Intensive Care Med (1988) 14:167–172

Intensive Care Medicine

Methods and devices

Sickness scoring and response to treatment as predictors of outcome from critical illness

J.F. Bion[1,*], T.C. Aitchison[2], S.A. Edlin[1,**] and I.McA. Ledingham[1]

[1] Clinical Shock Study Group, Western Infirmary and [2] Department of Statistics, Glasgow University, Glasgow, Scotland

Received: 12 January 1987; accepted: 5 June 1987

Abstract. A physiological sickness scoring system (SS), based on the APACHE II score, has been used to assess outcome from critical illness in 128 patients admitted to a general intensive care unit. Physiological data were collected on each patient from admission until death or discharge from the unit, and survival was recorded as survival to home. The admission SS correctly classified 80.6% of survivors, and 70.4% of non-survivors. Predictive power did not improve with time using the daily SS. However, when the proportional change in SS over time was included in the analysis, predictive power improved; and at day 4, 87.1% of survivors and 75% of non-survivors were correctly classified. At intermediate levels of sickness severity (admission score of 13–18), a reduction in SS of 30% by day 4 reduced the risk of death by 32%; at higher levels (>18) a similar reduction in SS was associated with a 47% reduction in death-risk. Failure to obtain a reduction in score by day 4 was associated with increased risk of death. Survivors consistently showed a greater fall in SS by day 4 than non-survivors. The APACHE score and its modifications provide an accurate, unitary measure of physiological disturbance. Correction of abnormal physiology, and the measurement of responsiveness to therapy are important components in the prediction of outcome from critical illness.

Key words: Intensive Care – Severity of illness – Outcome – Logistic regression

Present addresses:
* Department of Anaesthetics and Intensive Care, University of Birmingham, Birmingham, UK
** Department of Anaesthetics, Charing Cross Hospital, London, UK

Patients are admitted to intensive care units (ICU) because they have developed a degree of physiological disturbance which makes it difficult to care for them on an ordinary ward. It is the purpose of ICUs to contain and correct this disturbance. A unitary standard by which deranged physiology may be measured, is likely to be of value in the assessment of severity of illness, prognosis, and efficacy of treatment. Such a standard has recently been developed and validated in the USA for prognosis, the APACHE II score [1]; and the value of this method has been confirmed using modifications of the scoring system in France [2], and for transporting critically ill patients in the UK [3]. Previous studies have concentrated on the use of admission values alone, but the UK study was the first to examine the longitudinal use of a sickness scoring system, and we now report an extension of this approach to 128 patients admitted to an ICU.

Methods

One hundred and twenty-eight critically ill patients admitted to a general intensive care unit were recruited to the trial. Patients requiring over-night post-operative care without ventilator support were not included. Data for subsequent sickness scoring were collected when the patient was admitted to the unit, and on each day thereafter at 0800 hours. Two doctors were responsible for data collection. Outcome is defined as survival to home, or non-survival if the patient died in hospital or was discharged to long-term institutional care.

The sickness scoring system (SS) [Appendices 1 and 2] is a modification of the APACHE II score, and the reasons for these modifications have already been reported [3]; these modifications include: (1) conversion to SI units, (2) haemoglobin is used in preference to haematocrit, (3) oxygenation is assessed using a

ratio between the % inspired oxygen concentration and the arterial oxygen tension in kPa. In addition, we now include in the 'chronic disease' category any condition (including psychiatric) which is sufficiently severe to prevent independent self-care, or which confines the patient to home or an institution. The patient's Glasgow coma score was assumed to be normal in the presence of sedative drugs, unless clinical examination in the absence of sedation suggested neurological disease; in this case, the best score obtainable within a 12-h period was the one used for inclusion in the SS. Data for daily scoring were collected at 0800 hours, with the exception of cardiovascular variables (heart rate and mean arterial pressure) which were recorded and scored two-hourly, the scores summed, and then rescored as follows:

Sum of two-hourly scores	>21	16–20	11–15	5–10	0–4
Rescored value	4	3	2	1	0

The rescored values were obtained by constructing blood pressure and pulse charts for hypothetical patients in such a way that they could be broadly categorised into five bands, thereby following the method applied for the construction of the APACHE II score. The bands are drawn in such a way that occasional falls in blood pressure occasioned for example by the injudicious administration of sedative agents, will not score unless the effect occurs repeatedly.

These various modifications have been made after reviewing data from our pilot studies. Incapacity to provide for one's own care for whatever reason, appears to have an adverse effect on outcome from subsequent critical illness. The maximum weighting of 12 points for central nervous system disease makes the exclusion of drug effects important; this did not present any practical difficulties during the trial. The time at which data are collected for scoring, however, does present difficulties which have not yet been resolved: admission data are influenced by the level of physiological support offered on the wards before admission to the ICU; using a specific time each day risks collecting unrepresentative data; and collecting the worst values in 24 h is unphysiological in the sense that it presents a false image of illness severity. Comparisons are in progress, but our use of rescored cardiovascular variables is a compromise between using data at any one point during the day, and using the worst values obtainable.

Data have been analysed for the first 4 ICU days, on the presumption that, if action is to be taken on the basis of a prognostic scoring system, then it should be taken within this period for the score to be of practical value.

Variables examined in the analysis were:

1. The 'raw' SS on admission and for each day.
2. The categorised SS (CSS), in bands of 0–6, 7–12, 13–18, >18. These bands were chosen because they span the range of scores fairly evenly; our previous study [3] had shown that scores in excess of 18 were associated with subsequent death.
3. The acute physiology scores (APS). This excludes weighting for age and chronic health.
4. The patients' ages.
5. The proportional change in SS (PCh) from the admission value:

$$\text{i.e.: PCh Day 4} = \frac{\text{SS 4} - \text{SS Adm}}{\text{SS Adm}}.$$

To assess which, if any, of these variables was useful in predicting outcome, the technique employed was stepwise logistic regression using the BMDP statistical programme [4].

Predictive accuracy of the SS was compared with that of various members of the ICU staff, who gave a percentage chance of survival when the patient was admitted. Replies were categorised into 'less than 50%' to indicate probable death, and 'more than 50%' to indicate probable survival.

Results

The sample size declined from 128 patients at entry, to 121 on day 1, to 71 on day 4 [Table 1]. Fifty-three pa-

Table 1. Sample size from admission to day 4

Time	Survivors *n* (%)	Non-survivors *n* (%)	Total
Admission	75 (58.5)	53 (41.4)	128
Day 1	71 (58.8)	50 (41.4)	121
Day 2	63 (57.7)	46 (42.2)	109
Day 3	48 (55.1)	39 (44.8)	87
Day 4	38 (53.5)	33 (46.4)	71

Table 2. Main diagnostic categories of 128 patients

Diagnoses	75 Survivors No. (%)	53 Non-survivors No. (%)
Sepsis	13 (17.3)	25 (47.1)
ARDS	4 (5.3)	13 (24.5)
Trauma	9 (12.0)	4 (7.5)
Acute renal failure	6 (8.0)	20 (37.7)
Chronic renal failure	2 (2.6)	2 (3.7)
Pump failure	1 (1.3)	4 (7.5)
Aortic aneurysm	5 (6.6)	3 (5.6)
Pneumonia	5 (6.6)	6 (11.3)
COAD/Asthma	8 (10.6)	2 (3.7)
Pancreatitis	1 (1.3)	3 (5.6)
Self-poisoning	4 (5.3)	1 (1.8)
Spinal cord lesion	1 (1.3)	1 (1.8)

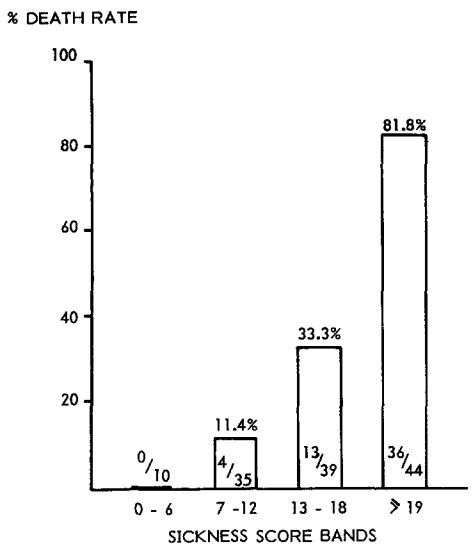

Fig. 1. Categorised admission sickness scores and actual death rate

tients died (41.4%) and 75 (58.6%) survived to home. Two patients (classified as non-survivors) were discharged to long-term institutional care. Table 2 lists the main diseases present during the patients' stay on the ICU, for survivors and non-survivors; 50 of the survivors (66.6%) and 34 of the non-survivors (64.1%) were post-operative admissions.

Analysis of predictors of outcome

Stepwise logistic regression was performed using the complete data for each day from admission until day 4. The variables included in the analysis for possible selection are as described in the methods section. The analysis concluded that the daily SS was the only worthwhile predictor of outcome and that none of the other variables had any additional predictive value. This is presented graphically in Figure 1, with the admission SS categorised into bands of increasingly severe physiological disturbance.

The analysis was then repeated using categorised SSs. The purpose here was to see if this simpler version of the SS would allow other useful variables into the predictor. This analysis showed that two variables provided the only worthwhile prediction of outcome, the categorised SS with (after day 1) the proportional change in SS from admission. A jack-knifed classification matrix was then employed to assess the predictive performance of first the raw SS, and second the combination of categorised SSs with the proportional change each day. The basic idea of 'jack-knifing' is to correct to a certain extent the apparent 'over-successful' estimation of prediction assured by re-employing the technique on the data which was used in the first place to estimate the model for prediction.

Table 3. Prognostic power of individual variables

Day	Variable 'chosen'	% Correct classification		
		Survivor	Non-survivor	Total
(Raw data)				
1.	SS Adm	80.6	70.4	76
2.	SS 2	75	78.4	76.1
3.	SS 3	74.5	75	74.7
4.	SS 4	73	70.6	71.8
(Categorised data)				
1.	CSS Adm + APS 1	81.8	70.9	76.9
2.	CSS Adm + PCh 2	80.7	69.2	75.2
3.	CSS Adm + PCh 3	84.6	70.8	77
4.	CSS Adm + PCh 4	87.1	75	80.3

The matrix expresses as a percentage the proportion of correct classifications based on predicted and actual outcome. The results of using the 'raw' and the categorised data are shown in Table 3.

Using the raw data, the percentage of patients correctly classified fell with time. When the categorised data was employed with the proportional change in SS from admission, this gave improved prognostic power as time passed. This suggests that prognostic power is increased by a scoring system which provides an index of response to treatment, as is given by the proportional change in SS from admission.

Effect of response to treatment on outcome

That responsiveness to treatment influences outcome, may be demonstrated by plotting estimated probability of death against the proportional change in SS by day 4 [Fig. 2]. Logistic regression analysis derives an

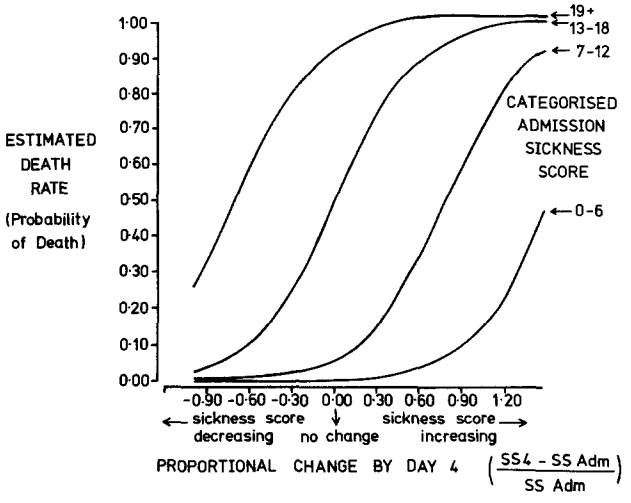

Fig. 2. Estimated risk of death using categorised admission sickness score and proportional change by day 4 (logistic regression model from 128 patients)

Table 4. % Reduction in sickness score by day 4 for survivors and non-survivors

Admission SS band	Survivors, day 4			Non-survivors, day 4			
	n	Mean fall	±SD	*n*	Mean fall	±SD	*p*
5–12	15	−11.4%	29.1	0	–	–	–
13–18	14	−22.3%	24.4	10	3.8%	23.5	<0.01
>18	7	−45.5%	13.6	22	−18.3%	24.8	<0.01

equation for those patients alive on day 4, from which estimated probabilities of death may be obtained [Appendix 3]. This allows construction of a model for the group, stratified by admission SS bands. The model, presented graphically in Figure 2, shows that the greater the initial severity of illness, the greater the reduction in score required to reduce the risk of death.

The extent of which outcome is affected by change in SS is shown in Table 4. The lowest admission SS band has been excluded because proportional change at these low levels of sickness is difficult to interpret. All the patients in the 5–12 SS band on day 4 survived to home. In the middle range, survivors show a mean fall in SS of 22.3%, while non-survivors show a small rise; at higher levels of physiological disturbance, both groups show a fall in SS, but this is much more marked in the survivors, with a mean reduction in score of 45.5%.

Figure 2 only provides a point-estimate of the probability of death; this is open to error, and the clinician needs some measure by which the accuracy of prediction can be gauged for a particular patient. An example of this measure, the interval estimate, is given for Figure 2, in Appendix 2.

The predictive power of the SS on admission was compared with that of various categories of ICU staff, who were asked when the patient was admitted, to give a percentage chance of survival from 1% to 99%. The 'raw' SS was the best predictor of outcome, followed closely by the staff nurses [Table 5].

Discussion

The decision to withdraw or continue treatment of critically-ill patients is often made on the basis of clinical judgement and previous experience. This partly intuitive process is likely to be enhanced by scoring systems which present the clinician with an accurate measure of physiological disturbance. The accuracy of such systems is generally assessed by their capacity to predict outcome from critical illness. This encourages medical staff to regard physiological scoring systems simply as predictors or crystal balls. It is important to realise that they are in fact descriptors and comparators rather than predictors. The predictive power of these systems is an index of their accuracy for describing physiological disturbance, and their use for prognosis depends on recognition of this fact. The SS should not be regarded as an absolute value, dispensing with the need for clinical judgement; it is a single value with an interval, or error, attached to it, as described in Appendix 2. The single value is an accurate descriptor of disturbed physiology, but the interval which surrounds it requires interpretative clinical skill.

Table 5. Prognostic accuracy of members of ICU staff

Source	% Total correctly classified
SS Adm	74.4
Staff nurse	73.9
Sister	70.3
Resident SHO	65.8
Consultant	65.8

Correction of deranged physiology is one of the main aims of intensive care. The capacity to respond to treatment is an index both of the patient's homeostatic reserve and of the specificity of therapy. It is reasonable to suggest that the more rapid and complete the restoration of normal physiology, the more likely it is that the patient will survive. Conversely, failure to respond to treatment is likely to be associated with a poor outcome. Knaus et al. [1] have shown that without weighting for diagnosis, APACHE II scores correlate poorly with outcome in diabetic ketoacidotic coma, and suggest that this is because there is a specific cure available, insulin. In such patients, measuring response to treatment rather than initial severity of illness, will give a better index of likely outcome, and may reduce the need to weight diagnostic groups when calculating risk of death.

To withold treatment on the basis of a single admission SS value is unwise for two reasons; first, the system is not infallible, second, to do so would bring a charge of therapeutic nihilism, that no matter how skilled the treatment, outcome was already decided. A more appropriate course of action would be to observe the response of the patient to treatment over a fixed period, an approach which accords with current practice. The purpose of our study was to identify whether response to treatment influences outcome, and to what extent it does so, thereby providing medical staff with an index of efficacy of treatment.

The trial confirms that measuring severity of illness with a physiological scoring system is an important method for assessing outcome from critical illness. Accuracy is improved by measuring responsiveness to treatment, as assessed by the proportional change in SS over time. Reduction in SS is associated with a reduction in risk of death, and failure to obtain

Appendix 1. Sickness scores: values for individual variables (adapted from Knaus et al. [7])

	Score								
	4	3	2	1	0	1	2	3	4
Core temperature (°C)	⩾41	39–40.9	–	38.5–38.9	36–38.4	34–35.9	32–33.9	30–31.9	⩽29.9
Mean blood pressure (mmHg)*	⩾160	130–159	110–129	–	70–109	–	50–69	–	⩽49
Heart rate/min	⩾180	140–179	110–139	–	70–109	–	55–69	40–54	⩽39
Respiratory rate/min	⩾50	35–49	–	25–34	12–24	10–11	6–9	–	⩽5
% Inspired O_2/PaO_2 (kPa)	⩾5.0	4.0–4.99	2.1–3.99	–	<2.09	–	–	–	–
Arterial pH	⩾7.7	7.6–7.69	–	7.5–7.59	7.33–7.49	–	7.25–7.32	7.15–7.24	⩽7.15
Creatinine (μmol/l)†	⩾600	300–599	180–299	130–179	50–129	–	⩽49	–	–
Sodium (mmol/l)	⩾180	160–179	155–159	150–154	130–154	–	120–129	111–119	⩽110
Potassium (mmol/l)	⩾7.0	6.0–6.9	–	5.5–5.9	3.5–5.4	3.0–3.4	2.5–2.9	–	<2.5
Haemoglobin (g/dl)	⩾18.0	–	15.0–17.9	14.0–14.9	9.0–13.9	–	6.1–8.9	–	⩽6.0
White cell count ($\times 10^9$/l)	⩾40.0	–	20.0–39.9	15.0–19.9	3.0–14.9	–	1.0–2.9	–	<1.0
Glasgow coma scale	Score as 15 – actual score								

* Mean arterial pressure (mmHg) = (2 diastolic + systolic)/3
† If acute renal failure has occurred, double score
Conversion: SI to traditional units – PaO_2: 1 kPa ≈ 7.5 mmHg. Creatinine: 1 μml/l ≈ 0.01 mg/100 ml. Sodium: 1 mmol/l = 1 mEq/l. Potassium: 1 mmol/l = 1 mEq/l

an improvement in physiology results in an increase in risk. Survival from intermediate levels of sickness severity is associated with a reduction in physiological disturbance by day 4 of around 20%; and from high levels with a reduction of around 45%.

The importance of severity scoring has been emphasized by recent leading articles [5, 6]. Our study demonstrates the value of measuring response to treatment as well as absolute severity of illness, and provides clinicians with an useful method for assessing the critically ill and the effectiveness of the treatment they receive.

Appendix 2. Scores for age and chronic disease (from Knaus et al. [7])

Age (years)	⩽44	45–54	55–64	65–74	⩾75
Score	0	2	3	5	6

Chronic disease score. If chronic disease history is positive: elective postoperative patients score 2; emergency postoperative or medical patients score 5

Chronic disease category. (1) Disease (*a*) must have been evident before this hospital admission; (*b*) must be of sufficient severity to prevent independent self care. This category includes chronic dialysis, documented cirrhosis, or portal hypertension and disease of other systems of severity which will generally confine patient to the house. (2) Immunosuppression: patients receiving chemotherapy, radiation long-term low dose steroids, or short-term high dose steroids; or malignant or other disease which is sufficiently advanced to impair resistance to infection.

Appendix 3. Prediction model for Figure 2

Our final estimated model for prediction of outcome on day 4 is that the log-odds on death equals:

$$-8.167+(2.644\times\text{CSS Adm})+(3.568\times\text{PCh4})$$

where CSS Adm = the categorised admission sickness score, ordered as (0–6) = 1, (7–12) = 2, (13–18) = 3, and (>18) = 4 and PCh4 = proportional change in SS from admission to day 4.

If the log-odds on death take the value 'c' for a particular patient, then the estimated probability of death is:

$$\exp(c)/[1+\exp(c)]$$

For example, if a patient has an admission SS of 16 falling to 12 by day 4, then CSS Adm = 3, and PCh4 = 12–16/16 = −0.25; log-odds on death is then:

$$-8.167+(2.644\times 3)+(3.568\times -0.25) = -1.13$$

The estimated probability of death is therefore:

$$\exp(-1.13)/[1+\exp(-1.13)] = 0.24, = 24\% \text{ risk of death.}$$

If we take into account the standard errors of our estimates of the coefficients in the logistic regression, then for this hypothetical patient the 95% confidence intervals for the true log-odds ratio of this patient are: [−1.13, 0.88], i.e.: [−2.01 to −0.25], and the corresponding interval estimate for the true probability of death for this

patient is [0.12, 0.44], that is a range of 12% to 44%. As this interval contains no values greater than 50% for the probability of death, survival is the more likely outcome. If, for any patient, the interval estimate were to contain 50% for the probability of death, the point estimate would be unreliable.

Acknowledgements. We are grateful to Dr. Knaus and his colleagues at the George Washington University Medical Centre, who have been most generous with their time and advice, and in exchanging unpublished data.

References

1. Wagner DP, Knaus WA, Draper EA (1986) Physiologic abnormalities and outcome from acute disease. Arch Intern Med 146:1389
2. Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, Mercier P, Thomas R, Villers D (1984) A simplified acute physiology score for ICU patients. Crit Care Med 12:975
3. Bion JF, Edlin SA, Ramsay G, McCabe S, Ledingham IMcA (1985) Validation of a prognostic score in critically ill patients undergoing transport. Br Med J 291:432
4. Dixon WJ (ed) (1983) BMDP-83, Biomedical Data Computer Programs, P-series. University of California Press, Los Angeles
5. Morgan CJ, Branthwaite MA (1986) Severity scoring in intensive care. Br Med J 292:1546
6. Anonymous (1986) TPN and APACHE. Lancet I:1478
7. Knaus WA, Draper EA, Wagner DP, Zimmerman JE (1985) Apache II: a severity of disease classification system. Crit Care Med 13:818

Dr. J. F. Bion
Department of Anaesthetics and Intensive Care
University of Birmingham
Birmingham B15 2TJ
UK