Re-evaluation and modification of the Stuivenberg Hospital Acute Renal Failure (SHARF) scoring system for the prognosis of acute renal failure: an independent multicentre, prospective study

R. L. Lins\textsuperscript{1}, M. M. Elseviers\textsuperscript{2}, R. Daelemans\textsuperscript{1}, P. Arnouts\textsuperscript{3}, J.-M. Billiouw\textsuperscript{4}, M. Couttenye\textsuperscript{2}, E. Gheuens\textsuperscript{1}, P. Rogiers\textsuperscript{5}, R. Rutsaert\textsuperscript{2}, P. Van der Niepen\textsuperscript{6} and M. E. De Broe\textsuperscript{2}

\textsuperscript{1}Department of Nephrology-Hypertension, ACZA Campus Stuivenberg, Lange Beeldekensstraat 267, B-2060 Antwerpen, \textsuperscript{2}Department of Nephrology-Hypertension, University Hospital Antwerp, Wilrijkstraat 10, B-2650 Edegem/Antwerpen, \textsuperscript{3}Department of Nephrology-Hypertension, Sint Jozefziekenhuis, Steenweg op Merksplas 44, B-2300 Turnhout, \textsuperscript{4}Department of Nephrology-Hypertension, O.L.-Vrouw Ziekenhuis, Moorselbaan 164, B-9300 Aalst, \textsuperscript{5}Department of Intensive Care, General Hospital Middelheim, Lindendreef 1, B-2020 Antwerpen and \textsuperscript{6}Department of Nephrology-Hypertension, Free University Brussels (VUB), Laarbeeklaan 101, B-1090 Brussels, Belgium

Abstract

**Background.** A prognostic scoring system for hospital mortality in acute renal failure (Stuivenberg Hospital Acute Renal Failure, SHARF score) was developed in a single-centre study. The scoring system consists of two scores, for the time of diagnosis of acute renal failure (ARF) and for 48 h later, each originally based on four parameters (age, serum albumin, prothrombin time and heart failure). The scoring system was now tested and adapted in a prospective study.

**Methods.** The study involved eight intensive care units. We studied 293 consecutive patients with ARF in 6 months. Their mortality was 50.5%. The causes of ARF were medical in 184 (63%) patients and surgical in 108 (37%). In the latter group, 74 (69%) patients underwent cardiac and 19 (18%) vascular surgery.

**Results.** As the performance of the original SHARF scores was much lower in the multicentre study than in the original single-centre study, we re-analysed the multicentre data to customize the original model for the population studied. The independent variables were the score developed in the original study plus all additional parameters that were significant on univariate analysis. The new multivariate analysis revealed an additional subset of three parameters for inclusion in the model (serum bilirubin, sepsis and hypotension). For the modified SHARF II score, $r^2$ was 0.27 at 0 and 0.33 at 48 h, respectively, the receiver operating characteristic (ROC) values were 0.82 and 0.83, and the Hosmer–Lemeshow goodness-of-fit $P$ values were 0.19 and 0.05.

**Conclusion.** After customizing and by using two scoring moments, this prediction model for hospital mortality in ARF is useful in different settings for comparing groups of patients and centres, quality assessment and clinical trials. We do not recommend its use for individual patient prognosis.

**Keywords:** acute renal failure; intensive care unit; mortality prediction; prognosis; prospective multicentre study; severity score

Introduction

In view of current trends in health care, especially with resources becoming increasingly scarce, the need to predict patient outcomes is evident [1]. After proper testing in various institutions and adaptation to different populations, prognostic scoring systems may aid physicians in different ways: to calculate the probabilities of outcomes in patient groups and in individual patients, advise patients and families regarding questions about continued life support, compare institutions and judge the quality of care and the performance of a medical care unit. Such predictive models can also be used to study new therapies, which might ameliorate outcomes in patients.

Acute renal failure (ARF) is a disease where risk modelling is important due to its high mortality rate of 50%, even higher in certain populations [2]. In general, however, scoring systems predict only the prognosis of a group of patients and are useful for research,
but they are unlikely to be useful in the assessment of the prognosis of a given individual [3]. In these scoring systems, the parameters were obtained only once after hospitalization or recognition of ARF. Although it is important for intensive care unit (ICU) physicians and nephrologists to predict mortality in the first 24 h after hospitalization, according to some authors, relying on a single scoring point in time can be misleading [4–7].

In the developmental phase of our scoring system, a prospective, cohort, single-centre study was performed, involving 197 adult patients admitted consecutively to a medical ICU. Relevant parameters were documented at 0 and 48 h, in order to develop a score usable in different centres and for different causes of ARF [8]. Thus we developed a scoring protocol with two measuring points (Stuivenberg Hospital Acute Renal Failure, SHARF), one applied at the diagnosis of ARF and the other 48 h later. The same parameters were included in the score at the two time points, but were assigned different weights: age, albumin, prothrombin time, respiratory support and heart failure. Their $r^2$ values were 0.36 and 0.44, respectively, which means that 36 and 44%, respectively, of the predicted mortality is explained by the model. The receiver operating characteristic (ROC) values at 0 and 48 h were 0.87 and 0.90 [8]. The model was next tested in the present prospective multicentre study in eight ICUs located in different medical and surgical settings and with different case mixes of patients.

**Methods**

To re-evaluate and improve the model, we included in this study all adult patients (age ≥18 years) with a presumptive diagnosis of ARF who were admitted consecutively during a 6 month study period (from 15 September 1997 to 15 March 1998) to any of eight participating ICUs. These units were in two university hospitals with mixed surgical–medical populations (573 and 700 hospital beds and 30 and 30 ICU beds, respectively) and in six general hospitals: three with mixed populations (294–558 hospital beds and 6–30 ICU beds), one with a predominantly surgical population (651 hospital beds and 12 ICU beds) and two with almost exclusively medical populations (651 and 957 beds and 12 and 36 ICU beds). Included in the eight was the medical unit used for the original single-centre study [8]. Also, the same clinical and laboratory parameters were documented at T0 and T48. The overall severity of illness was based on the APACHE II score determined on the day when the first data collection was done for the score (i.e. T0).

Relevant definitions of clinical parameters are given in Table 1. Organ failure was defined according to published criteria [10–12].

The outcome of hospital treatment was determined as ‘died’, ‘end-stage renal disease’ (ESRD), ‘partially recovered’ (hospital discharge with a serum creatinine >2 mg/dl) or ‘completely recovered’. Patients who were transferred to other departments were followed-up to their discharge or death in hospital, and their last known serum creatinine was recorded. From each of the eight participating institutions, every eligible patient was reported to the coordinating centre within the first day that the criteria for the definition of ARF were met, the day of diagnosis (T0). For patients referred later in the course of their illness to the ICU, the day of admission to the ICU was taken as the starting day (T0). The second measurement point was 48 h later (T48).

The time of diagnosis of ARF in relation to hospital admission was defined as ‘initial’ if the inclusion criteria were met at the time of admission to the ICU, and as ‘delayed’ if the criteria were met during the ICU stay or if ARF developed when the patient was an in-patient on a regular ward before transfer to the ICU. The type of ARF, the cause and the primary disease setting were determined exactly as in the original single-centre study [8]. Also, the same clinical and laboratory parameters were documented at T0 and T48. The overall severity of illness was based on the APACHE II score determined on the day when the first data collection was done for the score (i.e. T0).

The first set of data was collected on the first day that the criteria for the definition of ARF were met, the day of diagnosis (T0). For patients referred later in the course of their illness to the ICU, the day of admission to the ICU was taken as the starting day (T0). The second measurement point was 48 h later (T48).

The time of diagnosis of ARF in relation to hospital admission was defined as ‘initial’ if the inclusion criteria were met at the time of admission to the ICU, and as ‘delayed’ if the criteria were met during the ICU stay or if ARF developed when the patient was an in-patient on a regular ward before transfer to the ICU. The type of ARF, the cause and the primary disease setting were determined exactly as in the original single-centre study [8]. Also, the same clinical and laboratory parameters were documented at T0 and T48. The overall severity of illness was based on the APACHE II score determined on the day when the first data collection was done for the score (i.e. T0). Relevant definitions of clinical parameters are given in Table 1. Organ failure was defined according to published criteria [10–12].

The outcome of hospital treatment was determined as ‘died’, ‘end-stage renal disease’ (ESRD), ‘partially recovered’ (hospital discharge with a serum creatinine >2 mg/dl) or ‘completely recovered’. Patients who were transferred to other departments were followed-up to their discharge or death in hospital, and their last known serum creatinine was recorded.

From each of the eight participating institutions, every eligible patient was reported to the coordinating centre within

<table>
<thead>
<tr>
<th>Table 1. Definitions of clinical parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Heart failure (two or more criteria have to be present): clinical and X-ray signs of pulmonary congestion; lung oedema; PCW &gt;20 mmHg; CI &lt;2.2/min/m² with PCW &gt;12 mmHg; use of vasoactive drugs.</td>
</tr>
<tr>
<td>2. Myocardial infarction: within 48 h before T0, and documented by established enzymatic and ECG criteria.</td>
</tr>
<tr>
<td>3. Blood pressure: mean systolic and diastolic blood pressures over a period of 3 h during the day, measured up to three digits either invasively or non-invasively.</td>
</tr>
<tr>
<td>4. Hypotension: systolic blood pressure below 90 mmHg, one or more episodes.</td>
</tr>
<tr>
<td>5. CVP: central venous pressure measured invasively, mean of at least three consecutive measurements.</td>
</tr>
<tr>
<td>6. Need for inotropic drugs: any dosage and any inotropic drug—except dopamine used at a dose ≤2 µg/kg/min.</td>
</tr>
<tr>
<td>7. Respiratory support: either intubation with mechanical ventilation or assisted ventilation.</td>
</tr>
<tr>
<td>8. Narcosis: general anaesthesia before respiratory support.</td>
</tr>
<tr>
<td>9. Unconsciousness: unconscious without sedation. Patients under sedation are excluded for this parameter unless the status before sedation is known.</td>
</tr>
<tr>
<td>10. Glasgow Coma Scale: the sum of the score is collected, with the same remark as 9.</td>
</tr>
<tr>
<td>11. Sepsis (two or more criteria have to be present): (i) temperature &gt;38°C or &lt;36°C; (ii) heart rate &gt;90 beats/min; (iii) respiratory rate &gt;20 breaths/min or PaCO₂ &lt;32 mmHg; (iv) WBC &gt;12 000 cells/mm³, &lt;4000 cells/mm³ or &gt;10% immature band forms.</td>
</tr>
<tr>
<td>12. Gastrointestinal dysfunction, any of: intolerance of enteral feeding, GI bleeding requiring transfusion, acalculous cholecystitis, pancreatitis, liver dysfunction, diarrhoea, bowel perforation.</td>
</tr>
<tr>
<td>13. Liver dysfunction: bilirubin &gt;2 mg/dl.</td>
</tr>
</tbody>
</table>
24 h after diagnosis, preferably using an established reporting form after it had been filled with admission data. After completion of the T48 data, the form was sent to the study monitor within 1 week after diagnosis. Corrections and additional information were requested when necessary. All laboratory data were given in the units of the local laboratory, and were recalculated to the units of the coordinating centre.

**Statistical methods**

All the prospectively collected parameters were tested for significance with univariate analysis using the Student t-test and the \( \chi^2 \) test.

First, all recorded parameters were compared between centres, to evaluate the quality of data collection and differences in patient populations and interpretation of parameters.

Secondly, the performance of the original model with values for T0 and T48 (SHARF \(_0\) and SHARF \(_{48}\) [8]) was evaluated. The squared correlation coefficient \( \left( r^2 \right) \) of the linear regression analysis was used to test the explanatory power of the model. The areas under the ROC curves were used to judge the discrimination ability of our approach [13]. The degree of correspondence (‘fitting’) between score and outcome was determined comparing the observed with the expected calculated mortality using the Hosmer–Lemeshow goodness-of-fit test [15]. In this test, comparing the observed and predicted frequencies for up to 10 cells, a small \( P \)-value means that the predicted values do not fit the data.

Thirdly, since the performance of the model in this population clearly was weaker than its performance in the original population, we re-analysed the data in order to identify the best additional subset of variables to customise the original model for the new populations. The independent variables that were entered in the new multivariate analysis were the SHARF scores developed in the original single-centre study plus all other parameters that were significant on univariate analysis. Weighted scores were assigned to each parameter based on the regression coefficients of the linear regression analysis. Parameters with skewed distributions were divided into categories to neutralize the effect of outliers. Age was divided into decades to simplify the formulae for bedside calculation. The coefficients were multiplied by 100 and rounded to obtain whole numbers. The final adapted scores at T0 (SHARF \(_0\)) and at T48 (SHARF \(_{48}\)) were calculated by summing the weights associated with each parameter. In the linear model, the significant contributors to mortality were exactly the same as in the logistic model. The same had been observed in the original, single-centre study [8]. The probability of in-hospital mortality was calculated using the score as the single parameter in a logistic regression equation: \( \logit = b_0 + b_1 \text{score} \). The logit was then converted to a probability of hospital mortality as \( Pr(y = 1 | \logit) = \frac{e^{b_1 \text{score}}}{1 + e^{b_1 \text{score}}} \), where \( y \) is 1 for patients who died and 0 for patients who lived. \( Pr \) indicates probability and \( e \) indicates the base of the natural logarithm [14]. Risk ratios were calculated for the parameters found to be significant contributors. The adapted scores (SHARF \(_0\) and SHARF \(_{48}\)) were again tested using the \( r^2 \), ROC values and Hosmer–Lemeshow goodness-of-fit test [15]. In this test, comparing the observed and predicted frequencies for up to 10 cells, a small \( P \)-value means that the predicted values do not fit the data.

Fourthly, the model was also tested in individual centres with at least 30 patients included in the study and in certain subpopulations: pre-renal causes vs acute tubular necrosis; medical vs surgical patients; dialysed vs non-dialysed patients; and initial vs delayed diagnosis of ARF. The standardized mortality ratio (SMR; ratio of the observed number of deaths vs that predicted by the model) for different ICUs and subpopulations [16] and the ROC values were calculated with their 95% confidence intervals (CIs).

Finally, the original and the newly designed scores were compared with the Acute Tubular Necrosis Severity Index of Liano [17] and the APACHE II score [9].

**Results**

A total of 80 895 patients were admitted to the participating hospitals in the 6 month study period; of these, 5113 were admitted initially or after some time to the ICUs of those hospitals. A total of 293 patients (0.36% of hospital admissions and 5.73% of ICU admissions) suffered from ARF. In the ARF group, 148 patients (50.5%) died, 42 of them (28.3%) within 48 h after the onset of ARF. The causes of ARF were

<table>
<thead>
<tr>
<th>Table 2. Basic characteristics of the patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Total number</td>
</tr>
<tr>
<td>Age: mean (range)</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Length of hospital stay:</td>
</tr>
<tr>
<td>Time to ARF from hospital admission</td>
</tr>
<tr>
<td>Initial (at admission)</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>Type of ARF</td>
</tr>
<tr>
<td>Pre-renal</td>
</tr>
<tr>
<td>Renal</td>
</tr>
<tr>
<td>Post-renal</td>
</tr>
<tr>
<td>Acute on chronic disease</td>
</tr>
<tr>
<td>Specified renal causes of ARF</td>
</tr>
<tr>
<td>ATN</td>
</tr>
<tr>
<td>AGN</td>
</tr>
<tr>
<td>AIN</td>
</tr>
<tr>
<td>Systemic disease</td>
</tr>
<tr>
<td>Setting of ARF</td>
</tr>
<tr>
<td>Medical toxic</td>
</tr>
<tr>
<td>Medical other</td>
</tr>
<tr>
<td>Surgical</td>
</tr>
<tr>
<td>Obstetrical</td>
</tr>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td>Conservative</td>
</tr>
<tr>
<td>Haemodialysis</td>
</tr>
<tr>
<td>Peritoneal dialysis</td>
</tr>
<tr>
<td>CRRT</td>
</tr>
<tr>
<td>Outcome</td>
</tr>
<tr>
<td>Dead</td>
</tr>
<tr>
<td>ESRD</td>
</tr>
<tr>
<td>Partial recovery</td>
</tr>
<tr>
<td>Complete recovery</td>
</tr>
</tbody>
</table>

ARF = acute renal failure, ATN = acute tubular necrosis, AGN = acute glomerulonephritis, AIN = acute interstitial nephritis, CRRT = continuous renal replacement therapy, ESRD = end-stage renal disease.

*Group smaller than 10 patients: no mortality calculated.
medical in 184 (63.1%) and surgical in 108 (36.9%) patients. In the latter group, 74 (68.5%) had cardiac surgery and 19 (17.6%) vascular surgery. The average interval between ICU admission and diagnosis of ARF was 4.6 ± 18.3 days for those initially admitted to the ICU and 11.3 ± 16.0 for delayed admissions to the ICU (P-value of difference < 0.01). The basic descriptive characteristics of patients are summarized in Table 2.

Outcome was unaffected by gender, history of myocardial infarction, chronic renal failure, chronic obstructive lung disease, chronic liver disease, alcohol abuse, diabetes, underlying neoplasm, serum creatinine, glucose, sodium and haemoglobin, white blood cell count, urinary sodium or urine osmolality (a total of 15 parameters).

In univariate analysis, 20 other parameters were significantly different at T0 or T48 between survivors and non-survivors (Table 3). Using the original formulae, the $r^2$ decreased to values < 0.25 and the areas under the ROC curve to < 0.80.

In the present analysis of the data, the original SHARF score formula plus three additional parameters (bilirubin, sepsis and hypotension) contributed significantly to the predictive value of the hospital mortality model at 0 and 48 h.

The final scoring formulae were:

**SHARF II**

\[
\text{SHARF II}_0 = 3.0 \times \text{age decade} + 2.6 \times \text{serum albumin category T0} + 1.3 \times \text{prothrombin time category T0} + 16.8 \times \text{respiratory support T0} + 3.9 \times \text{heart failure T0} + 2.8 \times \text{serum bilirubin T0} + 27 \times \text{sepsis T0} + 21 \times \text{hypotension T0} - 17.
\]

**SHARF II**

\[
\text{SHARF II}_{48} = 3.9 \times \text{age decade} + 3.3 \times \text{serum albumin category T0} + 1.7 \times \text{prothrombin time category T0} + 23.7 \times \text{respiratory support T48} + 8.8 \times \text{heart failure 48 h} + 2.5 \times \text{serum bilirubin T48} + 24 \times \text{sepsis T48} + 17 \times \text{hypotension T0} - 28.
\]

The data on the parameters included in the model are shown in Table 4. The calculation of the score is

### Table 3. Parameters significantly different between survivors (n = 147) and non-survivors (n = 146)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Survivors Mean (SD)</th>
<th>Non-survivors Mean (SD)</th>
<th>P-value Student’s t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systolic blood pressure* (mmHg) T0</td>
<td>130 (26)</td>
<td>114 (28)</td>
<td>0.000</td>
</tr>
<tr>
<td>Systolic blood pressure* (mmHg) T48</td>
<td>133 (22)</td>
<td>121 (27)</td>
<td>0.000</td>
</tr>
<tr>
<td>Diastolic blood pressure* (mmHg) T0</td>
<td>67 (15)</td>
<td>58 (16)</td>
<td>0.000</td>
</tr>
<tr>
<td>Diastolic blood pressure* (mmHg) T48</td>
<td>68 (12)</td>
<td>60 (15)</td>
<td>0.000</td>
</tr>
<tr>
<td>Central venous pressure* (cm H2O) T0</td>
<td>9.8 (4.6)</td>
<td>12.4 (5.3)</td>
<td>0.000</td>
</tr>
<tr>
<td>Central venous pressure* (cm H2O) T48</td>
<td>9.8 (4.1)</td>
<td>11.9 (4.7)</td>
<td>0.001</td>
</tr>
<tr>
<td>Serum urea* (mg/dl) T0</td>
<td>1.2 (1.3)</td>
<td>2.7 (4.1)</td>
<td>0.000</td>
</tr>
<tr>
<td>Serum bilirubin (mg/dl) T0</td>
<td>1.2 (1.2)</td>
<td>3.5 (6.0)</td>
<td>0.000</td>
</tr>
<tr>
<td>Serum albumin* (g/l) T48</td>
<td>29 (7)</td>
<td>26 (6)</td>
<td>0.000</td>
</tr>
<tr>
<td>Serum albumin* (g/l) T48</td>
<td>29 (5)</td>
<td>25 (6)</td>
<td>0.000</td>
</tr>
<tr>
<td>Thrombocytes* (&gt;1000/mm3) T48</td>
<td>156 (88)</td>
<td>118 (86)</td>
<td>0.000</td>
</tr>
<tr>
<td>Prothrombin time* (%) T0</td>
<td>71 (19)</td>
<td>62 (22)</td>
<td>0.000</td>
</tr>
<tr>
<td>Prothrombin time* (%) T48</td>
<td>75 (20)</td>
<td>67 (21)</td>
<td>0.003</td>
</tr>
<tr>
<td>Urinary volume (l/24h) T48</td>
<td>1.7 (1.2)</td>
<td>1.1 (1.1)</td>
<td>0.000</td>
</tr>
<tr>
<td>Urinary urea* (g/24h) T0</td>
<td>8.0 (5.1)</td>
<td>5.9 (4.5)</td>
<td>0.002</td>
</tr>
<tr>
<td>Urinary urea* (g/24h) T48</td>
<td>10.3 (6.9)</td>
<td>7.7 (5.5)</td>
<td>0.005</td>
</tr>
<tr>
<td>Urinary creatinine (g/24h) T0</td>
<td>1.1 (0.7)</td>
<td>0.8 (0.6)</td>
<td>0.000</td>
</tr>
<tr>
<td>Urinary creatinine (g/24h) T48</td>
<td>0.7 (0.5)</td>
<td>0.6 (0.4)</td>
<td>0.044</td>
</tr>
<tr>
<td>Days on renal replacement therapy</td>
<td>6.9 (17.1)</td>
<td>3.7 (8.6)</td>
<td>0.044</td>
</tr>
<tr>
<td>Number of organs failed* T0</td>
<td>1.3 (1.2)</td>
<td>2.6 (1.2)</td>
<td>0.000</td>
</tr>
<tr>
<td>Number of organs failed* T48</td>
<td>0.9 (1.2)</td>
<td>2.4 (1.3)</td>
<td>0.000</td>
</tr>
<tr>
<td>APACHE II</td>
<td>19 (18–20)</td>
<td>24 (23–25)</td>
<td>0.000</td>
</tr>
<tr>
<td>Heart failure* T0</td>
<td>37.2</td>
<td>56.8</td>
<td>0.000</td>
</tr>
<tr>
<td>Heart failure* T48</td>
<td>27.8</td>
<td>49.1</td>
<td>0.000</td>
</tr>
<tr>
<td>Hypotension* T0</td>
<td>20.0</td>
<td>50.0</td>
<td>0.000</td>
</tr>
<tr>
<td>Hypotension* T48</td>
<td>06.3</td>
<td>25.2</td>
<td>0.000</td>
</tr>
<tr>
<td>Need for inotropic drugs* T0</td>
<td>43.1</td>
<td>71.6</td>
<td>0.000</td>
</tr>
<tr>
<td>Need for inotropic drugs* T48</td>
<td>29.2</td>
<td>65.5</td>
<td>0.000</td>
</tr>
<tr>
<td>Ventilation* T0</td>
<td>38.6</td>
<td>66.2</td>
<td>0.000</td>
</tr>
<tr>
<td>Ventilation* T48</td>
<td>20.1</td>
<td>67.2</td>
<td>0.000</td>
</tr>
<tr>
<td>Sepsis* T0</td>
<td>13.1</td>
<td>41.2</td>
<td>0.000</td>
</tr>
<tr>
<td>Sepsis* T48</td>
<td>09.7</td>
<td>43.1</td>
<td>0.000</td>
</tr>
<tr>
<td>Gastrointestinal failure* T0</td>
<td>11.0</td>
<td>24.3</td>
<td>0.005</td>
</tr>
<tr>
<td>Gastrointestinal failure* T48</td>
<td>9.0</td>
<td>19.8</td>
<td>0.020</td>
</tr>
<tr>
<td>Immunological failure* T0</td>
<td>5.7</td>
<td>15.2</td>
<td>0.015</td>
</tr>
<tr>
<td>Immunological failure* T48</td>
<td>5.7</td>
<td>17.5</td>
<td>0.003</td>
</tr>
</tbody>
</table>

*Significant in developmental phase during single-centre study.

*aFor definitions, see Table 1.
The mean values and 95% CIs of SHARF II scores were 36 (33–40) for survivors and 63 (59–67) for non-survivors at T0, and 32 (28–36) and 67 (61–72) at T48, respectively. The predicted probability of death for any possible score is given in Figure 1.

The $r^2$ values were 0.27 at T0 and 0.33 at T48. The discriminative power of the new scoring system showed ROC values of 0.82 and 0.83, respectively. The calibration, tested with the Hosmer–Lemeshow goodness-of-fit C statistic, was for T0, $C = 9.95$, df = 8, $P = 0.191$; and for T48, $C = 15.66$, df = 8; $P = 0.048$.

The SHARF score for the prediction of hospital mortality, first developed in a single-centre study [8], was tested in a new and different cohort of intensive care patients in eight centres, including the original centre. The aim of the study was to evaluate the original model in different populations and, if necessary, to modify it. Indeed, we found that there clearly was a different case mix in these centres compared with the original single centre, and that these variations can have significant impacts on the performance of the scoring model [18,19].

In the present multicentre study, the performance of the SHARF score developed in the original study deteriorated compared with the original single-centre results (ROC values of 0.67 and 0.78 compared with 0.87 and 0.90). Nevertheless, even these results compare favourably with those of similar efforts reported in the literature [1,20]. Douma et al. compared 11 mortality prediction models [20]. They found that, of all ARF scores, only the Liano model [17] had an ROC value of 0.78. Halstenberg et al., testing different models in their own study population, found a low performance [1]. In our series, the ROC values for the Liano model [17] and the APACHE II scores [9] were lower than for the SHARF scores. The goodness-of-fit was acceptable for the Liano scores but not for APACHE II. This lack of fit can be explained by the large number of cardiac surgery patients in our cohort ($n = 74$). As in the

**Table 4.** Details of variables included in SHARF II$_0$ and SHARF II$_{48}$

<table>
<thead>
<tr>
<th>Age decade (10–19 years = 1; 20–29 = 2)</th>
<th>INR values (30)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum albumin category at T0 (ALBCAT) g/l</td>
<td>&gt;45 = 1</td>
</tr>
<tr>
<td>41–45 = 2</td>
<td>36–40 = 3</td>
</tr>
<tr>
<td>31–35 = 4</td>
<td>26–30 = 5</td>
</tr>
<tr>
<td>21–25 = 6</td>
<td>≤20 = 7</td>
</tr>
<tr>
<td>Prothrombin time category at T0 (PTCAT) %</td>
<td>≤&lt;1.14</td>
</tr>
<tr>
<td>&gt;80 = 1</td>
<td>(1.28–1.14)</td>
</tr>
<tr>
<td>66–80 = 2</td>
<td>(1.43–1.30)</td>
</tr>
<tr>
<td>56–65 = 3</td>
<td>(1.66–1.46)</td>
</tr>
<tr>
<td>46–55 = 4</td>
<td>(2.00–1.67)</td>
</tr>
<tr>
<td>36–45 = 5</td>
<td>(2.58–2.03)</td>
</tr>
<tr>
<td>26–35 = 6</td>
<td>(3.90–2.68)</td>
</tr>
<tr>
<td>16–25 = 7</td>
<td>(&gt;3.90)</td>
</tr>
<tr>
<td>Respiratory support (vent):</td>
<td></td>
</tr>
<tr>
<td>no = 0, yes = 1 (criteria see Table 1)</td>
<td></td>
</tr>
<tr>
<td>Heart failure (CHF):</td>
<td></td>
</tr>
<tr>
<td>no = 0, yes = 1 (criteria see Table 1)</td>
<td></td>
</tr>
<tr>
<td>Serum bilirubin (Bili): μmol/l</td>
<td></td>
</tr>
<tr>
<td>&gt;20 = 7</td>
<td></td>
</tr>
<tr>
<td>20–30 = 6</td>
<td></td>
</tr>
<tr>
<td>16–25 = 5</td>
<td></td>
</tr>
<tr>
<td>10–15 = 4</td>
<td></td>
</tr>
<tr>
<td>≤&lt;10 = 3</td>
<td></td>
</tr>
<tr>
<td>Sepsis (Seps) no = 0, yes = 1 (criteria see Table 1)</td>
<td></td>
</tr>
<tr>
<td>Hypotension (Hypo):</td>
<td></td>
</tr>
<tr>
<td>no = 0, yes = 1 (criteria see Table 1)</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Values for INR are approximate.

**Fig. 1.** Probability of death curves at 0 and 48 h for SHARF II. The probability of in-hospital mortality was calculated by using the score as the single parameter in a logistic regression equation [14].

**Fig. 2.** Calibration curves for SHARF II$_0$ and SHARF II$_{48}$ for 10% probability intervals.

Figure 2 shows graphically the calibration of the observed and the predicted mortalities in probability intervals of 10%.

**Discussion**

The SHARF score for the prediction of hospital mortality, first developed in a single-centre study [8], was tested in a new and different cohort of intensive care patients in eight centres, including the original centre. The aim of the study was to evaluate the original model in different populations and, if necessary, to modify it. Indeed, we found that there clearly was a different case mix in these centres compared with the original single centre, and that these variations can have significant impacts on the performance of the scoring model [18,19].

In the present multicentre study, the performance of the SHARF score developed in the original study deteriorated compared with the original single-centre results (ROC values of 0.67 and 0.78 compared with 0.87 and 0.90). Nevertheless, even these results compare favourably with those of similar efforts reported in the literature [1,20]. Douma et al. compared 11 mortality prediction models [20]. They found that, of all ARF scores, only the Liano model [17] had an ROC value of 0.78. Halstenberg et al., testing different models in their own study population, found a low performance [1]. In our series, the ROC values for the Liano model [17] and the APACHE II scores [9] were lower than for the SHARF scores. The goodness-of-fit was acceptable for the Liano scores but not for APACHE II. This lack of fit can be explained by the large number of cardiac surgery patients in our cohort ($n = 74$). As in the
original APACHE II publication, these patients had better than expected prognoses and were not included in the analysis [9].

In an attempt to overcome these problems, we performed a new multivariate analysis of the multicentre data in order to identify the best additional subset of parameters that help to explain the mortality in this cohort. Accordingly, adding other parameters to those in the original scoring yielded clearly higher values for $r^2$, the ROC and the Hosmer–Lemeshow goodness-of-fit statistics. Because there were not enough patients to divide the subjects randomly into a developmental and a validation sample (the split-sample method), the method of subgroup analysis was used to test the new customized scoring [21]. Therefore, the same tests of calibration and discrimination were performed for individual centre populations and subpopulations according to types and settings of renal failure. For most of the centres and the subgroups, the discriminative powers, as expressed by ROC values, were comparable with the overall results. Additionally, the SMRs confirmed the good ‘fit’ of the predicted mortality with the observed mortality in most centres and subgroups, but there was a large degree of ‘uncertainty’ as expressed by the 95% CIs [22].

The question remains of whether or not other outcomes such as those related to long-term survival and quality of life issues, which probably are more relevant to individual patients, can be predicted by any model that has been developed up to now [20]. This can only be answered at a later stage. In our study, we confronted the SHARF II score with 1-year survival rate (manuscript in preparation).

The SHARF II score at T48 predicted mortality better than at diagnosis (T0), as had already been found in the developmental phase of the original model [8]. The discriminative power of the SHARF II in the individual centres and subgroups of this study confirms the improvement of its performance at T48. In general studies of outcomes in intensive care settings, some authors recently paid attention to repeated scores [23,24]. This novel approach of adding a second measuring point to ameliorate the accuracy of the prediction increases the possibility of using the score to make predictions for individual patients.

We conclude that our original scoring model, developed in our single-centre study, with two measuring points showed an acceptable performance in its multicentre trial. The premonitory value of T48 scores is as good as the best results published until now.

However, customizing the original scoring model for the multicentric population of this study makes this model for predicting hospital mortality in ARF useful in different ICU settings. The use of the score to compare centres, for quality assessment and for clinical trials, is certainly justified. Although the predictive abilities of this method are very good, we are reluctant to recommend its use for individual patients over the entire range of possibilities. The performance of the model in patients with very high or very low probability of death, however, makes the score suitable for prognostication in such patients.

Acknowledgements. The authors thank A. Rombaut for excellent clinical monitoring of the study. The study was made possible by the commitment and cooperation of the following investigators who, in addition to the authors, collected the data in the participating centres: J. Berwaerts, General Hospital Middelheim, J. Berrens, General Hospital Stuivenberg, C. Claesen, St Joseph Hospital Turnhout, H. Demey, University Hospital Antwerp, I. Demeyer, O.L. Vrouw Hospital, E. J. Huygens, Free University Brussels, J. Nagler, General Hospital Middelheim, G. Nollet, O.L. Vrouw Hospital, P. Peeters, Free University Brussels, D. Van Caesbroeck, St Joseph Hospital Turnhout and F. Zachee, General Hospital Stuivenberg. The study was supported by the departmental fund F001 of the Department of Nephrology-Hypertension.

Conflict of interest statement. None declared.

References

Appendix. Calculation of customized SHARF II scores

Original SHARF score [8]:

\[
SHARF_0 = (7 \times \text{agedec}) + (6 \times \text{albcat}) + (3 \times \text{PTcat}) + (39 \times \text{vent 0 h}) + (9 \times \text{CHF}_0 \text{ h}) + 52
\]

\[1a\]

\[
SHARF_{48} = (7 \times \text{agedec}) + (6 \times \text{albcat}) + (3 \times \text{PT}) + (43 \times \text{vent 48 h}) + (16 \times \text{CHF}_{48} \text{ h}) + 52
\]

\[1b\]

Age decade (agedec), albumin category (albcat) and prothrombin time category (PTcat) are expressed as categories, respiratory support (vent) and heart failure (CHF) are presented as absent (0) or present (1).

The result of the new multivariate analysis at T0 (diagnosis) was:

\[
SHARF_{II0} = 0.43 \times SHARF_0 + 2.8 \times \text{bili T0} + 27 \times \text{seps T0} + 21 \times \text{hypo T0} - 39
\]

\[2a\]

\[
SHARF_{II48} = 0.55 \times SHARF_{48} + 2.5 \times \text{bili T48} + 24 \times \text{seps T48} + 17 \times \text{hypo T0} - 57
\]

\[2b\]

Bilirubin (bili) is expressed as mg/dL, sepsis (seps) and hypotension (hypo) are presented as absent (0) or present (1).

Substituting \(SHARF_0\) in equation 2a by equation 1a and algebraic transformation gives the following result:

\[
SHARF_{II0} = 3.0 \times \text{agedec} + 2.6 \times \text{albcat T0} + 1.3 \times \text{PTcat T0} + 16.8 \times \text{vent T0} + 3.9 \times \text{CHF T0} + 2.8 \times \text{bili T0} + 27 \times \text{seps T0} + 21 \times \text{hypo T0} - 17
\]

\[3a\]

\[
SHARF_{II48} = 3.9 \times \text{agedec} + 3.3 \times \text{albcat T0} + 1.7 \times \text{PTcat T0} + 23.7 \times \text{vent T48} + 8.8 \times \text{CHF}_{48} \text{h} + 2.5 \times \text{bili T48} + 24 \times \text{seps T48} + 17 \times \text{hypo T0} - 28
\]

\[3b\]