BMJ open

Identifying admitted patients at risk of dying: a prospective observational validation of four biochemical scoring systems

Mikkel Brabrand,[1] Torben Knudsen,[1] Jesper Hallas[2]

To cite: Brabrand M, Knudsen T, Hallas J. Identifying admitted patients at risk of dying: a prospective observational validation of four biochemical scoring systems. *BMJ Open* 2013;**3**: e002890. doi:10.1136/bmjopen-2013-002890

▸ Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2013-002890).

Received 15 March 2013
Revised 13 May 2013
Accepted 21 May 2013

[1]Department of Medicine, Sydvestjysk Sygehus Esbjerg, Esbjerg, Denmark
[2]Research Unit of Clinical Pharmacology, University of Southern Denmark, OdenseC, Denmark

Correspondence to
Dr Mikkel Brabrand;
mbrabrand@health.sdu.dk

ABSTRACT

Objectives: Risk assessment is an important part of emergency patient care. Risk assessment tools based on biochemical data have the advantage that calculation can be automated and results can be easily provided. However, to be used clinically, existing tools have to be validated by independent researchers. This study involved an independent external validation of four risk stratification systems predicting death that rely primarily on biochemical variables.

Design: Prospective observational study.

Setting: The medical admission unit at a regional teaching hospital in Denmark.

Participants: Of 5894 adult (age 15 or above) acutely admitted medical patients, 205 (3.5%) died during admission and 46 died (0.8%) within one calendar day.

Interventions: None.

Main outcome measures: The main outcome measure was the ability to identify patients at an increased risk of dying (discriminatory power) as area under the receiver-operating characteristic curve (AUROC) and the accuracy of the predicted probability (calibration) using the Hosmer-Lemeshow goodness-of-fit test. The endpoint was all-cause mortality, defined in accordance with the original manuscripts.

Results: Using the original coefficients, all four systems were excellent at identifying patients at increased risk (discriminatory power, AUROC ≥0.80). The accuracy was poor (we could assess calibration for two systems, which failed). After recalculation of the coefficients, two systems had improved discriminatory power and two remained unchanged. Calibration failed for one system in the validation cohort.

Conclusions: Four biochemical risk stratification systems can risk-stratify the acutely admitted medical patients for mortality with excellent discriminatory power. We could improve the models for use in our setting by recalculating the risk coefficient for the chosen variables.

ARTICLE SUMMARY

Article focus

- Physicians staffing emergency departments and admission units are not comfortable predicting the risk of mortality for their patients.
- Several systems that can do this have been developed but not externally validated and should thus not yet be used in clinical practice.
- The aim of this article was to validate four existing biochemical risk stratification systems predicting mortality of acutely admitted patients.

Key messages

- The four risk prediction systems based on biochemical data are excellent at predicting mortality of acutely admitted medical patients.
- The precision of the predictions is low, but can be improved by adjusting the systems to the local environment by recalculating the scores.

Strengths and limitations of this study

- This is the largest study to validate biochemical-based risk stratification systems in a medical admission unit.
- This study has good external validity and a low risk of selection bias.
- The study is limited by missing data especially in two of the four scores and by the fact that it is a single centre study.

INTRODUCTION

An important part of the routine work of front-line personnel in emergency departments and admission units is to assess the risk of individual patients. However, many physicians feel inadequately trained,[1] and prognostication is not a mandatory part of medical education.[2] As a consequence, automated risk stratification could assist physicians attending to emergency patients. However, in a recent review,[3] none of the risk stratification tools for use in the emergency departments and admission units attained the highest level of evidence. Several systems have been developed, but only a few have been externally validated, even though this is an important part of the development process.[4]

Some of the existing risk stratification systems are based solely on vital signs and others on biochemical analyses. Systems

based on vital signs require manual collection of data, whereas systems based on biochemical analyses can be automated. Data can easily be extracted from the hospital computer systems and risk stratification can be performed in an automated process.

We performed the present study with the objective of validating existing risk stratification systems that predict mortality for medical patients based solely on biochemical data. Four systems based on multiple (more than two) routinely available variables (in our setting) and not restricted to selected groups of medical patients were included.

METHODS

We performed an external validation of existing biochemical risk stratification systems by applying the coefficients and ORs reported in the original papers. Furthermore, we validated the choice of variables in the original papers by recalculating the coefficients to fit our current patient population.

Setting

Sydvestjysk Sygehus is a 460-bed regional teaching hospital in the western part of Denmark with a contingency population of 220 000. All subspecialties of internal medicine are represented.

Patients can be admitted to the medical admission unit (MAU) by their general practitioner, out-of-hours emergency medical service, outpatient clinics, emergency department and ambulance services. Two attending physicians, one in internal medicine and one in cardiology, one senior resident and two interns staff the MAU.

Design and data

We conducted a prospective observational cohort study of all patients admitted through the MAU at our hospital. All consecutive adult patients (ages ≥15 years) admitted from 2 October 2008 until 19 February 2009 (first cohort) and from 23 February 2010 until 26 May 2010 (second cohort) were included in the study.

Upon admission, a nurse recorded the vital signs and registered these along with demographic information and the primary complaint on a form. After inclusion of all patients, we extracted blood test results from the hospital computer systems. No extra biochemical analyses were added as part of this study, and only analyses ordered by the admitting doctor were included. Most patients had the following biochemical standard panel taken: haemoglobin, leukocytes, platelets, C reactive protein, sodium, potassium, creatine, urea, total calcium, glucose and albumin. Almost all patients admitted to the cardiology section had troponin, amylase and total cholesterol measured as well. We included blood tests drawn 1 h prior to admission and within 6 h after admission. If a patient had multiple analyses of the same biochemical variable, only the first was included. In case of missing data on forms (or completely missing forms), data were extracted from an electronic copy of the nurse's notes or the chart. Inclusion of all patients was ensured by validation against the central hospital database. As we have no formalised classification system for primary complaints, one of the authors (MB) converted the primary complaint to a diagnosis according to the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)[5] and compiled these as admissions due to

- Infectious disorders (ICD-10 diagnoses A and B);
- Malignancy (ICD-10 diagnoses C and D);
- Endocrine disorders (ICD-10 diagnoses E);
- Circulatory disorders (ICD-10 diagnoses I);
- Pulmonary disorders (ICD-10 diagnoses J);
- Symptoms (ICD-10 diagnoses R);
- Observational reasons (ICD-10 diagnoses Z);
- Other reasons (ICD-10 diagnoses F, G, H, K, L, M, N, O, P, Q, S, T, X and Y).

We analysed the performance of four different risk stratification systems based on biochemical variables: the system introduced by Prytherch *et al*[6] required gender, mode of admission, age, urea, sodium, potassium, albumin, haemoglobin, white cell count and creatine. Froom and Shimoni[7] included age, albumin, alkaline phosphatase, aspartate aminotransferase, urea, glucose, lactate dehydrogenase, neutrophil count proportion and total leucocyte count. Loekito *et al*[8] required haemoglobin, haematocrit, total CO_2, leucocytes, albumin, bilirubin, creatine and urea. We estimated haematocrit from haemoglobin[9] and total CO_2 from bicarbonate.[10] The score by Asadollahi *et al*[11] required age, urea, haemoglobin, leucocytes, platelets, sodium and glucose. If the patient missed one or more of the biochemical variables required for a given risk assessment tool, the patient was excluded from the validation of that tool.

We defined the primary outcome as in the original articles, that is, in-hospital mortality for Prytherch *et al*[6] Asadolliahi *et al*[11] and Froom and Shimoni[7] and imminent death (ie, death within one calendar day after the blood was drawn) for Loekito *et al*[8] Data on this were extracted from the hospital computer systems after the inclusion was completed and all patients were either discharged or dead.

The study was approved by the Danish Data Protection Agency. Approval from an Ethics Committee was not required according to Danish law. The study is reported in accordance with the STROBE statement.[12]

Statistics

The sample size was dictated by another part of the study. In brief, the sample size was calibrated to develop and validate a risk-stratification system to predict 7-day all-cause mortality.

We calculated the predicted mortality using the coefficients presented in the original papers. To assess the ability of each system to identify patients at highest risk of dying (ie, the discriminatory power), we calculated the area under the receiver-operating characteristic curve (AUROC). AUROC is a summary measure of

Table 1 Demographics of patients

	Total, n=5894	First cohort, n=3046	Second cohort, n=2848
Female	2950 (50.1%)	1460 (47.9%)	1490 (52.3%)
Age (years)	65 (49–77)	66 (50–77)	64 (48–76)
Length of stay (days)	2 (1–6)	2 (1–6)	1 (1–5)
In-hospital mortality	205 (3.5%)	116 (3.8%)	89 (3.1%)
Imminent death	46 (0.8%)	26 (0.9%)	20 (0.7%)
Admitted due to infectious disorder	178 (3.0%)	82 (2.7%)	96 (3.4%)
Admitted due to malignant disorder	128 (2.2%)	50 (1.6%)	78 (2.8%)
Admitted due to endocrine disorder	307 (5.2%)	147 (4.8%)	160 (5.6%)
Admitted due to circulatory disorder	1375 (23.4%)	527 (17.3%)	848 (29.9%)
Admitted due to pulmonary disorder	972 (16.5%)	547 (18.0%)	425 (15.0%)
Admitted due to symptoms	1194 (20.3%)	719 (23.6%)	475 (16.7%)
Admitted due to observation	1012 (17.2%)	585 (19.2%)	427 (15.1%)
Admitted due to other reasons	718 (12.2%)	389 (12.8%)	329 (11.6%)

sensitivity and specificity at each possible cut-off and basically represents the probability that a patient who eventually dies will have a higher score than a patient who survives. An AUROC above 0.8 is said to represent excellent discriminatory power.[13] The calibration was assessed using the Hosmer-Lemeshow goodness-of-fit test. The calibration assesses if the observed mortality rate matches the expected rate, derived from the scoring systems. For this test, we divided the population into decentiles by expected event rate. A p value above 0.05 indicates acceptable calibration. A scoring system might show excellent discriminatory power and yet have poor calibration if, for example, it was developed on a population with low overall mortality and then applied to a population with high overall mortality.

As the predictive power would be expected to vary across populations, we calculated the AUROC of each of the original scores for patients presenting with the previously specified presenting complaints.

Finally, we attempted to optimise the models to our setting by recalculating the scoring coefficients; that is, we performed the multivariable analyses anew by using the variables included in the original models. We used the first cohort (collected from 2008 to 2009) for the development and the second cohort (collected in 2010) for validation of the recalculated coefficients.

As the Asadollahi score[11] is a set score (ranging from 0 to 20) and not a regression formula, we initially performed a new logistic regression using our development cohort. From the coefficients derived, we assigned a score (from 1

Table 2 Variables included in the scores and the level of missing data

Variable	Percentage of missing	Prytherch score[6]	Froom score[7]	Loekito score[8]	Asadollahi score[11]
Lactate dehydrogenase	76.6		•		
Bilirubin	75.1			•	
Alkaline phosphatase	75.0		•		
Bicarbonate	71.6			•	
Alanine aminotransferase	68.3		•		
Neutrophil count proportion	42.1		•		
Urea/creatine	13.0	•			
Urea	12.7	•	•	•	•
Albumin	7.5	•	•	•	
Platelets	7.1				•
Glucose	6.9		•		•
White cell count	6.0	•		•	•
Creatine	5.8	•	•	•	
Potassium	5.5	•			
Sodium	5.2	•			•
Haemoglobin	5.1	•		•	•
Haematocrit	5.1			•	
Age	0.0	•	•	•	•
Gender	0.0	•			
Mode of admission	0.0	•			

•Required in the score.

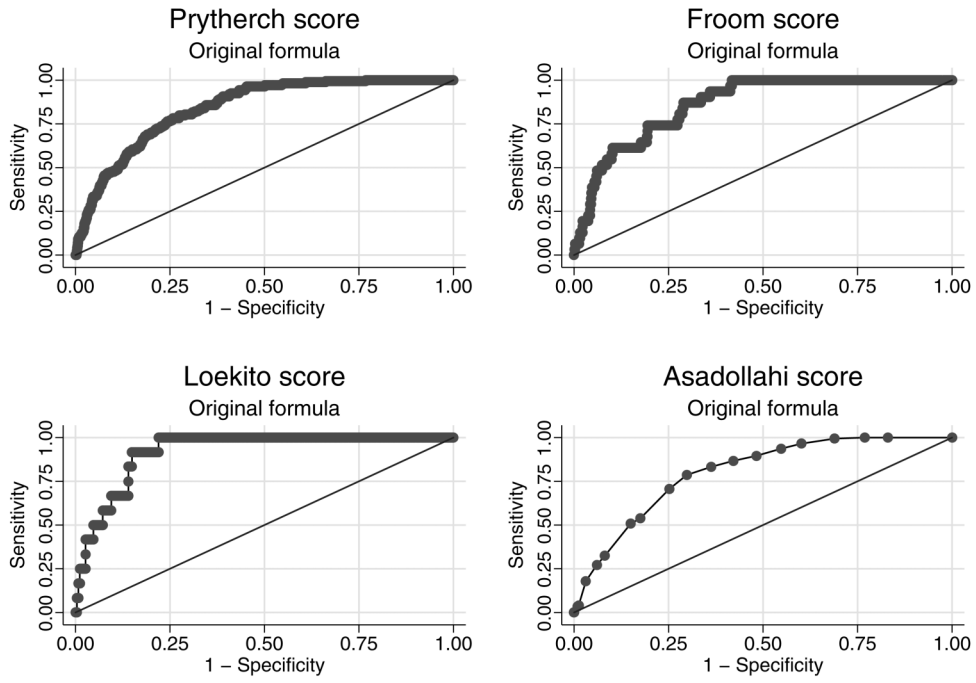

Figure 1 Discriminatory power of four risk stratification systems based on biochemical variables. Original coefficients were used to generate receiver-operating curves.

to 6) to each variable and recalculated the score for both cohorts. We tested calibration according to Seymour *et al*,[14] that is , we predicted the probabilities of the individual scores using logistic regression analysis and calculated the Hosmer-Lemeshow goodness-of-fit test.

Data are reported as median (IQR) or proportions whenever appropriate. Differences between patients with and without missing data were tested using the χ^2 test or Wilcoxon rank-sum test.

STATA V.12.1 (StataCorp, College Station, Texas, USA) was used for the analyses.

RESULTS

A total of 5894 patients were included in our study (see table 1 for details). Among these, 205 (3.5%) died during the admission, and 46 (0.8%) died within one calendar day.

Validation of the original scores

We could include 4925 patients (83.6% of the entire cohort) in the Prytherch score (table 2). Using the original formula, we found an AUROC of 0.842 (95% CI 0.818 to 0.865; figure 1 and table 3) and goodness-of-fit test, χ^2=419.63 (10 degrees of freedom), p<0.001. Thus, the Pryterch score showed a good ability to identify patients at high risk of dying, but failed in calibration, as fewer patients died than expected.

In calculating the Froom score,[7] we could include only 919 patients (15.6%; table 2). Using the ORs specified in the original article, we found an AUROC of 0.862 (95% CI 0.813 to 0.910; figure 1 and table 3). As the original paper did not provide the coefficient for the intercept, we were unable to reliably assess calibration. In an attempt to reduce selection bias, Froom and Shimoni[7] used imputation of the mean (by assigning the

Table 3 Performance of the model using the original coefficients and after recalculation

	Discriminatory power (AUROC)			Calibration (Hosmer-Lemeshow χ^2 test p value)		
		Recalculated model			Recalculated model	
Score	Original model	Development	Validation	Original model	Development	Validation
Prytherch score[6]	0.842 (0.818–0.865)	0.858 (0.827–0.889)	0.874 (0.841–0.907)	<0.001	0.59	0.66
Froom score[7]	0.862 (0.813–0.910)	0.930 (0.897–0.962)	0.882 (0.806–0.957)	–	0.93	0.009
Loekito score[8]	0.922 (0.879–0.965)	0.911 (0.819–1.000)	0.917 (0.823–1.000)	0.0007	0.79	1.00
Asadollahi score[11]	0.803 (0.776–0.829)	0.808 (0.774–0.842)	0.813 (0.772–0.854)	–	0.79	0.47

Area under receiver-operating curve (AUROC) above 0.8 represents good discriminatory power, and p value for calibration above 0.05 represents good calibration

value of 2.5 to all missing variables reduced into quartiles). Adapting this approach led to the inclusion of all 5894 patients with an AUROC of 0.814 (CI 0.788 to 0.841). Again, because of a missing coefficient for the intercept, we could not assess calibration. Thus, the Froom score was good at identifying patients at high risk, but we could not assess the level of precision.

As for the Loekito score,[8] we could include 540 patients (9.2%; table 2). Using the reported coefficients, we found an excellent discriminatory power (AUROC=0.922, CI 0.879 to 0.965, figure 1 and table 3). Calibration failed with a goodness-of-fit test, χ^2=30.7, p=0.0007. Thus, the Loekito score showed excellent discriminatory power but failed calibration.

We could include 4863 (82.5%) in the Asadollahi score[11] (table 2). We found a good calibration (AUROC=0.803; CI 0.776 to 0.829; figure 1 and table 3), but could not assess it because of the construction of the score in the original article.

The predictive ability of each score varied widely with each presenting compliant; however, within each complaint, the scores more or less had identical AUROCs (table 4). Overall, malignant, endocrine and pulmonary disorders had the lowest AUROC, while infectious disorders had the highest (table 4). Some of these calculations are based on limited numbers (as indicated by the CIs).

Recalculated coefficients

Performing the recalculation of the Prytherch score,[6] we achieved excellent AUROCs in both cohorts as well as acceptable calibration (figure 2 and table 3). Sex, urea, sodium, haemoglobin, creatine and potassium were not significantly associated with in-hospital mortality in our material, but because they were included in the original, we kept them in the analysis.

Recalculating the Froom score,[7] we achieved excellent AUROCs in both cohorts, but calibration failed in the validation cohort (figure 2 and table 3). Age, alkaline phosphatase, alanine aminotransferase, urea, white cell count and glucose were not significantly associated with in-hospital mortality, but were kept in the model.

When recalculating the Loekito score,[8] we found that urea, creatine, albumin, haemoglobin and white cell count were not significantly associated with the endpoint of 1-day mortality. We achieved excellent AUROCs in both cohorts as well as almost perfect calibration (figure 2 and table 3).

When recalculating the Asadollahi score,[11] we assigned a score of one each to haemoglobin, platelets and glucose (none of which were significantly associated with the endpoint), three to sodium, four each to age and white cell count and six to urea. AUROC was excellent in both cohorts and calibration acceptable (figure 2 and table 3).

In all four methods, the discriminatory power remained constant or improved when we compared it with the calculation based on the original coefficients and ORs.

Table 4 The predictive power (area under receiver-operating curve (95% CI)) of each score on patients with varying presenting complaints

Presenting complaint	Immediate death	In-hospital mortality	Prytherch score[6]	Froom score[7]	Loekito score[8]	Asadollahi score[11]
Infectious disorder	2 (1.1%)	8 (4.5%)	0.877 (0.772 to 0.982)	0.837 (0.738 to 0.935)	0.917 (0.760 to 0.100)	0.859 (0.739 to 0.979)
Malignant disorder	2 (1.6%)	5 (3.9%)	0.688 (0.502 to 0.874)	0.583 (0.300 to 0.867)	–	0.507 (0.342 to 0.672)
Endocrine disorder	2 (0.7%)	23 (7.5%)	0.789 (0.699 to 0.879)	0.650 (0.335 to 0.965)	0.718 (0.576 to 0.860)	0.694 (0.585 to 0.802)
Circulatory disorder	7 (0.5%)	33 (2.4%)	0.869 (0.809 to 0.929)	1.000 (1.000 to 1.000)	0.841 (0.665 to 1.000)	0.843 (0.767 to 0.919)
Pulmonary disorder	17 (1.8%)	52 (5.4%)	0.770 (0.708 to 0.832)	0.730 (0.601 to 0.860)	0.810 (0.741 to 0.878)	0.730 (0.662 to 0.798)
Symptoms	11 (0.9%)	56 (4.7%)	0.825 (0.773 to 0.877)	0.921 (0.857to 0.984)	0.967 (0.912 to 1.000)	0.766 (0.708 to 0.823)
Observation	4 (0.4%)	21 (2.1%)	0.848 (0.775 to 0.920)	0.674 (0.285 to 1.000)	0.800 (0.621 to 0.979)	0.875 (0.824 to 0.926)
Other	1 (0.1%)	7 (0.9%)	0.918 (0.858 to 0.977)	0.862 (0.821 to 0.903)	–	0.866 (0.729 to 1.000)

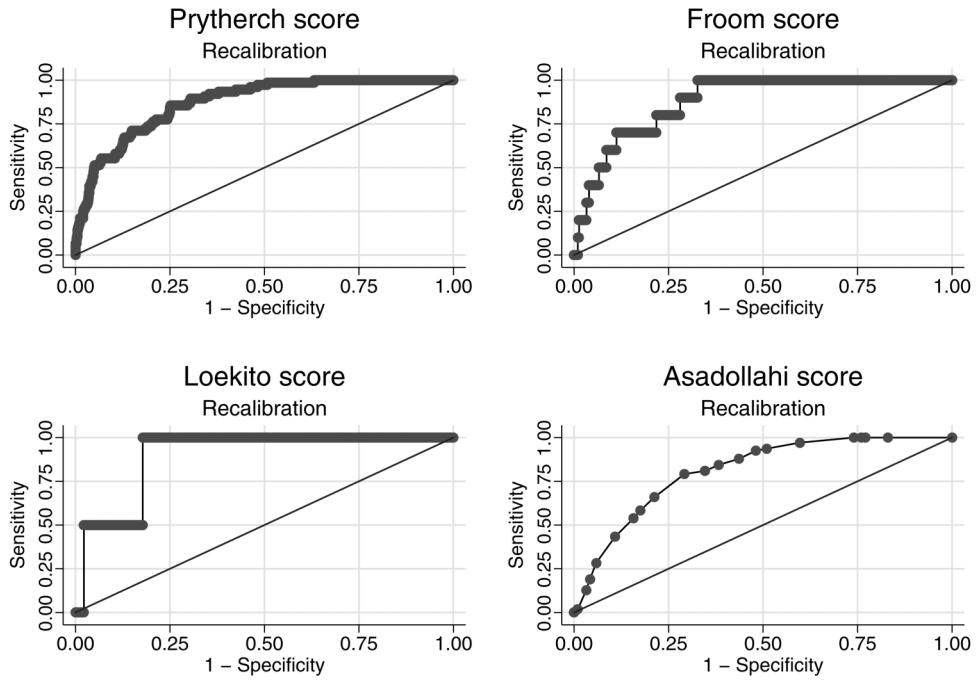

Figure 2 Discriminatory power after recalculation of new coefficients to match our setting.

Selection bias

For the Prytherch, Froom and Asadollahi scores, patients who were excluded because of missing values had the same mortality as those who were included (table 5). For the Loekito score, patients with missing data had significantly lower 1-day mortality (table 5).

DISCUSSION

Using four existing biochemical-based risk stratification systems, we could risk-stratify acutely admitted medical patients with excellent discriminatory power. We could only evaluate the calibration for two scores, the Prytherch score[6] and the Loekito score,[8] which both failed. When recalculating all four scores, both discriminatory power and calibration improved, except for the Froom score,[7] where calibration failed.

In the present article, we focused only on biochemical-based risk stratification systems. While systems based on vital signs can be calculated shortly after arrival, biochemical-based systems require the blood tests to be analysed first. On the other hand, for systems based only on biochemical data, interobserver or intraobserver variation is virtually eliminated. We have identified four systems with broad inclusion criteria that could potentially be used in emergency departments and MAUs. The systems included were developed in different settings, ranging from floor beds[6 8 11] to a medical emergency room.[7] One was internally validated using a split sample technique,[6] while the others were validated in external cohorts.[7 8 11] However, even if the systems were developed in a setting similar to ours and validated by the original authors, they still need to be externally validated in independent cohorts, as we now have performed, before they should be used in the clinical routine.[4]

Although all four systems had acceptable discriminatory power, two systems failed in calibration. One way of correcting poor calibration is to perform a recalculation. We have carried out so by performing a multivariable logistic regression in one cohort and then validating it in another. This approach generally improved the discriminatory power and made calibration acceptable. In fact, calibration became acceptable in both systems that previously failed. After recalculation, however, calibration failed in the Froom score,[7] a system for which we could not test calibration using the original formula. Our best explanation for this is differences in mortality because the Froom score[7] was developed and validated in cohorts with a higher mortality than ours (5.6% vs 3.5%).

The Prytherch score[6] seems to fit our setting best. The discriminatory power was excellent both before and after recalculation. Calibration failed before recalculation, but was acceptable afterwards. Most important, using our standard biochemical profile, we could include the majority of our patients. Both the Froom[7] and Loekito scores[8] performed better, but only marginally, and the Froom score[7] failed on calibration after recalculation; we could include only a few of our patients in both scores. However, the choice of score depends on several additional factors. Some hospitals might not routinely measure all investigations required by each score (eg, albumin) and some investigations are error prone (eg, haemolysis in potassium measurements). The Asadollahi score only relies on seven

Table 5 Data on potential selection biases of patients with missing data in the four scores

Score	Number of patients with missing data	Death (n (%))			Length of stay (days, median (IQR))			Age (years, median (IQR))		
		No missing data	Missing data	p Value	No missing data	Missing data	p Value	No missing data	Missing data	p Value
Prytherch[6]	969 (16. 4%)	174 (3.5%)	31 (3.2%)	0.60	2 (1–6)	1 (0–4)	<0.0001	66 (50–77)	63 (47–75)	0.0002
Froom[7]	4975 (84.4%)	31 (3.4%)	174 (3.5%)	0.85	2 (1–6)	2 (1–6)	0.92	57 (37–74)	66 (52–77)	<0.0001
Loekito[8]	5354 (90.8%)	12 (2.2%)	34 (0.6%)	<0.001	3 (1–8)	1 (1–6)	<0.0001	61 (39–76)	66 (50–77)	<0.0001
Asadollahi[11]	1031 (17.5%)	173 (3.6%)	32 (3.1%)	0.47	2 (1–6)	1 (0–4)	<0.0001	66 (50–77)	63 (47–75)	0.0003

parameters and could thus be easily obtained and perhaps less expensive to report on most patients. Also, it is not significantly inferior to the other scores and might therefore be more suitable for other settings.

Our study has limitations. First, we have a substantial amount of missing data. This absence is not a major problem when calculating the Prytherch[6] or Asadollahi score,[11] but it was for the Froom[7] and Loekito scores.[8] There is no doubt that this has introduced selection bias into our study. Although that we have not been able to demonstrate any selection bias for the Prytherch, Froom and Asadollahi scores looking at our primary endpoint of mortality,[6 7 11] we showed that patients with missing data in the Loekito score[8] had a significantly lower mortality. An apparent explanation is that bicarbonate is part of the formula. At our institution, bicarbonate is mostly analysed as part of arterial blood gas analyses and thus primarily measured in the most critically ill patients. Patients with missing data also had a significantly shorter length of stay, but were not uniformly older or younger than patients that could be included in each score (table 5). These indications of selection biases prompt us to question the external validity and generalisability of our findings, and we see this as an indication that further studies, where the risk of selection bias is minimised, are required. Second, the Loekito score[8] requires haematocrit (we estimated this using the haemoglobin level[9]) and total CO_2 (which we estimated using bicarbonate).[10] However, when performing our own logistic regression analyses of both systems, we had acceptable results, proving this to be of no concern. Third, this study still represents a single centre application of the scoring systems, and the results should be evaluated with this in mind. Fourth, we run a risk of overfitting[15–17] when performing recalculation. With only 26 imminent fatalities in the development cohort, overfitting is a potential risk for the Loekito score.[8] However, our validation proves that it was not an issue. As for the other three systems, we have enough fatalities for a valid recalculation.

We have found that four risk stratification systems based on biochemical data can identify patients at an increased risk of dying, although with limited precision. The models could be improved by recalculation, but the question remains if the use of these systems will improve clinical practice. In an ideal study, patients should be randomised to either be risk-stratified by a predefined system or be managed by clinical assessment alone, and the potential improvement in treatment should be measured. This approach is a complicated setup not previously performed for any of the present systems, but is the only way to show if the implementation of the system matters.

Contributors MB conceived and designed the study, collected, analysed and interpreted the data and wrote the report. TK and JH conceived and designed the study and assisted with analysis and interpretation of the data and writing of the report. All authors have had full access to all data and take

responsibility for the integrity of the data and the accuracy of the analyses. MB is the guarantor. All authors have read and approved the final manuscript.

Funding The study was funded by Sydvestjysk Sygehus, Karola Jørgensens Forskningsfond, Edith og Vagn Hedegaard Jensens Fond, AB Fonden and Johs M Klein og Hustrus Mindelegat. None of the funders have had influence on the design and conduct of the study; collection, management, analysis and interpretation of the data; or preparation, review or approval of the manuscript, as the researchers are independent from all sponsors.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Christakis NA, Iwashyna TJ. Attitude and self-reported practice regarding prognostication in a national sample of internists. *Arch Intern Med* 1998;158:2389–95.
2. Kellett J. Prognostication—the lost skill of medicine. *Eur J Intern Med* 2008;19:155–64.
3. Brabrand M, Folkestad L, Clausen NG, *et al*. Risk scoring systems for adults admitted to the emergency department: a systematic review. *Scand J Trauma Resusc Emerg Med* 2010;18:8.
4. McGinn TG, Guyatt GH, Wyer PC, *et al*. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA* 2000;284:79–84.
5. WHO. http://www.who.int/classifications/icd/en/ (accessed 6 May 2013).
6. Prytherch DR, Sirl JS, Schmidt P, *et al*. The use of routine laboratory data to predict in-hospital death in medical admissions. *Resuscitation* 2005;66:203–7.
7. Froom P, Shimoni Z. Prediction of hospital mortality rates by admission laboratory tests. *Clin Chem* 2006;52:325–8.
8. Loekito E, Bailey J, Bellomo R, *et al*. Common laboratory tests predict imminent death in ward patients. *Resuscitation* 2013;84:280–5.
9. Brown J, Theis L, Kerr L, *et al*. A hand-powered, portable, low-cost centrifuge for diagnosing anemia in low-resource settings. *Am J Trop Med Hyg* 2011;85:327–32.
10. Johnson CW, Timmons DL, Hall PE. *Essential laboratory mathematics: concepts and applications for the chemical and clinical laboratory technician*. 2nd edn. Clifton Park, NY: Delmar Learning, 2003.
11. Asadollahi K, Hastings IM, Gill GV, *et al*. Prediction of hospital mortality from admission laboratory data and patient age: a simple model. *EMA* 2011;23:354–63.
12. Vandenbroucke JP, Von Elm E, Altman DG, *et al*. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology* 2007;18:805–35.
13. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd edn. New York: John Wiley & Sons, 2000.
14. Seymour CW, Kahn JM, Cooke CR, *et al*. Prediction of critical illness during out-of-hospital emergency care. *JAMA* 2010;304:747–54.
15. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med* 1993;118:201–10.
16. Peduzzi P, Concato J, Feinstein AR, *et al*. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995;48:1503–10.
17. Peduzzi P, Concato J, Kemper E, *et al*. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–9.