BMJ Open

The ALERT scale: an observational study of early prediction of adverse hospital outcome for medical patients

Daniel Roberts,[1] Ward Patrick,[2] Julie Mojica,[1] Patricia Ostryzniuk,[3] Margaret Patrick,[4] Chris MacKnight,[5] Allen Kraut,[6] Leigh Anne Shafer[1]

ABSTRACT

Objectives: Some medical patients are at greater risk of adverse outcomes than others and may benefit from higher observation hospital units. We constructed and validated a model predicting adverse hospital outcome for patients. Study results may be used to admit patients into planned tiered care units. Adverse outcome comprised death or cardiac arrest during the first 30 days of hospitalisation, or transfer to intensive care within the first 48 h of admission.

Setting: The study took place at two tertiary teaching hospitals and two community hospitals in Winnipeg, Manitoba, Canada.

Participants: We analysed data from 4883 consecutive admissions at a tertiary teaching hospital to construct the Early Prediction of Adverse Hospital Outcome for Medical Patients (ALERT) model using logistic regression. Robustness of the model was assessed through validation performed across four hospitals over two time periods, including 65 640 consecutive admissions.

Outcome: Receiver-operating characteristic curves (ROC) and sensitivity and specificity analyses were used to assess the usefulness of the model.

Results: 9.3% of admitted patients experienced adverse outcomes. The final model included gender, age, Charlson Comorbidity Index, Activities of Daily Living Score, Glasgow Coma Score, systolic blood pressure, respiratory rate, heart rate and white cell count. The model was discriminative (ROC=0.83) in predicting adverse outcome. ALERT accurately predicted 75% of the adverse outcomes (sensitivity) and 75% of the non-adverse outcomes (specificity). Applying the same model to each validation hospital and time period produced similar accuracy and discrimination to that in the development hospital.

Conclusions: Used during initial assessment of patients admitted to general medical wards, the ALERT scale may complement other assessment measures to better screen patients. Those considered as higher risk by the ALERT scale may then be provided more effective care from action such as planned tiered care units.

Strengths and limitations of this study

- Multiple hospital sites with diverse populations and two several-year time periods were assessed to test the reliability and stability over time of the Adverse Hospital Outcome for Medical Patients (ALERT) scale predictor of hospital outcomes.
- ALERT scale elements are already routinely collected or easily available on admission at all hospitals.
- ALERT scale results may be used to assign patients to different levels of observation, depending on likelihood of an adverse event such as cardiac arrest, death or transfer to intensive care unit.
- ALERT scale predictor cut-off may be adjusted by individual hospitals in order to match the level and distribution of local resources.
- The need for an equation to generate the probability of an adverse event; the equation is available in the manuscript as well as a website to ease ALERT scale use (http://www.alertassessmenttool.ca).

and mortality in hospitalised patients. Unfortunately, initiatives such as the medical emergency team (MET) concept and the Modified Early Warning Score[1][2] have not consistently led to decreases in hospital mortality, cardiac arrest and transfer to the intensive care unit (ICU).[3–6] Litvak et al[7] have postulated that necessity for MET intervention implies either inadequate care or inappropriate triage to an inadequate care environment, suggesting that another approach is needed.

In most hospitals, patients requiring admission are triaged either to the ICU or a general ward. Mortality predictors such as Acute Physiology and Chronic Health Evaluation (APACHE) II and Simplified Acute Physiology Score (SAPS) II, have been used to stratify patients following ICU admission.[8][9] Risk stratification tools for nonsurgical patients have been proposed based on laboratory data and some physiological

INTRODUCTION

In recent years, a number of attempts have been made to reduce preventable morbidity

measurements from retrospective administrative databases.[10] [11] The addition of the Charlson Comorbidity Index to the Rapid Emergency Medicine score was shown to influence long term but not 3-day or 7-day mortality.[12] These scales may be limited because they do not include measures of functional status (ie, Activities of Daily Living (ADL)) which may influence morbidity and mortality.

Early stratification of patients admitted to non-ICU environments for risk of subsequent physiological collapse, cardiac arrest and death would help determine the optimal care environment to which the patient should initially be admitted.[10] [11] We propose this alternative approach, the Early Prediction of Adverse Hospital Outcome for Medical Patients (ALERT) scale using demographic, routine physiological data and easily obtainable measures of comorbidity and physical disability, available at or near the time of the decision to admit. The primary advantage of our model is that data needed to inform the model is likely to be already collected on admission in most hospitals or easily added to data collected on admission. We hypothesize that we would find a set of indicators which would guide the decision about which ward to assign hospital admissions.

METHODS
Study population
The model development population comprised 4883 patients consecutively admitted to a tertiary hospital (TH$_1$) in Winnipeg, Canada from 1 October 2003 to 30 September 2005. Admissions originated in emergency departments (79%), ICU (8%) or from other sources including rural hospitals, nursing stations and other hospital services (13%). Validation was performed using data collected from 65 640 admissions in four sites and two time periods. The validation sites included two community hospitals (CH$_1$ and CH$_2$), a second tertiary hospital (TH$_2$) and the same hospital in which the model was developed (TH$_{1B}$). The validation time periods were October 2005 to December 2008 and January 2009 to December 2012. To assess the generalisability of results to populations with varied characteristics, we validated the model for each of the four hospitals and two time periods.

Study design
We developed a model predictive of adverse outcome defined as any one of the following:
1. Death during hospitalisation within 30 days of admission.
2. Cardiopulmonary resuscitation (CPR) during hospitalisation attempted within 30 days of admission.
3. Transfer to an ICU within 48 h of admission.

The ALERT scale is a predicted probability of adverse outcome, ranging between 0 and 1.

In 2004, the University of Manitoba Department of Internal Medicine funded and implemented a permanent, comprehensive administrative and research database

for its 13 general medical ward services in four Winnipeg hospitals. Data collection for this database is performed by ICU trained data collection nurses through concurrent chart review. Collector training, data management, and collection and verification processes had already been developed through the establishment of a long-standing regional ICU database. A primary objective of this initiative was to develop a model-based outcome stratification tool for medical patients. All data used in development and testing of the ALERT scale was already being routinely collected on hospital admission before the administrative and research database was developed, but previous to the database, the data was not all recorded electronically.

Data elements for the model were selected by consensus in a series of investigator meetings. The final data set included patient demographics and all the elements of APACHE II, SAPS II, Charlson Comorbidity Index (CCI) and the Activities of Daily Living (ADL) Score.[8] [9] [13] [14] The objective of ALERT is to discriminate patients for level of care using variables routinely available at the time of admission.

The ADL Scoring system was modified to increase discrimination by adding an intermediate category. We assigned zero points to independent performance of individual activities, three points for an activity performed by the patient but requiring physical assistance and six points when the patient was unable to assist with the activity (see online supplementary appendix 1).

Exclusion criteria
Patients admitted for palliative care or those with an advance directive precluding CPR or ICU admission were excluded. Patients transferred to non-study hospitals or units within 30 days where outcome data could not be obtained were excluded.

Statistical analysis
To develop the model, logistic regression was performed with adverse outcome as the dependent variable. Generalised estimating equations (GEE) were used to account for more than one admission during the study period by some individuals. After assessing clinical relevance, the independent variables were selected via a stepwise procedure. Both forward and backward selection was used. To assess potential non-linear relationships between independent variables and the probability of an adverse outcome, quadratic terms of all continuous variables were considered.

To classify patients in the higher or lower-risk of adverse outcome groups, an optimal cut-off point in terms of predicted probability of an adverse outcome was chosen. As an improvement in sensitivity necessarily means decline in specificity, and vice versa, we chose the cut-off which ensured equal sensitivity and specificity in the development population. The same cut-off point was then applied to the eight validation populations—four

hospitals and two validation time periods. The impact of using different cut-offs from that which equated sensitivity to specificity was assessed descriptively.

Discriminatory power, the ability to distinguish between good and adverse outcomes, was determined by the area under the receiver-operating characteristic (ROC) curve[15] and by sensitivity and specificity analyses. The ROC is a plot of true-positive rate (sensitivity) versus false-positive rate (1-specificity), as its discrimination threshold is varied. In our context, it is a plot of the proportion of those that had an adverse outcome that were predicted to have one, versus the proportion that did not have an adverse outcome who were misclassified and predicted to have had one. Discrimination threshold is the cut-off probability used for classifying patients.

The calibration of the model was evaluated using admissions from each of the four hospital study sites and two time periods, using the Hosmer-Lemeshow goodness-of-fit test. The Hosmer-Lemeshow goodness-of-fit statistic ($\hat{C}$) approximates a χ^2 distribution. Larger $\hat{C}$ values and smaller p values indicate significant differences between predicted and actual adverse outcomes, and thus a lack of fit of the model.

Characteristics of the development and validation populations were compared using Analysis of Variance (ANOVA) and χ^2 tests. The association between each of the top 30 primary admission diagnoses and subsequent adverse outcome were assessed using Pearson χ^2.

RESULTS

During the study period 75 189 admissions occurred. Of these, 4666 (6.2%) were excluded for one or more of the reasons provided in our exclusion criteria. For the remaining 70 523 admissions all measurements were obtained in 99.7% of the entries. The median duration of admission was 7.4 days. The four validation populations and two validation time periods are compared to the development population in table 1, demonstrating differences in age, burden of chronic illness, physical disability and physiological variables.

Charlson Comorbidity Score, ADL Score, Glasgow Coma Score, age, heart rate, respiratory rate, systolic blood pressure and white cell count were predictive of adverse outcome. Gender was forced in the model because of its known association with survival. The final model is displayed in table 2. Two variables, age and systolic blood pressure, had significantly non-linear associations with adverse outcome.

Our model results in the development population are described in table 3A, indicating good calibration ($\hat{C}$=9.91) and good discrimination (ROC=0.831). Application of the model to the four validation populations in the 2005–2008 validation period produced ROC of 0.78 (TH$_2$), 0.80 (CH$_3$), 0.79 (CH$_4$) and 0.81 (TH$_{1B}$), indicating very similar discrimination to that obtained in the model development population. Similarly, in the 2009–2012 validation period, we found ROC of 0.77 (TH$_2$), 0.75 (CH$_3$), 0.80 (CH$_4$) and 0.79 (TH$_{1B}$). Figure 1 displays the ROC curves for the development

Table 1 Comparison of patient characteristics between model development population (TH1), each of the validation populations and each of the validation periods (Mean±SD)

Variable	TH1	Time	TH2	CH3	CH4	TH1B
N	4883		6646/8561*	6982/7765*	5762/6787*	10 069/13 068*
Age	60±18	2005–2008	67±18	74±16	74±16	60±18
		2009–2012	66±18	74±16	74±17	59±18
CCI	3.0±2.9	2005–2008	3.0±2.3	2.5±2.1	2.1±1.8	2.9±2.4
		2009–2012	3.0±2.4	2.8±2.1	2.9±2.1	3.1±2.5
ADLS	13.7±12.4	2005–2008	12.0±11.3	16.1±12.4	15.8±12.8	13.6±12.5
		2009–2012	12.4±11.6	16.9±13.9	17.9±12.5	9.4±12.4
GCS†	81%	2005–2008	79%	75%	70%	81%
		2009–2012	78%	72%	64%	83%
HR	86.1±20.7	2005–2008	84.8±19.8	84.1±20.01	83.4±20.5	85.6±19.8
		2009–2012	86.1±18.2	83.9±19.0	83.6±19.1	86.6±19.1
RR	20.4±4.3	2005–2008	20.3±4.4	20.6±4.3	19.6±3.8	19.9±4.2
		2009–2012	19.3±4.1	19.8±3.8	19.3±4.3	19.3±4.2
WCC	10.2±8.4	2005–2008	10.7±10.4	10.7±6.6	10.0±7.5	10.6±9.7
		2009–2012	11.2±10.3	11.2±8.7	10.7±8.0	11.4±12.2
SBP	129±31	2005–2008	130±27	131±27	135±29	133±28
		2009–2012	129±25	131±24	132±25	132±26
Male	52%	2005–2008	48%	47%	43%	53%
		2009–2012	49%	48%	45%	52%

Age in years.
*Sample sizes in the 2005–2008 and 2009–2012 validation periods.
†Percentage of patients with maximum GCS (15).
ADLS, Activities of Daily Living Score; CCI, Charlson Comorbidity Index; GCS, Glasgow Coma Score; HR, heart rate as beats per minute; RR, respiratory rate as breaths per minute; SBP, systolic blood pressure as mm of Hg; WCC, white cell count as cells per mL^3 divided by 10^3.

Table 2 Logistic Regression Model (GEE) predicting probability of an adverse outcome (30 day death or cardiac arrest, or 48 h ICU transfer) from the development population TH_1, N=4883

Predictor	Coefficient (SE)	OR (95% CI)*	p Value
Intercept	−3.839 (1.136)		0.0007
Male gender	0.157 (0.131)	1.170 (1.026 to 1.334)	0.1650
Charlson Comorbidity Score	0.071 (0.016)	1.074 (1.057 to 1.091)	<0.0001
Activities of Daily Living Score	0.053 (0.005)	1.054 (1.049 to 1.060)	<0.0001
Glasgow Coma Score	−0.172 (0.032)	0.842 (0.815 to 0.869)	<0.0001
Age	0.068 (0.023)		0.0027
Age^2†	−0.0004 (0.0002)		0.0269
Heart rate	0.018 (0.003)	1.018 (1.015 to 1.021)	<0.0001
Respiratory rate	0.063 (0.001)	1.065 (1.064 to 1.066)	<0.0001
White cell count	0.029 (0.006)	1.029 (1.023 to 1.036)	<0.0001
Systolic blood pressure (SBP)	−0.041 (0.011)		0.0002
SBP^2†	0.000 121 (0.000039)		0.0038

*The ORs for the two non-linear terms (age and SBP) could not be directly calculated as they vary.
†The Age^2 and SBP^2 terms allow non-linear associations between Age and SBP, and probability of adverse outcome.

population and each of the four validation sites and two validation periods.

Although the discrimination of the model was similar across the four validation sites and two validation periods, calibration of the model declined with time (data not shown in tables). During the 2005–2008 validation period, the Hosmer-lemeshow goodness-of-fit statistics ($\hat{C}$) were: 81.99 (p<0.01) in TH_2, 13.34 (p=0.22) in CH_3, 17.62 (p=0.06) in CH_4 and 20.66 (p=0.03) in TH_{1B}. During the 2009–2012 validation period, the corresponding $\hat{C}$-statistics were: 92.31 (p≤0.01) in TH_2, 55.93 (p≤0.01) in CH_3, 67.34 (p≤0.01) in CH_4 and 55.81 (p≤0.01) in TH_{1B}. The lower the $\hat{C}$-statistic and higher the p value, the better the model fits with the data.

A cut-off point of 0.088, or a predicted probability of adverse outcome of 8.8%, was used to classify patients into two groups; high-risk and low-risk of adverse outcome. Using this cut-off for each validation population resulted in sensitivity in 2005–2008 ranging from 62.5% to 76%, and specificity ranging from 66.2% to 77.1% (table 4). Corresponding sensitivity and specificity in each validation population in 2009–2012 ranged from 58.3% to 81.2% (sensitivity), and from 62.6% to 82.9% (specificity). These are very near the sensitivity and specificity of 75.1% and 75.2% found in the model development population (table 4).

Cut-offs other than 8.8% were examined. Adverse outcome was experienced by 454 (9.3%) of patients in the model development population (bottom of column 3, table 3B). Consequently, the majority had a predicted probability of adverse outcome of less than 10%. Using a cut-off of 8% would result in 32% of all patients classified as high risk (column 8, table 3B) and would provide 77.8% sensitivity (column 5, table 3B) and 72.7% specificity (column 6, table 3B). This demonstrates correct predictability of 77.8% of all adverse outcomes. However, the hospital may be required to provide additional monitoring to many with no adverse outcome.

With this cut-off, just 22.6% of those predicted at high risk had an adverse outcome (column 7, table 3B).

An alternative strategy may be to increase sensitivity at the expense of specificity. If an extreme cut-off such as 40% were used, then just 4.3% of all admissions would be classified as high risk (column 8, table 3B). Over half, (52.9%) of those classified as high risk would have an adverse outcome (column 7, table 3B) and 97.8% of those classified as low risk would be correctly classified (specificity, column 6, table 3B). This could save beds in a more intensely monitored hospital unit, but at a cost that only 24.2% of those with adverse outcome would be correctly classified (sensitivity, column 5, table 3B).

Incidence of one or more of adverse outcomes among the eight validation populations ranged from 6.3% to 10.6% and included death within 30 days (5.1–9.8%), CPR within 30 days (0.5–1.8%) and ICU admission within 48 h (1–1.9%). These events were not mutually exclusive. For example, most patients who suffered cardiopulmonary arrest died.

Among the 30 diagnoses most responsible for admission, four may be associated with an increased likelihood of adverse outcome (table 5). Among all admissions, 9% experienced one or more adverse outcome. However, 12.6% (688 out of 5452) admitted with pneumonia, 13.9% with cerebral vascular accident, 14.9% with pleural effusions and 28.9% with primary lung malignancy experienced adverse outcomes. With the exception of primary lung malignancy, however, adverse outcomes never occurred more than two times more frequently among those admitted with a specific diagnosis than among all admitted patients.

DISCUSSION

Given the frequency of undesired and often catastrophic outcomes on hospital wards, it is reasonable to question whether the available range of monitoring and surveillance modalities is appropriate for all patients. Our

Table 3 Model assessment in development population (TH$_1$)

(A) Hosmer-Lemeshow Goodness-of-Fit test ($\hat{C}$)

| Risk bands | Cases (N) | Mean predicted risk | Adverse outcome | | χ^2 |
			Observed	Expected	
0.00–0.04	2291	0.021	40	48	1.44
0.04–0.08	1029	0.057	61	59	0.09
0.08–0.10	266	0.089	24	24	0.00
0.10–0.12	237	0.109	32	26	1.61
0.12–0.16	289	0.138	42	40	0.11
0.16–0.20	174	0.179	31	31	0.00
0.20–0.30	264	0.241	62	64	0.05
0.30–0.40	125	0.347	52	43	2.64
0.40–0.60	138	0.486	65	67	0.12
0.60–1.00	70	0.745	45	52	3.84
Total	4883	0.093	454	454	9.91

$\hat{C}$ = 9.91 (p=0.276), 8 d.f.   ROC=0.831

(B) Model predicted versus actual adverse outcome

| Predicted probability of adverse outcome | Adverse outcome | | | Percent of adverse with this or higher prediction (sensitivity) | Percent non-adverse with lower than this prediction (specificity) | Percent with this or higher prediction who have adverse outcome | Number of patients with this or higher prediction (percent) |
	No	Yes	Total				
0–1%	1120	10	1130	100	0.0	9.3	4883 (100)
2–3%	1131	30	1161	97.8	25.3	11.8	3753 (76.9)
4–5%	593	31	624	91.2	50.8	16.0	2592 (53.1)
6–7%	375	30	405	84.4	64.2	19.5	1968 (40.3)
8–9%	242	24	266	77.8	72.7	22.6	1563 (32)
10–14%	404	66	470	72.5	78.1	25.4	1297 (26.6)
15–19%	191	39	230	57.9	87.3	31.8	827 (16.9)
20–39%	275	114	389	49.3	91.6	37.5	597 (12.2)
40–59%	73	65	138	24.2	97.8	52.9	208 (4.3)
60–79%	22	25	47	9.9	99.4	64.3	70 (1.4)
80–100%	3	20	23	4.4	99.9	87.0	23 (0.5)
Total	4429	454	4883				

study also demonstrated that a substantial proportion of patients are at low risk for such events. Initial identification of substantial proportions of the patient population at either extreme of the risk spectrum could lead to more effective and efficient allocation of nursing staff and monitoring modalities that would match resources to patient need in tiered care units.

A useful predictor of adverse hospital outcome should demonstrate good discrimination for heterogeneous patient populations in tertiary and community hospital settings. It should be independent of primary diagnosis due to the wide variety of possible diagnoses and often the difficulty at the time of admission to make an accurate diagnosis. As well, it should be simply derived from data routinely obtained at the time of initial evaluation. The ALERT scale model meets these requirements more clearly than those previously reported.[10 11]

We overcame several limitations of previous models by including admissions from multiple sources, thereby expanding the population to which the model may be generalised. Unlike Prytherch et al[10] and Olsson et al,[11] who used admissions from one emergency department, our validation procedure used admissions from four tertiary and community hospitals and two time periods, representing a broader spectrum of hospitalised medical patients. Our results demonstrate a better calibration than that reported for the Rapid Emergency Medical Score,[11] and better discrimination than those previously described.[10 11]

Including degree of comorbidity (CCI) and chronic disability (ADL Score), in addition to acute physiological data, may account for the increased robustness of our model. A report by Olsson suggested that combining the CCI with the Rapid Emergency Medical Score added to its prognostic value[12] for long-term outcomes. ALERT is the first to incorporate a functional measure, the ADL score, which is associated with mortality, into the predictive equation of short-term adverse hospital outcomes.

The ALERT scale is a better scoring system than those identified in previous studies because it includes easily collectable data at the time of admission, does not require diagnosis, which is often difficult to correctly identify at the time of admission and is the only one that

Figure 1 Receiver-operating characteristic ROC curves: Plot of sensitivity and 1-specificity for all possible cut points in the development group (TH1) and eight validation groups (namely, TH1B, TH2, CH3 and CH4 at two periods October 2005–December 2008 and January 2009–December 2012.
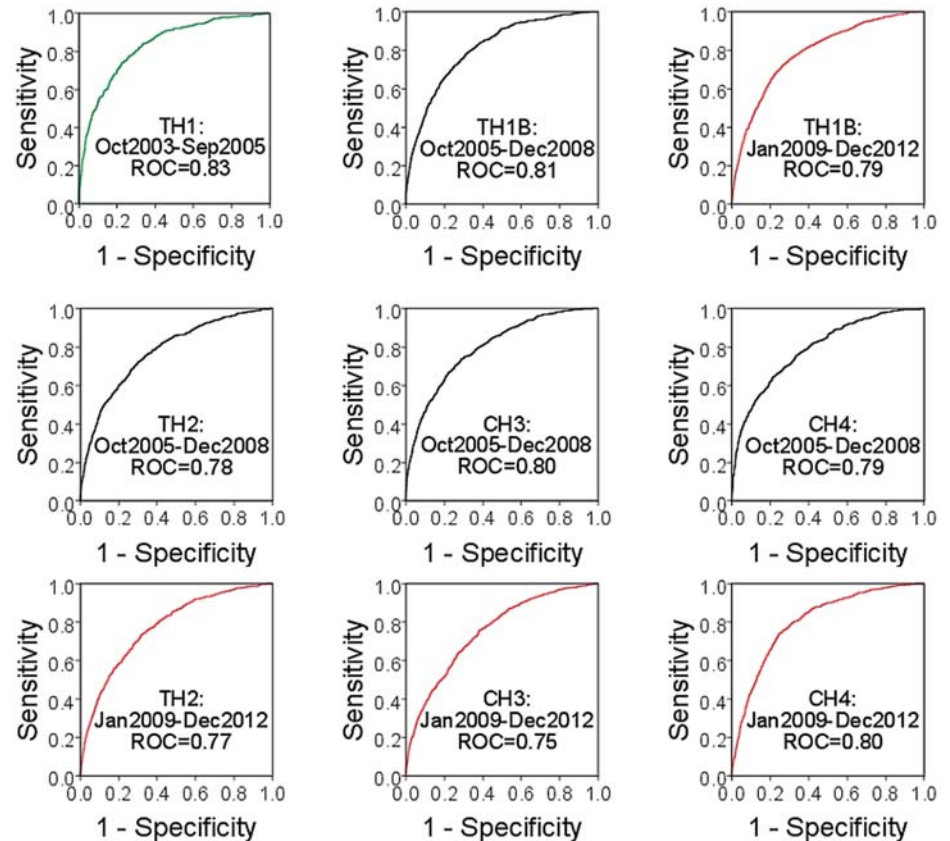
includes a measure of function, the ADL scale, which is independently associated with outcome. It could therefore be used as a practical decision aid in determining admission disposition. Further, the ALERT scale proved to be a good predictor of outcome over time, an 8-year period, as well as in teaching and community hospital settings. To the best of our knowledge this has not been replicated by other proposed models. The ALERT validation patient population was also larger than that used in other studies. In addition, the ALERT scale allows the user to set their own cut-off point based on local conditions to determine what level of risk should allow

admission to higher and lower monitoring inpatient setting. This will allow users to better utilise their resources. This flexibility is not something that developers of the other scales have suggested. Finally, our discrimination met or exceeded that found in other models. For example, while our c-statistic (ROC) was 0.83 in the model development site, and ranged from 0.75 to 0.81 in each of the eight validation periods/sites, the c-statistic in the three validation sites for the model described by Prytherch et al[10] was 0.78, 0.76 and 0.75. The c-statistic in the model described by Olsson et al[11] was 0.85, similar to ours, but their Hosmer-Lemeshow

Table 4 Sensitivity and specificity of the development (TH1) and validation populations at cut-off probability=0.088

		Observed adverse		Observed not-adverse			
	Study period	Predicted adverse N	Predicted not-adverse N	Predicted adverse N	Predicted Not-adverse N	Sensitivity Per cent	Specificity Per cent
TH1	Oct 2003–Sept 2005	341	113	1096	3333	75.1	75.3
TH2	Oct 2005–Dec 2008	431	259	1372	4584	62.5	77.0
	Jan 2009–Dec 2012	345	194	1987	6035	64.0	75.2
CH3	Oct 2005–Dec 2008	560	177	2109	4136	76.0	66.2
	Jan 2009–Dec 2012	579	208	2610	4368	73.6	62.6
CH4	Oct 2005–Dec 2008	404	171	1501	3686	70.3	71.1
	Jan 2009–Dec 2012	589	136	2106	3956	81.2	65.3
TH1B	Oct 2005–Dec 2008	561	253	2118	7137	68.9	77.1
	Jan 2009–Dec 2012	599	428	2058	9983	58.3	82.9
Total		4409	1939	16 957	47 218		

Table 5 Frequency of adverse outcome from top 30 admission diagnoses

Admission diagnosis	Patients*		Adverse outcome†	
	N	Per cent	N	Per cent
1 Pneumonia—community acquired	5452	7.7	688	10.8
2 Congestive heart failure	4702	6.7	438	6.9
3 Myocardial infarction	3609	5.1	330	5.2
4 COPD—exacerbation	3381	4.8	194	3.1
5 Cerebral vascular accident	2643	3.7	368	5.8
6 Upper GI bleed	2111	3.0	140	2.2
7 Cellulites	1980	2.8	63	1.0
8 Other neurological problems	1894	2.7	140	2.2
9 Septicemia	1893	2.7	190	3.0
10 Pneumonia	1614	2.3	223	3.5
11 Cystitis	1469	2.1	67	1.1
12 Renal insufficiency	1446	2.1	155	2.4
13 Delirium	1254	1.8	79	1.2
14 Tachyarrhythmia	1002	1.4	83	1.3
15 Falls	968	1.4	63	1.0
16 Unstable angina	945	1.3	26	0.4
17 Acute pyelonephritis	916	1.3	46	0.7
18 Seizures	895	1.3	51	0.8
19 Diabetic ketoacidosis	887	1.3	15	0.2
20 Gastroenteritis	872	1.2	31	0.5
21 Other GI problems	817	1.2	65	1.0
22 Pulmonary embolus	810	1.1	79	1.2
23 Acute pancreatitis	808	1.1	21	0.3
24 Acute coronary insufficiency	803	1.1	64	1.0
25 Lower GI bleed	748	1.1	46	0.7
26 Inflammatory bowel disease	740	1.0	8	0.1
27 Pleural effusion	734	1.0	109	1.7
28 Acute renal failure	716	1.0	89	1.4
29 Pain management	692	1.0	44	0.7
30 Primary lung malignancy	595	0.8	172	2.7
Total	43 919	62.3	4087	64.4

*Percentage calculated based on total study population N=70 523.
†Percentage calculated based on total observed adverse outcome N=6348.
COPD, chronic obstructive pulmonary disease; GI, gastrointestinal.

Goodness-of-Fit χ^2 (calibration) was 62, much higher than the goodness-of-fit χ^2 that we found in our development site and higher than the goodness-of-fit χ^2 in most of our validation sites. High goodness-of-fit χ^2 indicate a lack of fit of the model to data.

Arterial oxygen saturation is not routinely collected on non-ICU admissions and was only available on 5.3% of admissions. Its impact on the final model was not statistically significant (p=0.246) when we ran this model on the subset of admissions in which oxygen was recorded. For these reasons it was not included in the model development.

Although adverse outcome frequency is higher among those admitted for some primary diagnoses than others, our model provided good discrimination and calibration without using diagnosis as an input variable. This should not be surprising since the observation period was limited to 30 days, during which time factors that were included in model development such as physiological instability, comorbidity, physical disability and the level of consciousness may be more important determinants of short-term outcome. ALERT follows the precedent set by the SAPS II mortality predictor for ICU patients, which does not include admitting diagnosis.

In all nine study populations, the model development site and two validation periods times four validation sites, the 25% of patients included in the highest predicted risk bands accounted for more than two-thirds of actual adverse outcomes. Conversely, 33–47% of the patients were model-estimated to belong to the lowest risk band in each of the nine study populations. These observations are similar to those reported by Prytherch et al[10] and raise interesting alternatives. First, initial identification of the quartile accounting for the majority of potentially avoidable deaths would be prerequisite to developing focused preventative strategies. Second, these observations raise the possibility that some of the resources expended in the routine surveillance of low risk patients could be redirected to those in the highest risk quartile. Our findings and those of Olsson et al[11] challenge hospital planners and designers to consider reconfiguring the traditional ward environment in order

7

to allow better matching of resources to individual patient requirements as determined by early risk assessment.

Limitations

The limitation of our model is the need to use an equation to generate a probability from 0 to 1. Although a variety of probability cut-points can be used to stratify patients, in our institution, we have elected to use <0.04 as low, 0.04 to <0.088 as medium, 0.088 to <0.20 as high and 0.20 or higher as very high risk of adverse outcome. To facilitate calculation both an Excel file for downloading and a website are available http://www.alertassessmenttool.ca.

Whether calibration or discrimination should be prioritised when evaluating the performance of a model is an ongoing debate.[16–18] Calibration of our model declined with time. Indeed, as medical technology improves, one would expect that the risk of an adverse event would decline across all patients. Thus, one would expect that the calibration of a model, which classifies individuals into specific risk bands, would decline. Since the model discrimination remained stable over time, however, we feel that our model would be a valuable tool for many years. This is because, while the risk of adverse event among the entire patient population (hopefully) declines over time in most hospitals, we found that our model consistently (across time and site) distinguished the *relative* risk of adverse event across patients with different characteristics and medical histories.

CONCLUSION

Our investigation produced a practical early risk stratification tool for hospitalised medical patients. Whether or not early and focused attention to the high-risk quartile can reduce adverse events remains to be determined, however, early identification of these patients creates the possibility of resolving the question.

Author affiliations
[1]Department of Internal Medicine, University of Manitoba, Winnipeg, Manitoba, Canada
[2]Division of Critical Care, Medicine Queen Elizabeth II HSC, Dalhousie University, Halifax, Nova Scotia, Canada
[3]Internal Medicine Program, Winnipeg Regional Health Authority, Winnipeg, Manitoba, Canada
[4]Capital District Health Authority, Performance Excellence, Halifax, Nova Scotia, Canada
[5]Department of Geriatric Medicine, Dalhousie University Department of Medicine, Halifax, Nova Scotia, Canada
[6]General Internal Medicine and Community Health Sciences, Department of Internal Medicine, University of Manitoba, Winnipeg, Manitoba, Canada

REFERENCES

1. Bellomo R, Goldsmith D, Uchino S, *et al*. A prospective before-and-after trial of a medical emergency team. *Med J Aust* 2003;179:283–7.
2. Morgan R, Williams F, Wright M. An early warning scoring system for detecting developing critical illness. *Clin Intensive Care* 1997;8:100.
3. Buist MD, Moore GE, Bernard SA, *et al*. Effects of a medical emergency team on reduction of incidence of and mortality from unexpected cardiac arrests in hospital: preliminary study. *BMJ* 2002;324:387–90.
4. Parr M, Hadfield J, Flabouris A, *et al*. The medical emergency team: 12 month analysis of reasons for activation, immediate outcome and not-for-resuscitation orders. *Resuscitation* 2001;50:39–44.
5. DeVita MA, Bellomo R, Hillman K, *et al*. Findings of the first consensus conference on medical emergency teams. *Crit Care Med* 2006;34:2463–78.
6. Ranji SR, Auerbach AD, Hurd CJ, *et al*. Effects of rapid response systems on clinical outcomes: systematic review and meta-analysis. *J Hosp Med* 2007;2:422–32.
7. Litvak E, Pronovost PJ. Rethinking rapid response teams. *JAMA* 2010;304:1375–6.
8. Knaus W, Draper E, Wagner D, *et al*. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13:818–29.
9. Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993;270:2957–63.
10. Prytherch D, Sirl J, Schmidt P, *et al*. The use of routine laboratory data to predict in-hospital death in medical admissions. *Resuscitation* 2005;66:203–7.
11. Olsson T, Terént A, Lind L. Rapid Emergency Medicine Score: a new prognostic tool for in-hospital mortality in nonsurgical emergency department patients. *J Intern Med* 2004;255:579–87.
12. Olsson T, Terent A, Lind L. Charlson Comorbidity Index can add prognostic information to Rapid Emergency Medicine Score as a predictor of long-term mortality. *Eur J Emerg Med* 2005;12:220–4.
13. Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
14. Katz S, Downs TD, Cash HR, *et al*. Progress in development of the index of ADL. *Gerontologist* 1970;10:20–30.
15. Hosmer DW, Lemeshow S. *Applied logistic regression (Wiley series in probability and statistics)*. New York, NY: John Wiley & Sons, 2000.
16. Mallett S, Halligan S, Thompson M, *et al*. Interpreting diagnostic accuracy studies for patient care. *BMJ* 2012;345:e3999.
17. Baker SG. Putting risk prediction in perspective: relative utility curves. *J Natl Cancer Inst* 2009;101:1538–42.
18. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928–35.

Appendix 1

Activities of Daily Living (ADL) Scoring Scale

Activities	Unassisted	Assistance Required	Complete Dependancy
Bathing	0	3	6
Dressing	0	3	6
Toileting	0	3	6
Transfer	0	3	6
Continence	0	3	6
Feeding	0	3	6