Validation of a clinical score to determine the risk of acute renal failure after cardiac surgery

Daniel Heise*, Daniela Sundermann, Anselm Braeuer, Michael Quintel

Department of Anesthesiology, Emergency and Intensive Care Medicine, University Hospital Goettingen, Goettingen, Germany

Received 31 March 2009; received in revised form 6 July 2009; accepted 7 July 2009; Available online 27 August 2009

Abstract

Objective: Preoperative assessment of risk factors for postoperative acute renal failure (ARF) is a key point in the management of cardiac surgery patients. A reliable evaluation of individual risk permits the efficient application of prophylactic and therapeutic measures to patients with increased risk of ARF. In 2005, Thakar published a scoring system to determine the ARF risk prior to surgery. The score allowed calculation of the incidence of postoperative ARF requiring dialysis (ARF-D) with a high level of precision. Methods: In order to examine the reproducibility of the results on different patient populations, we retrospectively applied the scoring system to 3508 cardiac surgery patients at the Goettingen University Hospital. Results: We found a qualitative correlation between the score value and ARF-D, but prognostic significance, when judged by the area under the receiver operating characteristic (ROC) curve, was considerably reduced. Also, the incidence of ARF-D in the Goettingen population was remarkably higher than the incidence shown by Thakar for patients with the same score. Conclusions: The correlation between score values and incidence of ARF-D shows that the Thakar score is qualitatively transferable to cardiac surgery patients from other centres. Though the score allows the discrimination between patients with higher or lower risks of ARF-D within the Goettingen collective, it was not suitable to estimate the real incidence of postoperative ARF-D with sufficient precision.

Keywords: Acute renal failure; Prediction; Cardiac surgery

1. Introduction

Acute renal failure (ARF) is one of the most frequent complications encountered after cardiac surgery. This is especially true when a cardiopulmonary bypass (CPB) pump is employed [1,2]. An increased expenditure of resources [3] and, notably, an exceedingly increased mortality of up to 50% [4] illustrate the great impact of this problem for intensive care practice. In 2005, Thakar published a scoring system to estimate the individual risk of ARF requiring dialysis (ARF-D) after cardiac surgery [5]. Uni- and multivariate analyses of 15,000 patients showed a significant influence for 11 of 16 hypothetical preoperative risk factors for ARF-D after cardiac surgery. From the relative risk of each factor, a score was calculated that spanned a range from 0 to 17 points. This score was retrospectively validated in another 15,000 patients. In the validation group, the frequency of ARF-D at each score level fell within the 95% confidence interval (CI) of the predicted ARF incidence.

Despite the high level of precision shown for the correct prediction of postoperative ARF-D in the validation population, the authors mentioned that the score was developed based on a monocentric patient population as a possible limitation. Therefore, it remains to be verified whether calibration and discrimination of the score is similar in cardiac surgery patients from other centres. We, therefore, aimed to reassess the reproducibility of the scoring system developed by Thakar by applying it to a patient population of a second cardiac surgery centre. This also included the question of whether recalibration of the score (based on the relative risk values within our collective) could improve the power of discrimination between patients with and without postoperative ARF-D.

2. Material and methods

2.1. Patients

In order to calculate the Thakar score, we extracted the raw data of 3628 patients from the database of patients who underwent cardiac surgery between 2002 and 2005. The exclusion criteria were the preoperative need for dialysis, preoperative cardiac assist devices and irrecoverable missing...
data (when it rendered the calculation of the score value impossible). A total of 120 patients met at least one of these criteria so that the final analysis included 3508 patients.

We defined acute renal failure as the postoperative use of renal replacement therapy. According to the current protocols of our intensive care unit (ICU), indications for the initiation of renal replacement therapy were hyperkalaemia, severe fluid overload and uraemia leading to neurological deficits.

2.2. Data collection

Patient data was derived retrospectively from the GISI (Göttinger Informationssystem für Intensivmedizin), a patient data management system developed in the 1980s by the department for information systems at the University Hospital of Goettingen. Five of the eight ICUs at the hospital are running this system.

On admission to the ICU, the entry of most of the data needed to calculate the score was mandatory. Nevertheless, some data (like medical history) were entered as free text so that the completeness of this information could not be assessed in retrospect. Likewise, the majority of the data were objective values (e.g., gender and preoperative creatinine), whereas the presence of risk factors, like congestive heart failure, sometimes had to be assessed indirectly or from more subjective data such as medications. To avoid any interpretation bias, the evaluation of all patients regarding the presence of risk factors was done by one physician. After analysis of all the data, reliable assignment of risk factors was not possible in only 3.3% of all patients, causing their exclusion from the study.

2.3. Statistical analyses

First of all, we performed a univariate analysis on the distribution of risk factors between the score groups with and without ARF-D. We interpreted the distribution of the ordinal variables using $\chi^2$, whereas we used an unpaired $t$-test for continuous variables. Next, we performed a multivariate analysis using logistic regression on all factors that had a significantly imbalanced distribution between the groups. According to the original Thakar score, we calculated the score points for the predictors, which significantly increased the relative ARF-D risk by multiplying the regression coefficient of the logistic regression by two and rounding it to the nearest integer.

To assess the power of discrimination between patients with and without ARF-D, we calculated the area under the receiver operating characteristics (ROC) curve applying the original Thakar score to the Goettingen collective. As the score derived from the Goettingen collective differed slightly from the original score, we also assessed the sensitivity and specificity for the prediction of ARF-D in the Goettingen population for the modified score using ROC analysis.

MedCalc® Version 9.3 (MedCalc Software, Mariakerke, Belgium) was used to perform the statistical analyses. The limit of statistical significance probability for all tests was fixed at a $p$ value $<0.05$.

3. Results

3.1. The frequency of score values and related incidence of ARF requiring dialysis (Table 1)

Score levels of the Goettingen population ranged between 0 and 11 points (Cleveland: 0—13 points). Qualitatively, the relative frequency of the score values was similarly distributed in the Goettingen and Cleveland populations. This was true for single score levels as well as for risk groups comprising several score values (0—2, 3—5, 6—8 and 9—13 points). The largest group in both collections

<table>
<thead>
<tr>
<th>Score</th>
<th>Goettingen</th>
<th>Cleveland</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency of score values</td>
<td>ARF-dialysis</td>
</tr>
<tr>
<td>0</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>1</td>
<td>13.28</td>
<td>466</td>
</tr>
<tr>
<td>2</td>
<td>19.07</td>
<td>669</td>
</tr>
<tr>
<td>3</td>
<td>21.58</td>
<td>757</td>
</tr>
<tr>
<td>4</td>
<td>17.50</td>
<td>614</td>
</tr>
<tr>
<td>5</td>
<td>12.86</td>
<td>451</td>
</tr>
<tr>
<td>6</td>
<td>7.30</td>
<td>256</td>
</tr>
<tr>
<td>7</td>
<td>3.99</td>
<td>140</td>
</tr>
<tr>
<td>8</td>
<td>2.11</td>
<td>74</td>
</tr>
<tr>
<td>9</td>
<td>1.77</td>
<td>41</td>
</tr>
<tr>
<td>10</td>
<td>0.74</td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>0.31</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>0.09</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>0.00</td>
<td>0</td>
</tr>
</tbody>
</table>

Risk groups

| 0—2  | 53.93 | 1892 | 6.71 | 53.14 | 8416 | 0.4 | 0.28—0.56 |
| 3—5  | 37.66 | 1321 | 12.72 | 38.50 | 6097 | 1.8 | 1.5—2.2 |
| 6—9  | 7.27  | 255  | 29.02 | 7.46  | 1181 | 7.8 | 6.3—9.5 |
| 9—13 | 1.14  | 40   | 35    | 0.91  | 144  | 21.5 | 15.1—29.1 |
was formed by patients with 0–2 points and the maximum frequency was found at two points. Fig. 1a and b shows the relative frequency of the single score values and the four risk groups. In contrast, we found the incidence of ARF-D to not only be higher in the Goettingen population than in the Cleveland data set, but also to be outside the 95% CI range, as predicted by Thakar. This significant difference concerned both the single score levels and the four risk groups (Fig. 2a and b).

3.2. Univariate examination of the predictors (Tables 2 and 3)

Seven of the risk factors composing the score were distributed with a significant disparity between the two patient groups. Tables 2 and 3 show the distribution of the categorical and continuous predictors, respectively. In the Goettingen population, we did not detect a significant difference between the groups with and without ARF-D with respect to the variables of age (\( p = 0.068 \)), peripheral vascular disease (\( p = 0.15 \)), chronic obstructive pulmonary disease (\( p = 0.22 \)), congestive heart failure (\( p = 0.27 \)) or type of surgery (\( p = 0.78 \)).

3.3. Multivariate examination of the predictors (Table 4)

Using logistic regression, we examined the influence of the predictors that had been distributed with a significant difference between the two patient groups in the univariate analysis. We did not find a statistically significant influence on the development of ARF-D for the risk factors of previous cardiac surgery, ejection fraction <35% or non-insulin-dependent diabetes mellitus. Based on the regression coefficient, we derived a modified score from the remaining risk factors. This was done corresponding to the method used by Thakar. A comparison of the two scores is in Table 4.

3.4. ROC analysis of the original score (Fig. 3)

When we applied the Thakar score to predict ARF-D in our patients, we found that the sensitivity and specificity were diminished compared to Thakar’s validation data. The area under the ROC curve was 0.662 with a 95% CI of 0.646 to 0.678 (Thakar: 0.81).

3.5. ROC analysis of the modified score (Fig. 4)

Additionally, we carried out a separate assessment of the modified score that was created based on the Goettingen data. ROC analysis was used to determine the sensitivity and specificity for the prediction of ARF-D. The area under the curve was 0.671 (95%CI: 0.655–0.687). Thus, we did not find a statistical difference in the area under the curve derived from the application of the original score to the Goettingen population.
4. Discussion

4.1. Objectives

Preoperative evaluation of ARF-D risk is crucial for the treatment of cardiac surgery patients. Besides individual risk factors that are preoperatively fixed cannot usually be influenced by the attending physician, the probability of postoperative ARF-D depends on intra- and postoperative treatment strategies. For example, the incidence can be reduced by using off-pump techniques rather than CPB [6], by short CPB times [7], by optimising the haematocrit to between 21% and 25% [8] and by optimising oxygen delivery [9].

To avoid acute renal failure, evidence-based recommendations exist regarding haemodynamic parameters, for example, mean arterial pressure and cardiac output. This is also true for the management of fluid balance [10—12]. In contrast, pharmacological monotherapies with N-acetylcysteine [13,14] or the selective dopamine 1-receptor agonist Fenoldopam [13] for example, only led to a slight modification of renal function. Nevertheless, the evaluation of the ARF-D risk prior to surgery is imperative for systematically applying every possible therapeutic measure to patients in need.

Strikingly high death rates of 50% illustrate not only the need for a systematic application of therapeutic measures for high-risk patients, but also the requirement for further study.
clinical trials that investigate the prevention and treatment of ARF-D. To facilitate the application of innovative nephroprotective measures for high-risk patients and to evaluate the clinical effects, it is necessary to accurately quantify the supposed ARF-D risk. In this respect, the Thakar score is an essential contribution. This is why the question of reproducibility in other cardiac surgery patient populations is fundamental.

Since 2005, at least three more research groups have developed scoring systems for the prediction of ARF after cardiac surgery [15–17]. The main differences between them and Thakar’s study concerned definition of acute renal failure and the inclusion of some intra-operative risk factors such as cardiopulmonary bypass time. As collection and processing of our data was nearly completed when the other scoring systems were published, we decided not to evaluate them and confined our study to just the analysis of the Thakar score only. Consequently, we defined acute renal failure as the postoperative use of renal replacement therapy, being aware that acute renal failure is now defined much more objectively by the RIFLE or AKIN criteria. Therefore, we used the term ARF-D (acute renal failure requiring dialysis) instead of acute kidney injury.

4.2. Size of the patient population, quality of data

When sizing the patient population, several considerations were taken into account. We made an allowance for the basic idea that single risk factors would cease to be statistically significant if the group size was diminished. This assumption is affirmed by the results of Di Bella, who validated the Thakar score in 1642 patients [18]. Only five of 11 risk factors differed significantly in the univariate analysis between the patient groups with and without renal replacement therapy. On the other hand, an extended evaluation period can lead to systematic errors owing to changes in treatment standards over time. Furthermore, the Goettingen patient data management system was continually enhanced during the time period studied. Thus, the homogeneity of the data decreases with the increasing size

![Fig. 3. ROC curve of the Thakar score for the prediction of postoperative ARF-D in the Goettingen population. AUC = 0.662 (95% CI: 0.646–0.678).](image)

![Fig. 4. ROC curve of the modified Thakar score for the prediction of postoperative ARF-D in the Goettingen population. AUC = 0.671 (95% CI: 0.655–0.687).](image)
of the sample. In Goettingen, about 1000 patients annually undergo cardiac surgery involving CPB. Therefore, we considered the analysis of 3000–4000 patients to be a suitable compromise between the size of the patient population and the comparability of the data sets.

In principle, the data management system provided all necessary information required for the Thakar score. Differences in data quality were due to several factors. Most but not all information was obligatory at the admittance of new patients (e.g., type of surgery). The integrity of other data (e.g., previous medication) was not systematically verified. Furthermore, we did not dispose of objective, quantitative records (which were available, e.g., for ejection fraction) for all aspects of the patient data. Some features were gathered as entirely qualitative attributes in the form of free text (e.g., previous diseases). This is why the calculation of the individual score points are also dependent, although to a minor degree, on the evaluator’s interpretation. To minimise the interpretation bias as a source of systematic error, the interpretation of data and the allocation of risk factors were carried out by a single physician.

4.3. Results

While the incidence of ARF-D qualitatively increased concordantly with increasing score levels both in the Goettingen and Cleveland patient populations, we found that the quantitative results disagreed in three conditions: first, in the univariate analysis of the Goettingen population, not all risk factors differed significantly between the two patient populations. Some of these factors, which were therefore not taken into account for further analysis, failed to meet the level of significance by a small margin. Hence, a greater patient population would possibly have led to the inclusion of more risk factors in the multivariate analysis. Second, the discrimination of the score regarding ARF-D was also significantly lower than in Thakar’s validation data. A possible reason could be the loose definition of some risk factors. A number of the predictors cannot be defined in an entirely objective way (e.g., congestive heart failure or chronic obstructive pulmonary disease (COPD)). These factors are subject to the evaluator’s interpretation. In addition, the structure of the Goettingen data demanded, in the case of missing information, the indirect evaluation of some risk factors, for example, by means of previous medication. Third, the most striking difference was the inequality of the ARF-D incidence, which, in the case of the Goettingen population, lay outside of the 95% CI indicated by Thakar for all score levels. Candela-Toha and co-workers, who validated the Thakar score on 1780 patients, made the same observation [19]. They qualitatively found a significantly higher incidence of ARF-D than in the population studied by Thakar. Though this noticeable difference could be due to local differences in the balance between intra- and postoperative risk factors, we assume that the main reason for this result is the fact that acute renal failure was defined by the requirement for renal replacement therapy. Although this criterion can be objectively checked in retrospect, it nevertheless poses an important problem, namely that there is a great variance in the handling of the indications for renal replacement therapy [20]. The task of finding an objective definition of ARF is therefore not only fundamental for clinical practice [21] but especially for the interpretation of study results. Referring to this topic, the RIFLE criteria have recently been developed [22]. The AKIN classification [23], as an enhancement of the former, allows for a simple and objective assessment of ARF-D in a state when renal replacement therapy is not yet absolutely indicated. The RIFLE criteria were first published in 2005. This is why these definitions of different levels of acute kidney injury were only applied later in clinical practice.

Preoperative risk factors do have a decisive influence on the development of postoperative ARF-D. This finding is supported by the qualitative connection between the score level and the incidence of ARF-D even though renal replacement therapy was adopted significantly more often in the Goettingen population.

5. Limitations of the study

First, the expressiveness of our study is limited by the number of patients. For some of the risk factors, differences in the distribution missed the level of statistical significance only marginally. Thus, analysis of a larger collective probably would have led to a more precise identification of those factors, which have an influence on the development of ARF-D. Second, retrospective analysis of a database not always allowed the objective and direct rating of all risk factors. Although interpretation of the database content was done by one single physician, a prospective analysis based on the complete patient charts probably would lead to a higher quality of the data.

6. Conclusion

The Thakar score is convenient for monocentric use, provided it has been qualitatively calibrated on a large patient population by that centre. The results of monocentric studies could be more transferable if acute kidney injury is defined by the AKIN criteria. Furthermore, the results of larger, multicentric trials could possibly be interpreted with a higher grade of validity if the presence of risk factors is assessed in an objective, ideally prospective manner.

References


