected in our study. On the other hand, the hospitalization, mortality, and ICU-admission rates were steadily declining especially with the Omicron variants, so it can be assumed that the discriminatory ability of the scores decreases further. The strength of our study is a broad comparison of several widely used risk scores in a German population. Furthermore, we provide a comparison of the score performance spanning over the course of the pandemic with changing epidemiological and therapeutic conditions.

6. Conclusions

Interestingly, in our hospitalized cohort, the COVID-specific risk-scores COVID-GRAM and 4C were not superior to NEWS especially in predicting ICU admission, even though they use additional risk factors acknowledged for severe COVID-19 disease. As a simple bedside-use scoring system, NEWS has proven to be a useful tool to identify hospitalized patients at risk for critical illness in our study population. This finding is important especially in clinical real-life circumstances, and the usage of the NEWS score for the risk stratification of patients at hospital admission should be discussed. Decreasing mortality rates and ICU-admission rates affected the discriminatory ability of all scores, so further investigation will be needed to address the highly dynamic epidemiological evolution.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jpm12111775/s1>, Figure S1: Schematic workflow of study; Figure S2: Plots of area under receiver operating characteristics of (a) qSOFA, (b) COVID-GRAM, (c) NEWS, (d) CRB-65, and (e) 4C-Score with regard to ICU admission and in-hospital mortality in the first and second cohorts; Table S1: Missing variables in calculation of scores; Table S2: Characteristics of excluded patients; Table S3: Discriminatory indices at several cutoffs of scores.

Author Contributions: Conceptualization, L.H. and S.G.; data curation, L.H., C.E., K.S., S.B. (Stefanie Bunk) and S.G.; formal analysis, P.W., S.B. (Stephanie Biergans), S.L., M.d.A.B.H., M.K., B.P.G. and S.G.; investigation, L.H.; methodology, L.H. and B.P.G.; project administration, C.E., K.S., and N.P.M.; supervision, S.G.; writing—original draft, L.H. and S.G.; writing—review and editing, L.H., P.W., S.B. (Stephanie Biergans), S.L., M.d.A.B.H., C.E., K.S., M.K., N.P.M., S.B. (Stefanie Bunk), M.B., B.P.G. and S.G. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge the support by Open Access Publishing Fund of University of Tübingen. MH, SL, SB received funding from the Bundesministerium für Bildung und Forschung (BMBF) (DIFUTURE, 01ZZ1804D).

Institutional Review Board Statement: Ethical approval was obtained at the local ethic commission (Nr. 431/2020BO) of the University Hospital Tübingen. The study is being reported in accordance to the STROBE-Statement (Elm et al., 2007).

Informed Consent Statement: Patient consent was waived according to German law and the Institutional Review Board due to the importance of this study subject and the retrospective study design. Clinical routine data were retrospectively collected and pseudonymized for analysis.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. Since informed patient consent was waived according to German law, he data are not publicly available.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

The COVID-GRAM risk score summarizes 10 variables and was based on coefficients β of a logistic model. The COVID-GRAM score is based on the following formula: $\exp(\sum \beta \times X) / [1 + \exp(\sum \beta \times X)]$ using the 10 variables X-ray abnormality, age, hemoptysis, dyspnea, unconsciousness, number of comorbidities, cancer history, neutrophil/lymphocytes, lactate dehydrogenase, and direct bilirubin.

References

1. Cram, P.; Anderson, M.L.; Shaughnessy, E.E. All Hands on Deck: Learning to “Un-specialize” in the COVID-19 Pandemic. *J. Hosp. Med.* **2020**, *15*, 314–315. [[CrossRef](#)]
2. Liang, W.; Liang, H.; Ou, L.; Chen, B.; Chen, A.; Li, C.; Li, Y.; Guan, W.; Sang, L.; Lu, J.; 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.* **2020**, *180*, 1081–1089. [[CrossRef](#)] [[PubMed](#)]
3. Knight, S.R.; Ho, A.; Pius, R.; Buchan, I.; Carson, G.; Drake, T.M.; Dunning, J.; Fairfield, C.J.; Gamble, C.; Green, C.A.; et al. Risk stratification of patients admitted to hospital with COVID19 using the ISARIC WHO Clinical Characterisation Protocol: Development and validation of the 4C Mortality Score. *BMJ* **2020**, *370*, m3339. [[CrossRef](#)]
4. Armiñanzas, C.; Arnaiz de Las Revillas, F.; Gutiérrez Cuadra, M.; Arnaiz, A.; Fernández Sampedro, M.; González-Rico, C.; Ferrer, D.; Mora, V.; Suberviola, B.; Latorre, M.; et al. Usefulness of the COVID-GRAM and CURB-65 scores for predicting severity in patients with COVID-19. *Int. J. Infect. Dis.* **2021**, *108*, 282–288. [[CrossRef](#)]
5. De Socio, G.V.; Gidari, A.; Sicari, F.; Palumbo, M.; Francisci, D. National Early Warning Score 2 (NEWS2) better predicts critical Coronavirus Disease 2019 (COVID-19) illness than COVID-GRAM, a multi-centre study. *Infection* **2021**, *49*, 1033–1038. [[CrossRef](#)] [[PubMed](#)]
6. Jones, A.; Pitre, T.; Juneke, M.; Kapralik, J.; Patel, R.; Feng, E.; Dawson, L.; Tsang, J.L.Y.; Duong, M.; Ho, T.; et al. External validation of the 4C mortality score among COVID-19 patients admitted to hospital in Ontario, Canada: A retrospective study. *Sci. Rep.* **2021**, *11*, 18638. [[CrossRef](#)] [[PubMed](#)]
7. Covino, M.; de Matteis, G.; Burzo, M.L.; Russo, A.; Forte, E.; Carnicelli, A.; Piccioni, A.; Simeoni, B.; Gasbarrini, A.; Franceschi, F.; et al. Predicting In-Hospital Mortality in COVID-19 Older Patients with Specifically Developed Scores. *J. Am. Geriatr. Soc.* **2021**, *69*, 37–43. [[CrossRef](#)] [[PubMed](#)]
8. Satici, C.; Demirkol, M.A.; Sargin Altunok, E.; Gursoy, B.; Alkan, M.; Kamat, S.; Demirok, B.; Surmeli, C.D.; Calik, M.; Cavus, Z.; et al. Performance of pneumonia severity index and CURB-65 in predicting 30-day mortality in patients with COVID-19. *Int. J. Infect. Dis.* **2020**, *98*, 84–89. [[CrossRef](#)] [[PubMed](#)]
9. Gude-Sampedro, F.; Fernández-Merino, C.; Ferreira, L.; Lado-Baleato, Ó.; Espasandín-Domínguez, J.; Hervada, X.; Cadarso, C.M.; Valdés, L. Development and validation of a prognostic model based on comorbidities to predict COVID-19 severity: A population-based study. *Int. J. Epidemiol.* **2021**, *50*, 64–74. [[CrossRef](#)]
10. Doğanay, F.; Ak, R. Performance of the CURB-65, ISARIC-4C and COVID-GRAM scores in terms of severity for COVID-19 patients. *Int. J. Clin. Pract.* **2021**, *75*, e14759. [[CrossRef](#)]
11. Horby, P.; Lim, W.S.; Emberson, J.R.; Mafham, M.; Bell, J.L.; Linsell, L.; Staplin, N.; Brightling, C.; Ustianowski, A.; Elmahi, E.; et al. Dexamethasone in Hospitalized Patients with COVID19. *N. Engl. J. Med.* **2021**, *384*, 693–704. [[CrossRef](#)] [[PubMed](#)]
12. Boehm, E.; Kronig, I.; Neher, R.A.; Eckerle, I.; Vetter, P.; Kaiser, L. Novel SARS-CoV-2 variants: The pandemics within the pandemic. *Clin. Microbiol. Infect.* **2021**, *27*, 1109–1117. [[CrossRef](#)] [[PubMed](#)]
13. Kluge, S.; Janssens, U.; Welte, T.; Weber-Carstens, S.; Schälte, G.; Spinner, C.D.; Malin, J.J.; Gastmeier, P.; Langer, F.; Bracht, H.; et al. S3-Leitlinie—Empfehlungen zur Stationären Therapie von Patienten Mit COVID-19. Available online: https://www.awmf.org/uploads/tx_szleitlinien/113-001LGI_S3_Empfehlungen-zur-stationaeren-Therapie-von-Patienten-mit-COVID-19_2022-03.pdf (accessed on 14 April 2022).
14. Schilling, J.; Buda, S.; Fischer, M.; Goerlitz, L.; Grote, U.; Haas, W.; Hamouda, O.; Prahm, K.; Tolksdorf, K. *Retrospektive Phaseneinteilung der COVID-19-Pandemie in Deutschland bis Februar 2021*; Robert Koch-Institut: Berlin, Germany, 2021. [[CrossRef](#)]
15. Hippisley-Cox, J.; Coupland, C.A.; Mehta, N.; Keogh, R.H.; Diaz-Ordaz, K.; Khunti, K.; Lyons, R.A.; Kee, F.; Sheikh, A.; Rahman, S.; et al. Risk prediction of COVID19 related death and hospital admission in adults after COVID19 vaccination: National prospective cohort study. *BMJ* **2021**, *374*, n2244. [[CrossRef](#)] [[PubMed](#)]
16. Schmid-Küpke, N.K.; Neufeind, J.; Wichmann, O.; Siedler, A. COVIMO: COVID-19 Vaccination Rate Monitoring in Germany. Available online: <https://www.rki.de/EN/Content/infections/epidemiology/outbreaks/COVID-19/projects/covimo.html> (accessed on 28 April 2022).
17. Carpenter, J.; Bithell, J. Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Stat. Med.* **2000**, *19*, 1141–1164. [[CrossRef](#)]
18. DeLong, E.R.; DeLong, D.M.; Clarke-Pearson, D.L. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* **1988**, *44*, 837. [[CrossRef](#)]
19. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Müller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [[CrossRef](#)] [[PubMed](#)]

20. Hadley, W. *Ggplot2: Elegant Graphics for Data Analysis*, 2nd ed.; Springer: Basel, Switzerland, 2016; ISBN 978-3-319-24277-4.
21. Davison, A.C.; Hinkley, D.V. *Bootstrap Methods and Their Applications, Repr*; Cambridge University Press: Cambridge, UK, 1999; ISBN 0-521-57391-2.
22. Saito, T.; Rehmsmeier, M. Precrec: Fast and accurate precision-recall and ROC curve calculations in R. *Bioinformatics* **2017**, *33*, 145–147. [[CrossRef](#)] [[PubMed](#)]
23. Candel, B.G.; Duijzer, R.; Gaakeer, M.I.; ter Avest, E.; Sir, Ö.; Lameijer, H.; Hessels, R.; Reijnen, R.; van Zwet, E.W.; de Jonge, E.; et al. The association between vital signs and clinical outcomes in emergency department patients of different age categories. *Emerg. Med. J.* **2022**. [[CrossRef](#)] [[PubMed](#)]
24. Tobin, M.J.; Laghi, F.; Jubran, A. Why COVID-19 Silent Hypoxemia Is Baffling to Physicians. *Am. J. Respir. Crit. Care Med.* **2020**, *202*, 356–360. [[CrossRef](#)]
25. Karagiannidis, C.; Windisch, W.; McAuley, D.F.M.; Welte, T.; Busse, R. Major differences in ICU admissions during the first and second COVID-19 wave in Germany. *Lancet Respir. Med.* **2021**, *9*, e47–e48. [[CrossRef](#)]
26. Karagiannidis, C.; Mostert, C.; Hentschker, C.; Voshaar, T.; Malzahn, J.; Schillinger, G.; Klauber, J.; Janssens, U.; Marx, G.; Weber-Carstens, S.; et al. Case characteristics, resource use, and outcomes of 10 021 patients with COVID-19 admitted to 920 German hospitals: An observational study. *Lancet Respir. Med.* **2020**, *8*, 853–862. [[CrossRef](#)]
27. Long, B.; Carius, B.M.; Chavez, S.; Liang, S.Y.; Brady, W.J.; Koyfman, A.; Gottlieb, M. Clinical update on COVID-19 for the emergency clinician: Presentation and evaluation. *Am. J. Emerg. Med.* **2022**, *54*, 46–57. [[CrossRef](#)] [[PubMed](#)]