BMJ

RESEARCH

Derivation and validation of a risk adjustment model for predicting seven day mortality in emergency medical admissions: mixed prospective and retrospective cohort study

OPEN ACCESS

Steve Goodacre *professor of emergency medicine*, Richard Wilson *research associate*, Neil Shephard *medical statistician*, Jon Nicholl *professor of health service research*, on behalf of the DAVROS Research Team

School of Health and Related Research (ScHARR), University of Sheffield, Sheffield S1 4DA, United Kingdom

Abstract

Objectives To derive and validate a risk adjustment model for predicting seven day mortality in emergency medical admissions, to test the value of including physiology and blood parameters, and to explore the constancy of the risk associated with each model variable across a range of settings.

Design Mixed prospective and retrospective cohort study.

Setting Nine acute hospitals (n=3 derivation, n=9 validation) and associated ambulance services in England, Australia, and Hong Kong.

Participants Adults with medical emergencies (n=5644 derivation, n=13 762 validation) who were alive and not in cardiac arrest when attended by an ambulance and either were admitted to hospital or died in the ambulance or emergency department.

Interventions Data were either collected prospectively or retrospectively from routine sources and extraction from ambulance and emergency department records.

Main outcome measure Mortality up to seven days after hospital admission.

Results In the derivation phase, age, ICD-10 code, active malignancy, Glasgow coma score, respiratory rate, peripheral oxygen saturation, temperature, white cell count, and potassium and urea concentrations were independent predictors of seven day mortality. A model based on age and ICD-10 code alone had a C statistic of 0.80 (95% confidence interval 0.78 to 0.83), which increased to 0.81 (0.79 to 0.84) with the addition of active malignancy. This was markedly improved only when physiological variables (C statistic 0.87, 0.85 to 0.89), blood variables (0.87, 0.84 to 0.89), or both (0.90, 0.88 to 0.92) were added. In the validation phase, the models with physiology variables (physiology model) and all variables (full model) were tested in nine hospitals. Overall, the C statistics ranged across centres from 0.80 to 0.91 for the physiology

model and from 0.83 to 0.93 for the full model. The rank order of hospitals based on adjusted mortality differed markedly from the rank order based on crude mortality. ICD-10 code, Glasgow coma score, respiratory rate, systolic blood pressure, oxygen saturation, haemoglobin concentration, white cell count, and potassium, urea, creatinine, and glucose concentrations all had statistically significant interactions with hospital.

Conclusion A risk adjustment model for emergency medical admissions based on age, ICD-10 code, active malignancy, and routinely recorded physiological and blood variables can provide excellent discriminant value for seven day mortality across a range of settings. Using risk adjustment markedly changed hospitals' rankings. However, evidence was found that the association between key model variables and mortality were not constant.

Supplementary data appendix

Introduction

Around 5 million emergency hospital admissions occur each year in England, and about 4% of these result in death in hospital.[1] The UK Department of Health is developing performance indicators for emergency and urgent care that are intended to be clinically credible and evidence based outcome measures.[2] Mortality is undoubtedly an important outcome in emergency care, but comparison of crude mortality rates may be confounded by differences in case mix. Risk adjustment models may be used to assess the quality of emergency healthcare.[3 4] Observed mortality among emergency admissions can be compared with predicted risk adjusted mortality to determine whether the number of deaths exceeds the expected rate, with case mix taken into account.[5]

Case mix adjusted estimates of hospital mortality have been used to look for poor quality care by using routinely collected

Extra material supplied by the author (see http://www.bmj.com/content/344/bmj.e2904?tab=related#webextra)

data,[6-8] most notably at Mid Staffordshire NHS Trust.[9] However, these methods have been criticised as providing potentially misleading measures of quality of care.[10 11] The shortcomings of existing methods may be due to failure to adjust routine data adequately and reliably for differences in case mix.[12] Existing methods adjust for age, sex, and comorbidities but do not adjust for severity of illness, as indicated by physiological measures or blood tests.[6 7] This probably reflects a lack of available information systems to allow incorporation of these variables into risk adjustment models, as clinical risk prediction tools are typically based on physiological measures of severity of illness rather than on comorbidities.[13] In critical care, where information systems routinely collect physiological blood data, severity scoring methods such as the acute physiology and chronic health evaluation (APACHE) II and simplified acute physiology score (SAPS) II are used to produce risk adjusted estimates of mortality.[14 15]

The development of electronic data collection systems in emergency and pre-hospital care raises the potential for routine collection of measures of severity of illness and their incorporation in risk adjustment models. However, this may require additional data collection and the overcoming of substantial problems of data linkage, which can be justified only if additional variables improve risk adjustment. Furthermore, the addition of variables may not overcome another limitation of risk adjustment models, the constant risk fallacy,[16] whereby an association between a predictor variable and outcome is assumed to be constant whereas it actually varies between settings. For example, age is often included in risk adjustment because older age is associated with higher risk of death. Conventional risk adjustment models assume that the risk associated with age is constant, so the difference in risk associated with being 40 or 70 years old is the same in all settings. However, the risk associated with age may actually differ between healthy populations with long life expectancy and unhealthy populations with short life expectancy. Conventional risk adjustment models commit the fallacy of assuming that the difference in risk between 40 and 70 year olds is the same everywhere. Failure to recognise the constant risk fallacy can result in a model paradoxically increasing the effect of differences in case mix on the outcome rather than reducing it.[17]

We aimed to derive and validate a risk adjustment model for predicting seven day mortality in emergency medical admissions by using routinely collected data, pre-hospital and emergency department physiological data, and routine blood test results. We specifically aimed to determine the value of adding physiological data and blood data to basic risk adjustment models and to explore whether the risk associated with each model variable was constant across a range of settings or was subject to the constant risk fallacy.

Methods

Setting, participants, and data collection

The study took place in emergency departments in Sheffield, Barnsley, Rotherham, Hull, York, Leicester, and Northampton in the UK, and in Hong Kong and Melbourne, Australia. The first three hospitals each contributed two cohorts (derivation and validation), whereas the other hospitals each contributed a single validation cohort. Patients were eligible for inclusion if they were alive and not in cardiac arrest when attended by an emergency ambulance and then either were admitted to hospital or died in the ambulance or emergency department. We excluded children (under 16 years), women with obstetric emergencies,

adults with primarily mental health emergencies, and injured adults aged under 65. We identified patients at the UK sites retrospectively by review of hospital computer systems; patients in Hong Kong and Melbourne were identified prospectively by research staff working in the emergency department. We used different methods in different hospitals in response to differences in the ability of routine data systems to identify relevant cases and differences in availability of research staff for data collection.

We identified deaths up to seven days from hospitals' computer records, augmented by lists from local coroner's offices in the derivation phase. A researcher abstracted emergency department data, including patients' age, sex, physiological data (heart rate, respiratory rate, blood pressure, peripheral oxygen saturation, temperature, and Glasgow coma score), recorded comorbidities, and hospital admission within the previous 30 days, from hospital records. Paramedics routinely recorded physiological data in the ambulance on the standard patient report forms. Data from these forms were then either scanned into an electronic database or manually abstracted by a researcher. We then matched ambulance data to emergency department data by using the ambulance dispatch code. Wherever possible, we used the first physiological recording (that is, the ambulance recording). Where no physiological data were recorded in the ambulance or the cases could not be matched to the patient report form, we used the emergency department physiological data. Each patient had an ICD-10 (international classification of diseases, 10th revision) code attributed by hospital clerical staff as part of routine management, usually around two months after the initial presentation to hospital. We searched blood test data from the hospital laboratories (full blood count; urea, creatinine, potassium, sodium, and glucose concentrations) to identify the first blood result up to 24 hours after initial hospital attendance that matched with the hospital number of each patient. All data were entered on to a secure online database managed by the University of Sheffield Clinical Trials Research Unit.

Model derivation

To explore the univariable association between continuous variables and mortality, we plotted mortality against deciles of each variable. Age seemed to have a linear association with mortality, whereas other variables had more complex associations. We therefore categorised these variables into normal, low/high, and very low/high categories on the basis of their association with mortality and, where applicable, recognised clinical normal ranges. For peripheral oxygen saturation, we used different thresholds for low and very low saturation for recordings with and without supplemental oxygen. We initially grouped ICD-10 codes according to their chapter, but we then divided the two chapters with the largest number of cases (XI "Diseases of the digestive system" and X "Diseases of the respiratory system") into subgroups of diseases with similar mortality and amalgamated chapters with a small number of cases into a group of "others." The supplementary data appendix gives details of categorisations.

To explore missing data, we calculated the proportion of each variable that was missing among dead patients and survivors at seven days. A substantial proportion of patients, particularly among the survivors, had no blood test data. Given this high rate of missing blood data that seemed to be associated with outcome, we decided to develop two models: one without blood results using data from all patients (the physiology model) and one with blood results using data only from those with blood results (the full model). Because multivariable logistic regression excludes patients who do not have complete data for all variables

in the model, we investigated two methods for handling missing data under both models: multiple imputation and replacing missing values with sex specific means.

We analysed the univariable association between each variable and mortality by using logistic regression. We included only variables with a significant association with mortality at the 10% level (that is, P<0.1) in multivariable analyses, determining improvement of the model by using the likelihood ratio test with a 5% level of statistically significant improvement. To estimate how much each additional level of data contributed to the model, we developed a succession of multivariable models using increasing numbers of variables. The basic model used age and ICD-10 code only. Subsequent models included jointly predictive comorbidities, physiological variables, and blood results, either separately or in combination. We evaluated models incorporating blood tests only in cases with blood test data. We tested the other models separately in patients with and without blood tests to determine if differences between models were explained by selection of patients.

We fitted the risk score as an explanatory variable in a logistic regression model with mortality as the outcome. We used two standard criteria to assess the model's validity: log likelihood (testing whether additional variables improved the overall fit of nested models by using likelihood ratio tests); and sensitivity and specificity (using receiver operating characteristics curves and C statistics to quantify the sensitivity and specificity of the model).

Model validation

In the validation phase, we tested two models separately: the physiology model (age, ICD-10, active malignancy, and physiology) in all patients; and the full model (age, ICD-10, active malignancy, physiology, and bloods) only in patients with blood data. We decided to include all physiology and blood variables in the respective models, even if not all physiology and blood tests were predictors in the full model, because the logistics of collecting data meant that the non-predictive variables were automatically collected alongside the predictive variables and the distinction between a variable being an independent predictor or not often depended on the threshold for statistical significance as much as the strength of association.

We tested the model's validity in two ways in each setting to reflect the way the model can be used in practice: using coefficients from across the whole validation cohort, as might be used in research or national audit; and using separate coefficients for each validation site, as might be used in local audit. We assessed the model's performance by calculating the C statistic (with a 95% confidence interval) for each analysis.

We explored how the model would be implemented in practice by using it to estimate the expected number of deaths in each centre and calculate a standardised mortality ratio. We carried out this process three times using models of progressive complexity: the basic model consisting of age, ICD-10, and active malignancy; the physiology model outlined above; and the full model outlined above. We ranked the centres according to their observed death rate or standardised mortality ratio by using each model and derived 95% confidence intervals. We tested the rank correlation between different ranking methods to determine the extent to which the model changed centres' ranks. We repeated this process using an alternative method to estimate the effect of each centre on outcome. We included each centre as a covariate in the model and used the centre coefficient to estimate the effect of centre on mortality after adjustment for model covariates.

Finally, we tested for evidence of the constant risk fallacy by testing for interactions between the centre and each predictor variable in the model individually against outcome (death at seven days) to investigate whether the risk for a given factor was constant across centres, as evidence of interactions indicates that risk is not constant between centres.[16]

Results

Derivation phase

The derivation phase included 2381 eligible cases in Sheffield (11 February to 5 May 2008), 1626 cases in Barnsley (19 November 2007 to 24 February 2008), and 1637 cases in Rotherham (19 November 2007 to 25 February 2008). Overall seven day mortality was 311/5644 (5.5%). The mean age of the derivation cohort was 66.8 years, and 2687 (47.6%) were male. The supplementary data appendix gives details of missing data, univariable analysis, and multivariable analysis. Physiological variables had high rates of completeness, but around a third of patients were missing blood data. Dead patients had slightly higher rates of missing data. Comparison of the multiple imputation approach and the simpler method of imputing missing values as sex specific means showed no qualitative difference in the interpretation of the results between the two methods. Univariable analysis showed that age, ICD-10 code, active malignancy and chronic respiratory disease (comorbidities), steroid treatment, and all physiological and blood variables were significant predictors of mortality. Sex, diabetes, epilepsy, and heart disease (comorbidities) and recent hospital admission did not predict mortality.

We did two separate multivariable analyses—one including all patients but without blood data and the other limited to those with adequate blood data. In the first model age, ICD-10 code, active malignancy, and all physiological variables were important predictors of mortality. In the second model, heart rate and systolic blood pressure were less predictive, whereas white cell count and potassium concentration were important predictors of mortality, urea and creatinine concentration were marginal predictors, and haemoglobin, platelets, and sodium and glucose concentrations were poor predictors.

We then tested models of increasing complexity in patients with and without blood test data (table 1⇓). A model based on age and ICD-10 code alone had a C statistic of 0.80 (95% confidence interval 0.78 to 0.83). Adding active malignancy improved the discriminant value slightly (C statistic 0.81, 0.79 to 0.84), and adding physiological variables had a more marked effect (0.87, 0.85 to 0.89). The C statistics for these models were slightly higher when we limited analysis to patients with blood test data (0.81, 0.83, and 0.88). Adding blood variables to the basic model (age, ICD-10, and malignancy) improved the C statistic to 0.87 (0.84 to 0.89), whereas adding both physiological and blood variables (that is, the full model) improved the C statistic to 0.90 (0.88 to 0.92). The likelihood ratio tests showed that the improvement in the models' fit and the associated C statistics were statistically significant at the 5% level.

Validation phase

The validation phase included 13 762 patients across nine hospitals (n=1017-2305 per hospital) between 27 September 2008 and 25 July 2010. The supplementary data appendix gives details. Mean age varied across the hospitals from 64.3 to 75.6 years, and seven day mortality varied from 4.2% to 6.9%. The proportion with missing blood data varied markedly and was very low in hospitals B and E (0.6% and 1.5%), moderate in hospitals A, C, H, and I (13.0-16.3%), and high in hospitals D,

F, and G (45.5-70.0%). The variation among these sites reflects the success or failure to achieve record linkage.

Table 2⇓ shows the C statistics, goodness of fit, and log likelihood ratios for the physiology model (age, ICD-10, malignancy, and physiology) according to the source of the coefficients; table 3⇓ shows these statistics for the model including blood data. The discriminant value of the model is slightly higher when centre specific coefficients are used. Overall, the C statistics range from 0.80 to 0.91 for the physiology model and from 0.83 to 0.93 for the full model, suggesting that the models perform reasonably well in a variety of settings.

Table 4⇓ shows the expected number of deaths and the standardised mortality ratio that would be generated if the model were used to estimate risk adjusted mortality in each centre and the coefficient for each centre when included in the model. The standardised mortality ratios and coefficients for each centre were ranked from 1 (lowest ratio or coefficient) to 9 (highest). Table 5⇓ shows the Spearman correlation between ranks generated by observed mortality and ranks generated by different models. Correlations between mortality rates are shown in the bottom left corner of the table and correlations between coefficients are shown in the top right. We found greater correlation between ranks generated by different risk adjustment models than between rank based on observed mortality and ranks generated by the models. This suggests that risk adjustment markedly changes hospital ranking compared with ranking based on crude mortality but the specific model used does not markedly change hospital ranking.

Table 6⇓ shows the results of tests for interaction between centre and the association between each model variable and outcome. The table summarises whether an improvement in the model's fit between outcome and the denoted variable occurs if interaction between the two is included. Many of the variables used in the model have significant interactions with centre, suggesting that they may be subject to the constant risk fallacy.

Discussion

We have derived and validated a risk adjustment model for emergency medical admissions based on age, ICD-10 code, active malignancy, and routinely recorded physiological and blood variables that provided good discriminant value for seven day mortality in a variety of settings. This model could be used to estimate risk adjusted mortality as a quality indicator for emergency care. Age and ICD-10 code are routinely available for risk adjustment in emergency medical admissions and are key elements in models currently used to estimate hospital standardised mortality ratios.[9] We found that a model based on age and ICD-10 code alone had reasonable discriminant value, with a C statistic of 0.80. Sex and comorbidities are also routinely recorded, but these were poor predictors of mortality in our analysis.

Electronic recording of physiological variables and linkage between administrative, clinical, and laboratory databases would be needed to improve prediction to the degree suggested by our analysis, but this may be problematic. ICD-10 coding is not done until several weeks after hospital admission, thus delaying the time point at which risk adjustment can be done. We were unable to match a substantial proportion of admissions data to blood data, so we developed separate models with and without blood data. Our findings suggest that adding physiological variables and adding blood variables result in similar improvements to the model's prediction, but both need to be added to maximise prediction.

Risk adjustment markedly changed the ranking of hospitals from that based on observed mortality, but using a more complex model did not result in substantial further changes in ranking. The model had slightly better discriminant value when we used centre specific coefficients than when we used whole cohort coefficients. Centre specific coefficients would be appropriate for monitoring performance over time in a particular institution or service, whereas whole cohort coefficients would be appropriate for comparing performance across multiple sites. Research typically uses coefficients from a multivariable model to estimate the effect of each centre on outcome, whereas audit typically uses the model to generate a standardised mortality ratio for each centre. We found that the choice of method made a small difference to the ranking of centres with the more complex models.

Limitations

The main limitation of the model highlighted by our analysis was that many of the key variables in the model had significant interactions with centre, suggesting that they are subject to the constant risk fallacy.[16] In other words, the association between the variable and mortality varies between the study centres. Non-constant risk can arise because the variable reflects true differences in underlying risk in different populations (for example, the risk associated with age would differ between populations with different life expectancies), because the variable is measured or recorded differently (for example, at different times in different centres such as in the ambulance or later in the emergency department), because differences in quality of care between centres differ between subgroups (such as patients with minor or serious conditions), or because of a combination of these factors. Using the model to assess hospitals' performance by comparing risk adjusted mortality can result in misleading conclusions being drawn if evidence of non-constant risk exists.

Comorbidities and elective/emergency admission may be recorded in different ways in different centres, so these variables may be subject to the constant risk fallacy.[17] "Service" related variables (such as type of admission) or variables that are highly dependent on coding practices (such as number of comorbidities) might be hypothesised to be more prone to variation between centres than are biological variables, particularly blood variables for which measurement is automated. However, our study has shown that physiology and blood also exhibit non-constant risk. We can only speculate as to why this may be. Variation in the risk associated with physiological variables could be explained by the second type of variation outlined above (that is, differences in the timing, technique, or interpretation of measurement at different centres). Variation in the risk associated with blood measures could be explained by the first type of variation (that is, true population differences) or the third type of variation (differences between centres in the care provided to patients with different blood results).

Conclusions and implications for policy

Interest is increasing in using outcome measures to evaluate quality of emergency care.[18] The UK Department of Health is developing performance indicators for emergency and urgent care that are intended to be clinically credible and evidence based outcome measures.[2] Mortality rates in emergency admissions have been used to draw conclusions about the quality of emergency care,[19 20] and hospital standardised mortality ratios have been developed to evaluate risk adjusted mortality across emergency and elective admissions.[6 9]

Our data suggest that a risk adjustment model based on age, ICD-10 code, active malignancy, physiological variables, and (if available) blood variables can be used to produce risk adjusted estimates of mortality with good discriminant value in a variety of different settings. Our model can be used in a system of emergency care (hospital/ambulance service) to produce repeated estimates of risk adjusted mortality over time and thus monitor performance. If risk adjusted mortality were seen to increase, this might raise concerns about quality of care and prompt more detailed investigation. However, interpretation of any change in risk adjusted mortality would need to take into account the possibility of random error or failure of risk adjustment to adequately adjust for changes in case mix and illness severity.

Our model can also be used to compare risk adjusted mortality between different systems of emergency care and draw inferences about their relative performance. Risk adjusted estimates of mortality can be used in this way to produce hospital league tables or identify apparently poorly performing services or institutions. However, this potential use of risk adjustment is controversial and subject to additional limitations (other than random error and failure to adequately adjust). We found evidence that the constant risk fallacy affects key model variables. If risk adjustment is done using variables that have a non-constant association with outcome, then differences in mortality due to case mix or severity of illness may be exaggerated by risk adjustment rather than being accounted for. Conclusions about the relative performance of services or institutions based on risk adjusted mortality may then be very misleading.

The policy implications of our study are that risk adjusted estimates of mortality from our model can provide useful insights into the performance of a system of emergency care over time. However, risk adjusted mortality cannot be reliably used to compare the performance of systems of emergency care or to draw conclusions about relative quality of care. Analysis of risk adjusted mortality can provide valuable insights when used with appropriate caution,[9] but it may be damaging if erroneous conclusions are drawn on the basis of misleading analysis.[11] [12]

1   Blunt I, Bardsley M, Dixon J. Trends in emergency admissions in England 2004-2009. The Nuffield Trust, 2011.
2   Department of Health. Revision to the operating framework for the NHS in England 2010/11. Department of Health, 2010.
3   Stukenborg GJ, Wagner DP, Harrell FE Jr, Oliver MN, Heim SW, Price AL, et al. Which hospitals have significantly better or worse than expected mortality rates for acute myocardial infarction patients? Improved risk adjustment with present-at-admission diagnoses. *Circulation* 2007;116:2960-8.
4   Iezzoni LI, Shwartz M, Ash AS, Hughes JS, Daley J, Mackiernan YD. Using severity-adjusted stroke mortality rates to judge hospitals. *Int J Qual Health Care* 1995;7:81-94.
5   Iezzoni LI. Reasons for risk adjustment. In: Iezzoni LI, ed. Risk adjustment for measuring health care outcomes. Health Administration Press, 2003:1-5.
6   Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999;318:1515-20.
7   Shahian DM, Wolf RE, Iezzoni LI, Kirle L, Normand S-LT. Variability in the measurement of hospital-wide mortality rates. *N Engl J Med* 2011;363:2530-9.
8   Miyata H, Hashimoto H, Horiguchi H, Matsuda S, Motomura N, Takamoto S. Performance of in-hospital mortality prediction models for acute hospitalization: hospital standardized mortality ratio in Japan. *BMC Health Serv Res* 2008;8:229.
9   Bottle A, Jarman B, Aylin P. Strengths and weaknesses of hospital standardised mortality ratios. *BMJ* 2011;342:c7116.
10 Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 2010;340:955-7.
11 Black N. Assessing the quality of hospitals: hospital standardised mortality ratios should be abandoned. *BMJ* 2010;340:933-4.
12 Lilford RJ, Brown CA, Nicholl J. Use of process measures to monitor the quality of clinical practice. *BMJ* 2007;335:648-50.
13 Challen K, Goodacre S. Predictive scoring in non-trauma emergency patients: a scoping review. *Emerg Med J* 2011;28:827-37.
14 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13:818-29.
15 Le Gall J-R, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993;270:2957-63.
16 Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *J Epidemiol Community Health* 2007;61:1010-3.
17 Mohammed MA, Deeks JJ, Girling A, Rudge G, Carmalt M, Stevens AJ, et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ* 2009;338:b780.
18 Cameron PA, Schull MJ, Cooke MW. A framework for measuring quality in the emergency department. *Emerg Med J* 2011;28:735-40.
19 Sprivulis CP, Da Silva JA, Jacobs IG, Frazer AR, Jelinek GA. The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments. *Med J Aust* 2006;184:208-12.
20 Guttmann A, Schull MJ, Vermeulen MJ, Stukel TA. Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from Ontario, Canada. *BMJ* 2011;342:d2983.

Cite this as: *BMJ* 2012;344:e2904

What is already known on this topic

Quality of emergency care is assessed mainly by comparing process measures, such as times to treatment, rather than outcomes, such as mortality

Risk adjustment models using routine administrative data have been used to compare hospital standardised mortality rates for all admissions (elective and emergency)

Physiological and blood variables have been used to predict mortality in clinical practice

What this study adds

A risk prediction model for mortality in emergency medical admissions based on age, ICD-10 code, active malignancy, and physiological and blood variables has good discriminant value across a range of settings

Linkage of routine hospital admission data to physiological and blood data improves risk prediction for quality assessment in emergency care

Key predictor variables have a non-constant association with mortality, so differences between hospitals in risk adjusted mortality must be interpreted with caution

Tables

Table 1| **Summary statistics for models tested in derivation phase**

Model	Subset	C statistic (95% CI)	$-2 \times$ log likelihood	Likelihood ratio test χ^2 (df) and P value
Age and ICD-10 alone	All patients (n=5644)	0.80 (0.78 to 0.83)	2028.91	–
+ active malignancy		0.81 (0.79 to 0.84)	1991.76	35.03 (3); P<0.001
+ active malignancy + physiology		0.87 (0.85 to 0.89)	1720.2	267.12 (15); P<0.001
Age and ICD-10 alone	Those with blood test data (n=3720)	0.81 (0.78 to 0.84)	1232.96	–
+ active malignancy		0.83 (0.80 to 0.85)	1208.22	24.75 (3); P<0.001
+ active malignancy + physiology		0.88 (0.86 to 0.91)	1023.56	184.65 (15); P<0.001
+ active malignancy + bloods		0.87 (0.84 to 0.89)	1085.59	122.062 (20); P<0.001
+ active malignancy + physiology + bloods		0.90 (0.88 to 0.92)	963.16	75.23 (20); P<0.001

Table 2| **Validation phase C statistics, Hosmer-Lemeshow goodness of fit, and log likelihood for physiology model**

Centre	Whole validation cohort coefficients			Centre specific validation coefficients		
	C statistic (95% CI)	Goodness of fit (df=10)	Log likelihood	C statistic (95% CI)	Goodness of fit (df=10)	Log likelihood
A	0.88 (0.84 to 0.93)	6.58 (P=0.58)	−2065.82	0.90 (0.86 to 0.94)	5.71 (P=0.68)	−111.56
B	0.80 (0.76 to 0.84)	9.90 (P=0.27)		0.82 (0.77 to 0.86)	13.18 (P=0.11)	−269.73
C	0.87 (0.83 to 0.91)	6.86 (P=0.55)		0.90 (0.86 to 0.93)	4.80 (P=0.78)	−185.24
D	0.88 (0.84 to 0.92)	2.81 (P=0.94)		0.91 (0.87 to 0.95)	3.88 (P=0.87)	−122.25
E	0.83 (0.79 to 0.88)	2.72 (P=0.95)		0.85 (0.81 to 0.90)	6.84 (P=0.55)	−232.07
F	0.84 (0.80 to 0.88)	7.85 (P=0.45)		0.87 (0.84 to 0.91)	7.29 (P=0.51)	−216.94
G	0.86 (0.82 to 0.89)	9.41 (P=0.31)		0.87 (0.84 to 0.91)	3.72 (P=0.88)	−245.44
H	0.86 (0.83 to 0.90)	8.30 (P=0.41)		0.86 (0.83 to 0.90)	11.98 (P=0.15)	−281.77
I	0.83 (0.78 to 0.88)	6.39 (P=0.60)		0.84 (0.80 to 0.89)	8.63 (P=0.37)	−220.51

Table 3| Validation phase C statistics, Hosmer-Lemeshow goodness of fit, and log likelihood for full model

Centre	Whole validation cohort coefficients			Centre specific validation coefficients		
	C statistic (95% CI)	Goodness of fit (df=10)	Log likelihood	C statistic (95% CI)	Goodness of fit (df=10)	Log likelihood
A	0.91 (0.87 to 0.95)	3.03 (P=0.93)	−1977	0.92 (0.88 to 0.97)	11.15 (P=0.19)	−93.4
B	0.83 (0.79 to 0.87)	10.07 (P=0.26)		0.84 (0.80 to 0.88)	9.96 (P=0.27)	−249.1
C	0.87 (0.83 to 0.91)	12.53 (P=0.13)		0.92 (0.89 to 0.95)	5.82 (P=0.67)	−170.42
D	0.89 (0.85 to 0.93)	6.49 (P=0.59)		0.93 (0.89 to 0.96)	9.57 (P=0.30)	−107.32
E	0.87 (0.83 to 0.91)	9.84 (P=0.28)		0.92 (0.89 to 0.95)	8.67 (P=0.37)	−184.65
F	0.85 (0.81 to 0.89)	7.93 (P=0.44)		0.90 (0.87 to 0.93)	5.73 (P=0.68)	−198.68
G	0.87 (0.84 to 0.90)	15.16 (P=0.06)		0.89 (0.86 to 0.92)	7.41 (P=0.49)	−229.09
H	0.88 (0.85 to 0.91)	9.41 (P=0.31)		0.90 (0.86 to 0.93)	5.71 (P=0.68)	−254.06
I	0.86 (0.82 to 0.91)	3.34 (P=0.91)		0.90 (0.86 to 0.93)	7.18 (P=0.52)	−186.46

Table 4| Comparison of observed and expected death rates (and SMR) and centre coefficients for each model

Centre	Observed		Age, ICD-10, and malignancy			Physiology model			Full model		
	Deaths/patients (%); rank	Centre coefficients; rank	Expected (%)	SMR (95% CI); rank	Centre coefficients; rank	Expected (%)	SMR (95% CI); rank	Centre coefficients; rank	Expected/N* (%)	SMR (95% CI); rank	Centre coefficients; rank
A	43/1008 (4.3); 1	0.00; 1	45 (4.4)	96.4 (90.4 to 102.3); 4	0.00; 4	49 (4.9)	87.5 (82.1 to 92.9); 2	0.00 3	41/980 (4.2)	103.7 (97.2 to 110.2); 5	0.00; 4
B	95/1515 (6.3); 9	0.41; 9	103 (6.8)	92.6 (87.9 to 97.2); 3	−0.06; 3	93 (6.1)	101.6 (96.5 to 106.8); 4	0.25; 5	106/1497 (7.1)	88.3 (83.8 to 92.8); 2	−0.06; 1
C	70/1592 (4.4); 2	0.03; 2	63 (4.0)	111.2 (105.8 to 116.7); 6	0.11; 6	62 (3.9)	112.6 (107.8 to 118.2); 8	0.31; 8	48/1268 (3.8)	118.0 (111.5 to 124.5); 7	0.32; 6
D	56/1043 (5.4); 6	0.24; 6	50 (4.8)	112.9 (106.1 to 119.8); 7	0.16; 7	54 (5.2)	103.3 (97.0 to 109.6); 5	0.22; 4	16/311 (5.1)	124.1 (110.3 to 137.9); 9	0.47; 8
E	83/1501 (5.5); 7	0.27; 7	74 (4.9)	112.9 (107.2 to 118.7); 8	0.17; 8	80 (5.3)	103.9 (98.6 to 109.1); 6	0.42; 9	78/1476 (5.3)	99.0 (93.9 to 104.0); 3	0.25; 5
F	77/1478 (5.2); 5	0.21; 5	76 (5.1)	101.9 (96.7 to 107.0); 5	0.04; 5	71 (4.8)	108.3 (102.8 to 113.8); 7	0.29; 7	39/776 (5.0)	111.7 (103.8 to 119.5); 6	0.37; 7
G	91/1494 (6.1); 8	0.38; 8	74 (5.0)	123.2 (116.9 to 129.4); 9	0.25; 9	81 (5.4)	112.9 (107.1 to 118.6); 9	0.28; 6	29/620 (4.5)	120.0 (110.6 to 129.5); 8	0.49; 9
H	104/2048 (5.1); 3	0.18; 3	118 (5.8)	87.9 (84.1 to 91.7); 2	−0.18; 2	116 (5.7)	89.3 (85.4 to 93.2); 3	−0.02; 2	94/1703 (5.5)	88.1 (83.9 to 92.3); 1	−0.03; 3
I	73/1434 (5.1); 4	0.19; 4	91 (6.3)	80.3 (76.1 to 84.5); 1	−0.32; 1	84 (5.9)	85.8 (81.4 to 90.3); 1	−0.06; 1	66/1204 (5.5)	99.9 (94.2 to 105.5); 4	−0.03; 2

SMR=standardised mortality ratio.

*N=subset of patients with blood data available and included in the model.

Table 5| Spearman correlation of mortality rates (upper right triangle) and centre coefficients (lower left triangle)

	Observed	Age, ICD-10, and malignancy	Physiology model	Full model
Observed	–	0.38	0.48	0.00
Age, ICD-10, and malignancy	0.38	–	0.82	0.63
Physiology model	0.32	0.73	–	0.55
Full model	0.12	0.83	0.47	–

Table 6| **Summary of statistical effect of interactions between covariates and centre**

Covariate	df	P value
Active malignancy	8	0.071
ICD-10 classification	84	0.000
Glasgow coma score	16	0.001
Heart rate	24	0.302
Respiratory rate	12	0.004
Systolic blood pressure	24	0.006
Oxygen saturation	16	0.008
Temperature	19	0.101
Haemoglobin concentration	21	0.043
White cell count	22	0.010
Platelet count	16	0.244
Sodium concentration	15	0.063
Potassium concentration	15	0.002
Urea concentration	16	0.000
Creatinine concentration	23	0.001
Blood glucose concentration	20	0.027