Which Models Can I Use to Predict Adult ICU Length of Stay? A Systematic Review*

Ilona Willempje Maria Verburg, MSc[1]; Alireza Atashi, MSc[2,3]; Saeid Eslami, PhD[1,4,5]; Rebecca Holman, PhD[1,6]; Ameen Abu-Hanna, PhD[1]; Everet de Jonge, MD[7]; Niels Peek, PhD[8]; Nicolette Fransisca de Keizer, PhD[1]

Objective: We systematically reviewed models to predict adult ICU length of stay.

Data Sources: We searched the Ovid EMBASE and MEDLINE databases for studies on the development or validation of ICU length of stay prediction models.

Study Selection: We identified 11 studies describing the development of 31 prediction models and three describing external validation of one of these models.

Data Extraction: Clinicians use ICU length of stay predictions for planning ICU capacity, identifying unexpectedly long ICU length of stay, and benchmarking ICUs. We required the model variables to have been published and for the models to be free of organizational characteristics and to produce accurate predictions, as assessed by R^2 across patients for planning and identifying unexpectedly long ICU length of stay and across ICUs for benchmarking, with low calibration bias. We assessed the reporting quality using the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies.

Data Synthesis: The number of admissions ranged from 253 to 178,503. Median ICU length of stay was between 2 and 6.9 days. Two studies had not published model variables and three included organizational characteristics. None of the models produced predictions with low bias. The R^2 was 0.05–0.28 across patients and 0.01–0.64 across ICUs. The reporting scores ranged from 49 of 78 to 60 of 78 and the methodologic scores from 12 of 22 to 16 of 22.

Conclusion: No models completely satisfy our requirements for planning, identifying unexpectedly long ICU length of stay, or for benchmarking purposes. Physicians using these models to predict ICU length of stay should interpret them with reservation. (*Crit Care Med* 2017; 45:e222–e231)

Key Words: benchmarking; intensive care units; length of stay; prediction; review

*See also p. 379.

[1]Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

[2]Student Research Committee, Mashhad University of Medical Sciences, Mashhad, Iran.

[3]Cancer Informatics Department, Breast Cancer Research Center, ACECR, Tehran, Iran.

[4]Pharmaceutical Research Center, School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran.

[5]Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.

[6]Clinical Research Unit, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

[7]Department of Intensive Care, Leiden University Medical Center, Leiden, The Netherlands.

[8]Health e-Research Centre, Division of Imaging, Informatics and Data Sciences, School of Health Sciences, The University of Manchester, Manchester, United Kingdom.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccmjournal).

Dr. Verburg received grant support (The National Intensive Care Evaluation [NICE] foundation pays the department of Medical Informatics for maintaining the national database, providing feedback reports, and doing analyses. She is an employee of the department of Medical Informatics). Her institution received support for participation in review activities. Dr. Holman's institution received grant support (The NICE foundation pays the department of Medical Informatics for maintaining the national database, providing feedback reports, and doing analyses. She is an employee of the department of Medical Informatics) and received support for participation in review activities. She disclosed employment (She is also employed by the Clinical Research Unit of the Academic Medical Center, Amsterdam, The Netherlands). Dr. de Keizer's institution received grant support from the NICE foundation (The department of medical informatics receive funding for the maintenance of the Dutch NICE registry. This literature review is related to a project on benchmarking length of stay), served as a board member for the NICE foundation (She is a member of the NICE board), and is employed by the NICE foundation. The remaining authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: i.w.verburg@amc.uva.nl

DOI: 10.1097/CCM.0000000000002054

ICUs provide complex and expensive care and hospitals face pressure to improve efficiency and reduce costs (1, 2). Since costs are strongly related to ICU length of stay (LoS), shorter ICU LoS generally equates to lower costs (2, 3). Hence, models predicting ICU LoS can play an important role in examining the efficiency of ICU care. We identified three main reasons for clinicians to predict ICU LoS (4, 5): 1) planning the number of beds and members of staff required to fulfill demand for ICU care within a given hospital or geographic area; 2) identifying individual patients or groups of patients with unexpectedly long

ICU LoS to drive direct quality improvement; and 3) enabling case-mix correction when comparing average LoS between ICUs (benchmarking). The requirements of an ICU LoS prediction model differ between these situations. A model for planning purposes or to identify individuals or groups with unexpectedly long ICU LoS needs to predict ICU LoS reliably for individual patients. When benchmarking the quality or efficiency of ICU care and benchmark reports are based on summary measures of differences between expected and observed LoS (6), a prediction model needs to predict total ICU LoS accurately across ICUs.

A range of models to predict case-mix–adjusted ICU LoS have been published. However, their clinical utility is unclear (6, 7), and there is no consensus on which is the best (8–10). Predicting ICU LoS accurately is difficult for three reasons. First, statistical methods often assume a Gaussian distribution, but ICU LoS is generally right skewed (11). Second, patients admitted to an ICU form a heterogeneous group with a wide range of complex health issues, each of which may have a different association with ICU LoS (7). Third, the association between severity of illness and ICU LoS differs for ICU survivors and ICU nonsurvivors (11). If not correctly addressed, these points could lead to wildly inaccurate or biased predictions of ICU LoS, thus negating their utility (11).

In this article, we systematically review reporting and methodologic quality of models for predicting ICU LoS and assess their suitability for planning ICU resources, identifying unexpectedly long ICU LoS, and benchmarking. We examine characteristics most relevant to clinicians assessing the suitability of a published model to predict ICU LoS in their own hospital or group of hospitals.

TABLE 1. Country and Period of Data Collection, Outcome, Number of Predictors, and Sample Size for Model Development Studies

Reference	Data Collection Country	Data Collection Period	No. of ICUs Included	Percentage of Patients Excluded, %	No. of Included Admissions	Mean (Median) ICU Length of Stay (Days)	No. of Variables Estimated
Clermont et al (25)	11 European countries (ESICM)[a]	1995	49	–	989	–	12
Perez et al (22)	Colombia	1997 to 1998	20	6	1,528	20 (13)[b]	1
Zimmerman et al (8)	USA	2002 to 2003	104	12	69,652	3.9 (2.0)	131
Rothen et al (19)	35 countries (ESICM)[c]	2002	275	–	16,560	2.0	1
Moran et al (23)	Australia and New Zealand	1993 to 2003	99	12	178,503	3.6 (2.9)	79
Moran and Solomon (7), model 1–12	Australia and New Zealand	2008 to 2009	131	4	89,330	3.6 (2.9)	26
Niskanen et al (6), model 1–2	Finland	2000 to 2005	23	22	37,718	3.9 (2.0)	9
Vasilevskis et al (10), model 2[d]	USA	2001 to 2004	35	–	6,684	4.0 (2.0)	24
Vasilevskis et al (10), model 3	USA	2001 to 2004	35	–	6,684	4.0 (2.0)	2
Kramer and Zimmerman (24)	USA	2002 to 2007	83	10	12,640	4.2 (2.14)[e]	106
Al Tehewy et al (26)	Egypt	2004 to 2005	3	42	253	5.1 (4)	1
Verburg et al (11), model 1–8	Netherlands	2011	83	1[f]	32,667	4.2 (1.7)	151

ESICM = European Society of Intensive Care Medicine.

[a]Eleven unspecified countries participating in the European Society of Intensive Care Medicine.

[b]Hospital days after first day of ICU admission.

[c]List of 35 countries (27): Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Cuba, Czech Republic, Denmark, France, Germany, Greece, Hong Kong, Hungary, India, Ireland, Israel, Italy, Luxembourg, Mexico, Netherlands, Norway, Poland, Portugal, Russian Federation, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States.

[d]Vasilevskis et al (10): model 1: second-order recalibration of the Acute Physiology and Chronic Health Evaluation (APACHE) IV model for length of stay; model 2: covariates included from Mortality Probability Models 0-III; model 3: covariates included from simplified.

[e]Mean predicted remaining ICU stay after day 5 was 6.87 for the subgroup of admissions with ICU length of stay longer than 5 d.

[f]Percentage of excluded admissions after using the APACHE IV inclusion criteria.

Dashes indicate the result was not reported. The order of the studies in this table is based on "study period."

TABLE 2. Summary of Patient Exclusion Criteria for Model Development Studies and Predictor Variables Included in the Models, for Development Studies

	Patient Exclusion Criteria								
Reference	Based on Outcome	Based on Age	ICU Readmissions	Transfers From Another ICU	Transfers to Another ICU	Survival Status	Burns	Cardiac Population Surgery	Other[a]
Clermont et al (25)	x								
Perez et al (22)		x	x						
Zimmerman et al (8)	x	x	x	x	x		x	x	x
Rothen et al (19)			x						
Moran et al (23)	x	x	x						x
Moran and Solomon (7), model 1–12	x	x	x						x
Niskanen et al (6), model 1[d]	x	x		x				x	x
Niskanen et al (6), model 2									
Vasilevskis et al (10), model 2[e]									
Vasilevskis et al (10), model 3									
Kramer and Zimmerman (24)	x	x		x			x	x	x
Al Tehewy et al (26)		x				x	x	x	x
Verburg et al (11), model 1–8	x	x	x	x	x		x	x	x

[a]Other subgroups consist of as severity of illness, admission type, admission diagnose, unknown discharge location or date, trauma, dialysis, and unknown Glasgow Coma Score.

[b]Overall severity of illness consist of Acute Physiology Score (APS), Acute Physiology and Chronic Health Evaluation (APACHE) II score, APACHE III score, Simplified APS (SAPS) II score, SAPS III score, SAPS probability of mortality, and Mortality Probability Models (MPM).

[c]Interaction with age (coma, systolic blood pressure, cirrhosis, metastatic neoplasm, cardiac dysrhythmia, intracranial mass effect, and cardiopulmonary resuscitation before ICU), APACHE III score, SAPS II score, Therapeutic Intervention Scoring System score, mechanical ventilation, gender, calendar year, hospital type, yearly number of admissions, and ICU death.

[d]Niskanen et al (6): model 1: outcome measure truncated at 30 d; model 2: log-transformed outcome measure used.

[e]Vasilevskis et al (10): model 1: second-order recalibration of the APACHE IV model for length of stay; model 2: covariates included from MPM0-III; model 3: covariates included from simplified.

The order of the studies in this table is based on "study period." 'x' indicates patient exclusion criterium is used for this study or predictor is considered for inclusion for this study.

METHODS

Search Strategy and Inclusion and Exclusion Criteria

We searched the Ovid EMBASE and Ovid MEDLINE databases from database inception until October 31, 2014, by searching all fields and including citations in progress, which are not indexed with MeSH headings. The search query consisted of three subqueries, with synonyms and combined with "and," on intensive care, LoS, and prediction. We present the detailed search strategy in **Table S1** (Supplemental Digital Content 1, http://links.lww.com/CCM/C203). We included all original articles describing the development and/or validation of a prediction model for ICU LoS in adult patients. We excluded duplicate studies, articles not written in English and studies that were later updated by the same research group.

When deciding whether to include an article, the authors (I.W.M.V., A.A.) classified the articles by reading the title and abstract, then compared and discussed their results until they reached consensus about the eligibility of the article based on the title and abstract. Dr. Verburg manually reviewed the references in articles found. The authors (I.W.M.V., A.A.) read the full text of the included articles and independently scored the items for each model. If necessary, disagreement in all steps was reconciled by discussion with other authors (S.E., N.F.d.K.).

Assessment of Methodologic and Reporting Quality

We used the consensus-based Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (12) to assess the quality of reporting in the studies. This checklist was based on previously published reporting guidelines, systematic reviews, methodologic literature, and pilot versions discussed within the Cochrane Prognosis Methods Groups. Although the checklist is relatively new, it has already been used in scientific reviews (13–15). We extended the checklist with three items of importance in prediction models (16) and four items on how the prediction model handled four specific subgroups of ICU patients: 1) patients readmitted to the ICU; 2) patients transferred from or to another ICU; 3) patients who survived their ICU stay versus patients who died on the ICU; and 4) patients who underwent cardiac surgery. Although the checklist presents items to

Predictors Considered for Inclusion in the Model										
Overall Severity of Illness [b]	Admission Source	Age	Mechanical Ventilation	Glasgow Coma Score	Comorbidities	Length of Stay Hospital Before ICU Admission	Organizational Predictors	Locale (Country) Specific	Interaction Terms[c]	Other Predictors
x		x							x	x
	x									
x	x	x			x	x				
x										
x		x	x				x	x	x	x
							x	x		
x		x					x		x	x
x		x					x		x	x
		x	x		x				x	x
x										
x	x	x	x	x	x	x				x
x										
x	x	x	x	x	x					x

report on, no score and no weight per item were reported. To circumvent this, we assess the methodologic quality of prediction models using 11 items that we considered as important, based on literature (12, 16). We assigned item scores of zero (not), one (partly), or two (yes) points and calculated a total score for reporting and methodologic quality. There were 39 items in the reporting score for a total of 78 possible points, and 11 items in the methodologic score for a total of 22 possible points. All items are presented in **Table S2** (Supplemental Digital Content 2, http://links.lww.com/CCM/C204).

Definition of Utility in Predicting ICU LoS

In order to assess the models, we defined four requirements on prediction models in advance. The first three were applicable to models for all purposes. We regarded a model as suitable if: 1) the variables required to predict ICU LoS have been published; 2) it does not include any organizational characteristics; and 3) it has a low level of bias, demonstrate by at least "moderate" calibration (17). Moderate calibration is achieved if the mean observed values equal the mean predicted values for groups of patients with similar predictions (17). Moderate calibration is important when benchmarking to avoid distortion of benchmarks if mean predicted ICU LoS differs greatly between hospitals. Our fourth requirement was that a model produces accurate predictions. Our definition of "accurate predictions" is different for models used to plan resources or identify patients with unexpectedly long ICU LoS and for models used for case-mix correction in benchmarking. We regard a model as suitable for planning resources or identifying patients with unexpectedly long ICU LoS if it has an R^2 of at least 0.36 (strong) across patients. We regard a model as suitable for case-mix correction in benchmarking if it has an R^2 of at least 0.36 (strong) across ICUs (18).

RESULTS

We identified 3,818 unique articles in EMBASE and MEDLINE and selected 25 (0.7%) based on the title and abstract. Following inspection of the full text, we included 13 studies (52.0%) and identified one additional study from the references of previously included articles (19). We excluded three studies (9, 20, 21) because a more recent version of the model was available (8). Hence, we included 11 studies describing the development of 31 models (6–8, 10, 11, 19, 22–26). We present the inclusion process in **Figure S1** (Supplemental Digital Content 3, http://links.lww.com/CCM/C205; legend: flowchart representing the result of search and the number of articles excluded and eligible for review). Most studies presented one model, but four (6, 7, 10, 11) each presented between two and 12 models. Three studies (10, 11, 24) described the external validation and one a second-order recalibration (10) of the Acute Physiology And Chronic Health Evaluation (APACHE) IV (8) model for ICU LoS.

We present information on geographic location, time period of data collection, observed mean and median ICU LoS, number of variables estimated of ICUs, and included patients in **Table 1**. Data were collected between 1995 (25) and 2011 (11) over periods of 1 month (25) to 11 years (23) in Europe (6, 11, 25), United States (8, 10, 24), Southern America (22), Egypt (26), and worldwide (19). The data come from 253 (26) to 178,503 (23) admissions to three (26) to 275 (19) ICUs. The number of variables estimated ranged from one (19, 22) to 151 (11) and the average number of admissions per variable estimated from 82 (25) to 16,560 (19).

We present patient exclusion criteria and candidate predictors considered for inclusion in **Table 2**. Studies used exclusion based on observed value of ICU LoS, such as incomplete or unknown ICU LoS, ICU LoS less than 4 (7, 8, 10, 11, 20, 21, 23, 24) or 6 hours (6), and ICU LoS greater than 48 hours (25) or 60 days (7, 23). Three studies included ICU organizational characteristics, such as geographic location, ICU level, number of beds, or teaching hospital status (6, 7, 23).

In **Table 3**, we present how researchers handled patients, who were readmitted to the ICU, transferred between ICUs, underwent cardiac surgery, or died before ICU discharge. The researchers' strategies for handling readmissions were as follows: 1) including only a patient's first admission to the ICU (7, 8, 10, 11, 19, 23); 2) including readmission as predictor in the model (24); 3) defining ICU LoS as the sum of all ICU LoS of a patient's ICU admission (22) and; 4) including readmissions as separate data records (6). Three studies excluded patients based on transfer to (11, 24) or from another ICU (8, 11, 24) and six excluded one or more groups of cardiac surgery patients (6, 8, 10, 11, 24, 26). Researchers handled patients, whose ICU stay ended in death by: including death as a predictor (6, 7, 23), excluding patients dying within an hour of ICU admission (26), or developing different models for ICU survivors and nonsurvivors (11). Two studies related ICU LoS to in-hospital mortality (6, 19), presenting expected and observed ICU LoS separately for survivors and nonsurvivors (8, 24). Five studies neglected mortality in their model.

We present results on model evaluation and performance reported by the original authors in **Table 4**. The authors used a random sample from the development dataset of (6–8, 10, 22–24) bootstrap methods (7) or data from different time periods (26) or ICUs (25). The total number of admissions for validation ranged from 460 (25) to 46,517 (8) and the average number per variable from 35 (10) to 2,843 (6). The Pearson correlation coefficients ranged from 0.05 (10, 26) to 0.28 (6) across patients and from 0.01 (10) to 0.62 (8) across ICUs. Differences between the mean observed and mean predicted ICU LoS ranged between 0.01 and 4.7 days and were not statistically significant for seven models. Two studies test these differences for subgroups of the covariates included in their model (10, 28). Four studies presented a calibration curve (6, 8, 10, 24) but none regression coefficients.

Table 5 described the suitability of models according to our predefined requirements. No models met all of our requirements for models for planning the number of beds and members of staff required or identifying individual patients or groups of patients with unexpectedly long ICU LoS to drive direct quality improvement. The APACHE IV model (8) and a second-order recalibration with updated model variables of this model (10) fulfilled most of our requirements for models for benchmarking (Table S3, Supplemental Digital Content 4, http://links.lww.com/CCM/C206). However, the requirement for moderate calibration was not fulfilled.

We present external validation studies on the APACHE IV model in Table S3 (Supplemental Digital Content 4, http://links.lww.com/CCM/C206). The number of included admissions ranged between 4,611 and 32,667 admissions to between 35 and 83 ICUs in the United States and The Netherlands between 2001 and 2011. The average number of patients per variable was between 35 and 249. The squared Pearson Coefficient across patients was moderate (0.16–0.18) and strong (0.43–0.44) across ICUs. The difference in days between the observed and predicted mean LoS was larger than for the internal validation of the model.

We present the scores assigned to the methodologic quality items in Table S2 (Supplemental Digital Content 2, http://links.lww.com/CCM/C204). The overall reporting quality scores ranged from 49 (26) to 60 (6) and median score of 55. The overall methodologic quality scores ranged from 12 (7, 19, 23, 25) to 16 (22) points and a median score of 14. Further items extracted from each study are described in **Table S4** (Supplemental Digital Content 5, http://links.lww.com/CCM/C207).

DISCUSSION

In this systematic review, we focussed on the utility of models for predicting ICU LoS and assessed their suitability for planning ICU resources, identifying unexpectedly long ICU LoS, and benchmarking ICUs. We included eleven studies on model development and three studies externally validating the APACHE IV model (8). We concluded that no models fulfilled all of our requirements for planning ICU resources or identifying patients with unexpectedly long ICU LoS. The original (8) and a second-order recalibration of the (10) APACHE IV model fulfilled most of our requirements for benchmarking. However, these models did not fulfill our requirement for moderate calibration. As no models fulfilled our requirements, physicians choosing to use them to predict ICU LoS should interpret the predications with reservation. Benchmarking incorrect predictions can have large consequences, especially when benchmarking results are published and those without specialist statistical

TABLE 3. The Handling of Subgroups of Patients in Each of the Models for Predicting ICU Length of Stay

Reference	Treatment of Patients Readmitted to the ICU	Exclusion of Patients Transferred Between ICUs	Excluding Patients Admitted to the ICU Following Cardiac Surgery[a]	Patients, Who Died Before ICU Discharge
Clermont et al (25)	–	–	–	–
Perez et al (22)	Summed over[b]	–	–	Death as one of the states
Zimmerman et al (8)	Excluded	From another ICU	CABG	–
Rothen et al (19)	Excluded	–	–	–
Moran et al (23)	Excluded	–	–	Predictor: death
Moran and Solomon (7), model 1–12	Excluded	–	–	Predictor: death
Niskanen et al (6), model 1–2	Included	–	Open heart	Predictor: death
Vasilevskis et al (10), model 2[c]	Excluded	–	CABG	–
Vasilevskis et al (10), model 3	–	–	CABG	–
Kramer and Zimmerman (24)	Predicted	To and from another ICU	CABG	–
Al Tehewy et al (26)	–	–	CABG; cardiac valve or heart transplant	Excluded: deaths within first hour; deaths cardiopulmonary arrest within 4 hr
Verburg et al (11), model 1–8	Excluded	To and from another ICU	CABG; all elective	Model for whole group and separate models for survivors and nonsurvivors

CABG = coronary artery bypass grafting.

[a]The sum of ICU length of stay of all admissions of a patient.

[b]Vasilevskis et al (10): model 1: second-order recalibration of the Acute Physiology and Chronic Health Evaluation IV model for length of stay; model 2: covariates included from Mortality Probability Models 0-III; model 3: covariates included from simplified.

[c]Coronary artery bypass grafting; elective = elective surgery; open heart, cardiac valve, heart transplant all surgical admission type.

The order of the studies in this table is based on “study period.” Dashes indicate the result was not reported.

knowledge use them to judge hospitals. As patient characteristics in individual hospitals over time may remain more similar than between hospitals, benchmarking ICU LoS to an historical benchmark may be acceptable with these models. In addition, as healthcare and hospital policies differ between countries and over time, we recommend validating a model using recent local data before using it to predict individual or to benchmark hospitals on ICU LoS.

Four main aspects of reporting and methodologic quality in studies reporting on the development of prediction models for ICU LoS could be improved. The first is the exclusion of patients based on observed ICU LoS. Excluding a few extreme values might enable researchers to obtain a model with a reasonable fit for the majority of patients. However, suboptimal patient care could lead to prolonged ICU LoS and, when benchmarking, truncating ICU LoS or excluding patients with prolonged ICU LoS can lead to biased performance results (29). The second is the handling of ICU nonsurvivors. The association between severity of illness and ICU LoS differs for ICU survivors and nonsurvivors (11). It seem sensible to develop different models for these two groups. However, higher mortality rates could lead to shorter average ICU LoS, but it is undesirable to reduce ICU LoS at the cost of increasing mortality. Third, for benchmarking, composite indicators incorporating information on ICU LoS and mortality may be preferable to LoS as a single outcome measure. Fourth, some researchers used ordinary linear regression to predict ICU LoS. This can lead to predictions of ICU LoS which are negative and, as such, conceptually incorrect (11). Logarithmic or other transformations of ICU LoS or other regression models can overcome these problems.

Our study has two main strengths over previous reviews of prediction models for ICU LoS. First, this is the first systematic review of ICU LoS prediction models for the general ICU population. Second, we systematically assessed the studies in light of three clinical applications and using established frameworks (12, 16, 30, 31). Previously, researchers compared the APACHE IV (8), mortality probability (32), and Simplified Acute Physiology Score (33) models and found that the APACHE IV scoring system is most frequently used to predict ICU LoS (8). A systematic review in cardiac surgery patients identified several models for unexpectedly long ICU LoS (34), but we found no studies which defined ICU LoS in this way for the general ICU population. This may be because cardiac surgery is

TABLE 4. Validation of Prediction Model Performance by the Original Authors

	Model Validation		Model Performance			
Reference	Validation Method	No. of Patients Included in Validation Set	R^2 Across ICU (Validation Set)	R^2 Across Patients (Validation Set)	Difference Mean Observed and Mean Predicted Length of Stay (Validation Set) in Days (Bias)	Recalibration Plot Presented
Clermont et al (25)	12 other ICUs	460	–	–	0.50[a]	No
Perez et al (22)	Random sample (pps) of 50% stratified by Intensive Care National Audit and Research Centre coding method code	1,531	–	–	0.01–6.69[b,a]	No
Zimmerman et al (8)	Simple random sample of 40%	46,517	0.62	0.22	0.08[a]	Yes[c]
Rothen et al (19)	–	–	–	–	0.40[d]	No
Moran et al (23)	Random sample (pps) of 20% stratified by year of admission	44,625	–	0.18	–[e]	No
Moran and Solomon (7), model 1–12	Random sample (pps) of 20% stratified by year of admission	22,333	–	0.18–0.20	0.2–4.7	Yes
Niskanen et al (6), model 1[f]	Simple random sample of 40%	25,586	0.57	0.27	0.01[a]	Yes[c]
Niskanen et al (6), model 2	Simple random sample of 40%	25,586	0.64	0.28	0.76	Yes
Vasilevskis et al (10), model 2[g]	Simple random sample of 40%	4,611	0.28	0.10	0.01[a]	Yes[c]
Vasilevskis et al (10), model 3	Simple random sample of 40%	4,611	0.01	0.05	0.02[a]	Yes
Kramer and Zimmerman (24)	Simple random sample of 50% and different time period	12,904	0.43[d]	0.18[d]	0.02[a] and 0.61[h]	Yes
Al Tehewy et al (26)	Two ICUs and different time period	–	–	0.05	–	No
Verburg et al (11), model 1–8	Bootstrap (100×)	32,667	–	0.09–0.15	–	Yes

pps = probability proportional to size sampling.

[a]Difference tested as not significant different from 0, if tested by *t* test, Mann-Whitney *U* test or chi-square goodness-of-fit test. Used test and reported *p* value are presented in Supplemental **Table S3**, Supplemental Digital Content 4, http://links.lww.com/CCM/C206.

[b]Presented per day, per system: 0.01 (1 d) to 6.69 (30 d), not significant for three body anatomic systems.

[c]Recalibration reported as accurate, good, or prefect. Reported conclusion presented in Supplemental Table S3, Supplemental Digital Content 4, http://links.lww.com/CCM/C206.

[d]Performance was based on the development set, and presented as a median.

[e]**Figure 1**, presents raw and predicted mortality per year.

[f]Niskanen et al (6): model 1: outcome measure truncated at 30 d; model 2: log-transformed outcome measure used.

[g]Vasilevskis et al (10): model 1: second-order recalibration of the Acute Physiology and Chronic Health Evaluation IV model for length of stay; model 2: covariates included from Mortality Probability Models 0-III; model 3: covariates included from simplified.

[h]Performance for this study reported for included subgroup of admissions with ICU length of stay > 5 d. Difference was 0.02 for internal validation and 0.61 for external validation.

The order of the studies in this table is based on "study period." Dashes indicate the result was not reported.

performed in relatively small number of specialist hospitals, and patients generally have a shorter and less variable ICU LoS than other patients.

Our study also has five main weaknesses. First, we did not require studies to have a minimum sample size. Little consensus exists on the sample size required to prevent overfitting when constructing a prediction model for a continuous variable (35–37). The minimum mean number of admissions per predictor is 82, and hence sufficient, according to the sparse literature. Second, we did not place restrictions on the period in which the data were collected; during the ICU stay, patient characteristics were measured or the point from which ICU LoS was measured. Two studies report on data collected in or before 1999 (22, 25), two (24, 25) used characteristics that changed over time, and one (24) predicted prolonged ICU LoS beyond a threshold value. These models may not be suitable for current implementation, but throw light on patient characteristics that should be included in prediction models for

TABLE 5. Suitability of Models Predicting ICU Length of Stay for Planning Resources; Identifying Individual Patients or Groups of Patients With Unexpectedly Long ICU Length of Stay; and Benchmarking, Based on the Defined Requirements

Reference	Model Variables Published	No Organizational- or Country-Specific Predictors	Only Predictors Available at Admission or First 24 Hr Included	Moderate Calibration (for All Subgroups)	Strong R^2 Across ICUs	Strong R^2 Across Patients	Suitable for Planning and Identifying Unexpectedly Long ICU Length of Stay	Suitable for Benchmarking
Clermont et al (25)	Yes	Yes	No	No	No	No	No	No
Perez et al (22)	Yes	Yes	Yes	No	No	No	No	No
Zimmerman et al (8)	Yes	Yes	Yes	No[a]	Yes	No	No	No
Rothen et al (19)	Yes	Yes	Yes	No	No	No	No	No
Moran et al (23)	Yes	No	Yes	No	No	No	No	No
Moran and Solomon (7), model 1–12	No	No	Yes	No	No	No	No	No
Niskanen et al (6), model 1[b]	Yes	No	Yes	No	Yes	No	No	No
Niskanen et al (6), model 2	Yes	No	Yes	No	Yes	No	No	No
Vasilevskis et al (10), model 2[c]	Yes	Yes	Yes	No[a]	No	No	No	No
Vasilevskis et al (10), model 3	Yes	Yes	Yes	No[a]	No	No	No	No
Kramer and Zimmerman (24)	Yes	Yes	No	No	Yes	No	No	No
Al Tehewy et al (26)	Yes	Yes	Yes	No	No	No	No	No
Verburg et al (11), model 1–8	No	Yes	Yes	No	No	No	No	No

[a]Moderate recalibration was tested, and differences were not significant for several subgroups.
[b]Niskanen et al (6): model 1: outcome measure truncated at 30 d; model 2: log-transformed outcome measure used.
[c]Vasilevskis et al (10): model 1: second-order recalibration of the Acute Physiology and Chronic Health Evaluation IV model for length of stay; model 2: covariates included from Mortality Probability Models 0-III; model 3: covariates included from simplified.
The order of the studies in this table is based on "study period."

ICU LoS. Although these models are substantially different to the other models included in this review, we believe that their inclusion does not influence our final conclusions because we do not recommend them for either purpose. Third, we did not examine differences in performance according to whether ICU LoS is defined in fractional or billing days. Fourth, we only included studies that developed or validated prediction models for ICU LoS. We did not include all studies evaluating associations between patient characteristics, such as Sequential Organ Failure Assessment (SOFA) score, physiologic or laboratory values, and ICU LoS. Fifth, we did not consider the utility of models for the early identification of individuals with a high risk of excessively long ICU LoS.

Constructing a good model for ICU LoS would require specialized statistical and clinical knowledge (8, 30, 31, 38–40). For instance, as several studies (25) have reported an association between daily SOFA scores and ICU LoS, researchers could explore using these data for day-on-day planning. They could also examine novel statistical methods, such as joint modelling (41) and competing risks (42) to predict ICU LoS (43–46). However, we do not expect that these methods will result in substantially better models for predicting ICU LoS if they only consider patient characteristics. We expect that data on hospital characteristics and ICU policies and practice will also be required. However, including this type of characteristics will make it more difficult to use the resulting models for benchmarking purposes.

CONCLUSION

No previously published models satisfy our three general requirements for prediction models for ICU LOS or our specific requirements for models to plan resource allocation and to identify patients with unexpectedly long ICU LoS or our specific requirements for models for benchmarking purposes. Physicians considering using these models to predict ICU LoS

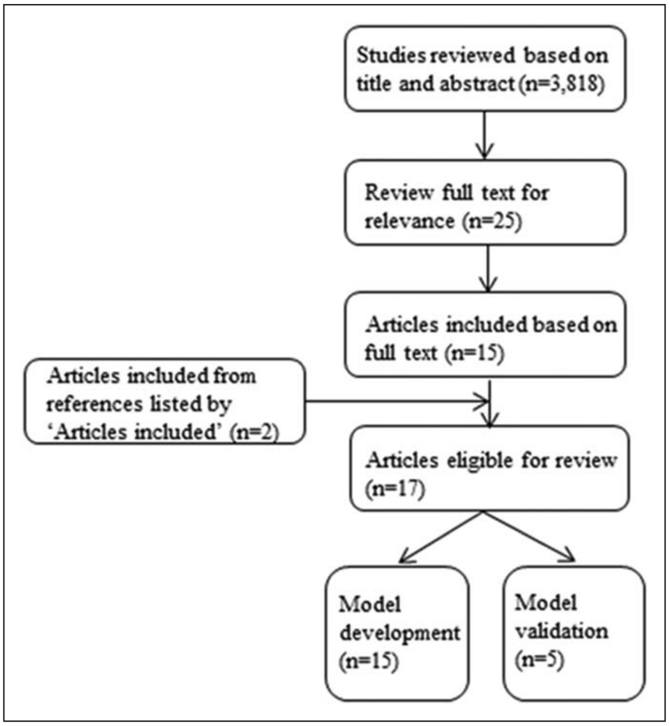

Figure 1. Flowchart representing the result of search and the number of articles excluded and eligible for review.

should interpret them with reservation until a validation study using recent local data has shown that they obtain moderate calibration and produce accurate predictions.

REFERENCES

1. Halpern NA, Pastores SM: Critical care medicine beds, use, occupancy, and costs in the United States: A methodological review. *Crit Care Med* 2015; 43:2452–2459
2. Halpern NA, Pastores SM, Greenstein RJ: Critical care medicine in the United States 1985-2000: An analysis of bed numbers, use, and costs. *Crit Care Med* 2004; 32:1254–1259
3. Kahn JM, Rubenfeld GD, Rohrbach J, et al: Cost savings attributable to reductions in intensive care unit length of stay for mechanically ventilated patients. *Med Care* 2008; 46:1226–1233
4. Marik PE, Hedman L: What's in a day? Determining intensive care unit length of stay. *Crit Care Med* 2000; 28:2090–2093
5. Rapoport J, Teres D, Zhao Y, et al: Length of stay data as a guide to hospital economic performance for ICU patients. *Med Care* 2003; 41:386–397
6. Niskanen M, Reinikainen M, Pettilä V: Case-mix-adjusted length of stay and mortality in 23 Finnish ICUs. *Intensive Care Med* 2009; 35:1060–1067
7. Moran JL, Solomon PJ; ANZICS Centre for Outcome and Resource Evaluation (CORE) of the Australian and New Zealand Intensive Care Society (ANZICS): A review of statistical estimators for risk-adjusted length of stay: Analysis of the Australian and New Zealand Intensive Care Adult Patient Data-Base, 2008-2009. *BMC Med Res Methodol* 2012; 12:68
8. Zimmerman JE, Kramer AA, McNair DS, et al: Intensive care unit length of stay: Benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV. *Crit Care Med* 2006; 34:2517–2529
9. Woods AW, MacKirdy FN, Livingston BM, et al: Evaluation of predicted and actual length of stay in 22 Scottish intensive care units using the APACHE III system. Acute Physiology and Chronic Health Evaluation. *Anaesthesia* 2000; 55:1058–1065
10. Vasilevskis EE, Kuzniewicz MW, Cason BA, et al: Mortality probability model III and simplified acute physiology score II: Assessing their value in predicting length of stay and comparison to APACHE IV. *Chest* 2009; 136:89–101
11. Verburg IW, de Keizer NF, de Jonge E, et al: Comparison of regression methods for modeling intensive care length of stay. *PLoS One* 2014; 9:e109684
12. Moons KG, de Groot JA, Bouwmeester W, et al: Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. *PLoS Med* 2014; 11:e1001744
13. Wallace E, Uijen MJ, Clyne B, et al: Impact analysis studies of clinical prediction rules relevant to primary care: A systematic review. *BMJ Open* 2016; 6:e009957
14. Walsh ME, Horgan NF, Walsh CD, et al: Systematic review of risk prediction models for falls after stroke. *J Epidemiol Community Health* 2016; 70:513–519
15. Kunath F, Spek A, Jensen K, et al: Prognostic factors for tumor recurrence in patients with clinical stage I seminoma udergoing surveillance-protocol for a systematic review. *Syst Rev* 2015; 4:182
16. Medlock S, Ravelli AC, Tamminga P, et al: Prediction of mortality in very premature infants: A systematic review of prediction models. *PLoS One* 2011; 6:e23441
17. Van Calster B, Nieboer D, Vergouwe Y, et al: A calibration hierarchy for risk models was defined: From utopia to emperical data. *J Clin Epidemiol* 2016; 74:167–176
18. Divaris K, Vann WF Jr, Baker AD, et al: Examining the accuracy of caregivers' assessments of young children's oral health status. *J Am Dent Assoc* 2012; 143:1237–1247
19. Rothen HU, Stricker K, Einfalt J, et al: Variability in outcome and resource use in intensive care units. *Intensive Care Med* 2007; 33:1329–1336
20. Knaus WA, Wagner DP, Zimmerman JE, et al: Variations in mortality and length of stay in intensive care units. *Ann Intern Med* 1993; 118:753–761
21. Rosenberg AL, Zimmerman JE, Alzola C, et al: Intensive care unit length of stay: Recent changes and future challenges. *Crit Care Med* 2000; 28:3465–3473
22. Perez A, Chan W, Dennis RJ: Predicting the length of stay of patients admitted for intensive care using a first step analysis. *Health Serv Outcomes Res Methodol* 2006; 6:127–138
23. Moran JL, Bristow P, Solomon PJ, et al; Australian and New Zealand Intensive Care Society Database Management Committee (ADMC): Mortality and length-of-stay outcomes, 1993-2003, in the binational Australian and New Zealand intensive care adult patient database. *Crit Care Med* 2008; 36:46–61
24. Kramer AA, Zimmerman JE: A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay. *BMC Med Inform Decis Mak* 2010; 10:27
25. Clermont G, Kaplan V, Moreno R, et al: Dynamic microsimulation to model multiple outcomes in cohorts of critically ill patients. *Intensive Care Med* 2004; 30:2237–2244
26. Al Tehewy M, El Houssinie M, El Ezz NA, et al: Developing severity adjusted quality measures for intensive care units. *Int J Health Care Qual Assur* 2010; 23:277–286
27. Metnitz PG, Moreno RP, Almeida E, et al; SAPS 3 Investigators: SAPS 3–From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. *Intensive Care Med* 2005; 31:1336–1344
28. Public Health England: Public Health Observatories (PHOs) 2013. Available at: http://www.apho.org.uk/resource. Accessed October 7, 2016
29. Steyerberg EW: Study design for prediction models. *In:* Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Gail M, Krickeberg K, Sarnet J, et al (Eds). New York, Springer, 2009, pp 33–52
30. Mallett S, Royston P, Waters R, et al: Reporting performance of prognostic models in cancer: A review. *BMC Med* 2010; 8:21
31. Mallett S, Royston P, Dutton S, et al: Reporting methods in studies developing prognostic models in cancer: A review. *BMC Med* 2010; 8:20
32. Lemeshow S, Teres D, Avrunin JS, et al: Refining intensive care unit outcome prediction by using changing probabilities of mortality. *Crit Care Med* 1988; 16:470–477

33. 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
34. Messaoudi N, De Cocker J, Stockman B, et al: Prediction of prolonged length of stay in the intensive care unit after cardiac surgery: The need for a multi-institutional risk scoring system. *J Card Surg* 2009; 24:127–133
35. Austin PC, Steyerberg EW: The number of subjects per variable required in linear regression analyses. *J Clin Epidemiol* 2015; 68:627–636
36. Babyak MA: What you see may not be what you get: A brief, nontechtnical introduction to overfitting in regression-type models. *Psychosom Med* 2004; 66:411–421
37. Kelley K, Maxwell SE: Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychol Methods* 2003; 8:305–321
38. Steyerberg EW: Restrictions on candidate predictors. *In:* Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Gail M, Krickeberg K, Sarnet J, et al (Eds). New York, Springer, 2009, pp 175–187
39. Peduzzi P, Concato J, Feinstein AR, et al: Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995; 48:1503–1510
40. Harrell FE Jr, Lee KL, Mark DB: Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15:361–387
41. Henderson R, Diggle P, Dobson A: Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; 1:465–480
42. Wolbers M, Koller MT, Witteman JC, et al: Prognostic models with competing risks: Methods and application to coronary risk prediction. *Epidemiology* 2009; 20:555–561
43. Zhou Z: Ensemble Methods: Foundation and Algorithms. London, Chapman & Hall, 2012
44. Shawe-Taylor J, Cristiano N: Kernel Methods for Pattern Analysis. Cambridge, Cambridge University Press, 2004
45. Koenker R, Hallock KF: Quantile regression an introduction. *J Econ Perspect* 2001; 14:123–148
46. Koenker R, Bassett G: Regression quantiles. *Econometrica* 1978; 46:33–50