Journal of Critical Care 28 (2013) 935–941

Contents lists available at ScienceDirect

Journal of Critical Care

journal homepage: www.jccjournal.org

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☆,☆☆

Eldho Paul, MSc [a,*], Michael Bailey, PhD, MSc, BSc (hons) [a], David Pilcher, MRCP, FRACP, FCICM [a,b,c]

[a] Australian and New Zealand Intensive Care Research Centre, Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
[b] Australian and New Zealand Intensive Care Society Centre for Outcome and Resource Evaluation, Carlton South, Victoria 3053, Australia
[c] Department of Intensive Care Medicine, The Alfred Hospital, Prahran, Victoria 3004, Australia

ARTICLE INFO

Keywords:
Mortality prediction
SMR
Intensive care
ANZROD

ABSTRACT

Purpose: The purpose of this study is to develop and validate a new mortality prediction model (Australian and New Zealand Risk of Death [ANZROD]) for Australian and New Zealand intensive care units (ICUs) and compare its performance with the existing Acute Physiology and Chronic Health Evaluation (APACHE) III-j.
Materials and Methods: All ICU admissions from 2004 to 2009 were extracted from the Australian and New Zealand Intensive Care Society Adult Patient Database. Hospital mortality was modeled using logistic regression with training (two third) and validation (one third) data sets. Predictor variables included APACHE III score components, source of admission to ICU and hospital, lead time, elective surgery, treatment limitation, ventilation status, and APACHE III diagnoses. Model performance was assessed by standardized mortality ratio, Hosmer-Lemeshow C and H statistics, Brier score, Cox calibration regression, area under the receiver operating characteristic curve, and calibration curves.
Results: There were 456 605 patients available for model development and validation. Observed mortality was 11.3%. Performance measures (standardized mortality ratio, Hosmer-Lemeshow C and H statistics, and receiver operating characteristic curve) for the ANZROD and APACHE III-j model in the validation data set were 1.01, 104.9 and 111.4, and 0.902; 0.84, 1596.6 and 2087.3, and 0.885, respectively.
Conclusions: The ANZROD has better calibration; discrimination compared with the APACHE III-j. Further research is required to validate performance over time and in specific subgroups of ICU population.
© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Adult prognostic models have many applications in intensive care medicine, including clinical research, benchmarking, performance monitoring, and accreditation [1]. The Acute Physiology and Chronic Health Evaluation (APACHE) [2-5], Simplified Acute Physiology Score (SAPS) [6-9], and Mortality Probability Model (MPM) [10-13] have been used internationally to predict the risk of hospital mortality for adult patients admitted to intensive care units (ICUs). These models provide risk adjustment to account for differences in the case mix of different hospitals, including patients' comorbidities and illness severity while benchmarking intensive care outcomes between organizations [14]. Furthermore, they help explain regional variation in mortality and understand changes in outcome over time.

The Australian and New Zealand Intensive Care Society (ANZICS) [15] Centre for Outcome and Resource Evaluation (CORE) has adopted the APACHE methodology [1] for benchmarking outcomes in Australasian ICUs. The APACHE is the most widely used international benchmark, and it has a strong research foundation. The APACHE II was introduced in Australia and New Zealand [16] in the late 1990s. The APACHE III-j model ("j" being the tenth iteration) was implemented in Australasian hospitals subsequent to its release into the public domain in 2002 [17]. This was implemented without customization as it had good calibration at that time. Over the years, performance of this model has deteriorated, possibly due to changes in case mix and clinical practice. Improvement in outcomes such as reduction in mortality over time may also have contributed to the worsening performance of this model. Although this system currently provides the benchmark for Australian and New Zealand (ANZ) ICUs, it no longer provides a true reflection of the "real" risk of death for Australasian ICU population. The deteriorating performance of APACHE III-j model in ANZ ICUs has been reported in greater detail

☆ Conflicts of interest: None to declare.
☆☆ Funding: Eldho Paul is a PhD student funded by ANZICS CORE and the ANZIC RC.
* Corresponding author. Department of Epidemiology and Preventive Medicine, Monash University, The Alfred Centre, Melbourne Victoria 3004, Australia. Tel.: +61 3 9903 0599; fax: +61 3 9903 0556.
E-mail address: eldho.paul@monash.edu (E. Paul).

0883-9441/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jcrc.2013.07.058

previously [18]. Because the American-based APACHE III-j consistently and progressively overpredicts mortality in Australasia, a new model has become a necessity for the ANZ ICU population.

Newer versions of major adult prognostic models have been developed overseas [19] to overcome performance degradation of older generation models. The APACHE has been upgraded to APACHE IV [5]; SAPS and MPM have been upgraded to SAPS3 [8,9] and MPM III [10], respectively. Although these new generation models are reported to perform well overseas [19], they cannot be implemented in ANZ ICUs because the data required for these models are not routinely collected by ANZICS.

Mortality prediction models derived from hospital administrative data have also been used to effectively benchmark performance at a local level. Critical care outcome prediction equation [1] and hospital outcome prediction equation [20] are examples of such models. However, these models may not perform as well as models derived using clinical data sets if applied to broader populations [14]. Although Bohensky et al [14] compared 180-day mortality post hospital discharge for critically ill patients, it is notable that most of these deaths had occurred in the hospital, and hence, prediction models based on administrative data may not be appropriate for implementation in ANZ ICUs for risk adjustment.

A valid outcome prediction model should include all patients and accurately predict outcomes in large and heterogeneous populations. Such a model, therefore, must take into account the complexity and diversity of patient case mix and illness severity. Major adult prognostic models developed overseas have used different exclusion criteria [21] for selection of an appropriate patient data set for model development, and the effect of these exclusions on model performance over time cannot be determined. Therefore, the present study aimed to develop and validate a new mortality prediction model called "the Australian and New Zealand Risk of Death (ANZROD) model" for the entire ICU population in Australasia with minimum exclusions. More specifically, we sought to develop a mortality prediction model tailored to ANZ ICUs, which would perform well in various subgroups of the ICU population. In this article, we present the methodology used to develop ANZROD, the results supporting its validity, and the potential uses of ANZROD.

2. Methods

2.1. Study setting

Data for the study were extracted from the ANZICS Adult Patient Database (APD) [16], which is one of the largest ICU data sets in the world with more than 1.4 million ICU admissions. This high-quality binational database records demographic, severity of illness, and outcome data from adult ICU admissions, beginning in 1990. The APD contains sufficient raw physiologic data to calculate the popular APACHE II and APACHE III scores along with their corresponding diagnostic categories. Data are collected at individual ICUs throughout all jurisdictions in ANZ and sent to ANZICS CORE for processing and benchmarking. They are submitted on behalf of each ICU director, and each hospital allows subsequent use as appropriate under the ANZICS CORE standing procedures and in compliance with the ANZICS CORE terms of reference.

2.2. Study subjects

All adult admissions between January 1, 2004, and December 31, 2009, whose data were submitted to the ANZICS APD were initially included in this study. We excluded patients younger than 16 years, patients missing an acute physiology score on ICU day 1, patients admitted for palliative care, and those with missing hospital outcomes. To avoid counting more than 1 hospital outcome for a patient during the same hospital stay, we included only each patient's first ICU admission. Access to the data was granted by the ANZICS CORE Management Committee in accordance with standing protocols. All data were deidentified, and the study was conducted with approval of the Monash University Human Research Ethics Committee.

2.3. Study variables

The data extracted from the ANZICS APD for each patient are summarized in Table E1. The data included age, chronic health conditions, and physiologic data required to calculate an acute physiology score (APS) of APACHE III. The APS is based on the worst measurement for 16 physiologic components on ICU day 1. Each of these components and the composite APACHE III score were extracted from the APD. We also extracted additional variables such as treatment limitation [22], hospital source of admission, lead time, elective admission, APACHE II chronic illnesses, and others, as presented in Table E1 [22].

To assess performance of the new model across different hospitals, hospital classification was also extracted from the ANZICS APD. Hospital classifications are self-reported but mutually agreed by hospitals, jurisdictions and by ANZICS. Tertiary hospitals represent larger university-affiliated public hospitals with specialist surgical services such as cardiothoracic and neurosurgery. Metropolitan hospitals are generally public hospitals in city regions, which provide full intensive care services such as invasive mechanical ventilation and hemofiltration. Rural/regional ICUs are generally smaller ICUs outside of major city areas, sometimes with co-located coronary care units. Private ICUs are those within private hospitals and may encompass a variety of ICUs from those similar to smaller rural/regional hospitals to large ICUs resourced in a similar fashion to tertiary public ICUs.

2.4. Model development

We randomly split the extracted data into derivation (67%) and validation (33%) data sets. The derivation data set was used for prediction models construction, whereas the validation data set was used to test the performance of the developed models. Both derivation and validation data sets were mutually exclusive. In developing the new model, we sought to improve upon the risk prediction generated from the internationally validated APACHE III [4] model by reevaluating the selection and weighting of variables. We used multivariate logistic regression analyses to determine the relationship between hospital mortality and each of the predictor variables listed in Table E1. Several models were constructed by adding or deleting individual variables, and their performance was compared. The model that displayed best discrimination (measured by area under the receiver operating characteristic curve [AUROC] [23]) with the least difference between observed and predicted outcomes (using Hosmer-Lemeshow C and H statistics [24]) was selected as the final model. Model development was undertaken at 4 different stages as described below.

2.4.1. Stage I (existing APACHE III methodology plus reweighting ANZ diagnostic categories)

The APACHE is an internationally validated scoring system used in the ICUs, which has been well described previously [25]. We used the existing APACHE III methodology [4] to develop a new model for Australasian ICUs. The predictor variables included in this model are an APACHE score composing of age score, chronic health status score, and APS plus coefficients for source of admission to ICU, postemergency surgery, and disease categories. Age score, chronic health status score, and APS score were objectively derived using the weightings recommended by Knaus et al [4]. The ICUs in ANZ use 94 original APACHE III diagnoses plus an additional 30 locally derived diagnoses, which were not in the original APACHE III prognostic system (Appendix E1). Using logistic regression analysis,

we reweighted all 124 diagnoses and estimated model coefficients for APACHE III score, source of admission to ICU, and postemergency surgery. Model performance was assessed after reweighting 124 diagnostic categories.

2.4.2. Stage II (reweighting components of APACHE III score)

The APACHE III score consists of weightings for age, chronic health status, and APS, which is made up of 16 physiological components. We reweighted each of these 16 components of APS, age score, and chronic health score by putting the total score for each component into the model at stage II of model development. Logistic regression model included 16 components of APS plus age score, chronic health status score, source of admission to ICU, postemergency surgery, and ICU diagnoses as predictor variables. Model coefficients were estimated for each predictor variable included in the regression. Model performance was assessed after reweighting components of the APACHE III score.

2.4.3. Stage III (addition of new variables)

At this stage, we tested new variables such as treatment limitation, hospital admission source, lead time, APACHE II chronic illnesses, gender, smoking status and ventilation (Table E1) in the logistic regression model. These variables were added along with the predictor variables included at stage II to maximize explanatory power of the new model. Furthermore, we unbundled the chronic health score and reweighted each of the 7 chronic illnesses at stage III. Several models were constructed by adding the new variables individually and in combination. Given the size of our data set, only variables with a $P < .001$ were retained in the final model (Appendix E1). Model coefficients were estimated for each predictor variable included in the regression. Calibration and discrimination of models were assessed with the addition of new predictor variables.

2.4.4. Stage IV (individual models for each major diagnostic group)

At stage IV of model development, we explored the presence of statistical interactions between APACHE III score, APS, components of APS, and ICU diagnoses. As the interactions were significant, separate logistic regression models were fitted for each major diagnostic group to improve overall mortality prediction. All significant variables identified at stage III were incorporated into each of these models. Major diagnostic groups were derived from the major system groupings used in the APACHE III scoring system (cardiovascular,

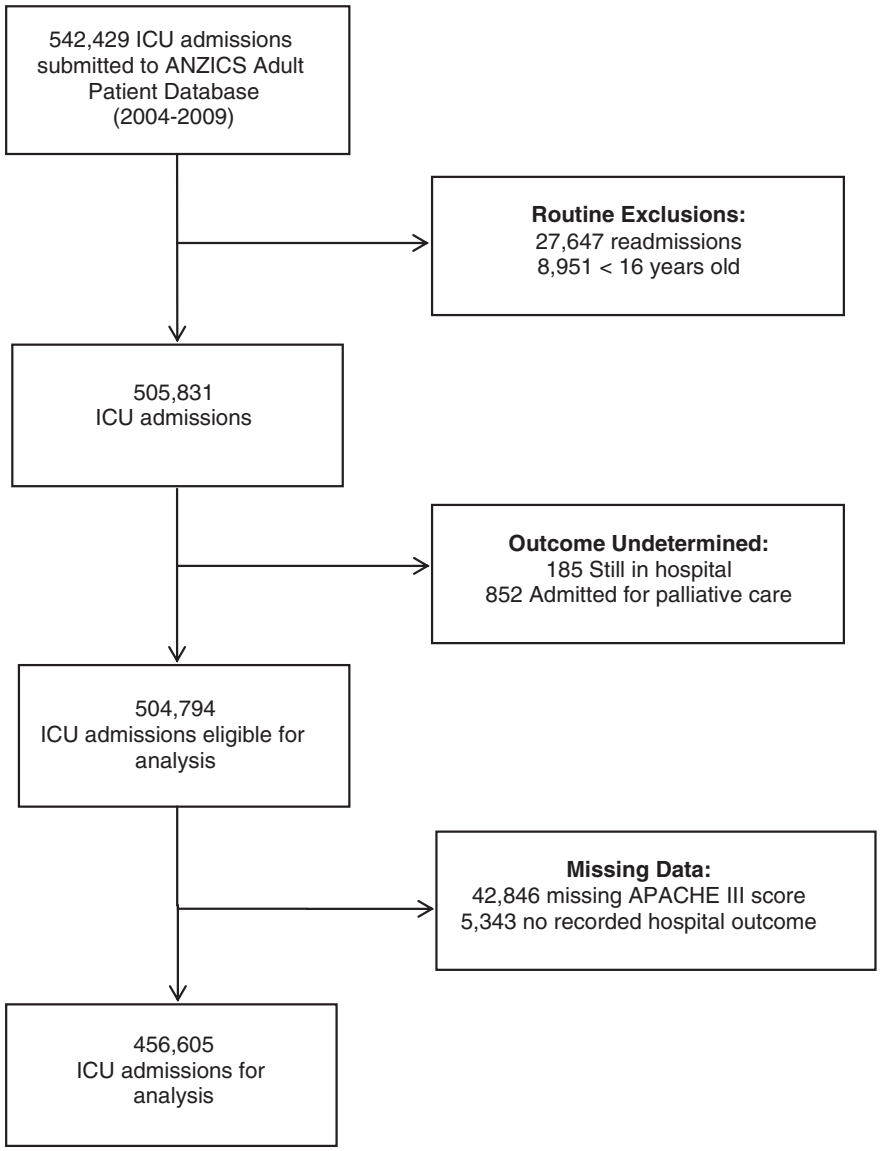

Fig. 1. Flowchart of patient inclusion criteria.

respiratory, etc) and combined operative and nonoperative diagnoses together in each system with the exception of cardiovascular. Table E2 lists the major diagnostic groups for which separate models were fitted. Model coefficients were estimated for each predictor variable included in the regression. Model calibration and discrimination were assessed for each individual model as well as a combined model.

2.5. Statistical analyses

All analyses were performed using SAS version 9.2 (SAS Institute, Cary, NC) or Stata version 11 (StataCorp, College Station, TX). Patient survival at hospital discharge was the primary outcome variable. Multivariate logistic regression analyses were used to determine the relationship between hospital mortality and predictor variables listed in Table E1. Model coefficients were estimated using the derivation data set and then applied to the validation data set. To facilitate a more robust prediction model, only variables with $P < .001$ were included in the final model. Models were developed and validated at each of the four stages described above. Model performance was assessed by evaluating calibration and discrimination parameters. The discriminatory power of models was assessed by computing the AUROC. Model fit was primarily assessed by using the Hosmer-Lemeshow C and H statistics. However, as these 2 statistics may be insufficient for comprehensive decisions on goodness of fit, we also included Brier's score [26], standardized mortality ratio (SMR) [27], and Cox's calibration regression [28] as additional parameters for comparing the overall agreement between predicted and observed outcomes (Table E3). For visual inspection of goodness of fit of models, calibration curves were drawn by plotting observed mortality against predicted mortality in equal sized groups based on quantiles of predicted mortality.

3. Results

During the study period, data pertaining to 542 429 ICU admissions were submitted to the ANZICS APD. Of these, 85 824 admissions were excluded from the analysis due to age below 16 years, readmission during the same hospital stay, admitted for palliative care, and unknown hospital outcomes or APACHE III scores (Fig. 1). Patient characteristics and main outcomes are displayed in Table 1. The average age of patients was 61 years, and 59% were males. Forty-four percent were planned admissions to ICUs following elective surgery, and 21% had 1 or more chronic illness. Intensive care unit and hospital mortality were 7.1% and 11.3%, respectively.

3.1. Performance of models

Table 2 shows performance assessment of the current APACHE III-j and new models in the derivation and validation data sets. Overall, the new models have better calibration and discrimination in the validation sample. In particular, the model based on separate equations for major diagnostic groups (ANZROD) has the highest AUROC and smallest Hosmer-Lemeshow C and H statistics, both indicating better performance compared with other models.

The ANZROD model has excellent discrimination (AUROC, 0.902), and aggregate predicted hospital mortality (11.3%) is very close to observed mortality (11.4%) for the validation data set. Calibration is good with substantial reductions in Hosmer-Lemeshow statistics (C and H) despite the large validation sample size (Table 2). The Cox calibration regression statistics are nonsignificant, Brier score is small, and SMR is 1. The difference between observed and mean predicted hospital mortality across risk deciles is 0.02% to 0.5%, except for the 70% to 80% decile (1.1%), 80% to 90% decile (2.2%), and 90% to 100% decile (1.6%). Calibration curve (Fig. 2) shows nearly perfect calibration at the lower end of the curve where most cases are placed.

4. Discussion

The ANZROD model is a novel outcome prediction tool for adult patients admitted to ICUs across Australia and New Zealand. It has been derived from a large binational clinical data set through detailed customization of the APACHE III specifically tailored to the ANZ ICU population. It provides a method for risk adjustment and a contemporary nationwide benchmark for the assessment of clinical performance.

The present study is an attempt to develop a well-calibrated model to benchmark ICU performance more accurately in Australasia. Because models developed in other countries vary in performance when applied to other settings, locally developed models are more desirable. Reid et al [29] developed the AusSCORE by modifying an international score to an Australian data set for similar benchmarking reasons.

The ANZROD model has been developed as an alternative to the current APACHE III-j by addressing many of its deficiencies reported previously [18]. The model has been developed using a locally derived data set specifically tailored to an Australasian population. The required resources and logistics needed to collect the data set already exist in most hospitals in ANZ. Its variables are derived from a comprehensive and regularly audited high-quality data set (ANZICS APD) that is available with ANZICS CORE. The ANZROD model has been designed in such a way that it can be recalibrated periodically (biennially), to allow for changes in case mix of patients and to maintain its validity as a performance monitoring tool. As the ANZICS APD is updated regularly, the ANZROD model could be used for intermittent benchmarking of outcomes (eg, with annual SMRs) and for continuous monitoring of mortality outcomes with process control charts (eg, exponentially weighted moving average charts).

Valid prediction models should accurately predict outcomes in the entire as well as particular subgroups of the population. The ANZROD model provides not only an overall assessment of hospital performance but also the ability to evaluate specific subgroups such as diagnoses, age and hospital type between hospitals. This dual perspective provides a comprehensive view of hospital performance. Hospital-wide poor outcomes may have a different etiology than poor outcomes in particular subgroups. Although the former may arise from systemic resource deficiencies due to hospital access or staffing levels, the latter may arise from the lack of more specific diagnostic or therapeutic interventions [20].

We found that the ANZROD model performs well in its “home” ICU population (validation data set), which is consistent with previous

Table 1
Patient characteristics and main outcomes for the entire study cohort

No. of hospitals	147
Total admissions (January 1, 2004, to December 31, 2009)	542429
No. of patients analyzed	456605
Number included in development and validation data sets	304149 and 152456
Age (y), mean (SD)	61.2 (18.1)
Male gender (%)	59.1
Elective surgery (%)	44.1
Hospital length of stay (h), median (IQR)	221.5 (125.1-421.8)
One or more chronic illnesses (%)	21.2
APACHE III score, mean (SD)	50.9 (27.8)
ICU mortality (%)	7.1
Hospital mortality (%)	11.3
APACHE III-j predicted risk of death (%), mean (SD)	14.3 (21.7)
Source of admission to ICU (%)	
Operating theater or recovery room	52.4
Accident and emergency	25.8
Hospital ward	13.6
Other hospital	7.3
Other ICU, same hospital	0.3
Other hospital ICU	0.5
Unknown	0.1

IQR indicates interquartile range.

Table 2
Comparison of model performance in derivation data set (a) and validation data set (b)

Model performance measure	Ideal value	APACHE III-j	Model 1	Model 2	Model 3	ANZROD
(a) Derivation data set						
Hosmer-Lemeshow						
C statistic	0	3404.2	497.3	367.6	399.0	189.5
H statistic	0	4359.4	534.6	374.8	405.9	174.1
Brier score	0	0.074	0.069	0.068	0.066	0.065
SMR	1	0.83	1	1	1	1
Cox calibration regression						
Intercept	0	−0.49	0	0	0	0
Slope	1	0.88	1	1	1	1
AUROC	1	0.887	0.892	0.896	0.902	0.905
(b) Validation data set						
Hosmer-Lemeshow						
C statistic	0	1596.6	259.7	196.7	228.3	104.9
H statistic	0	2087.3	255.9	198.5	208.9	111.4
Brier score	0	0.075	0.07	0.069	0.067	0.066
SMR	1	0.84	1.01	1.02	1.01	1.01
Cox calibration regression						
Intercept	0	−0.48	0.02	0.02	0.01	−0. 01
Slope	1	0.88	1	1	0.99	0.98
AUROC	1	0.885	0.891	0.895	0.90	0.902

Predictor variables included in different models were as follows:

APACHE III-j: APS + age score + chronic health score + ICU admission source + elective surgery + 94 diagnoses.
Model 1: APACHE III score + ICU admission source + elective surgery + 124 diagnoses.
Model 2: Components of APACHE III APS + APACHE III age score + APACHE III chronic health score + ICU admission source + elective surgery + 124 diagnoses.
Model 3: Components of APACHE III APS + APACHE III age score + APACHE III chronic health variables + ICU admission source + elective surgery + 124 diagnoses + treatment limitation + lead time + hospital admission source + APACHE II chronic respiratory disease + ventilation.
ANZROD: Same as model 3 but a separate model for each major disease group listed in Table E2.

studies conducted in North America [19,30] and Europe [21,31]. All the major ICU prognostic models such as APACHE, SAPS, and MPM are reported to perform well in the countries where they were originally developed. Performance deterioration occurred when these models were applied to other ICU populations possibly due to changes in case mix and clinical practices, thereby suggesting the need for external validation before implementation to heterogeneous ICU populations. Consequently, ANZROD model will also require prospective validation before its application to ICUs outside Australasia.

Our analysis of predictive accuracy over diagnoses showed that performance of the ANZROD model in the ANZ ICU population varied in different diagnostic subgroups (Table E4). This is in agreement with previous research [32-34], which suggested that prognostic models can underpredict or overpredict mortality in specific patient subgroups, which might not have been well represented in the original case mix used to develop the model. In the present study, the ratio of observed to mean predicted hospital mortality ranged between 0.9 and 1.10 for over half of the diagnoses, and 92% of the SMRs within disease groups are not significantly different from 1.0 because the confidence intervals (CIs) for SMRs included 1. Approximately 40% of diagnoses in our validation sample had sample size below 500, and this smaller sample size may have caused a low power of detection for some disease groups [34]. Thus, the accuracy of ANZROD over diagnoses, especially in smaller diagnostic groups, will require further investigation.

Performance of ANZROD across hospital types highlighted some important differences (Table E5). The model appears to perform well in rural and metropolitan hospitals, as indicated by SMRs and the corresponding CIs. In contrast, the SMRs for tertiary and private hospitals were significantly different from 1, suggesting a difference between observed and predicted mortality, although the magnitude of this difference was small. This finding is consistent with 2 regional studies conducted by Duke et al [1,20], which also reported differences in model performance according to hospital type. Although the SMR for regional hospital was not significantly different from 1, SMRs for tertiary and metropolitan hospitals were different in both studies. Such differences seen in model performance according to hospital type are most likely due to patient case mix, organizational structure, and financing of health care [31]. Furthermore, patient factors such as social, economic and remoteness could play a role and require further research.

Although the relationship between mortality and age is complex, our analysis found that age treated as a continuous linear variable could not improve the goodness of fit in comparison with originally derived APACHE III age scores. When comparing ANZROD performance over age groups, we found that it performed well in most age

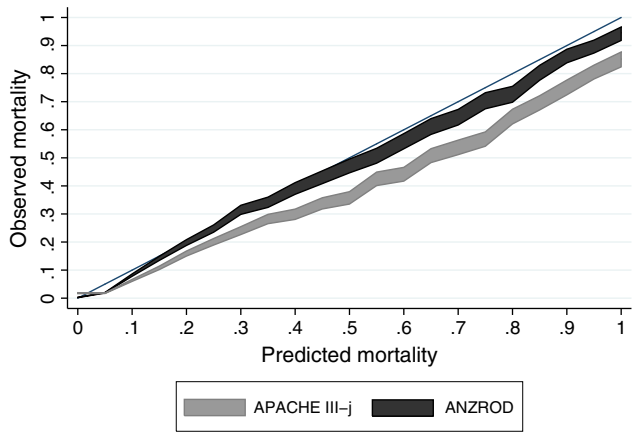

Fig. 2. Calibration curve of ANZROD (shaded in black) vs APACHE III-j (shaded in gray) in the validation data set. Diagonal line represents the line of perfect calibration. Shaded areas represent 95% CIs.

groups (Fig. 3). The 95% CI for SMRs included 1 in all the age groups except 35 to 44, 65 to 74, and 75 to 84 years. Although the model underestimated mortality in 2 age groups, it overpredicted mortality in 65 to 74 years. Although these differences in model performance across age groups are most likely the result of underlying case mix of patients, it could also arise from unmeasured confounding as a result of local differences in patient characteristics [35].

Popular ICU prognostic models developed overseas have used different exclusion criteria for selection of patients for model development [21]. Although some are standard exclusions (eg, readmissions within the same hospital stay), others are model specific. The APACHE III model excluded patients younger than 16 years, whereas SAPSII and MPMII have excluded patients younger than 18 years. Other exclusions have been transfers from [5] or to another ICU [6,12] and specific diagnoses (eg, burns and coronary artery bypass graft in earlier versions of APACHE) for each of these models, and their effect on model performance has not been assessed. In contrast, the ANZROD model has been developed with minimum exclusions, thereby covering larger ICU population in ANZ. The main exclusions are patients younger than 16 years and those admitted for palliative care. This not only increases overall performance of ANZROD but also facilitates evaluation of its performance in more patient subgroups.

This study had several strengths. It is the first study to develop and validate a new mortality prediction model for critically ill adult patients in Australasia. Although ICU scoring systems were implemented here over a decade ago, no model was developed specifically for ANZ, and this current study fills that void. The ANZICS APD is one of the largest ICU databases in the world with more than one and a quarter million ICU admissions. The magnitude of this database not only allowed us to accurately develop ANZROD but also aids in evaluating its performance in specific patient subgroups such as age, diagnoses and hospital type. Case-mix adjustment was improved by expanding the number of diagnostic coefficients to 124. In addition, the final model was developed incorporating 8 separate models for major diagnostic groups, and we believe that this was a major factor in improving predictive accuracy. Finally, we added new variables such as treatment limitations, hospital admission source, lead time and APACHE II chronic illnesses to improve performance of our model.

Our study has potential limitations. The "local" nature of the study limits its external validity, and our results may not be applicable to ICUs outside Australasia where patient case mix, care models, and admission criteria are different. Although the data were extracted from a high-quality database (ANZICS APD), nearly 8% of eligible admissions could not be used for model development due to missing data with no recorded hospital outcomes or no APACHE III scores, and their subsequent effect on model performance cannot be determined. Fortunately, ongoing attention and monitoring of this database have consistently improved the quality and consistency of the data, and it is anticipated that future versions of ANZROD will be constructed with less than 3% of eligible admissions being missing. Furthermore, although data were collected prospectively and stored in the APD, it was necessary to retrospectively extract data for a specific period and, as such, carries with it some historical bias. Prognostic models lose calibration over time, and the performance of ANZROD in this patient cohort may not reflect current performance. A further potential limitation is the complexity of the ANZROD model, with a separate model for each major diagnostic group. Future research could focus on assessing performance of ANZROD over specific patient subgroups.

5. Conclusions

The ANZROD model is a novel outcome prediction tool for critically ill adult patients in ANZ ICUs. Its predictions of hospital mortality have excellent discrimination and good calibration and will be useful for benchmarking ICU performance in Australasian hospitals. Although further research is undoubtedly needed to assess performance of this model over time and in specific patient subgroups, ANZROD will provide an evolving platform from which to base further research.

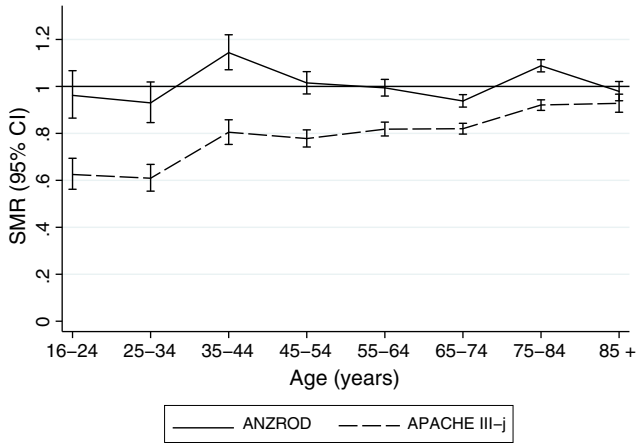

Fig. 3. Performance of ANZROD vs APACHE III-j over age strata in the validation data set. The horizontal line represents ideal value for SMR.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jcrc.2013.07.058.

References

[1] Duke GJ, Santamaria J, Shann F, Stow P, Pilcher D, Ernest D, et al. Critical care outcome prediction equation (COPE) for adult intensive care. Crit Care Resusc 2008;10(1):35–41.
[2] Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. Crit Care Med 1981;9(8):591–7.
[3] Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985;13(10):818–29.
[4] Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest 1991;100(6):1619–36.
[5] 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(5):1297–310.
[6] 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(24):2957–63.
[7] Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. A simplified acute physiology score for ICU patients. Crit Care Med 1984;12(11):975–7.
[8] Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, et al. 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(10):1336–44.
[9] Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, 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(10):1345–55.
[10] Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). Crit Care Med 2007;35(3):827–35.
[11] Lemeshow S, Klar J, Teres D, Avrunin JS, Gehlbach SH, Rapoport J, 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(9):1351–8.
[12] 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(20):2478–86.
[13] Lemeshow S, Teres D, Pastides H, Avrunin JS, Steingrub JS. A method for predicting survival and mortality of ICU patients using objectively derived weights. Crit Care Med 1985;13(7):519–25.
[14] Bohensky MA, Jolley D, Pilcher DV, Sundararajan V, Evans S, Brand CA: Prognostic models based on administrative data alone inadequately predict the survival outcomes for critically ill patients at 180 days post-hospital discharge. J Crit Care 2012, 27(4):422 e411-421.
[15] Australian and New Zealand Intensive Care Society. Centre for Outcome and Resource Evaluation Annual Report 2010 – 2011, Melbourne, Australia.
[16] Stow PJ, Hart GK, Higlett T, George C, Herkes R, McWilliam D, et al. Development and implementation of a high-quality clinical database: the Australian and New Zealand Intensive Care Society Adult Patient Database. J Crit Care 2006;21(2):133–41.

[17] Pilcher DV, Hoffman T, Thomas C, Ernest D, Hart GK. Risk-adjusted continuous outcome monitoring with an EWMA chart: could it have detected excess mortality among intensive care patients at Bundaberg Base Hospital? Crit Care Resusc 2010;12(1):36–41.
[18] Paul E, Bailey M, van Lint A, Pilcher DV. Performance of APACHE III over time in Australia and New Zealand: a retrospective cohort study. Anaesth Intensive Care 2012;40(6):980–94.
[19] Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. Crit Care Med 2011;39(1):163–9.
[20] Duke GJ, Graco M, Santamaria J, Shann F. Validation of the hospital outcome prediction equation (HOPE) model for monitoring clinical performance. Intern Med J 2009;39(5):283–9.
[21] Harrison DA, Brady AR, Parry GJ, Carpenter JR, Rowan K. Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. Crit Care Med 2006;34(5):1378–88.
[22] Godfrey G, Pilcher D, Hilton A, Bailey M, Hodgson CL, Bellomo R. Treatment limitations at admission to intensive care units in Australia and New Zealand: prevalence, outcomes, and resource use. Crit Care Med 2012 Jul;40(7):2082–9.
[23] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143(1):29–36.
[24] Hosmer DW, Lemeshow S. Goodness of fit tests for the multiple logistic regression model. Commun Stat 1980;A9:1043–69.
[25] Zimmerman JE, Kramer AA. Outcome prediction in critical care: the Acute Physiology and Chronic Health Evaluation models. Curr Opin Crit Care 2008; 14(5):491–7.
[26] Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev 1950;75:1–3.
[27] Armitage P, Berry G, Matthews J. Statistical methods in medical research. 4th ed. Malden, Massachusetts: Blackwell Science; 2002.
[28] Cox DR. Two further applications of a model for binary regression. Biometrika 1958;45:562–5.
[29] Reid C, Billah B, Dinh D, Smith J, Skillington P, Yii M, et al. An Australian risk prediction model for 30-day mortality after isolated coronary artery bypass: the AusSCORE. J Thorac Cardiovasc Surg 2009;138(4):904–10.
[30] Higgins TL, Kramer AA, Nathanson BH, Copes W, Stark M, Teres D. Prospective validation of the intensive care unit admission Mortality Probability Model (MPM0-III). Crit Care Med 2009;37(5):1619–23.
[31] Beck DH, Smith GB, Pappachan JV, Millar B. External validation of the SAPS II, APACHE II and APACHE III prognostic models in South England: a multicentre study. Intensive Care Med 2003;29(2):249–56.
[32] Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, de Jonge E, Bosman RJ, Peelen L, et al. External validation of Acute Physiology and Chronic Health Evaluation IV in Dutch intensive care units and comparison with Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II. J Crit Care 2011;26(1): 105e111–108.
[33] Nierman DM, Schechter CB, Cannon LM, Meier DE. Outcome prediction model for very elderly critically ill patients. Crit Care Med 2001;29(10):1853–9.
[34] Zhu BP, Lemeshow S, Hosmer DW, Klar J, Avrunin J, Teres D. Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. Crit Care Med 1996;24(1): 57–63.
[35] Mann SL, Marshall MR, Woodford BJ, Holt A, Williams AB. Predictive performance of Acute Physiological and Chronic Health Evaluation releases II to IV: a single New Zealand centre experience. Anaesth Intensive Care 2012;40(3):479–89.