



Call, chosen, HA₂T₂, ANDC: validation of four severity scores in COVID-19 patients

Selina Wolfsberg¹ · Claudia Gregoriano¹ · Tristan Struja¹ · Alexander Kutz¹ · Daniel Koch¹ · Luca Bernasconi² · Angelika Hammerer-Lercher² · Christine Mohr³ · Sebastian Haubit^{1,3} · Anna Conen^{1,3} · Christoph A. Fux^{1,3} · Beat Mueller^{1,4} · Philipp Schuetz^{1,4}

Received: 27 August 2021 / Accepted: 4 November 2021 / Published online: 19 November 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany 2021

Abstract

Purpose To externally validate four previously developed severity scores (i.e., CALL, CHOSEN, HA₂T₂ and ANDC) in patients with COVID-19 hospitalised in a tertiary care centre in Switzerland.

Methods This observational analysis included adult patients with a real-time reverse-transcription polymerase chain reaction or rapid-antigen test confirmed severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) infection hospitalised consecutively at the Cantonal Hospital Aarau from February to December 2020. The primary endpoint was all-cause in-hospital mortality. The secondary endpoint was disease progression, defined as needing invasive ventilation, ICU admission or death.

Results From 399 patients (mean age 66.6 years ± 13.4 SD, 68% males), we had complete data for calculating the CALL, CHOSEN, HA₂T₂ and ANDC scores in 297, 380, 151 and 124 cases, respectively. Odds ratios for all four scores showed significant associations with mortality. The discriminative power of the HA₂T₂ score was higher compared to CALL, CHOSEN and ANDC scores [area under the curve (AUC) 0.78 vs. 0.65, 0.69 and 0.66, respectively]. Negative predictive values (NPV) for mortality were high, particularly for the CALL score (≥ 6 points: 100%, ≥ 9 points: 95%). For disease progression, discriminative power was lower, with the CHOSEN score showing the best performance (AUC 0.66).

Conclusion In this external validation study, the four analysed scores had a lower performance compared to the original cohorts regarding prediction of mortality and disease progression. However, all scores were significantly associated with mortality and the NPV of the CALL and CHOSEN scores in particular allowed reliable identification of patients at low risk, making them suitable for outpatient management.

Keywords COVID-19 · Validation study · Risk scores · Switzerland

Background

The coronavirus disease 2019 (COVID-19) pandemic, with its overwhelming resource use, has been a major challenge for clinicians and health care institutions worldwide. Identifying patients at high risk of disease progression may help allocating resources more efficiently. Since presentation and course of the infection can vary considerably (including asymptomatic cases), no single trait is sufficient to appropriately categorise patients [1–9]. Thus, several scores have attempted to improve identification of patients at high risk of progression or death of COVID-19. Among these scores, the CALL, CHOSEN, HA₂T₂ and the ANDC score have generated much interest [10–13].

✉ Philipp Schuetz
schuetzph@gmail.com

¹ Medical University Department of Medicine, Kantonsspital Aarau, Tellstrasse, 5001 Aarau, Switzerland

² Institute of Laboratory Medicine, Kantonsspital Aarau, Aarau, Switzerland

³ Department of Infectious Diseases and Hospital Hygiene, Kantonsspital Aarau, Aarau, Switzerland

⁴ Medical Faculty, University of Basel, Basel, Switzerland

The CALL score (Comorbidity, Age, Lactate dehydrogenase (LDH) and Lymphocyte count) showed great discriminatory potential for disease progression with an area under the curve (AUC) of 0.91 (95%-CI 0.86–0.94) in its derivation cohort [10]. Disease progression was defined as respiratory rate ≥ 30 breaths per minute (bpm), peripheral oxygen saturation (SpO_2) $\leq 93\%$, arterial partial oxygen pressure (PaO_2)/fraction of inspired oxygen (FiO_2) ≤ 300 mmHg, mechanical ventilation or worsening of lung computer tomography (CT) findings [10]. The CHOSEN score used age, FiO_2 and albumin to predict progression defined as requiring supplemental oxygen, admission to the intensive care unit (ICU) or death [11]. The authors reported a good discriminative capacity for their score with an AUC of 0.89 (95%-CI 0.87–0.91) in their derivation and 0.87 (95%-CI 0.81–0.93) in their validation cohort [11]. The HA_2T_2 score was used to predict all-cause in-hospital mortality in COVID-19 patients based on need for supplemental oxygen, age and troponin [12]. It showed good discriminative power in both their derivation (AUC 0.83, 95%-CI 0.79–0.88) and their validation cohort (AUC 0.78, 95%-CI 0.72–0.84) [12]. The ANDC score, based on age, neutrophil-to-lymphocyte ratio (NLR), D-dimer and C-reactive protein (CRP), predicted all-cause in-hospital mortality with an excellent AUC of 0.92 (95%-CI 0.84–0.97) in their derivation and 0.98 (95%-CI 0.95–1.00) in their validation cohort [13].

So far, only the CALL score has undergone external validation, with the score performing markedly worse than in the original cohort (AUC 0.62 vs. 0.91) [14]. Thus, before wide-spread implementation, independent external validation of all these scores is mandatory. Herein, we validated four severity scores (i.e., the CALL, CHOSEN, HA_2T_2 and ANDC scores) in patients with COVID-19 hospitalised in a tertiary care centre in Switzerland.

Methods

Study design and participants

This retrospective observational analysis included all consecutive adult patients (≥ 18 years) with a confirmed Severe Acute Respiratory Syndrome Corona Virus type 2 (SARS-CoV-2) infection that required hospitalisation for at least 24 h at the Medical University Clinic of the Cantonal Hospital Aarau (Switzerland) between February 26, 2020 and April 30, 2020 (first wave) and between October 1, 2020 and December 31, 2020 (second wave). In this tertiary care centre with 130 medical ward beds, indications for in-hospital treatment of COVID-19 were respiratory distress with need for oxygen supplementation, high fever or relevant clinical deterioration. This study was approved by the local ethics committee (EKZN, 2020-01306).

Detailed description of the study methodology has been reported previously [6, 15]. A confirmed SARS-CoV-2 infection was defined as a combination of typical clinical symptoms (e.g., respiratory symptoms with or without fever, and/or pulmonary infiltrates and/or anosmia/dysgeusia) and a positive real-time reverse-transcription polymerase chain reaction (RT-PCR) test, obtained from nasopharyngeal swabs or lower respiratory tract samples, according to guidance by the World Health Organization (WHO) [16, 17]. Data for the second wave also included patients with positive rapid-antigen tests. However, due to their lower positive predictive value, we excluded asymptomatic patients unless their rapid-antigen results were confirmed by a positive RT-PCR test. We further excluded patients from the analysis if they did not provide general informed consent or if they had not yet been discharged when data collection was closed (January 20, 2021). This study adheres to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement for reporting of prediction models.

Data collection

All analysed data were collected as part of the clinical routine during the hospitalisation (from admission to discharge or death). We performed chart reviews and automatic export from electronic health records (EHR), including vital signs and clinical characteristics upon admission as well as sociodemographic factors, comorbidities based on pre-existing diagnoses and home medication. COVID-19-specific inpatient medication was assessed until hospital discharge or death and exported from the EHR. Experimental treatment was offered to all suitable patients according to ongoing clinical trials and WHO guidelines [16–18]. During the second wave, this also included the application of high-dose glucocorticoids [19]. The age-adjusted Charlson comorbidity index (ACCI) [20] and the Clinical Frailty Scale score (CFS) [21] were calculated for all patients as part of the clinical routine or through chart review. Laboratory values were available according to clinical routine and derived from the first blood draw obtained within 7 days from admission.

Definition of endpoints

All-cause in-hospital mortality was defined as the primary endpoint. The secondary endpoint, disease progression, had different definitions in the original studies. For easier comparability between the scores, we defined disease progression as needing invasive ventilation, ICU admission or death in our own analysis. Originally, the CALL score defined progression as respiratory rate ≥ 30 bpm, $\text{SpO}_2 \leq 93\%$, $\text{PaO}_2/\text{FiO}_2 \leq 300$ mmHg, requiring mechanical ventilation or worsening of lung CT findings. CT findings were not available

for our analysis and thus not considered. The definition of progression for the CHOSEN score was requirement of supplemental oxygen, admission to the ICU or death. Validation results were based on these original definitions.

Statistical analysis

Discrete variables are expressed as frequency (percentage) and continuous variables as medians with interquartile ranges (IQR, for skewed data) or mean with standard deviation (SD, for normally distributed data). We used the Wilcoxon rank-sum test to compare continuous variables and the Pearson's chi-squared test to compare categorical or binary variables. Odds ratios (OR) were calculated with corresponding 95% confidence intervals (CI) as measures of association. We assessed calibration for mortality numerically by tabulating the observed risks against those reported in the original studies. These were not available for the CALL and CHOSEN scores. We considered a two-sided *p*-value of <0.05 significant and calculated the unadjusted area under the receiver operating characteristic curve (AUC) as a measure of discrimination. Statistical analysis was performed as a complete-case-analysis based on the original regression coefficients using Stata 15.1 (StataCorp, College Station, TX, USA).

Results

Figure 1 provides an overview of the study flow and Table 1 shows overall patient demographics, comorbidities, laboratory values and vital signs on admission as well as stratified according to the individual score cohorts. In total, 399 patients hospitalised with a confirmed SARS-CoV-2 infection were included in this analysis (mean age 66.6 years \pm 13.4 SD, 68% male). Complete data sets to allow for the calculation of the CALL and CHOSEN score were available in 297 and 380 patients, respectively. Fewer

patients had all values necessary to calculate the HA₂T₂ (*n* = 151) and ANDC score (*n* = 124). There were several noticeable differences between the score cohorts, for example, transfer rates from other hospitals (range from 14.5% for ANDC to 28.5% for HA₂T₂), supplemental oxygen (29.8% for CALL to 45.7% for HA₂T₂), obesity (30.8% for CHOSEN to 41.7% for ANDC) and ICU admission (19.5% for CHOSEN to 46.4% for HA₂T₂). However, overall comorbidity and frailty were similar.

Table 2 shows the discriminative power of each score for mortality and disease progression (defined as requiring invasive ventilation, ICU admission or death for all scores for easier comparability). For mortality, the HA₂T₂ performed best (AUC 0.78, 95%-CI 0.70–0.85). For progression, overall discriminative capacity was lower, with the CHOSEN score performing slightly better than the others (AUC 0.66, 95%-CI 0.72–0.60). All scores were associated with mortality.

Sensitivity and specificity as well as positive and negative predictive value for each proposed cut-off are summarised in Table 3 and visualised in Fig. 2. The negative predictive value of the CALL score was highest (≥ 6 points: 100%, 95%-CI 75.3–100), while the highest positive predictive value was found for the HA₂T₂ score (≥ 3 points: 58.6%, 95%-CI 38.9–76.5).

The direct comparison with the original outcomes can be found in Table 4. Only the HA₂T₂ score performed similarly with an AUC of 0.78 (95%-CI 0.72–0.84) in the original validation cohort and an AUC of 0.78 (95%-CI 0.70–0.85) in our sample. The discriminative power for all other scores was markedly worse in comparison with their respective original cohorts. These results persisted when performed in the cohort with full data sets for all scores (*n* = 67, data not shown).

The calibration assessment for mortality for the HA₂T₂ and ANDC scores can be found in the additional files 1 and 2 (Tables S1 and S2). Overall, calibration was poor, with the ANDC score performing slightly better (overprediction

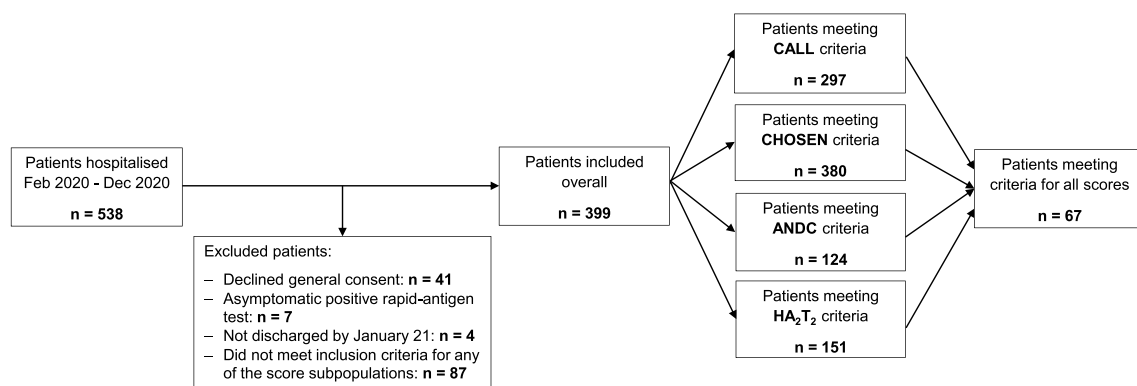


Fig. 1 Overview of study flow. In total, 399 patients were included in the final analysis, 67 of whom had complete data sets available

Table 1 Baseline characteristics and treatment of patients hospitalised with confirmed SARS-CoV-2 infection

Factor	Overall	CALL	CHOSEN	HA ₂ T ₂	ANDC
<i>N</i>	399	297	380	151	124
<i>Pre-admission history</i>					
Age (years), mean (SD)	66.6 (13.4)	66.2 (13.0)	66.4 (13.4)	65.9 (12.4)	65.1 (12.6)
≥ 65 years	232 (58.1%)	167 (56.2%)	219 (57.6%)	84 (55.6%)	72 (58.1%)
Sex, male	271 (67.9%)	206 (69.4%)	260 (68.4%)	111 (73.5%)	90 (72.6%)
Transfer from other hospital	75 (18.8%)	46 (15.5%)	67 (17.6%)	43 (28.5%)	18 (14.5%)
Time from symptom onset to admission [days], median (IQR)	7 (4, 9)	7 (4, 9)	7 (4, 9)	7 (4, 9)	7 (4, 9)
<i>Presentation to emergency department</i>					
Supplemental oxygen administered	103 (30.4%)	81 (29.8%)	96 (30.0%)	69 (45.7%)	43 (37.7%)
FiO ₂ (%), mean (SD)	65.6 (28.4)	68.2 (28.5)	64.6 (28.5)	68.4 (28.4)	72.9 (27.9)
SpO ₂ (%), median (IQR)	93.7 (89.4, 96.0)	93.1 (88.5, 95.7)	93.7 (89.4, 96.0)	92.7 (88.2, 95.0)	91.9 (86.9, 94.9)
Heart rate (bpm), mean (SD)	90 (18)	91 (19)	90 (18)	90 (21)	92 (18)
Respiratory rate (bpm), mean (SD)	21 (8)	21 (8)	21 (8)	21 (10)	21 (10)
Temperature (°C), mean (SD)	37.7 (1.0)	37.7 (1.0)	37.7 (1.0)	37.6 (0.9)	37.7 (0.9)
<i>Laboratory values</i>					
Lymphocyte count (10 ³ /mm ³), median (IQR)	0.9 (0.6, 1.2)	0.9 (0.6, 1.2)	0.9 (0.6, 1.2)	0.8 (0.6, 1.2)	0.9 (0.7, 1.2)
Neutrophil-lymphocyte ratio, median (IQR)	5.8 (3.7, 10.4)	6.0 (3.6, 10.9)	5.8 (3.6, 10.3)	7.0 (4.5, 11.7)	6.4 (4.1, 11.1)
C-reactive protein (mg/L), median (IQR)	81.5 (33.8, 140.0)	89.8 (42.0, 145.0)	81.5 (36.7, 140.0)	89.5 (48.3, 152.0)	101.0 (61.9, 158.5)
Lactate dehydrogenase (IU/L), median (IQR)	322 (245, 449)	325 (250, 449)	325 (245, 449)	346 (268, 520)	333 (269, 452)
Albumin (g/L), median (IQR)	29.8 (26.8, 33.3)	29.8 (27.1, 33.2)	29.8 (26.8, 33.3)	29.3 (26.4, 32.7)	29.8 (27.1, 33.1)
D-dimer (mg/L), median (IQR)	1.0 (0.6, 1.6)	0.9 (0.6, 1.6)	1.0 (0.5, 1.6)	1.1 (0.6, 1.9)	1.0 (0.6, 1.6)
Troponin (ng/L), median (IQR)	18 (10, 55)	17 (9, 40)	17 (10, 48)	18 (9, 55)	16 (9, 31)
<i>Comorbidities</i>					
ACCI, median (IQR)	3 (2, 5)	3 (2, 5)	3 (2, 5)	3 (2, 5)	3 (2, 4)
≥ 4 points	194 (48.6%)	137 (46.1%)	183 (48.2%)	67 (44.4%)	58 (46.8%)
CFS, median (IQR)	3 (2, 5)	3 (2, 4)	3 (2, 5)	3 (2, 4)	3 (2, 4)
≥ 4 points	142 (35.6%)	94 (31.6%)	136 (35.8%)	44 (29.1%)	29 (23.4%)
Smoker	34 (12.1%)	24 (11.0%)	33 (12.1%)	10 (9.9%)	10 (10%)
Obesity (BMI > 30 kg/m ²)	119 (30.9%)	97 (33.7%)	113 (30.6%)	56 (38.6%)	50 (41.7%)
Diabetes mellitus	119 (29.8%)	88 (29.6%)	113 (29.7%)	54 (35.8%)	44 (35.5%)
Hypertension	237 (59.4%)	171 (57.6%)	225 (59.2%)	97 (64.2%)	80 (64.5%)
Coronary artery disease	82 (20.6%)	60 (20.2%)	75 (19.7%)	47 (31.1%)	28 (22.6%)
Chronic heart failure (LVEF < 40%)	11 (2.8%)	7 (2.4%)	11 (2.9%)	6 (4.0%)	3 (2.4%)
Bronchial asthma	26 (6.5%)	20 (6.7%)	26 (6.8%)	10 (6.6%)	5 (4.0%)
COPD	30 (7.5%)	19 (6.4%)	28 (7.4%)	11 (7.3%)	5 (4.0%)
OSAS	39 (9.8%)	31 (10.4%)	36 (9.5%)	21 (13.9%)	16 (12.9%)
Solid organ transplant	9 (2.3%)	6 (2.0%)	9 (2.4%)	2 (1.3%)	3 (2.4%)
Active rheumatic disease	12 (3.0%)	10 (3.4%)	10 (2.6%)	8 (5.3%)	2 (1.6%)
Cancer	46 (11.5%)	33 (11.1%)	46 (12.1%)	9 (6.0%)	12 (9.7%)
Chronic kidney disease	86 (21.6%)	59 (19.9%)	79 (20.8%)	36 (23.8%)	26 (21.0%)
<i>SARS-CoV-2 infection treatment</i>					
Experimental (antiviral) treatment	71 (17.8%)	53 (17.8%)	66 (17.4%)	34 (22.5%)	23 (18.5%)
Antibiotic treatment	94 (23.6%)	71 (23.9%)	88 (23.2%)	47 (31.1%)	34 (27.4%)
High-dose glucocorticoids	258 (64.7%)	206 (69.4%)	245 (64.5%)	106 (70.2%)	106 (85.5%)
<i>Outcomes</i>					
All-cause in-hospital mortality	80 (20.1%)	62 (20.9%)	77 (20.3%)	43 (28.5%)	33 (26.6%)
Time to death (days), median (IQR)	9.0 (4.0, 17.0)	9.0 (4.0, 17.0)	9.0 (4.0, 17.0)	11.0 (5.0, 19.0)	10.0 (5.0, 19.0)

Table 1 (continued)

Factor	Overall	CALL	CHOSEN	HA ₂ T ₂	ANDC
ICU admission	80 (20.1%)	67 (22.6%)	74 (19.5%)	70 (46.4%)	51 (41.1%)
Time to ICU (days), median (IQR)	1.0 (0.0, 3.0)	0.5 (0.0, 3.5)	1.0 (0.0, 3.0)	1.0 (0.0, 3.0)	0.0 (0.0, 3.0)
Invasive ventilation	57 (14.3%)	44 (14.8%)	51 (13.4%)	50 (33.1%)	31 (25.0%)
Disease progression ^a	129 (32.3%)	101 (34.0%)	120 (31.6%)	84 (55.6%)	61 (49.2%)
LOS (days), median (IQR)	7 (4.0, 12.0)	7.0 (4.0, 12.0)	6.5 (4.0, 12.0)	10.0 (5.0, 18.0)	8.5 (5.0, 13.5)
Discharge status ^b					
Home care	147 (36.8%)	88 (29.6%)	108 (28.4%)	48 (31.8%)	41 (33.1%)
Rehabilitation care	115 (28.8%)	109 (36.7%)	141 (37.1%)	38 (25.2%)	40 (32.3%)
Other hospital	43 (10.8%)	8 (2.7%)	12 (3.2%)	5 (3.3%)	2 (1.6%)
Nursing facility	13 (3.3%)	29 (9.8%)	41 (10.8%)	17 (11.3%)	8 (6.5%)
Unknown	1 (0.3%)	1 (0.3%)	1 (0.3%)	n.a	n.a

ACCI age-adjusted Charlson comorbidity index, BMI body mass index, bpm beats/ breaths per minute, CFS clinical frailty scale, COPD chronic obstructive pulmonary disease, CRP C-reactive protein, FiO₂ fraction of inspired oxygen, ICU intensive care unit, IQR interquartile range, LVEF left ventricular ejection fraction, n.a. not applicable, OSAS obstructive sleep apnoea syndrome, SARS-CoV-2 severe acute respiratory syndrome coronavirus type 2, SD standard deviation, SpO₂ peripheral oxygen saturation

^aDefined as invasive ventilation, ICU admission or death

^bOther than death

Table 2 Score values stratified by survivorship with corresponding OR and AUC

Score (range)	Survivors	Non-Survivors	<i>p</i> -value	OR ^a [95% CI], <i>p</i> -value Mortality	AUC [95% CI] Mortality	Progression ^b
CALL (4–13 points)	<i>n</i> (%) 235 (79%)	62 (21%)		1.30 [1.12– 1.50], <0.01	0.65 [0.58–0.71]	0.59 [0.52–0.65]
	Median (IQR) 10 (8, 12)	11.5 (10, 13)	<0.01			
CHOSEN (0–55 points ^c)	<i>n</i> (%) 303 (80%)	77 (20%)		0.92 [0.89– 0.96], <0.01	0.69 [0.76–0.62]	0.66 [0.72–0.60]
	Median (IQR) 39 (30, 43)	30 (29, 39)	<0.01			
HA ₂ T ₂ (0–5 points)	<i>n</i> (%) 108 (72%)	43 (28%)		2.38 [1.68– 3.38], <0.01	0.78 [0.70–0.85]	0.59 [0.50–0.68]
	Median (IQR) 1 (0, 2)	2 (1, 3)	<0.01			
ANDC (0–ca. 300 points)	<i>n</i> (%) 91 (73%)	33 (27%)		1.01 [1.00–1.02], 0.01	0.66 [0.56–0.77]	0.63 [0.54–0.73]
	Median (IQR) 85.4 (74.6, 99.6)	95.2 (85.7, 111.6)	<0.01			

AUC area under the curve, CI confidence interval, IQR interquartile range, OR odds ratio

^aOR per point increase

^bDefined as invasive ventilation, ICU admission or death

^cMore points = progression less likely

up to 18 percentage points) than the HA₂T₂ score (under-prediction up to 30 percentage points). Calibration for the CALL and CHOSEN scores were not possible due to lacking published data.

Discussion

In this validation study, four currently available scores to predict mortality and disease progression in COVID-19 patients performed markedly worse in patients hospitalised at a Swiss tertiary care centre than in their original cohorts. The HA₂T₂ score showed the best discrimination

for mortality (AUC 0.78, 95%-CI 0.70–0.85) and the only results similar to the derivation cohort.

Some loss of predictive ability can be explained by the differences between our study population and the original derivation cohorts. This is most apparent when comparing age, which has been recognised as an important risk factor for worse outcomes [22] and is included in all four scores. Mean age ranged from 44 to 65 years for the CALL, CHOSEN, HA₂T₂ and ANDC scores in the original publications whereas the mean age in our population was 67 years. However, even when comparing the scores among the 67 patients who had all parameters required for all scores, the HA₂T₂ score showed the best discriminative power (data not

Table 3 Sensitivity, specificity, positive and negative predictive values for mortality and disease progression for all scores and their original cut-offs

Score	Cut-off	n (%)	Sensitivity [95%-CI]	Specificity [95%-CI]	PPV [95%-CI]	NPV [95%-CI]
<i>Mortality</i>						
CALL	≥ 6 points	284 (96%)	100% [94.2–100]	5.5% [3.0–9.3]	21.8% [17.2–27.1]	100% [75.3–100]
	≥ 9 points	219 (74%)	93.5% [84.3–98.2]	31.5% [25.6–37.8]	26.5% [20.8–32.9]	94.9% [87.4–98.6]
CHOSEN	≤ 30 points	135 (36%)	62.3% [50.6–73.1]	71.3% [65.8–76.3]	35.6% [27.5–44.2]	88.2% [83.4–91.9]
HA ₂ T ₂	≥ 3 points	29 (19%)	39.5% [25.0–55.6]	88.9% [81.4–94.1]	58.6% [38.9–76.5]	78.7% [70.4–85.6]
ANDC	< 59 points	11 (9%)	0.0% [0.0–10.6]	87.9% [79.4–93.8]	0.0% [0.0–28.5]	70.8% [61.5–79.0]
	59–101 points	79 (64%)	57.6% [39.2–74.5]	34.1% [24.5–44.7]	24.1% [15.1–35.0]	68.9% [53.4–81.8]
	> 101 points	34 (27%)	42.4% [25.5–60.8]	78.0% [68.1–86.0]	41.2% [24.6–59.3]	78.9% [69.0–86.8]
<i>Disease progression^a</i>						
CALL	≥ 6 points	284 (96%)	98.0% [93.0–99.8]	5.6% [2.8–9.8]	34.9% [29.3–40.7]	84.6% [54.6–98.1]
	≥ 9 points	219 (74%)	84.2% [75.6–90.7]	31.6% [25.2–38.6]	38.8% [32.3–45.6]	79.5% [68.8–87.8]
CHOSEN	≤ 30 points	135 (36%)	53.3% [44.0–62.5]	72.7% [66.8–78]	47.4% [38.8–56.2]	77.1% [71.4–82.2]
HA ₂ T ₂	≥ 3 points	29 (19%)	23.8% [15.2–34.3]	86.6% [76.0–93.7]	69.0% [49.2–84.7]	47.5% [38.4–56.8]
ANDC	< 59 points	11 (9%)	4.9% [1.0–13.7]	87.3% [76.5–94.4]	27.3% [6.0–61.0]	48.7% [39.2–58.3]
	59–101 points	79 (64%)	60.7% [47.3–72.9]	33.3% [22–46.3]	46.8% [35.5–58.4]	46.7% [31.7–62.1]
	> 101 points	34 (27%)	34.4% [22.7–47.7]	79.4% [67.3–88.5]	61.8% [43.6–77.8]	55.6% [44.7–66.0]

CI confidence interval, NPV negative predictive value, PPV positive predictive value

^aDefined as invasive ventilation, ICU admission or death

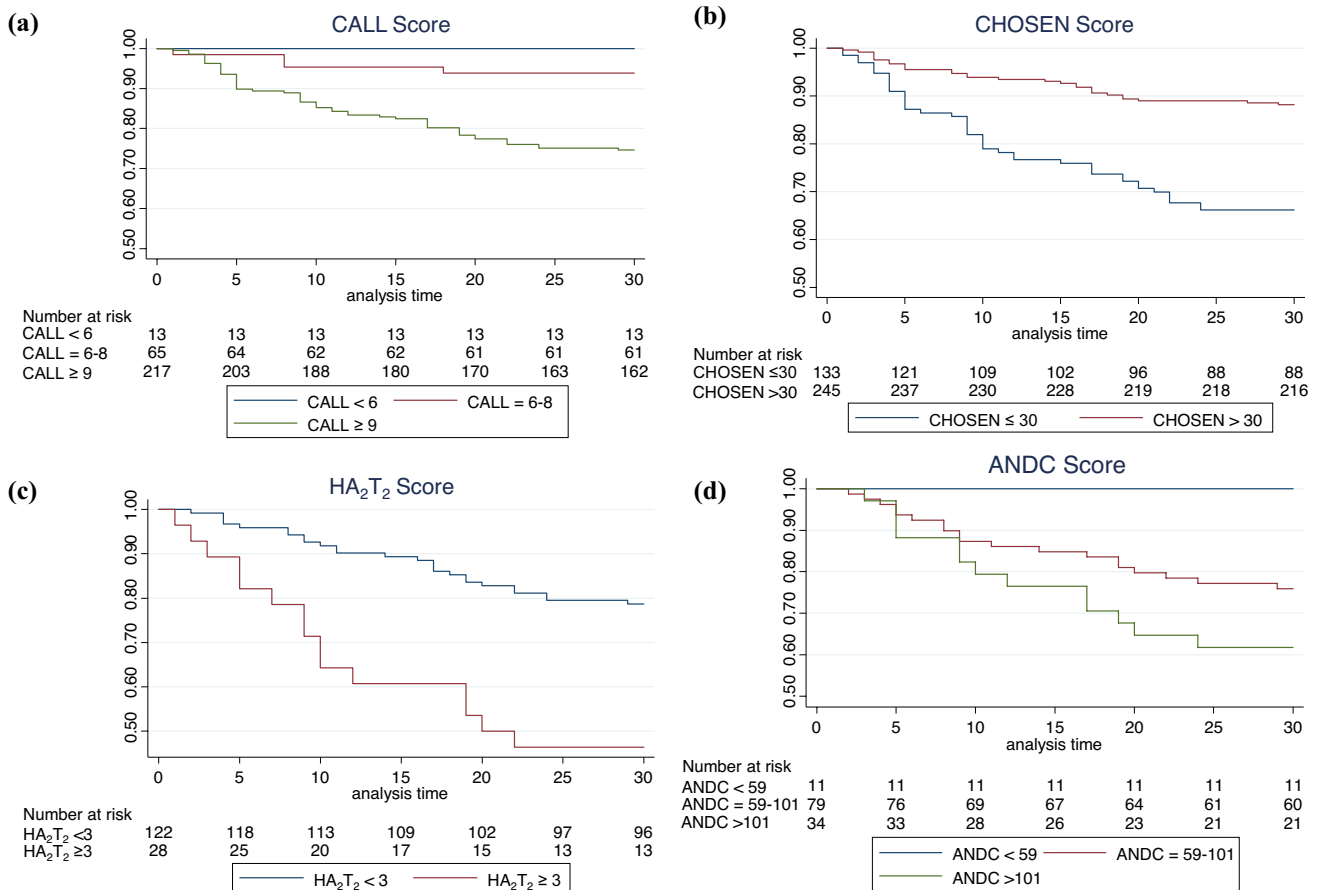


Fig. 2 Survival time analysis for **a** CALL score, **b** CHOSEN score, **c** HA₂T₂ score, **d** ANDC scores and their respective cut-off subgroups

Table 4 Comparison of current analysis with original study results and outcomes

Score	Reference, country	Included predictors	Original outcome(s)	AUC (95% confidence interval) ^b	
				Original publication	Current analysis
CALL	Ji et al., China [10] Grifoni et al., Italy [14]	Comorbidity, age, LDH, lymphocyte count	Respiratory rate ≥ 30 bpm, SpO ₂ $\leq 93\%$, PaO ₂ / FiO ₂ ≤ 300 mmHg, mechanical ventilation, worsening of lung CT findings ^a	0.91 (0.86–0.94)	0.61 (0.55–0.68)
				External validation 0.62 (0.53–0.69)	
CHOSEN	Levine et al., United States [11]	Age, FiO ₂ , albumin	Hypoxia, ICU admission, death (within 14 days)	0.89 (0.87–0.91)	0.65 (0.59–0.71)
HA ₂ T ₂	Manocha et al., United States [12]	Supplemental oxygen, age, troponin	All-cause in-hospital mortal- ity	0.83 (0.79–0.88)	0.78 (0.70–0.85)
				Validation cohort 0.78 (0.72–0.84)	
ANDC	Weng et al., China [13]	Age, NLR, D-dimer, CRP	All-cause in-hospital mortal- ity	0.92 (0.84–0.97)	0.66 (0.56–0.77)
				Validation cohort 0.98 (0.95–1.00)	

AUC area under the curve, *bpm* breaths per minute, *CRP* C-reactive protein, *CT* computer tomography, *FiO₂* fraction of inspired oxygen, *ICU* intensive care unit, *LDH* lactate dehydrogenase, *NLR* neutrophil-to-lymphocyte ratio, *PaO₂* arterial partial pressure of oxygen, *SpO₂* peripheral oxygen saturation

^aCT findings not included in our results, data not available

^bAll results calculated for original outcomes

shown). Apart from the small sample size, further limitations in this comparison arise from the fact that the study populations were also different in their origins. The CALL and ANDC scores were based on Chinese patients while the CHOSEN and the HA₂T₂ score were derived in US American patients. Interestingly, the other currently available external validations of the CALL score in Italian and Turkish patients resulted in AUCs that were very similar to our own (original AUC for disease progression 0.91 vs. Italian AUC 0.62, Turkish AUC 0.59, our AUC 0.61) [14, 23]. Hence, it seems that compatibility and comparability of these scores for different populations cannot be assumed.

Further difficulties are rooted in the novelty of COVID-19. Much is still unknown about the disease including which factors best predict progression or mortality. This is reflected in the very different factors included in the scores. Still, these more recent approaches are already an improvement to initial scores which included up to 12 different items, making them difficult to use in a clinical setting [24]. However, in a busy environment such as the emergency department, ease of use is crucial. The scores discussed here all use no more than four variables that are relatively readily available in middle- to high-income countries. There also exists a simplified version of the CHOSEN score that does not rely on laboratory values but did also not perform as well in the original cohort [11].

All scores were significantly associated with mortality and their respective discriminative capacities were moderate

to good but calibration was poor due to considerable population differences. Furthermore, the negative predictive value of the CALL score was particularly high and could thus help identify patients who are not at risk. The CHOSEN score, whose explicit aim was to differentiate between patients who needed hospitalisation and those who could be sent home safely, also had a high negative predictive value and, in addition, showed a relatively balanced relation between sensitivity and specificity, making it a potentially valuable tool for risk stratification. Since we did not include outpatients in our study, our results are likely to underestimate the true value of the CHOSEN score.

Limitations

There are certain limitations to our study. First, our findings are limited to hospitalised patients in a single centre in Switzerland, limiting generalisability. In addition, baseline parameters of our population were markedly different from the original study populations including ethnicity and important predictors such as age. Unfortunately, regression coefficients could not be updated based on the available data. Similarly, we could not calculate calibration for the CALL and CHOSEN score. Internal validity is also limited due to the retrospective design, which meant that a considerable proportion of patients had to be excluded from certain score cohorts because the required data were missing. Additional validation analyses should be conducted in larger data sets.

Furthermore, troponin and D-dimer values (required for the HA₂T₂ and ANDC scores, respectively) were usually available for sicker patients who reached the primary and secondary endpoints more often, which not only limited study population sizes but also comparability between scores. Finally, we had to exclude four patients due to missing outcome data, thus increasing the risk for selection bias.

Conclusions

In our independent validation, the four analysed scores performed worse than in their original cohorts regarding prediction of mortality and disease progression. However, all scores were significantly associated with mortality. While the HA₂T₂ score identified high risk patients, the negative predictive values of the CALL and CHOSEN scores allowed reliable identification of patients at low risk, which may make them suitable for outpatient management.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s15010-021-01728-0>.

Acknowledgements We thank all participating patients, their families and all healthcare workers at the Cantonal Hospital Aarau for their help and dedication to reduce the burden of the ongoing pandemic.

Author contributions PS and SW conceived of the study and its design. SW performed the statistical analysis and wrote the first draft of the paper. SW, CG, DK, LB, AH, CH, CM and SH collected and compiled the data. CG, PS, AK, CF, TS and BM critically revised the manuscript. All authors read and approved the final manuscript.

Funding This study was funded by the Research Council KSA (Kantonsspital Aarau). The funding agency had no bearing on the study design, data collection and analysis or writing of the manuscript.

Data availability The datasets used during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethics approval and consent to participate This study was approved by the local ethics committee (EKZN, 2020-01306).

Consent for publication Not applicable.

References

- Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020;382:1708–20.
- Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, The Northwell C-RC, Barnaby DP, Becker LB, Chelico JD, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. *JAMA*. 2020;323:2052–9.
- Haase N, Plovsing R, Christensen S, Poulsen LM, Brochner AC, Rasmussen BS, Helleberg M, Jensen JUS, Andersen LPK, Siegel H, et al. Characteristics, interventions, and longer term outcomes of COVID-19 ICU patients in Denmark-A nationwide, observational study. *Acta Anaesthesiol Scand*. 2021;65:68–75.
- Allameh SF, Nemati S, Ghalehtaki R, Mohammadnejad E, Aghili SM, Khajavirad N, Beigmohammadi MT, Salehi M, Mirfazaelian H, Edalatifard M, et al. Clinical characteristics and outcomes of 905 COVID-19 patients admitted to imam khomeini hospital complex in the capital city of Tehran, Iran. *Arch Iran Med*. 2020;23:766–75.
- Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, Cereda D, Coluccello A, Foti G, Fumagalli R, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy. *JAMA*. 2020;323:1574–81.
- Gregoriano C, Koch D, Haubitz S, Conen A, Fux CA, Mueller B, Bernasconi L, Hammerer-Lercher A, Oberle M, Burgermeister S, et al. Characteristics, predictors and outcomes among 99 patients hospitalised with COVID-19 in a tertiary care centre in Switzerland: an observational analysis. *Swiss Med Wkly*. 2020;150:w20316.
- Moon SS, Lee K, Park J, Yun S, Lee YS, Lee DS. Clinical characteristics and mortality predictors of COVID-19 patients hospitalized at nationally-designated treatment hospitals. *J Korean Med Sci*. 2020;35: e328.
- Thompson JV, Meghani NJ, Powell BM, Newell I, Craven R, Skilton G, Bagg LJ, Yaqoob I, Dixon MJ, Evans EJ, et al. Patient characteristics and predictors of mortality in 470 adults admitted to a district general hospital in England with Covid-19. *Epidemiol Infect*. 2020;148: e285.
- Golozar A, Lai LY, Sena AG, Vizcaya D, Schilling LM, Huser V, Nyberg F, Duvall SL, Morales DR, Alshammari TM et al. Baseline phenotype and 30-day outcomes of people tested for COVID-19: an international network cohort including > 3.32 million people tested with real-time PCR and > 219,000 tested positive for SARS-CoV-2 in South Korea, Spain and the United States. *medRxiv*. 2020. <https://doi.org/10.1101/2020.10.25.20218875>
- Ji D, Zhang D, Xu J, Chen Z, Yang T, Zhao P, Chen G, Cheng G, Wang Y, Bi J, et al. Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. *Clin Infect Dis*. 2020;71:1393–9.
- Levine DM, Lipsitz SR, Co Z, Song W, Dykes PC, Samal L. Derivation of a clinical risk score to predict 14-day occurrence of hypoxia, ICU admission, and death among patients with coronavirus disease 2019. *J Gen Intern Med*. 2021;36:730–7.
- Manocha KK, Kirzner J, Ying X, Yeo I, Peltzer B, Ang B, Li HA, Lerman BB, Safford MM, Goyal P, et al. Troponin and other biomarker levels and outcomes among patients hospitalized with COVID-19: derivation and validation of the HA2T2 COVID-19 mortality risk score. *J Am Heart Assoc*. 2021;10: e018477.
- Weng Z, Chen Q, Li S, Li H, Zhang Q, Lu S, Wu L, Xiong L, Mi B, Liu D, et al. ANDC: an early warning score to predict mortality risk for patients with Coronavirus Disease 2019. *J Transl Med*. 2020;18:328.
- Grifoni E, Valoriani A, Cei F, Vannucchi V, Moroni F, Pelagatti L, Tarquini R, Landini G, Masotti L. The CALL score for predicting outcomes in patients with COVID-19. *Clin Infect Dis*. 2021;72:182–3.
- Wolfisberg S, Gregoriano C, Struja T, Kutz A, Koch D, Bernasconi L, Hammerer-Lercher A, Mohr C, Haubitz S, Conen A, et al. Comparison of characteristics, predictors and outcomes between

- the first and second COVID-19 waves in a tertiary care centre in Switzerland: an observational analysis. *Swiss Med Wkly.* 2021;151:20569.
16. Clinical management of severe acute respiratory infection when novel coronavirus (nCoV) infection is suspected: interim guidance. 2020. <https://apps.who.int/iris/handle/10665/330854>. Accessed 1 July 2021.
 17. Clinical management of COVID-19: interim guidance. 2020. <https://apps.who.int/iris/handle/10665/332196>. Accessed 1 July 2021.
 18. W. H. O. Solidarity Trial Consortium, Pan H, Peto R, Henao-Restrepo AM, Preziosi MP, Sathiyamoorthy V, Abdool Karim Q, Alejandria MM, Hernandez Garcia C, Kieny MP, et al. Repurposed antiviral drugs for Covid-19—Interim WHO solidarity trial results. *N Engl J Med.* 2021;384:497–511.
 19. Corticosteroids for COVID-19, Living Guidance 2. 2020. <https://www.who.int/publications/i/item/WHO-2019-nCoV-Corticosteroids-2020.1>. Accessed 1 July 2021.
 20. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol.* 1994;47:1245–51.
 21. Juma S, Taabazuing MM, Montero-Odasso M. Clinical frailty scale in an acute medicine unit: a simple tool that predicts length of stay. *Can Geriatr J.* 2016;19:34–9.
 22. Yadaw AS, Li YC, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *Lancet Digit Health.* 2020;2:e516–25.
 23. Erturk Sengel B, Tukenmez Tigen E, Ilgin C, et al. Application of CALL score for prediction of progression risk in patients with COVID-19 at university hospital in Turkey. *Int J Clin Pract.* 2021;75: e14642.
 24. Galloway JB, Norton S, Barker RD, Brookes A, Carey I, Clarke BD, Jina R, Reid C, Russell MD, Snee R, et al. A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: an observational cohort study. *J Infect.* 2020;81:282–8.