

HS&DR 09/2000/65

Ensuring comparisons of health care providers are fair: development and validation of risk prediction models for critically ill patients

David A Harrison¹, Paloma Ferrando-Vivas¹, Jason Shahin^{1,2}, Kathryn M Rowan¹

¹ Clinical Trials Unit, Intensive Care National Audit & Research Centre (ICNARC), London, UK

² Department of Medicine, Respiratory Division and Department of Critical Care, McGill University, Montreal, Quebec, Canada

Corresponding author:

Dr David Harrison

Senior Statistician

Intensive Care National Audit & Research Centre

Napier House

24 High Holborn

London WC1V 6AZ

Tel: 020 7831 6878

Fax: 020 7831 6879

Email: david.harrison@icnarc.org

Competing interests: None declared.

Keywords: Cardiopulmonary Resuscitation; Critical Care; Heart Arrest; Hospital Mortality; Intensive Care; Models, Statistical; Outcome Assessment (Health Care); Prognosis; Risk Adjustment.

Word count (main body of report): 25,600

Important

A 'first look' scientific summary is created from the original author-supplied summary once the normal NIHR Journals Library peer and editorial review processes are complete. The summary has undergone full peer and editorial review as documented at NIHR Journals Library website and may undergo rewrite during the publication process. The order of authors was correct at editorial sign-off stage.

A final version (which has undergone a rigorous copy-edit and proofreading) will publish as part of a fuller account of the research in a forthcoming issue of the Health Services and Delivery Research journal.

Any queries about this 'first look' version of the scientific summary should be addressed to the NIHR Journals Library Editorial Office - journals.library@nihr.ac.uk

The research reported in this 'first look' scientific summary was funded by the HS&DR programme or one of its predecessor programmes (NIHR Service Delivery and Organisation programme, or Health Services Research programme) as project number 09/2000/65. For more information visit

<http://www.nets.nihr.ac.uk/projects/hsdr/09200065>

The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HS&DR editors have tried to ensure the accuracy of the authors' work and would like to thank the reviewers for their constructive comments however; they do not accept liability for damages or losses arising from material published in this scientific summary.

This 'first look' scientific summary presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not

necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health.

FIRST LOOK DRAFT

Scientific summary

Background

High quality care is at the centre of the NHS. Assessing outcomes of health care providers requires comparison against other providers—comparative audit—using high quality clinical data to put the outcome of the particular provider in context and enable benchmarking. National clinical audit has a key role to play in ensuring high quality care, particularly in areas of health care, such as emergency and critical care, where patient choice does not, and cannot, play a significant part.

However, quality of care is only one of many factors that will contribute to a patient's outcome and, if crude outcomes were to be compared between health care providers, any effect of quality would likely be overwhelmed by variation in patient demographics, underlying health status, acute conditions, and severity of the acute illness (collectively termed case mix). When comparing outcomes between health care providers, it is therefore essential to take the differing case mix of the providers into account to be able to make fair comparisons. Sophisticated and accurate risk prediction models are therefore required to adjust for patient case mix in national clinical audits.

The Intensive Care National Audit & Research Centre (ICNARC) is an independent charitable organisation that co-ordinates two national clinical audits: the Case Mix Programme (CMP)—the national clinical audit for adult critical care; and the National Cardiac Arrest Audit (NCAA)—the national clinical audit for in-hospital cardiac arrest. Both national clinical audits are underpinned by the need and ability to report accurate risk-adjusted results.

Risk prediction models for adult, general critical care are well established, but ongoing improvement work is essential to further improve accuracy. In 2006, ICNARC published a validation of four existing models and concluded that there was little difference in performance among the models, but there was scope for further improvement. While retaining the APACHE II model for international comparability, ICNARC developed and validated the ICNARC model, which underpins the risk-

adjusted outcomes reported for the CMP. However, we have identified a number of areas where we have the potential to improve our modelling.

Prior to this project, there was no validated risk prediction model for predicting outcomes following in-hospital cardiac arrest. Initial comparative reporting for NCAA was based on stratifying patients according to single risk factors.

Objectives

The aim of the current project was to improve risk prediction models to underpin quality improvement programmes for the critically ill (patients receiving general or specialist adult critical care or experiencing an in-hospital cardiac arrest).

We set out to address this aim through the following objectives:

1. To improve current risk prediction models for critically ill patients, to include:
 - a. external validation of current models in critical care units in Scotland;
 - b. introduction of new important variables;
 - c. improved modelling of interactions between physiological parameters;
 - d. improved handling of missing data; and
 - e. improved modelling of reasons for admission to/diagnosis on admission to critical care.
2. To develop and validate new risk prediction models for critically ill patients, to include:
 - a. models for cardiothoracic critical care;
 - b. models for patients experiencing an in-hospital cardiac arrest; and
 - c. models for critical care units admitting lower risk patients (ultimately addressed within objective 1).
3. Immediate translation of improved risk prediction models into practice, through:
 - a. adoption into routine comparative outcome reporting for national clinical audits; and
 - b. communication of research output to providers, managers, commissioners, policy makers and academics in critical care.

Methods and results

External validation of the current ICNARC model in Scottish critical care units

Data were extracted from the Scottish Intensive Care Society Audit Group (SICSAG) database for the years 2007 to 2009. Recoding and mapping of variables was performed, as required, to apply the ICNARC model (2009 recalibration) to the SICSAG data. The performance of the ICNARC model was assessed for discrimination, calibration and overall fit and compared with that of the APACHE II model.

There were 29,626 admissions to 24 adult, general critical care units in Scotland between 1 January 2007 and 31 December 2009. After exclusions, 23,269 admissions were included in the analysis. The ICNARC model outperformed APACHE II on measures of discrimination (c index 0.848 versus 0.806), calibration (Hosmer-Lemeshow chi-squared statistic 18.8 versus 214) and overall fit (Brier's score 0.140 versus 0.157; Shapiro's R 0.652 versus 0.621). Model performance was consistent across the three years studied.

Development and validation of a risk prediction model for admissions to cardiothoracic critical care units

Data were extracted from the CMP database for admissions to cardiothoracic critical care units for the years 2010 to 2012 (development) and for January 2013 to June 2014 (validation). Risk prediction models were fitted using logistic regression to predict mortality before discharge from acute hospital. Missing data on predictors were imputed using multiple imputation by fully conditional specification. Alternative functional forms were considered for modelling continuous predictors in univariable analyses. A full multivariable model was fitted including all potential predictors and simplified by removing non-significant terms and the functional form re-examined. The model was then simplified further by removing predictors in a stepwise approach to select a model balancing parsimony with performance. The final parsimonious model was further improved by considering additional factors specific to patients admitted following cardiac surgery.

There were 17,002 admissions to five cardiothoracic critical care units between 1 January 2010 and 31 December 2012. Of these, 1,881 (11.1%) died before discharge from acute hospital. The optimal approach to modelling most predictors was with restricted cubic splines, except for blood lactate, which was found to have a linear relationship with outcome. The full multivariable model with 17 predictors had a c index of 0.914 and a Brier's score of 0.064; all predictors were statistically significant. Following the stepwise procedure, ten predictors were retained (in order of importance): location prior to critical care unit admission/surgical urgency; blood lactate; Glasgow Coma Score (GCS); age; arterial pH; platelet count; prior dependency; mean arterial pressure; white blood cell count; and creatinine. The resulting simplified model had a c index of 0.895 and Brier's score of 0.066. The model was improved by introducing interactions between admission following cardiac surgery and physiological predictors (blood lactate, platelet count and creatinine). The resulting final model had a c index of 0.904 and Brier's score of 0.055 in external validation data.

Development and validation of the new ICNARC model for prediction of acute hospital mortality for admissions to adult critical care units

Data were extracted from the CMP database for admissions to adult (general and specialist) critical care units during January to December 2012 (development) and January to September 2013 (validation). A set of 21 physiological and 15 non-physiological candidate predictors were selected a priori based on the previous ICNARC model, published studies, and expert knowledge. Alternative approaches were considered for imputation of missing predictors and compared with using complete case data for model development in terms of bias and loss of precision. The optimal functional form for continuous predictors was considered in univariable analyses. A full physiology model was fitted using logistic regression including main terms for all the physiological candidate predictors. Non-significant predictors were removed from the model, and continuous predictors were tested for linearity. A simplified physiology model was developed by backward elimination. Starting from this simplified physiology model, a full multivariable model was fitted by adding non-physiological predictors. Reason for admission to the critical care unit was modelled

based primarily on the combination of body system and pathological/physiological process, making use of the hierarchical approach to coding. The full model was again refined and simplified using a similar approach to the physiology model. Potentially important interactions between the candidate predictors were identified by an expert group of clinicians. These interactions were introduced one-by-one into the model and significant interactions ($p < 0.05$) were retained. The full model including all such interactions was then fitted and interaction terms were retained if they were significant at $p < 0.001$ to avoid overfitting. Model performance was assessed in terms of discrimination, calibration and goodness of fit and compared with that of the current ICNARC model using reclassification techniques. In addition, the performance of the new model was compared with the current ICNARC model (and, where relevant, recent recalibrations to specific unit types) for subgroups defined by patient characteristics and critical care unit types.

There were 155,239 admissions to 232 adult critical care units between 1 January 2013 and 31 December 2013. Use of complete case data was found to have minimal impact on the model selection process and so the model was developed using data from 121,573 admissions with complete data for all candidate predictors, with multiple imputation using fully conditional specification applied in parallel at important steps in the process, including to estimate the final coefficients of the model. The optimal functional form for continuous predictors was found to be best modelled with either restricted cubic splines or right-restricted cubic splines. The simplified physiology model retained all 12 physiological predictors from the current ICNARC model (systolic blood pressure, temperature, heart rate, respiratory rate, $\text{PaO}_2/\text{FiO}_2$, arterial pH, urine output, creatinine, urea, sodium, white blood cell count and GCS/sedation) and also PaCO_2 , blood lactate and platelet count. Non-physiological predictors included in the final model were age, dependency prior to admission, severe conditions in the past medical history (severe liver disease, metastatic disease, haematological malignancy), cardiopulmonary resuscitation (CPR) within 24 hours prior to admission, location prior to admission (in combination with surgical urgency and planned versus unplanned admission) and primary reason for admission (56 system/process combinations and 16 individual conditions). In

addition, 19 interactions between reasons for admission and physiological predictors and five interactions between past medical history and physiological predictors were included. The final model had a c index of 0.891 and Brier's score of 0.103 in the development dataset. Performance of the new model was similar in the validation dataset of 90,017 admissions to 216 critical care units between January and September 2013 (c index 0.885, Brier's score 0.108) and slightly better than the most recent recalibration of the current ICNARC model (c index 0.869, Brier's score 0.115). Net reclassification improvement for the new model was 19.9. Performance of the new model was similar or improved across all types of specialist critical care units when compared with coefficients for the current ICNARC model specifically recalibrated to these unit types.

Development and validation of risk prediction models to predict outcomes following in-hospital cardiac arrest

Data were extracted from the NCAA database for patients (aged 28 days or over) who received chest compressions and/or defibrillation following an in-hospital cardiac arrest and were attended by the hospital-based resuscitation team in response to an emergency (2222) call between April 2011 and March 2013. Risk prediction models were developed for two outcomes—return of spontaneous circulation (ROSC) for greater than 20 minutes and survival to hospital discharge. For each outcome, a full model was fitted and then simplified by testing for nonlinearity, combining categories and stepwise reduction. Finally, interactions between predictors were considered. Models were assessed for discrimination, calibration and accuracy in data from the same hospitals over time and in new hospitals that had recently joined NCAA.

A total of 22,479 in-hospital cardiac arrests in 143 hospitals were included (14,688 development, 7,791 validation). The final risk prediction model for ROSC>20 minutes included: age (nonlinear); sex; prior length of stay in hospital; reason for attendance; location of arrest; presenting rhythm; and interactions between presenting rhythm and location of arrest. The model for hospital survival included the same predictors, excluding sex. Both models had acceptable performance across the range of

measures, although discrimination for hospital survival exceeded that for ROSC > 20 minutes (c index 0.81 versus 0.72 in the validation dataset).

Conclusions

We have demonstrated that the current ICNARC model retains similar performance to that reported from previous validation within the CMP when externally validated using independently collected data from critical care units in Scotland. Nevertheless, we identified a number of areas where the current risk prediction model could be improved. The first related to its performance in specialist critical care units. We therefore developed a specific risk prediction model for admissions to cardiothoracic critical care units, which had excellent performance. As well as providing a specific model, tailored to the unique case mix of these units, this model also served as a baseline to be able to assess the performance of the new ICNARC model in cardiothoracic critical care units, serving as an assessment of the ability of a generic model to work across different types of units.

In developing the new ICNARC model, we also addressed further areas for improvement, including handling of missing data, continuous nonlinear modelling of physiological predictors and making better use of the available data within the hierarchical coding of reasons for admission to the critical care unit. The resulting risk prediction model performed well not only in the full validation dataset but also when evaluated in specific patient subgroups and specific types of critical care unit.

Finally, using data from NCAA we developed risk prediction models to predict two important outcomes following in-hospital cardiac arrest – the immediate outcome of ROSC greater than 20 minutes and the slightly longer-term outcome of hospital survival. Based on only a small number of predictors, the model for hospital survival had good discrimination and validated well on subsequent data. The performance of the model for ROSC greater than 20 minutes was less good, possibly reflecting inter-hospital variation in resuscitation practice.

Implications for healthcare

The newly developed risk prediction models have been, or are being, introduced into routine comparative reporting for the CMP and NCAA. For the CMP, this will enable

fairer comparison across critical care units including, for the first time, across different types of critical care units, underpinning annual public reporting of critical care unit outcomes. For NCAA, the models permit genuine risk-adjusted comparisons across hospitals for the first time and will enable NCAA to also move towards public reporting of results.

Recommendations for research

Recommendation 1: Further research should be conducted by linking with death registrations to evaluate mortality at fixed time points and using time to event analyses.

Recommendation 2: Further research in this field should make better use of data linkage across national clinical audits.

Recommendation 3: Further research in this field should make better use of other routinely-collected datasets.

Recommendation 4: Future research should consider the necessity for specific data collection to support national clinical audit compared with benchmarking providers using routinely-collected data alone.

Word count: 2385