Mortality prediction models in the adult critically ill: A scoping review

Britt E. Keuning1 | Thomas Kaufmann2 | Renske Wiersema1 | Anders Granholm3 | Ville Pettilä4 | Morten Hylander Møller3,5 | Christian Fynbo Christiansen6 | José Castela Forte1,7 | Harold Snieder8 | Frederik Keus1 | Rick G. Pleijhuis9 | Iwan C. C. van der Horst1,10 | HEALICS consortium

1Department of Critical Care, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
2Department of Anesthesiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
3Department of Intensive Care, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
4Division of Intensive Care Medicine, Department of Anesthesiology, Intensive Care and Pain Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
5Centre for Research in Intensive Care, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
6Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark
7Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Groningen, The Netherlands
8Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
9Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
10Department of Intensive Care, Maastricht University Medical Center+, Maastricht University, Maastricht, The Netherlands

Background: Mortality prediction models are applied in the intensive care unit (ICU) to stratify patients into different risk categories and to facilitate benchmarking. To ensure that the correct prediction models are applied for these purposes, the best performing models must be identified. As a first step, we aimed to establish a systematic review of mortality prediction models in critically ill patients.

Methods: Mortality prediction models were searched in four databases using the following criteria: developed for use in adult ICU patients in high-income countries, with mortality as primary or secondary outcome. Characteristics and performance measures of the models were summarized. Performance was presented in terms of discrimination, calibration and overall performance measures presented in the original publication.

Results: In total, 43 mortality prediction models were included in the final analysis. In all, 15 models were only internally validated (35%), 13 externally (30%) and 10 (23%) were both internally and externally validated by the original researchers. Discrimination was assessed in 42 models (98%). Commonly used calibration measures were the Hosmer-Lemeshow test (60%) and the calibration plot (28%).

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. Acta Anaesthesiologica Scandinavica published by John Wiley & Sons Ltd on behalf of Acta Anaesthesiologica Scandinavica Foundation.
1 | INTRODUCTION

Outcome prediction models, severity scales and risk scores are prognostic tools to estimate the probability for a pre-specified outcome. These prognostic tools use variables (eg about the severity of illness) to predict outcome, often mortality, in a specific patient population such as the critically ill. In the intensive care unit (ICU), mortality prediction models may be applied to stratify patients in different risk categories and to facilitate benchmarking using standardized mortality rates. An accurate mortality prediction model provides a stratification of the risk of an outcome at a population level. These models generally provide a numerical estimate of that risk based on estimates from previous populations. Per definition, all mortality prediction models are best suited for use at a population level and not for individual prognostication, as uncertainty for individual patients remains high.

Several models are widely known and broadly applied such as the Acute Physiology and Chronic Health Evaluation (APACHE) I-IV, the Mortality Prediction Model (MPM) and the Simplified Acute Physiology Score (SAPS) I-III, whereas others like the Intensive Care National Audit & Research Centre (ICNARC) are used solely in one country. Previous literature has only reviewed commonly used models, models with different outcome than mortality or disease- or organ-specific prognostic models. To the best of our knowledge, no study has systematically assessed which mortality prediction models have been developed and validated for broad cohorts of adult critically ill patients.

1.1 | Rationale and objective

The objective of this study was to provide an overview of available mortality prediction models in adult critically ill patients as a step-up towards future head-to-head comparison of model performance through systematic external validation.

2 | METHODS

2.1 | Protocol and registration

This scoping review was performed following our protocol (Appendix S1) and was reported in accordance with the PRISMA-ScR checklist.

2.2 | Search strategy

We conducted a systematic search of MEDLINE, EMBASE, Web of Science and The Cochrane Central Register of Controlled Trials (CENTRAL) to identify relevant ICU mortality prediction models (Appendix S1). Mortality was chosen as the outcome of interest, as prediction models were originally developed to identify patients with high mortality risk. For all databases, except the CENTRAL database, the search period encompassed a period starting from the 1st January 2008 to the 21st April 2019. We used snowballing, that is, searching references and related articles, to identify additional prediction models that were published before 2008.

One author ran the search, after which the screening of records and data extraction were performed in duplicate. All records were screened based on title and/or abstract. Papers clearly irrelevant to the purpose were excluded. The remaining articles were screened for eligibility. Consulting a third opinion solved disagreements. More detailed information is presented in the protocol (Appendix S1).

2.3 | Eligibility criteria

To be considered eligible, mortality prediction models had to meet the following criteria: (a) originally developed specifically for use in adult critically ill patients as defined by the included studies, (b) representing...
broad groups of ICU patients (with large diversity of admission diagnoses, eg non-diabetic patients, medical admissions, surgical admissions, etc), (c) availability of the original article in English and (d) mortality at any time as (primary or secondary) outcome of interest.

Prediction models were excluded (a) when developed for low- or middle-income countries, as characteristics of ICU patients in these countries often substantially differ from those in high-income countries and, epidemiological data from low-income countries have been frequently unavailable,10,11 (b) when developed as a digital model or derived from a machine-learning algorithm, since code and data availability are not requirements in all journals. Since our utmost goal is to make a head-to-head comparison of available mortality prediction models using an independent external validation cohort, the code or data necessary to retrieve the underlying prediction model formula are required to reproduce the prediction models. (c) When the development of multiple customized prediction models was described in one article, but no final model was proposed, the prediction models were excluded. Finally, (d) we excluded prediction models specifically developed for subgroups of intensive care patients such as those with sepsis, trauma, cardiac and neurological patients. Studies not specifying inclusion of these subgroups within a wider, general ICU population were considered to be eligible. Prediction models developed in a medical or surgical ICU were included.

2.4 Data extraction

If multiple mortality outcomes (eg at different time points) were used, we used the primary outcome in the original publication (or the first mortality outcome if the primary outcome was not mortality) to describe the performance of the prediction model.

Details on the development process of the mortality prediction models included were shown, as well as the number of variables included in the prediction models, mortality rate in each development setting and method of handling of missing data. To give an overview of the performance of all mortality prediction models, for example, values from discrimination, calibration and overall performances measures12 for mortality were presented for development and internal or external validation cohorts in the original publication (if available).

The discrimination measure presented was the C-statistic (area under the receiver operating characteristic curve [AUROC]), calibration measures presented were goodness-of-fit tests like the Hosmer-Lemeshow (HL) test, calibration plot and calibration slope, and the overall performance measures presented were the Nagelkerke’s $R^2$ and the Brier score.12

Preferable values from external validation were presented if both internal and external validation values were present in the
original publication. If not available, values of internal validation cohorts were presented. External validation was defined as using a separate individual dataset for validation of the mortality prediction model (ie no split sampling of a dataset also used for the development of the model).

Citations of original publications were screened for internal and/or external validation articles and shown as being present (+) or absent (−). A list of variables sought for in the identified articles can be found in Appendix S1.

3 | RESULTS

The selection of sources of evidence can be found in the flowchart (Figure 1). Articles evidently developed for specific groups of patients (ie sepsis, trauma, cardiac, neurological patients) were excluded based on the title and/or abstract. Evaluating 99 full-text articles for eligibility resulted in exclusion of another 39 articles, leaving 60 articles that were screened for original publications. Eventually, 43 relevant mortality prediction models reported in 38 publications were extracted and included in the final analysis.

3.1 | Characteristics of the included mortality prediction models

Characteristics of the mortality prediction models and underlying derivation cohorts are presented in Table 1. In all, 19 mortality prediction models (44%) were developed using prospectively collected data specifically gathered for the development of the prediction model, whereas 24 (56%) were developed using either retrospective data or prospective data previously collected for other purposes. The start of data collection for the development cohorts spanned 36 years (1979-2015), and the duration of the cohort studies varying from 2 months up to 10 years for each cohort. Two mortality prediction models (4.7%) did not report the timespan during which their development cohort was assembled. In all, 31 mortality prediction models (74%) were developed in a single country, whereas 24 (56%) were developed using either retrospective data or prospective data previously collected for other purposes. The number of patients included in the development databases ranged from 232 to 731,611 patients with a median of 4,895 (IQR 528-35878). The minimum age at which patients were included was 15 years (2.3%), In all, 11 mortality prediction models (26%) did not specify age. The number of variables included in the mortality prediction models varied from 5 up to 5695, with a median of 16 (IQR 9-24).

3.2 | Outcome measures

The timing of mortality outcome varied between the studies. Hospital mortality was the most frequently used primary outcome in 29 (67%) mortality prediction models. Other primary outcome variables were ICU mortality (7%), 28-day mortality (4.7%), 90-day mortality (4.7%), 3- to 28-day mortality (4.7%), 30-day mortality (2.3%), 180-day mortality (2.3%), 6-month mortality (2.3%), 15-year mortality (2.3%), and 6- and 12-month mortality (2.3%).

Secondary outcomes were 1-month mortality after ICU admission (4.7%), hospital mortality (4.7%), ICU mortality (2.3%), 3-month mortality after ICU admission (2.3%), 6-month mortality after ICU admission (2.3%), 9-month mortality (2.3%), 1-year mortality (2.3%) and length of stay (2.3%). Of the 43, 37 mortality prediction models (86%) did not prognosticate any secondary outcome.

For 21 mortality prediction models (49% of 43), data were collected within the first 24 hours after patient admission to the ICU, whereas for the remaining prediction models data timing varied from 24 days before admission up to 5 days after patient admission to the ICU.

Handling of missing data was not reported in 11 mortality prediction models (26%), 23,25,26,31,33,38,41,43,46,49 20 prediction models (47% of 43) excluded records with missing data, whereas for the remaining prediction models data timing varied from 24 days before admission up to 5 days after patient admission to the ICU.

Discrimination, calibration and overall performance measures

Discrimination, calibration and overall performance measures are presented in Table 2. Of the 43 mortality prediction models, 15 (35%) were only internally validated, whereas 13 (30%) only externally, 16,19-21,25,36,42,43,47 10 (23%) were both internally and externally validated, and 5 prediction models (12%) were not validated at all. In all, 15 prediction models (35%) included a description of an external validation in their original publication.

Discrimination was expressed as the AUROC in 42 of the 43 mortality prediction models. Only the APACHE II model did not report an AUROC value in the original publication. In the development cohorts, the lowest discrimination was AUROC 0.72 (95% CI 0.71-0.74), and the highest AUROC 0.91 (95% CI not specified). In the validation cohorts, the lowest AUROC was 0.58 (95% CI not specified), and the highest AUROC 0.95 (0.91-0.99).

Calibration measures were expressed by various statistical measures. The HL goodness-of-fit test was used in 26 mortality prediction models (47% of 43) excluding patients when more than a certain percentage of the data was missing (>5% or >25%).
| Mortality prediction model | Year published | Development database | Cohort assembly period | ICU population | Number of variables<sup>a</sup> | Outcome | Hospital mortality rate in each development setting | Data collection | Handling of missing data |
|----------------------------|----------------|----------------------|------------------------|----------------|---------------------------------|---------|-----------------------------------------------|----------------|------------------------|
| ICNARC-Harrison et al<sup>6</sup> | 2007 | 216 626 Prospective | December 1995-August 2003 | General, adult patients in England, Wales and Ireland | 16 | Hospital mortality | Not reported | Worst values and total urine output in initial 24 h in ICU | Exclusion |
| ICNARC-II Ferrando-Vivas et al<sup>13</sup> | 2017 | 155 239 Prospective | 01/01/2012-31/12/2012 | General, adult patients in England, Wales and Ireland | 23 | Hospital mortality | 32 064/155 239 (20.7%) | Worst values and total urine output in initial 24 h in ICU | No missing data |
| APACHE IV Zimmerman et al<sup>14</sup> | 2006 | 66 270 Prospective | 01/01/2002-31/12/2003 | General, adult (≥16 y) patients in the USA | 142 | Hospital mortality | 9,013/66,270<sup>b</sup> (13.6%) | Worst values in initial 24 h in ICU | Exclusion |
| SAPS III Moreno et al<sup>15</sup> | 2005 | 13 428<sup>b</sup> Prospective | 14/10/2002-15/12/2002 | General, adult (≥16 y) patients worldwide | 20 | Hospital mortality | Not reported | ICU admission ± 1 h | Imputation of normal values |
| MPM<sup>24</sup>-II Lemeshow et al<sup>16</sup> | 1993 | 12 610 Prospective | 17/04/1989-31/07/1990 (dataset I) and 30/09/1991-27/12/1991 (dataset II) | General, adult (≥18 y) patients in Europe and the USA | 15 | Hospital mortality | 2632/12 610 (20.9%) | ICU admission | Exclusion |
| MPM<sup>24</sup>-II Lemeshow et al<sup>16,21</sup> | 1993 | 10 357 Prospective | 17/04/1989-31/07/1990 (dataset I) and 30/09/1991-27/12/1991 (dataset II) | General, adult (≥18 y) patients in Europe and the USA | 13 | Hospital mortality | 2261/10 357 (21.8%) | At 24 h in ICU | Exclusion |
| SAPS II Le Gall et al<sup>17</sup> | 1993 | 8369 Prospective | 30/09/1991-28/02/1992 | General, adult (≥18 y) patients in Europe and North-America | 17 | Hospital mortality | 1824/8369<sup>b</sup> (21.8%) | Worst values in initial 24 h in ICU | Imputation of normal values |
| APACHE III Knaus et al<sup>18</sup> | 1991 | 7848<sup>b</sup> Prospective | May 1988-November 1989 | General, adult (≥16 y) patients in the USA | 26 | Hospital mortality | Not reported | Worst values in initial 24 h in ICU | Imputation of normal values |
| APACHE II Knaus et al<sup>19</sup> | 1985 | 5030 Prospective | 1979-1982 | General, adult (≥16 y) patients in the USA | 18 | Hospital mortality | 993/5030 (19.7%) | Worst values in initial 24 h in ICU | Exclusion |

(Continues)
| Mortality prediction model | Year published | Development database | Cohort assembly period | ICU population | Number of variables \(^a\) | Outcome | Hospital mortality rate in each development setting | Data collection | Handling of missing data |
|----------------------------|----------------|----------------------|-----------------------|----------------|-----------------|---------|---------------------------------|----------------|-------------------------|
| SUPPORT Knaus et al\(^{20}\) | 1995 | 4301 Prospective | June 1989-June 1991 | General, adult (≥18 y) patients in the USA | 15 | Primary 180-day mortality | – | 2072/4301 (48.2%) | After 3 days | Imputation of normal values, missing data at day 3 were imputed with day 1 values |
| MPM\(_{48}\)-II Lemeshow et al\(^{21}\) | 1994 | 2049 Prospective | 17/04/1989-31/07/1990 | General, adult (≥18 y) patients in the USA | 13 | Hospital mortality | – | 307/2049\(^b\) (15.0%) | At 48 h in ICU | Exclusion |
| MPM\(_{72}\)-II Lemeshow et al\(^{21}\) | 1994 | 1497 Prospective | 17/04/1989-31/07/1990 | General, adult (≥18 y) patients in the USA | 13 | Hospital mortality | – | 418/1497\(^b\) (27.9%) | At 72 h in ICU | Exclusion |
| TRIOSTimsi et al\(^{22}\) | 2001 | 893 Prospective | Not reported (validation dataset in March 1999) | General, adult (≥16 y) patients, hospitalized >48 h in France | 32 | Hospital mortality | – | 268/893 (30.0%) | First 3 days in ICU | Imputation of normal values |
| Mortality Risk Score Dóleri-Moreno et al\(^{23}\) | 2016 | 844 Prospective | January 2013-April 2014 | General, adult patients in Spain | 6 | ICU mortality | – | 91/844 (10.8%) | ICU admission | Not reported |
| Mortality Multifactor Model Lieta\(^{24}\) | 2017 | 500 Prospective | 01/03/2014-30/04/2014 | General, adult (≥18 y) patients in China | 36 | Hospital mortality | Mortality 30 days after ICU admission, LOS | 102/500 (20.4%) | First 24 h in ICU | Exclusion |
| Mortality Prognostic Model Hadique et al\(^{25}\) | 2017 | 500 Prospective | November 2013-April 2014 | Medical, adult patients in the USA | 44 | 6-month mortality | – | 180/500 (36.0%) | ICU admission, SQ within 12-24 h of admission | Not reported |
| Mortality Prediction Model Fika et al\(^{26}\) | 2018 | 400 Prospective | January 2012-July 2013 | General, adult (≥18 y) patients in Greece | 12 | ICU mortality | – | 131/400 (33.8%) | Worst values in initial 24 h in ICU | Not reported |
| APACHE II-APM Nematifard et al\(^{27}\) | 2018 | 304 Prospective | June 2014-November 2016 | General, adult (≥16 y) patients in Iran | 19 | Hospital mortality | – | 96/304 (31.6%) | Worst values in initial 24 h in ICU | Exclusion |

(Continues)
| Mortality prediction model | Year published | Development database | Cohort assembly period | ICU population | Number of variables | Outcome | Hospital mortality rate in each development setting | Data collection | Handling of missing data |
|---------------------------|----------------|---------------------|-----------------------|----------------|------------------|----------|--------------------------------------------------|----------------|------------------------|
| APACHE III-APM Nematifard et al<sup>27</sup> | 2018 | 304 | Prospective June 2014-November 2016 | General, adult (≥16 y) patients in Iran | 27 | Hospital mortality | 96/304 (31.6%) | Worst values in initial 24 h in ICU | Exclusion |
| ANZROD Paul et al<sup>28</sup> | 2017 | 731 611 | Retrospective 01/01/2006-31/12/2015 | General, adult (≥16 y) patients in Australia and New Zealand | 11 | Hospital mortality | 69 503/731 611<sup>b</sup> (9.5%) | ICU admission | Exclusion |
| MMI Min et al<sup>29</sup> | 2017 | 354 154<sup>b</sup> | Retrospective January 2003-December 2013 | Medical, veteran ICU patients in the USA | 5695 | All-cause mortality at 6- and 12-months post-hospital discharge | Not reported | Worst values of 24 h before and 24 h after admission | Imputation of mean values |
| ANZROD Paul et al<sup>30</sup> | 2013 | 304 149 | Retrospective 01/01/2004-31/12/2009 | General, adult (≥16 y) patients in Australia and New Zealand | 38 | Hospital mortality | 34 369<sup>b</sup> (11.3%) | Worst values in initial 24 h in ICU | Exclusion |
| Customized APACHE IV Brinkman et al<sup>21</sup> | 2013 | 77 616 | Retrospective 01/01/2008-01/07/2011 | Non-CABG, adult critically ill patients in the Netherlands | 142 | Hospital mortality | 12 186/77 616<sup>b</sup> (15.7%) | First 24 h in ICU | Not reported |
| MPM<sub>0</sub>-III Higgins et al<sup>32</sup> | 2005 | 74 578 | Retrospective October 2001-March 2004 | General, adult (≥18 y) patients in the USA, Canada and Brazil | 16 | Hospital mortality | 10 292/74 578 (13.8%) | ICU admission | Exclusion |
| NQF-ICOMmorth Philip R. Lee Institute<sup>33</sup> | 2016 | 40 395 | Retrospective Not reported | General, adult (≥18 y) patients in the USA | 17 | Hospital mortality | Not reported | 1 h prior to ICU admission to 1 h after admission | Not reported |
| OASIS Johnson et al<sup>34</sup> | 2013 | 39 070 | Retrospective 01/01/2007-15/09/2011 | General, adult (≥16 y) patients in the USA | 10 | ICU mortality | 4571/39 070<sup>b</sup> (11.7%) | Worst values and total urine output in initial 24 h in ICU | Exclusion |
| COPE-4 Duke et al<sup>35</sup> | 2013 | 35 878 | Retrospective 01/07/2004-30/06/2006 | General, adult (≥15 y) patients in Australia | 6 | Hospital mortality | 4415/35 878 (12.3%) | ICU admission (mechanical ventilation during ICU admission) | No missing data |
| Mortality prediction model | Year published | Development database | Cohort assembly period | ICU population | Number of variables | Outcome | Hospital mortality rate in each development setting | Data collection | Handling of missing data |
|---------------------------|----------------|----------------------|------------------------|----------------|---------------------|---------|--------------------------------------------------|----------------|--------------------------|
| RDW-SAPS                  | 2012           | 17 922               | January 2001-December 2008 | General, adult (≥18 y) patients in the USA | 15 | Hospital mortality | ICU mortality, 1-year mortality | 2007/17 922 (11.2%) | ICU admission, Not reported |
| COPE                      | 2008           | 17 880               | 01/07/2004-30/06/2005    | General, adult patients in Australia | 5 | Hospital mortality | – | 2186/17 880 (12.1%) | ICU admission (mechanical ventilation during ICU admission), No missing data |
| PREDICT                   | 2008           | 11 930               | 1989-2002               | General, adult (≥16 y) patients in Australia | 6 | 15-year mortality | – | 829/11 930 (6.9%) | First 5 days in ICU, No missing data |
| High-Risk Selection System| 2006           | 8248                 | October 1994-February 1995 | General, adult patients (>24 h in ICU) in Europe | 16 | Hospital mortality | – | 1617/8248 (19.6%) | ICU admission, Not reported |
| GV-SAPS II                | 2016           | 4895                 | 2001-2008               | Non-diabetic, adult (≥18 y) patients in the USA | 20 | 30-day mortality | 9-month mortality | 649/4895 (13.3%) | First 24 h in ICU, When >5% exclusion, <5% not reported |
| MODS/NEMS                 | 2016           | 4321                 | 01/01/2009-30/11/2012   | General, adult patients in Canada | 32 | Hospital mortality | – | 986/4321 (22.8%) | First 24 h in ICU, Not reported |
| SMS-ICU                   | 2018           | 4086                 | 23/12/2009-30/06/2016   | General, adult (≥18 y), acutely admitted patients worldwide | 7 | 90-day mortality | – | 1403/4086 (34.3%) | Worst values in initial 24 h in ICU, Multiple imputations, exclusion when >25% |
| P-model                   | 2010           | 3505                 | 01/01/2007-31/12/2007   | General, adult (≥20 y) patients in Japan | 10 | Mortality at 28 days after the first ICU day | – | 336/3505 (9.6%) | First 24 h in ICU, Not reported |
| BCV model                 | 2013           | 1624                 | 01/01/2006-01/12/2008   | General, adult (≥18 y) patients in Taiwan | 6 | Daily probability of mortality from day 3 to day 28 post-ICU admission | – | Not reported | Daily complete blood count, Exclusion |

(Continues)
| Mortality prediction model | Year published | Development database | Cohort assembly period | ICU population | Number of variables | Outcome | Hospital mortality rate in each development setting | Data collection | Handling of missing data |
|----------------------------|----------------|----------------------|------------------------|----------------|---------------------|---------|---------------------------------------------------|----------------|---------------------|
| BCV/APACHE II model        | 2013           | 1624 Retrospective   | 01/01/2006-01/12/2008  | General, adult (≥18 y) patients in Taiwan | 24       | Daily probability of mortality from day 3 to day 28 post-ICU admission | Not reported | Daily complete blood count, APACHE II score in the first 24 h in ICU |
| CREEK                      | 2008           | 528 Retrospective    | April 2003-January 2004| Medical, adult (≥18 y) patients in Germany | 8        | Hospital mortality | 87/528 (16.5%) | ICU admission | Not reported |
| SAPS-R                     | 1991           | 351 Retrospective    | 01/01/1986-31/10/1988 | General, adult patients in France | 5        | Hospital mortality | Not reported | Worst values in initial 24 h in ICU | Exclusion |
| SAPS-E                     | 1991           | 351 Retrospective    | 01/01/1986-31/10/1988 | General, adult patients in France | 7        | Hospital mortality | Not reported | Worst values in initial 24 h in ICU | Exclusion |
| 25OHD Deyo-Charlson Comorbidity Index | 2016       | 310 Retrospective    | 01/06/2012-30/05/2015 | General, adult (≥18 y) patients in the USA | 18       | 90-day mortality after ICU admission | 59/310 (19.0%) | First 24 h in ICU | Not reported |
| DELAWARE                   | 2008           | 271 Retrospective    | April 2003-January 2004| Surgical, adult (≥18 y) patients in Germany | 9        | Hospital mortality | 67/271 (24.7%) | ICU admission | Exclusion |
| Simplified Mortality Score | 2018           | 232 Retrospective    | June 2015-February 2016| Medical, adult (≥18 y) patients in Korea | 8        | 28-day mortality | 72/232 (31.1%) | Within 24 h of ICU admission | Exclusion |

Abbreviations: ANZROD, Australian and New Zealand Risk Of Death; APACHE, Acute Physiology and Chronic Health Evaluation; APM, adductor pollicis muscle; BCV, blood cell variability; COPE, critical care outcome prediction equation; CREEK, critical risk evaluation by early keys; DELAWARE, Dense Laboratory Whole Blood Applied Risk Estimation; GV, glucose variability; ICNARC, Intensive Care National Audit Research Centre; ICU, intensive care unit; LOS, length of stay; MMI, multi-morbidity index; MODS, multiple organs dysfunctional score; MPM, mortality prediction model; NEMS, nine equivalents nursing manpower use score; NQF-ICMmort, national quality forum ICU outcomes model(mortality); OASIS, oxford acute severity of illness score; PREDICT, predicted risk, existing diseases and intensive care therapy; RDW, red cell distribution width; SAPS, simplified acute physiology score; SMS-ICU, simplified mortality score for the intensive care unit; SQ, surprise question; SUPPORT, study to understand prognoses and preferences for outcomes and risks of treatments; TRIOS, three-day recalibrating ICU outcomes.

*When (parts of) other mortality prediction models were used as variables in a mortality prediction model (eg the Charlson Comorbidity Index and APACHE III as variable in the Mortality Prognostic Model), variables included in these specific mortality prediction models were also taken into account.

**Estimated based on information in original publication.
TABLE 2 Performance of the 43 mortality prediction models

| Mortality prediction model | Validated? | AUROC (95% CI) Development cohort | Calibration Development cohort | Overall performance Development cohort | AUROC (95% CI) Validation cohort | Calibration Validation cohort | Overall performance Validation cohort |
|---------------------------|-----------|-----------------------------------|--------------------------------|----------------------------------------|---------------------------------|---------------------------------|----------------------------------------|
| ICNARC                    | +         | 0.89 (0.89-0.89)                  | −                              | −                                      | 0.89 (0.88-0.89)                | Calibration plot present         | Brier score: 0.132                   |
| Harrison et al            | +         | Data splitting                    | −                              | −                                      | Internal validation dataset     | −                              | Brier score: 0.108                   |
| ICNARC-II                 | +         | Bootstrapping                     | +                              | −                                      | 0.89 (0.89-0.89)                | Brier score: 0.103               | −                                      |
| Ferrando-Vivas et al      | +         | Original publication              | −                              | −                                      | 0.89 (0.88-0.89)                | −                              | −                                      |
| APACHE IV                 | +         | Data splitting                    | +                              | −                                      | 0.88 (n.s.)                     | −                              | −                                      |
| Zimmerman et al          | +         | Original publication              | −                              | −                                      | 0.88 (n.s.)                     | −                              | −                                      |
| SAPS III                  | +         | Cross-validation                  | −                              | −                                      | 0.85 (n.s.)                     | −                              | −                                      |
| Moreno et al             | +         | Internal validation dataset       | −                              | −                                      | 0.85 (n.s.)                     | −                              | −                                      |
| MPM I                    | −         | +                                 | 0.84 (n.s.)                     | −                                      | 0.82 (n.s.)                     | −                              | −                                      |
| Lemeshow et al            |           | Original publication              | −                              | −                                      | 0.82 (n.s.)                     | −                              | −                                      |
| MPM IV                    | −         | +                                 | 0.84 (n.s.)                     | −                                      | 0.84 (n.s.)                     | −                              | −                                      |
| Lemeshow et al            |           | Original publication              | −                              | −                                      | 0.84 (n.s.)                     | −                              | −                                      |
| MPM IV                    | −         | +                                 | 0.84 (n.s.)                     | −                                      | 0.84 (n.s.)                     | −                              | −                                      |
| Le Gall et al             | +         | Data splitting                    | +                              | −                                      | 0.86 (0.84-0.88)                | −                              | −                                      |
| APACHE III                | +         | Data splitting                    | +                              | −                                      | 0.90 (n.s.)                     | −                              | −                                      |
| Knaus et al               | +         | Data splitting                    | +                              | −                                      | 0.90 (n.s.)                     | −                              | −                                      |
| APACHE II                 | −         | +                                 | −                              | −                                      | −                              | −                              | −                                      |
| Knaus et al               |           | Original publication              | −                              | −                                      | −                              | −                              | −                                      |
| SUPPORT                   | −         | +                                 | 0.79 (n.s.)                     | −                                      | 0.78 (n.s.)                     | Calibration plot present         | −                                      |
| Knaus et al               |           | Original publication              | −                              | −                                      | −                              | −                              | −                                      |
| MPM V                    | −         | +                                 | 0.81 (n.s.)                     | −                                      | 0.81 (n.s.)                     | −                              | −                                      |
| Lemeshow et al            |           | Original publication              | −                              | −                                      | −                              | −                              | −                                      |
| MPM VI                    | −         | +                                 | 0.79 (n.s.)                     | −                                      | 0.79 (n.s.)                     | −                              | −                                      |
| Lemeshow et al            |           | Original publication              | −                              | −                                      | −                              | −                              | −                                      |

(Continues)
| Mortality prediction model           | Validated? | AUROC (95% CI) Development cohort | Calibration Development cohort | Overall performance Development cohort | Type of validation cohort in original publication | AUROC (95% CI) Validation cohort | Calibration Validation cohort | Overall performance Validation cohort |
|-------------------------------------|------------|-----------------------------------|--------------------------------|----------------------------------------|-----------------------------------------------|----------------------------------|---------------------------------|----------------------------------|
| TRIOS Timsit et al22                | +          | 0.79 (0.77-0.82)                 | HLC-statistic: 5.6 \( (P = .70) \) | -                                      | External validation dataset                   | 0.83 (0.78-0.87)                 | -                              | -                                |
| Mortality Risk Score Dólera-Moreno et al23 | +          | 0.84 (0.80-0.87)                 | HLC-statistic: 12.3 \( (P = .14) \) | -                                      | Internal validation dataset                   | 0.95 (0.91-0.99)                 | Likelihood ratio test X\(^2\): 296.8\(^c\) | -                                |
| Mortality Multifactor Model Li et al24 | -          | 0.83 (0.80-0.87)                 | HLC-statistic: 6.5 \( (P = .59) \) | -                                      | Internal validation dataset                   | 0.85 (0.73-0.97)                 | HLC-statistic: 9.2 \( (P = .33) \) | -                                |
| Mortality Prognostic Model Hadique et al25 | +          | 0.85 (0.85-0.86)                 | HLC-statistic: 459.3               | -                                      | Internal validation dataset                   | 0.85 (0.85-0.85)                 | HLC-statistic: 264.9 Calibration plot present | Brier score: 0.069 Adjusted Brier score: 0.190 |
| APACHE II-APM Nematifard et al27    | -          | 0.85 (0.81-0.90)                 | -                                | -                                      | -                              | -                              | -                                |
| APACHE III-APM Nematifard et al27   | -          | 0.87 (0.82-0.91)                 | -                                | -                                      | -                              | -                              | -                                |
| ANZRODO Paul et al28                | +          | 0.85 (0.85-0.86)                 | HLC-statistic: 459.3               | -                                      | Internal validation dataset                   | 0.85 (0.95-0.85)                 | Brier score: 0.069 Adjusted Brier score: 0.190 |
| MMI Min et al29                     | +          | 0.86 (0.85-0.86)                 | HLC-statistic: 259.3               | -                                      | Internal validation dataset                   | 6-month mortality: 0.86 (0.85-0.86) | 12-month mortality: 0.84 (0.83-0.84) | -                                |
| Mortality prediction model | Validated\(a\) | AUROC (95% CI) Development cohort\(b\) | Calibration Development cohort\(b\) | Overall performance Development cohort\(b\) | Type of validation cohort in original publication | AUROC (95% CI) Validation cohort | Calibration Validation cohort | Overall performance Validation cohort |
|----------------------------|-----------------|--------------------------------------|--------------------------------------|---------------------------------------------|-----------------------------------------------|-----------------------------|---------------------------------|----------------------------------|
| ANZROD Paul et al\(^{30}\) | + Data splitting | 0.91 (n.s.) | HL C-statistic: 189.5 | Brier score: 0.065 | Internal validation dataset | 0.90 (n.s.) | HL C-statistic: 104.9 | Brier score: 0.066 |
| Customized APACHE IV Brinkman et al\(^{31}\) | + Bootstrapping | 0.88 (0.88-0.88) | Calibration plot present | Brier score: 0.09 | Internal validation dataset | – | – | – |
| MPM\(_{0-III}\) Higgins et al\(^{32}\) | + Data splitting | 0.83 (0.82-0.83) | HL statistic: 11.5 \((P = .17)\) | – | Internal validation dataset | 0.82 (0.82-0.83) | HL statistic: 11.6 \((P = .31)\) | – |
| NQF-ICOMmortality Philip R. Lee Institute\(^{33}\) | + Data splitting | – | – | – | Internal validation dataset | 0.82 (0.81-0.83) | HL C statistic: 12.0 \((P = .28)\) | – |
| OASIS Johnson et al\(^{34}\) | + Data splitting + Original publication | – | – | – | External validation dataset | 0.90 \((P < .0003)^{c}\) | HL \(X^2\): 19.6 | Brier score: 0.048 \(^{c}\) |
| COPE-4 Duke et al\(^{35}\) | – + Original publication | – | – | – | External validation dataset | – (0.82-0.83) | HL H-statistic: 14.8 \((P = .06)\) | – |
| RDW-SAPS Hunziker et al\(^{30}\) | – | – | 0.77 (n.s.) | Quasi Likelihood under the Independence model Criterion \((QIC) \ X^2\): 1.83 | – | – | – | – | – |

\(\text{Continues}\)
| Mortality prediction model | Validated? | AUROC (95% CI) | Calibration Development cohort | Overall performance Development cohort | Type of validation cohort in original publication | AUROC (95% CI) Validation cohort | Calibration Validation cohort | Overall performance Validation cohort |
|---------------------------|------------|----------------|---------------------------------|----------------------------------------|----------------------------------------|-----------------------------|---------------------------------|----------------------------------|
| COPE Duke et al 36         | -          | -              | Original publication            | -                                      | -                                      | -                           | -                               | -                                |
| PREDICT Ho et al 37        | +          | -              | Bootstrapping                   | -                                      | -                                      | Internal validation dataset  | 0.76 (0.75-0.77)                | Calibration plot present         | Nagelkerke's $R^2$: 0.255       |
| High-Risk Selection System Iapichino et al 46 | +          | -              | Data splitting                  | 0.81 (n.s.)                            | -                                      | Internal validation dataset  | 0.81 (n.s.)                    | HL $X^2$: n.s. ($P = .22$)       |                                      |
| GV-SAPS II Liu et al 37    | -          | +              | Original publication            | 0.83 (0.81-0.84)                       | -                                      | External validation dataset  | 0.82 (0.81-0.83)               | -                                | -                                |
| MODS/NEMS Kao et al 48     | +          | -              | Bootstrapping                   | 0.79 (n.s.)                            | -                                      | Internal validation dataset  | 0.76 (n.s.)                    | HL $X^2$: 5.48 ($P = .32$)       | -                                |
| SMS-ICU Granholm et al 48  | +          | +              | Bootstrapping                   | 0.72 (0.71-0.74)                       | NLX $X^2$: 9.0 ($P = .34$)              | Nagelkerke's $R^2$: 0.191    | Internal validation dataset  | 0.73 (n.s.)                    | Calibration slope: 0.99          | Nagelkerke's $R^2$: 0.193       |
| P-model Umegaki et al 39   | +          | -              | Cross-validation                | 0.87 (0.85-0.90)                       | HL $X^2$: 14.5 ($P = .07$)              | Internal validation dataset  | 0.90 (0.88-0.92)               | HL $X^2$: 13.5 ($P = .10$)       | -                                |
| BCV model Huang et al 40   | +          | -              | Data splitting                  | 0.79 (0.76-0.81)                       | HL $X^2$: 8.7 ($P = .37$)               | Internal validation dataset  | 0.76 (0.71-0.81)               | HL $X^2$: 11.1 ($P = .19$)       | -                                |
| BCV/APACHE II model Huang et al 40 | +          | -              | Data splitting                  | 0.80 (0.78-0.83)                       | HL $X^2$: 6.2 ($P = .63$)               | Internal validation dataset  | 0.78 (0.73-0.83)               | HL $X^2$: 5.4 ($P = .72$)        | -                                |
| CREEK Stachon et al 45     | +          | -              | Cross-validation                | 0.86 (n.s.)                            | Brier score: 0.096                     | Internal validation dataset  | 0.832 (n.s.)                   | -                                | -                                |
| SAPS-R Viviand et al 42    | -          | +              | Original publication            | -                                      | -                                      | External validation dataset  | 0.76 (n.s.)                    | -                                | -                                |

(Continues)
TABLE 2  (Continued)

| Mortality prediction model | Validated? | AUROC (95% CI) Development cohort<sup>a</sup> | Calibration Development cohort<sup>b</sup> | Overall performance Development cohort<sup>b</sup> | Type of validation cohort in original publication | AUROC (95% CI) Validation cohort | Calibration Validation cohort | Overall performance Validation cohort |
|---------------------------|-----------|---------------------------------------------|---------------------------------------------|---------------------------------------------|-------------------------------------------------|---------------------------------|-----------------------------------|-----------------------------------|
| SAPS-E Viviand et al<sup>42</sup> | - | - | - | External validation dataset | 0.79 (n.s.) | - | - | - |
| 25OHD Deyo-Charlson Comorbidity Index Mahato et al<sup>49</sup> | - | - | 0.75 (0.67-0.83) | - | - | - | - | - |
| DELAWARE Stachon et al<sup>53</sup> | - | + Original publication | 0.86 (0.80-0.91) | HL statistic: n.s. (P = .28) | External validation dataset | 0.81 (0.75-0.87) | HL statistic: 0.44 (P = n.s.) | Calibration plot present | - |
| Simplified Mortality Score Goag et al<sup>44</sup> | + Data splitting | - | - | - | Internal validation dataset | 0.58 (n.s.) | - | - |

Abbreviations: ANZROD, Australian and New Zealand Risk Of Death; APACHE, Acute Physiology and Chronic Health Evaluation; APM, adductor pollicis muscle; AUROC; area under the receiving operating curves; BCV, Blood Cell Variability; CI, confidence interval; COPE, Critical care Outcome Prediction Equation; CREEK, Critical Risk Evaluation by Early Keys; DELAWARE, Dense Laboratory Whole Blood Applied Risk Estimation; GV, glucose variability; HL, Hosmer-Lemeshow; ICNARC, Intensive Care National Audit Research Centre; ICU, intensive care unit; MMI, Multi-morbidity Index; MODS, Multiple Organs Dysfunctional Score; MPM, mortality prediction model; NEMS, Nine Equivalents Nursing Manpower use Score; NQF-ICOMmort, National Quality Forum ICU outcomes model (mortality); n.s., not specified; OASIS, Oxford Acute Severity of Illness Score; PREDICT, Predicted Risk, Existing, Diseases and Intensive Care Therapy; RDW, red cell distribution width; SAPS, Simplified Acute Physiology Score; SMS-ICU, Simplified Mortality Score for the Intensive Care Unit; SUPPORT, Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments; TRIOS, Three-day Recalibrating ICU Outcomes.

<sup>a</sup>Citations of original publications were screened on internal and/or external validation articles and shown as being present (+) or not present (−). When internal validation was present, the method of internal validation used in the original publication was presented. When external validation in the original publication was present, original publication was added in the column.

<sup>b</sup>Development cohort indicates the cohort in whom the prediction model was developed, sometimes also referred to as training cohort.

<sup>c</sup>Not clear whether the value was derived from the development or validation dataset in the original publication, or value was derived from the development and validation dataset together.

<sup>d</sup>Not clear whether this value is calculated for the 6-month mortality outcome or 12-month mortality.

<sup>e</sup>Not clear whether the value was derived from the internal or external validation dataset in the original publication.
models (60%). Calibration plot was expressed for 12 prediction models (28%), and two prediction models (4.7%) presented the calibration slope value. Finally, one prediction model (2.3%) used the likelihood ratio test chi-squared value, and one prediction model (2.3%) used the Quasi likelihood under the Independence Criterion. In 11 prediction models (26%), calibration was not assessed. Overall performance was expressed as the Brier score in eight mortality prediction models (19%), and as Nagelkerke’s R² in two prediction models (4.7%).

4 | DISCUSSION

4.1 | Main findings

In this scoping review, we presented a contemporary overview of 43 mortality prediction models used in adult ICU patients in high-income countries. We found varying methodology, and the validation and performance of individual prediction models differ. Only 23 mortality prediction models of the 43 (53%) were externally validated. This overview provides a basis for head-to-head comparison of existing mortality prediction models through systematic external validation, with the ultimate goal to identify the most suitable prediction model for a certain cohort of patients.

4.2 | Summary of evidence

In previous literature, the maximum number of ICU mortality prediction models reviewed was 12, which is considerably less than the 43 prediction models identified by this review. Where we included all developed prediction models specifically designed to assess mortality, other reviews regarding ICU mortality prediction models focused mainly on commonly used models like the APACHE, SAPS and MPM, or identified models with different outcome than mortality (eg organ dysfunction) or disease- or organ-specific prognostic models. Additionally, only Siontis et al and Strand et al applied a systematic search to identify the models and discussed the validation of the models. Where we included all developed mortality prediction models, Strand et al did only include prediction models when the search for the specific scoring system yielded more than 50 citations. Siontis et al conducted an evaluation of validated tools for hospitalized patients to predict all-cause mortality. However, their analysis included specific patient groups (eg heart or liver patients) rather than general ICU patients as included in the current review.

Model performance is affected by the choice of outcome. Most mortality prediction models used hospital mortality as outcome measure. In general, longer fixed-time outcome measures used in some models are currently recommended. To elaborate, hospital mortality is dependent on discharge practices and availability of post-ICU care, and is therefore a subjective measure. Furthermore, critical illness affects patients after hospital discharge.

The time span during which the mortality prediction models gathered their data varied from short (eg upon ICU admission or during the first initial hour of admission to the ICU) to long (eg during the first 24 hours of admission). Concerning complexity (time consumption) and missing data problems, it may be better in some situations to use a simpler model with less missing data than a more complex model built from a dataset with more missing data which achieves a slightly better performance. Longer collection periods may lead to more complete data, as incompleteness is often substantial for biochemical variables for patients with short-duration admissions (ie less than 24 hours). However, sampling rate affects predictions. This limitation is considered less important in models with shorter data collection. Similarly, the treatments administered during the first 24 hours in the ICU obviously also affect predictions.

4.3 | Comparison of performance

We reported the performance of mortality prediction models in terms of discrimination, calibration and overall performance values. Direct comparison of prediction models predictive performances is not possible, as the development cohorts differed substantially from one another. As a consequence, prediction models cannot be considered interchangeable. Comparisons that are not done head-to-head in external samples independent of all models developed are at high risk of being misleading and may lead to inappropriate conclusions and resource use.

Of 43, 26 (60%) mortality prediction models used the HL goodness-of-fit test for calibration. The HL test is commonly used, despite being frequently non-significant for small data cohorts and nearly always significant for large data cohorts. When only the HL test is reported without any calibration plot or table comparing predicted and observed outcome frequencies, inadequate information regarding calibration is provided. Many ICU mortality prediction models are available and comparatively assessing their performance is a crucial task. In all, 25 articles compared the performance of the new model with existing models but used the same cohort of patients that was used in the development of the ‘novel’ model. This methodology is inherently biased in favor of the ‘novel’ model. Comparisons between prediction models should therefore only be executed in independent external validation samples not used to develop any of the models.

4.4 | Machine-learning algorithms

Mortality prediction models developed as an electronic model or derived from a machine-learning algorithm such as AutoTriage are currently recommended.
were excluded in our manuscript since code and data availability are not requirements in all journals and this is necessary to reproduce the specific prediction model. However, code availability appears to be a rising trend.\textsuperscript{59} Machine-learning-based prediction models seem to achieve increasingly higher accuracies and are becoming more dynamic,\textsuperscript{60} although they still have to include a sufficiently large development and validation cohort to adequately assess performance and the risk of overfitting. However, a recent systematic review concluded that machine learning did not have superior performance over logistic regression for clinical prediction models.\textsuperscript{61} The association between mortality and variables may have changed since the original mortality prediction models were developed, for example, as a result of advancements in diagnostics and therapeutics.\textsuperscript{62} Mortality alone however is rarely the only outcome measure for interventional studies in ICU patients, and many trials, especially in sepsis, include an organ dysfunction score as part of ongoing patient assessment so that effects on morbidity can also be evaluated.\textsuperscript{3} Misuse of mortality prediction models can lead to inappropriate use of resources and potentially even mismanagement of patient care due to incorrect stratification.\textsuperscript{57} Awareness of the differences in model design, the variance of predictions across different ICU settings and the effect of heterogeneity in populations are of utmost importance.

\subsection*{4.5 \textbf{Limitations}}

Some limitations of this study need to be addressed. First, having restricted our search to the period from 2008, relevant mortality prediction models might have been overlooked. Even though some of the most widely used mortality prediction models precede the screening period, we identified 16 prediction models that were published before 2008, but optimally searches have no time limit.\textsuperscript{63} Second, we only included mortality prediction models originally developed for use in the ICU. Mortality prediction models not originally developed for mortality prediction in the ICU could still be valuable clinically. Third, in some original publications, it was unclear whether the presented discrimination, calibration and/or overall performance values were derived from the development cohort or from the validation dataset. We aimed to clarify these, but certain values might reflect another dataset from the original publication. Fourth, we only provided a systematic overview of all developed mortality prediction models in adult critically ill patients. We did not perform a systematic review of every retrieved model complete with all consecutive internal and external validations, as results from different external validations in different cohorts are not directly comparable due to differences in populations, case-mix and settings. We restricted the scope of this review to only identify whether internal or external validation had been performed as a measure of thoroughness of development of the identified models. For this reason, only screening of citations of the original articles was done to identify internal and/or external validation articles. Therefore, we should address that our assessment on mortality prediction models not being internally and/or externally validated might be incomplete if validation in different publications was missed. A systematic search specifically designed for retrieving validation papers is advised when systematically reviewing the internal and external validations of mortality prediction models.\textsuperscript{64}

\subsection*{4.6 \textbf{Unanswered research questions}}

Although we retrieved many developed mortality prediction models that can be used as a step towards future head-to-head comparison, with the results of this scoping review it is not possible to make a recommendation on what mortality prediction models to use and it was not our intention to do so. External validation involving direct head-to-head comparisons in independent cohorts is needed to unravel the comparable performance of individual models. Although we provide a systematic overview of mortality prediction models and describe whether these were internally and/or externally validated, it was not desirable to give an overview of all external validations of the prediction models since this would require a specific search strategy for each model. Moreover, we would have liked to assess risk of bias using the recently developed PROBAST score.\textsuperscript{1} However, this was not feasible because of the number of prediction models.

\section*{5 \textbf{FUTURE PERSPECTIVES}}

To identify the most suitable mortality prediction model for a certain patient cohort, ideally a head-to-head comparison of available models should be performed through systematic external validation using prospectively obtained datasets and appropriate statistical methods. The eventual aim will be to use this review to identify, update and implement the best performing mortality prediction models in daily practice. We are in the process of validating the found prediction models in independent contemporary cohorts to provide external validation of these models. Second, the process should be performed in different cohorts as heterogeneity of ICU patients exists on multiple levels, that is, patient level, hospital level, region and country level.\textsuperscript{65} The best mortality prediction model in one setting is not necessarily the best performing prediction model in another setting. Third, it is worth mentioning that ICU patients have reduced long-term survival and impaired quality of life after ICU discharge compared to the general population.\textsuperscript{66} Future research should also look at determinants of poor outcomes in ICU survivors to help guide long-term follow-up.\textsuperscript{67}

\section*{6 \textbf{CONCLUSIONS}}

In this review, 43 mortality prediction models have been studied. The validation and performance of individual prediction models
differ and the best prediction models for guiding clinical care and research is still to be established.

COMPETING INTERESTS/DISCLOSURES

AG and MHM were involved in the development of one of the mortality prediction models included. RGP reports shares in Evidencio BV, an online platform aiming to facilitate the creation, validation and implementation of clinical prediction models. Evidencio was not involved in the development of any of the prediction models mentioned nor is expecting to be affected financially by publication of this scoping review.

ORCID

Britt E. Keuning [https://orcid.org/0000-0002-6222-7783]
Thomas Kaufmann [https://orcid.org/0000-0003-0589-8879]
Renske Wiersema [https://orcid.org/0000-0002-2413-2852]
Anders Granholm [https://orcid.org/0000-0001-5799-7655]
Ville Pettilä [https://orcid.org/0000-0002-3921-4423]
Morten Hylander Møller [https://orcid.org/0000-0002-6378-9673]
Christian Fynbo Christiansen [https://orcid.org/0000-0002-0727-953X]
José Castela Forte [https://orcid.org/0000-0001-9273-0702]
Harold Snieder [https://orcid.org/0000-0003-1949-2298]
Frederik Keus [https://orcid.org/0000-0003-1516-1475]
Rick G. Pleijhuis [https://orcid.org/0000-0003-4856-019X]
Iwan C. C. van der Horst [https://orcid.org/0000-0003-3891-8522]

REFERENCES

1. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med. 2019;170:1-33.
2. Lemeshow S, Le J-R. Modeling the severity of illness of ICU patients: a systems update. JAMA. 1994;272:1049-1055.
3. Vincent J-L, Moreno R. Clinical review: scoring systems in the critically ill. Crit Care Med. 2010;40:207. https://doi.org/10.1161/cc28204
4. Bouch C, Thompson J. Severity scoring systems in the critically ill. Contin Educ Anaesth Crit Care Pain. 2008;8:181-185.
5. Strand K, Flaatten H. Severity scoring in the ICU: a review. Acta Anaesthesiol Scand. 2008;52:467-478 https://doi.org/10.1111/j.1399-6576.2008.01586.x
6. Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K. A new risk prediction model for critical care: The Intensive Care National Audit & Research Centre (ICNARC) model*. Crit Care Med. 2007;35:1091-1098.
7. Rapsang AG, Shyam DC. Scoring systems in the intensive care unit: a compendium. Indian J Crit Care Med. 2014;18:220-228.
8. Siontis GCM, Tzoulaki I, Ioannidis JPA. Predicting death: an empirical evaluation of predictive tools for mortality. Arch Intern Med. 2011;171:1721-1726.
9. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation The PRISMA-ScR Statement. Ann Intern. Med. 2018;169:467-473.
10. Riviello ED, Kiviri W, Fowler RA, et al. Predicting mortality in low-income country ICUs: The Rwanda Mortality Probability Model (R-MPM). PLoS ONE. 2016;11:e0155858.
11. Adhikari NKJ, Fowler RA, Bhagwanjee S, Rubenfeld GD. Critical care and the global burden of critical illness in adults. The Lancet. 2010;376:1339-1346.
12. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21:128-138.
13. Ferrando-Vivas P, Jones A, Rowan KM, Harrison DA. Development and validation of the new ICNARC model for prediction of acute hospital mortality in adult critical care. J Crit Care. 2017;38:335-339.
14. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today’s critically ill patients*. Crit Care Med. 2006;34:1297-1310.
15. Moreno RP, Metnitz P, Almeida E, et al. SAPS 3–From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med. 2005;31:1345-1355.
16. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. JAMA. 1993;270:2478-2486.
17. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA. 1993;270:2957-2963.
18. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Study to understand prognoses and preferences for outcomes and risks of treatments. Ann Intern Med. 1995;122:191-203.
19. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13:818-829.
20. Knaus WA, Harrell FEJ, Lynn J, et al. The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults. Study to understand prognoses and preferences for outcomes and risks of treatments. Ann Intern Med. 1996;125:943-950.
21. Lemeshow S, Klar J, Teres D, et al. Mortality probability models for patients in the intensive care unit for 48 or 72 hours: a prospective, multicenter study. Crit Care Med. 1994;22:1351-1358.
22. Timsit JF, Fosse JP, Troche G, et al. Accuracy of a composite score using daily SAPS II and LOD scores for predicting hospital mortality in ICU patients hospitalized for more than 72 h. Intensive Care Med. 2001;27:1012-1021.
23. Dólöra-Moreno C, Palazón-Bru A, Colomina-Climent F, Gil-Guillén VF. Construction and internal validation of a new mortality risk score for patients admitted to the intensive care unit. Int J Clin Pract. 2016;70:916-922.
24. Li Z, Cheng B, Wang J, et al. A multifactor model for predicting mortality in critically ill patients: a multicenter prospective cohort study. J Crit Care. 2017;42:18-24.
25. Hadique S, Culp S, Sangani RG, et al. Derivation and validation of a prognostic model to predict 6-month mortality in an intensive care unit population. Am J Respir Crit Care Med. 2017;196:1556-1561.
26. Fika S, Nanas S, Baltopoulos G, Charitidou E, Myrianthefs P. A novel mortality prediction model for the current population in an adult intensive care unit. Heart Lung. 2018;47:10-15.
27. Nematifard E, Ardehali SH, Shahbazi S, Eini-Zinab H, Vahdat A. New risk prediction model for critical care: The Intensive Care National Audit & Research Centre (ICNARC) model*. Crit Care Med. 2007;35:1091-1098.
28. Riviello ED, Kiviri W, Fowler RA, et al. Predicting mortality in low-income country ICUs: The Rwanda Mortality Probability Model (R-MPM). PLoS ONE. 2016;11:e0155858.
29. Min H, Avramovic S, Wojtusiak J, et al. A comprehensive multimorbidity index for predicting mortality in intensive care unit patients. J Palliat Med. 2017;20:35-41.
30. Paul E, Bailey M, Pilcher D. Risk prediction of hospital mortality for adult patients admitted to Australian and New Zealand intensive care units: development and validation of the Australian and New Zealand Risk of Death model. J Crit Care. 2013;28(6):935-941. https://doi.org/10.1016/j.jcrc.2013.07.058
31. Brinkman S, Abu-Hanna A, de Jonge E, de Keizer NF. Prediction of long-term mortality in ICU patients: model validation and assessing the effect of using in-hospital versus long-term mortality on benchmarking. Intensive Care Med. 2013;39:1925-1931.
32. Higgins TL, Teres D, Copes W, Nathanson B, Stark M, Kramer R. Lee Institute for Health Policy Studies. Summary of NQF-46. Duke GJ, Santamaria J, Shann F, et al. Critical care outcome prediction and validation of the Simplified Mortality Score for risk prediction in adult patients based on routine laboratory parameters. J Intens Care. 2016;10:41. https://doi.org/10.1097/CCM.0b013e3182a24fe
33. Duke GJ, Barker A, Rasekaba T, Hutchinson A, Santamaria JD. Development and validation of the critical care outcome prediction equation, version 4. Crit Care Resusc. 2013;15:191-197.
34. Ho KM, Knuiman M, Finn J, Webb SA. Estimating long-term survival of critically ill patients: the PREDICT model. PLoS ONE. 2008;3:e3226.
35. Kao R, Priestap F, Donner A. To develop a regional ICU mortality prediction model during the first 24 h of ICU admission utilizing MODS and NEMS with six other independent variables from the Critical. Care Information System (CCIS) Ontario, Canada. J Intens Care. 2016;4:16.
36. Umegaki T, Sekimoto M, Hayashida K, Imanaka Y. An outcome prediction model for adult intensive care. Crit Care Resusc. 2010;12:96-103.
37. Huang YC, Chang KY, Lin SP, et al. Development of a daily mortality probability prediction model for Intensive Care Unit patients using a discrete-time event history analysis. Comput Methods Programs Biomed. 2013;111:280-289.
38. Stachon A, Segbers E, Hering S, Kempf R, Holland-Letz T, Krieg M. A laboratory-based risk score for medical intensive care patients. Clin Chem Lab Med. 2008;46:855-862.
39. Viviani X, Gouvertent J, Granthil C, Francois G. Simplification of the SAPS by selecting independent variables. Intensive Care Med. 1991;17:164-168.
40. Stachon A, Becker A, Holland-Letz T, Friese J, Kempf R, Krieg M. Estimation of the mortality risk of surgical intensive care patients based on routine laboratory parameters. Eur Surg Res. 2008;40:263-272.
41. Goag EK, Lee JW, Roh YH, et al. A simplified mortality score using delta neutrophil index and the thrombotic microangiopathy score for prognostication in critically ill patients. Shock. 2018;49:39-43.
42. Hunziker S, Cell LA, Lee J, Howell MD. Red cell distribution width improves the simplified acute physiology score for risk prediction in unselected critically ill patients. Crit Care. 2012;16: https://doi.org/10.1186/cc11351
43. Iapichino G, Miraletti G, Cordella D, et al. Scoring system for the selection of high-risk patients in the intensive care unit. Crit Care. 2006;34:1039-1043.
44. Liu W-Y, Lin S-G, Zhu G-Q, et al. Establishment and validation of the Simplified Mortality Score for risk prediction in critically ill patients. PLoS ONE. 2016;11:e0166085 https://doi.org/10.1371/journal.pone.0166085
45. Granholm A, Perner A, Krag M, et al. Development and internal validation of the Simplified Mortality Score for the Intensive Care Unit (SAPS-ICU). Acta Anaesthesiol Scand. 2018;62:336-346.
46. Mahato B, Otero TMN, Holland CA, et al. Addition of 25-hydroxyvitamin D levels to the Deyo-Charlson Comorbidity Index improves 90-day mortality prediction in critically ill patients. J Intens Care. 2016;4:40 https://doi.org/10.1186/s40560-016-0165-0
47. Rydenfelt K, Engerström L, Walther S, Sjöberg F, Strömberg U, Samuelsson C. In-hospital vs. 30-day mortality in the critically ill - a 2-year Swedish intensive care cohort analysis. Acta Anaesthesiol Scand. 2015;59:846-858.
48. Jammer I, Wickboldt N, Sander M, et al. Standards for definitions and use of outcome measures for clinical effectiveness research in perioperative medicine: European Perioperative Clinical Outcome (EPOC) definitions: a statement from the ESA-ESICM joint taskforce on perioperative outcome measures. Eur J Anaesthesiol. 2015;32:88-105.
49. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. BMJ. 2009;338: https://doi.org/10.1136/bmj.b604.
50. Suistomaa M, Kari A, Ruokonen E, Takala J. Sampling rate causes bias in APACHE II and SAPS II scores. Intensive Care Med. 2000;26:1773-1778.
51. Labaree J, Renaud B, Fine MJ. How to derive and validate critical prediction models for use in intensive care medicine. Intensive Care Med. 2014;40:513-527.
52. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pennica MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2014;74:167-176.
53. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York, NY: Springer; 2009.
54. Meiring C, Dixit A, Harris S, et al. Optimal intensive care outcome prediction over time using machine learning. PLoS ONE. 2018;13:e020682.
55. Shickel B, Loftus TJ, Adhikari L, Ozrazgat-Baslanti T, Bihorac A, Rashidi P. DeepSOFA: a continuous acuity score for critically ill patients using clinically interpretable deep learning. Sci Rep. 2019;9:1879.
56. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for critical prediction models. J Clin Epidemiol. 2019;110:12-22.
57. Poncet A, Perneger TV, Merleau P, Capuzzo M, Combesure C. Determinants of the calibration of SAPS II and SAPS 3 mortality scores in intensive care: a European multicenter study. Crit Care. 2017;21:85.
58. Kitchenham B. Procedures for Performing Systematic Reviews. Keele, UK: Keele University. 2004;33:1-26.
59. Geersing G-J, Bouwmeester W, Zuiithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic
prediction studies in medline to enhance systematic reviews. PLoS ONE. 2012;7:e32844.

65. Prin M, Wunsch H. International comparisons of intensive care: informing outcomes and improving standards. Curr Opin Crit Care. 2012;18:700-706.

66. Winters BD, Eberlein M, Leung J, Needham DM, Pronovost PJ, Sevransky JE. Long-term mortality and quality of life in sepsis: a systematic review. Crit Care Med. 2010;38:1276-1283.

67. Gayat E, Cariou A, Deye N, et al. Determinants of long-term outcome in ICU survivors: results from the FROG-ICU study. Crit Care. 2018;22:8.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Keuning BE, Kaufmann T, Wiersema R, et al; HEALICS consortium. Mortality prediction models in the adult critically ill: A scoping review. Acta Anaesthesiol Scand. 2020;64:424–442. [https://doi.org/10.1111/aas.13527](https://doi.org/10.1111/aas.13527)