Assessment of SOFA Score as a Diagnostic Indicator in Intensive Care Medicine

James N. Hewett*, Geoffrey W. Rodgers***
J. Geoffrey Chase*, Aaron J. Le Compte*, Christopher G. Pretty* and Geoffrey M. Shaw**

*Dept. of Mechanical Engineering, University of Canterbury, Christchurch, New Zealand 8140
(Tel: 64-3-3642596; e-mail: geoff.rogers@canterbury.ac.nz, geoff.chase@canterbury.ac.nz, aaron.lecompte@canterbury.ac.nz, christopher.pretty@pg.canterbury.ac.nz).

**Dept. of Orthopaedic Surgery and Musculo-Skeletal Medicine, University of Otago Christchurch, New Zealand (e-mail: geoff.shaw@otago.ac.nz)

Abstract: Intensive care unit mortality is strongly associated with organ failure rate and severity. The sequential organ failure assessment (SOFA) score is assessed to evaluate its efficacy as a diagnostic indicator. Statistical analyses investigate the SOFA score distributions in the days leading up to patient mortality and patient discharge. It is found that the SOFA score is not an effective predictor of patient mortality, but it is a useful tool for prediction of patient discharge from the Intensive Care Unit (ICU). The distribution of overall SOFA score was observed not to change notably in the days leading up to patient death. However, the SOFA score distribution was observed to have a trend towards lower SOFA scores in the days leading up to patient discharge. Finally, assessment of the individual components of the overall SOFA score indicated that the coagulation and cardiovascular scores showed the highest correlation to mortality and are therefore the most useful individual groups to be used as diagnostic indicators.

Keywords: Biomedical Systems, Critical Care, Sequential Organ Failure Assessment (SOFA), Patient Testing, Intensive Care Unit, Intensive Care Medicine.

1. INTRODUCTION

The development of reliable prognostic models is the subject of ongoing research in many fields of medicine. Their potential and use in clinical management of patients can have significant health, social and economic impact (Beck et al. 2003; Lin et al. 2007; Lilja et al. 2008; Juneja et al. 2011). Hence, the search for better, more accurate, and better discriminating models that require only the most certain and measurable of input data is an ongoing task (Beck et al. 2003; Rowan et al. 2007; Minne et al. 2008).

A primary goal is to ensure these models are adequately validated in development. A second major goal in this search is to ensure utility of the prognostic model through the need for minimal or straightforward input data. In particular, an ideal model would require minimal, easily available data that had minimal error or variability, while providing a highly accurate prediction.

The use of prognostic models for mortality has a long history in critical care. One main use has looked towards managing patient care and treatment decisions (Lilja et al. 2008). A greater use has been for delineating and comparing Intensive Care Unit (ICU) performance or quality (Timsit et al. 2001; Glance et al. 2002; Bakhshi-Raiez et al. 2007). Their use is becoming more evident as critical care resources become increasingly stressed economically and in terms of nursing or clinician time. A good prognostic model could be used to better allocate resources in these situations, which would lead to more efficient care without reducing outcomes in the consistency or quality of care given.

After the first 2-3 days of patient stay, mortality in the ICU and in-hospital are strongly associated with, and/or attributable to, organ failure and sepsis (Messahel et al. 1989; Tran et al. 1990; Sakr et al. 2008). In particular, a lack of organ failure resolution over a patient’s stay is associated with increased morbidity and mortality, as commonly measured by the sequential organ failure assessment (SOFA) score (Vincent et al. 1996; Vincent et al. 1998; Vincent 2006). However, the specific mechanisms linking organ failure to mortality are not necessarily fully understood (Van den Berghe 2004; Ellger et al. 2008; Langouche et al. 2008; Koch et al. 2009), particularly in terms of the specific physiological effects of organ failure leading to mortality. That said, it is equally well understood that the cascade of stress on major organ and system functions due to organ failure leads to increased risk of infection and decreased ability to recover, eventually resulting in death (Cryer 2000; Jeschke et al. 2004).

This investigation examines a set of patient data that contains two cohorts of patients, with 371 patients (3356 days) on the SPRINT tight glycemic control algorithm (August 2005 – April 2007) and 413 retrospective patients from 2 years prior that pre-dated the introduction of SPRINT (3211 days). The data forms the basis of a prior study into the effects of the SPRINT protocol. The differences in the two cohorts that underwent different glycemic control and the effects on mortality are presented in (Chase et al. 2010). In this study, SOFA score is assessed as a prognostic indicator and the type of glycemic control provided to the patients is not considered. Therefore, both patient cohorts are grouped and overall conclusions are drawn across the entire 784 patient cohort.
2. METHODS

2.1 SPRINT Glycemic Control Protocol

SPRINT is a model-derived (Lonergan et al. 2006; Lonergan et al. 2006) TGC protocol developed from clinically validated computer models used for real-time control in the ICU (Chase et al. 2005; Lonergan et al. 2006; Wong et al. 2006). Implemented at the Christchurch Hospital Department of Intensive Care in August 2005 (Chase et al. 2008), SPRINT has been used on over 1,000 patients. In a clinical comparison to statistically matched retrospective cohorts, the SPRINT TGC intervention reduced hospital mortality for those patients staying 3-5 days in the ICU by 25-40% (Chase et al. 2008). For the purposes of this investigation, the SPRINT and pre-SPRINT cohorts are considered together. It is assumed that if patients treated with the SPRINT protocol have an improved prognosis, that this will be reflected in improved SOFA scores.

2.2 Patient Data

This study uses data from 371 patients treated on SPRINT (August 2005 – May 2007) and 413 patients from (January 2003 – August 2005) prior to SPRINT, as in the original study. In this study, both cohorts are combined to analyse the impact of organ failure and outcome. The Upper South Regional Ethics Committee New Zealand granted ethics approval for the audit, analysis and publication of this data.

2.3 Organ Failure Assessment

Hospital records were examined for all patients and each day of ICU stay. The total SOFA score (Vincent et al. 1996; Vincent et al. 1998; Moreno et al. 1999) was calculated daily for each patient, taking the most abnormal value for each parameter in each 24hr period of ICU stay. Where a data point was missing or not available for a component, a value was interpolated from surrounding data. In this study, the Glasgow Coma score reflecting central nervous system function was excluded due to its reported lack of robustness and unreliability (Hellawell et al. 2000; Udekwu et al. 2004; Arts et al. 2005; Kerby et al. 2007; Tallgren et al. 2009), and it is thus not consistently recorded in Christchurch Hospital. Other studies have made a similar exclusion (Zygun et al. 2006). The remaining 5 SOFA component scores are each directly related to organ function or failure, and thus yield a maximum score of 20 (0-4 per metric). The parameters used assess: renal, cardiovascular, liver, respiratory function, and blood coagulation. A high SOFA score indicates a high level of organ dysfunction. Further details are in (Chase et al. 2010).

2.4 Analysis and Statistics

The primary goal of this study is to investigate the SOFA score of patients during their stay in the ICU and examine whether the SOFA score can be used as a prognostic indicator. First, the probability of mortality against SOFA score is investigated to determine the overall correlation of increasing SOFA score and increase chance of mortality. The SOFA score distributions in the days before mortality are examined for trends and predictive ability. Conversely, the SOFA score distributions in the days leading up to patient discharge are also examined to determine if SOFA score is an effective tool to identify a positive prognosis. Finally, the probability of mortality is examined for individual SOFA score components to investigate if specific components are better indicators of patient outcome than total SOFA score.

3. RESULTS

3.1 Probability of Mortality vs. SOFA Score

The probability of mortality for different SOFA score ranges and for different days in the ICU are presented in Fig. 1. Fig. 1a presents the analysis for each of the first seven days of ICU stay and Fig. 1b presents the same analysis but for three-day groups. It can be seen in Fig. 1 that the higher SOFA score bands are significantly more likely to lead to mortality in that given day. Beyond seven days there is a slight decline in the probability of death for the higher SOFA score groups. While the results in Fig. 1 show a strong correlation between SOFA score and mortality, the SOFA bands (0-5, 5-10 and 10-20) are broad and more detailed analysis is necessary.

![Fig. 1. Probability of death vs. SOFA score for 3-day groups of ICU stay.](image)

3.2 SOFA Score as a Prognostic Indicator

To assess the SOFA score as a prognostic indicator over the course of the patient stay, the SOFA score distribution is plotted in the seven days leading up to death. This analysis is restricted to the subset of patients that died during their stay.
in the ICU and is presented in Fig. 2. It is evident in Fig. 2 that there is no clear trend in the SOFA score leading up to patient mortality. It might be expected that there would be an overall trend towards increasing SOFA score, but this behavior is not evident in the results. To provide an indication of the statistical significance of any difference in the distributions, p-values are calculated using the Wilcoxon rank sum and Kolmogorov-Smirnov tests and presented in Tables 1 and 2. It is evident that there is little statistically significant difference between distributions of SOFA scores in the seven days prior to death.

Fig. 2. SOFA score distributions for patients in the seven days leading up to death.

Table 1. P-values of distributions in Fig. 2 calculated using the Wilcoxon rank sum test for equal medians.

<table>
<thead>
<tr>
<th>Day</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6577</td>
<td>0.2880</td>
<td>0.1782</td>
<td>0.1383</td>
<td>0.6372</td>
<td>0.3517</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.5477</td>
<td>0.3506</td>
<td>0.2467</td>
<td>0.9423</td>
<td>0.5675</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.6753</td>
<td>0.5108</td>
<td>0.6898</td>
<td>0.9064</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.8180</td>
<td>0.4476</td>
<td>0.7858</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.3319</td>
<td>0.6544</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.6549</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. P-values of distributions in Fig. 2 calculated using the Kolmogorov-Smirnov test of distribution spread.

<table>
<thead>
<tr>
<th>Day</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7925</td>
<td>0.1408</td>
<td>0.1304</td>
<td>0.2576</td>
<td>0.3557</td>
<td>0.1204</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.8010</td>
<td>0.7219</td>
<td>0.9105</td>
<td>0.9824</td>
<td>0.6453</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.9945</td>
<td>0.9320</td>
<td>1.0000</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.9985</td>
<td>0.8931</td>
<td>0.9541</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.7674</td>
<td>0.9280</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.9998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Given that SOFA score is a poor indicator of impending mortality, SOFA scores are also assessed as an indicator of survival in the seven days leading to patient discharge. This analysis is presented in Figure 3. It is evident that there are much more pronounced trends in the SOFA score distributions in the days leading up to patient discharge. Specifically, SOFA > 7 decreases significantly, as expected, before discharge, and thus SOFA is a more effective indicator of positive prognosis than of negative outcomes. Tables 3-4 present the p-values comparing the distributions in Fig. 3, which reflect this outcome.

Fig. 3. SOFA score distributions for patients in the seven days leading up to discharge from the ICU.

Table 3. P-values of the distributions in Fig. 3 calculated using the Wilcoxon rank sum test for equal medians.

<table>
<thead>
<tr>
<th>Day</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.7126</td>
<td>0.6432</td>
<td>0.6376</td>
<td>0.0048</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.8827</td>
<td>0.8973</td>
<td>0.0203</td>
<td>0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.9752</td>
<td>0.0303</td>
<td>0.0028</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.0365</td>
<td>0.0039</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.4309</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. P-values of the distributions in Fig. 3 calculated using the two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis test.

<table>
<thead>
<tr>
<th>Day</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.4205</td>
<td>0.7208</td>
<td>0.9449</td>
<td>0.0105</td>
<td>0.0027</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.9341</td>
<td>0.9847</td>
<td>0.0359</td>
<td>0.0107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.9435</td>
<td>0.0430</td>
<td>0.0136</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.1630</td>
<td>0.0628</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.9725</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To further investigate SOFA score distribution and assess it as a prognostic metric, the odds ratio of death (death / discharge) and its inverse, the odds ratio of survival, are calculated for each of the 7 days. Results are shown in Fig. 4, which is effectively the ratios between Fig. 2 and Fig. 3.

Fig. 4 shows significant discrimination is achieved between the number of patients that are discharged to those that die at low SOFA ≤ 4 based on a nominal 4:1 threshold line shown. At SOFA ≥ 10 the inverse relation (mortality odds ratio) is also significant in its ability to discriminate. Thus, low SOFA ≤ 4 are a good indicator of patient discharge, and SOFA ≥ 10 are a good prognostic of impending mortality.

However, 5 ≤ SOFA ≤ 9 is a critical range of modest organ dysfunction in which a majority of initial and maximum SOFA scores occur in a medical ICU (Chase et al. 2010). It is also a range in which mortality is also potentially quite high, as seen in Fig. 2. It is the lack of prognostic ability in this range that makes using SOFA score difficult despite its obvious potential link to a main cause of ICU mortality.

3.3 Individual SOFA score components.

In all of the previous analyses only the overall aggregate SOFA score has been considered. This overall score out of 20 comprises components that assess renal, cardiovascular, liver, and respiratory function, and blood coagulation, each with a score of 0-4. To investigate whether any individual components may provide improved prognostic capability, the probability of death is calculated for each component individually. Fig. 5 presents these results.
It can be seen in Fig. 5 that the cardiovascular and respiratory scores show the strongest correlation with patient mortality across the first seven days of ICU stay. The respiratory SOFA score also shows a strong correlation between patient mortality and increasing score, but does not reach the high probability of mortality seen for the cardiovascular and coagulation groups, matching clinical observation. There are very high mortality percentages (=100%) for the liver-bilirubin group on days 2 and 3. However, this observation is due to very low patient numbers with scores of 3 or 4 in this category. The renal creatinine score in Fig. 5e shows some unexpected trends with SOFA score. However, this category should be used with caution, as the data shows that the distribution of SOFA scores within this group tends to be very bi-modal. There are very few patients with renal SOFA scores of 1 or 2, with most patients having a score of 0, 3 or 4. The large variation in probabilities seen in Fig. 5e can be attributed to these low patient numbers.

3. DISCUSSION

The analysis of the SOFA scores of a cohort of 784 medical ICU patients has reaffirmed that patient mortality is correlated with overall SOFA score. However, there is no clear, statistically significant trend observed in the days leading to death. It might be expected that the distribution moves towards a higher SOFA score in the days leading up to mortality, but this trend was not confirmed. This result indicates that the SOFA score alone is not a good prognostic indicator of negative patient outcome.

Conversely, if the distribution of SOFA scores is examined in the days leading up to patient discharge, a more evident and expected downwards trend in SOFA is observed. Analysis of p-values for these distributions indicate that this shift in the distributions is statistically significant. Therefore, the SOFA score can be an effective prognostic tool for positive patient outcomes, but not for impending mortality.

When considering the SOFA score ranges where patient outcomes could be predicted, very low total SOFA ≤ 4 are likely to lead to patient discharge. Likewise, very high SOFA ≥ 10 are more likely to lead to death. However, it is the middle band that is critical, as these prior outcomes are more likely expected, and there is no clear discrimination in this range, limiting the use of SOFA as a prognostic marker.

Finally, individual SOFA cardiovascular and coagulation components show strong and reliable trends between increasing SOFA score and probability of death. The other components were less effective and suffered from low numbers despite a large cohort, making further conclusions difficult. It is this set of components that may cause the loss of clarity in the analysis of total SOFA score. Equally, cardiovascular and coagulation components are associated with cardiac failure and sepsis, which are leading causes of ICU mortality, and may be better prognostic markers.

Overall, it has been shown that the aggregate SOFA score is only effective as an indicator of positive patient outcome if the full range of SOFA scores in considered. If the SOFA score are grouped into broad bands then there is a strong correlation between increasing SOFA score and increased chance of patient death. However, these trend are much less clear when the full range of SOFA scores are considered.

4. CONCLUSIONS

This manuscript presents the evaluation of the SOFA score as a prognostic indicator. Overall, SOFA score has limited use as an indicator of negative patient outcome, but it is stronger as an indicator of positive patient outcomes. More generally, organ failure alone is well correlated with mortality but there is no specific linkage particularly due to in ability to discriminate at intermediate SOFA score values. Hence, while organ failure is strongly associated with mortality, prognostic capability for a well respected organ failure score is minimal in the most clinically useful ranges.

REFERENCES


using the Glasgow Outcome Scale." Scand J Rehabil Med 

"Insulin treatment improves the systemic inflammatory

Juneja, D., Gopal, P. B., Kapoor, D., Raya, R. and
Sathyarayanan, M. (2011). "Profiles and Outcome of
Patients With Liver Cirrhosis Requiring Mechanical

Kerby, J. D., MacLennan, P. A., Burton, J. N., McGwin, G.,
Jr. and Rue, L. W., 3rd (2007). "Agreement between
prehospital and emergency department glasgow coma
scores." J Trauma 63(5): 1026-1031.

Koch, A., Gressner, O. A., Sanson, E., Tacke, F. and
patients are associated with inflammation, organ
dysfunction and metabolism and may predict survival of

Langouche, L., Meerssman, W., Vander Perre, S., Milants,
I., Wouters, P. J., Hermans, G., Gjedsted, J., Hansen, T. K.,
Arnout, J., Wilmer, A., Schetz, M. and Van den Bergh, G.

"Utilization of intensive care unit resources in severe acute

Lin, C. Y., Tsai, F. C., Tian, Y. C., Jenq, C. C., Chen, Y. C.,
outcome scoring systems for patients on extracorporeal

Lonergan, T., Compte, A. L., Willacy, M., Chase, J. G.,
Shaw, G. M., Hann, C. E., Lotz, T., Lin, J. and Wong, X.

Lonergan, T., LeCompte, A., Willacy, M., Chase, J. G.,
Shaw, G. M., Wong, X. W., Lotz, T., Lin, J. and Hann, C.
Glycemic Control in Critical Illness: Development and

surgical intensive care unit--the role of sepsis and organ

"Evaluation of SOFA-based models for predicting

Moreno, R., Vincent, J. L., Matos, R., Mendonca, A.,
Cantraine, F., Thijs, L., Takala, J., Sprung, C., Antonelli,
M., Bruining, H. and Willatts, S. (1999). "The use of

"The use of artificial neural networks to stratify the length
of stay of cardiac patients based on preoperative and initial postoperative factors." Artif Intell Med 40(3): 211-221.

Sakr, Y., Vincent, J-L., Ruokonen, E., Pizzamiglio, M.,
and organ system failure are major determinants of post-


Timsit, J. F., Fosse, J. P., Troche, G., De Lassence, A.,
Alberti, C., Garrouse-Orgeas, M., Azoulay, E., Chevret, S.,

Tran, D. D., Groeneveld, A. B., van der Meulen, J., Nauta, J.

Udekwu, P., Kromhout-Schiro, S., Vaslef, S., Baker, C. and
Oller, D. (2004). "Glasgow Coma Scale score, mortality,


Vincent, J. L., de Mendonca, A., Cantraine, F., Moreno, R.,
Takala, J., Suter, P. M., Sprung, C. L., Colardyn, F. and
Blecher, S. (1998). "Use of the SOFA score to assess the
incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working

Vincent, J. L., Moreno, R., Takala, J., Willatts, S., De
Mendonca, A., Bruining, H., Reinhart, C. K., Suter, P. M. and
Thijs, L. G. (1996). "The SOFA (Sepsis-related Organ
Failure Assessment) score to describe organ

Vincent, J. L., de Mendonca, A., Cantraine, F., Moreno, R.,
Takala, J., Suter, P. M., Sprung, C. L., Colardyn, F. and
Blecher, S. (1998). "Use of the SOFA score to assess the
incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working

Wong, X. W., Singh-Levitt, I., Hollingsworth, L. J., Shaw,
G. M., Hann, C. E., Lotz, T., Lin, J., Wong, O. S. and

Zygyn, D., Berthiaume, L., Laupland, K., Kortbeek, J. and
Doig, C. (2006). "SOFA is superior to MOD score for the