

Primary research

Prospectively validated predictions of shock and organ failure in individual septic surgical patients: the Systemic Mediator Associated Response Test

Gus J Slotman

University of Medicine and Dentistry of New Jersey, Camden, New Jersey, USA

Received: 23 December 1999 2000

Revisions requested: 16 May 2000

Revisions received: 17 July 2000

Accepted: 31 July 2000

Published: 8 September 2000

Crit Care 2000, 4:319–326

© Current Science Ltd (Print ISSN 1364-8535; Online ISSN 1466-609X)

Abstract

Introduction: Clinically useful predictions of end-organ function and failure in severe sepsis may be possible through analyzing the interactions among demographics, physiologic parameters, standard laboratory tests, and circulating markers of inflammation. The present study evaluated the ability of such a methodology, the Systemic Mediator Associated Response Test (SMART), to predict the clinical course of septic surgery patients from a database of medical and surgical patients with severe sepsis and/or septic shock.

Patients and methods: Three hundred and three patients entered into the placebo arm of a multi-institutional sepsis study were randomly assigned to a model-building cohort ($n = 200$; 119 surgical) or to a predictive cohort ($n = 103$; 55 surgical). Using baseline and baseline plus serial measurements of physiologic data, standard laboratory tests, and plasma levels of IL-6, IL-8, and granulocyte colony-stimulating factor (GCSF), multivariate models were developed that predicted the presence or absence of pulmonary edema on chest radiography, and respiratory, renal, coagulation, hepatobiliary, or central nervous system dysfunction and shock in individual patients. Twenty-eight-day survival was predicted also in baseline plus serial data models. These models were validated prospectively by inserting baseline raw data from the 55 surgical patients in the predictive cohort into the models built on the comprehensive training cohort, and calculating the area under the curve (AUC) of predicted versus observed receiver operator characteristic (ROC) plots.

Results: SMART predictions of physiologic, respiratory, metabolic, hepatic, renal, and hematologic function indicators were validated prospectively, frequently at clinically useful levels of accuracy. ROC AUC values above 0.700 were achieved in 30 out of 49 (61%) of SMART baseline models in predicting shock and organ failure up to 7 days in advance, and in 30 out of 54 (56%) of baseline plus serial data models.

Conclusion: SMART multivariate models accurately predict pathophysiology, shock, and organ failure in individual septic surgical patients. These prognostications may facilitate early treatment of end-organ dysfunction in surgical sepsis.

Keywords: cytokines, multiple organ failure, prognostication, severe sepsis, shock

Introduction

Conventional outcomes research in surgical sepsis has focused on scoring systems that predict grouped risk of mortality, utilization of resources, and, sometimes, the development of broadly defined multiple organ failure conditions. For example, the Sepsis Score [1], the Multiple Organ System Dysfunction Score [2], and the Sepsis-related Organ Failure Assessment [3] were attempts to predict group percentage of risk of death among septic patients. In the general adult intensive care unit population, the Mortality Probability Model II score [4], the Simplified Acute Physiology Score II [5], and the Acute Physiology and Chronic Health Evaluation (APACHE) III score [6] are the best known predictive engines. For trauma patients, the Injury Severity Score, the Trauma and Injury Severity Score, and A Severity Characterization of Trauma, among others, also predict mortality risk [7–11]. Others have related the duration of systemic inflammatory conditions to organ dysfunction, duration of hospital stay, and mortality [12].

In septic surgical patients, increased circulating cytokines, prostaglandins, complement, and other inflammatory response mediators have been associated with poor outcome and the development of acute end-organ dysfunction [13–15]. However, conventional scoring systems have nevertheless grouped patients with disparate pathophysiologies together on the basis of similar probabilities of dying. Ultimately, then, these methods forecast only grouped percentage risks of hospital death, and possibly consumption of health care resources [13]. The clinically important pathophysiologic events that actually define the risk of mortality in the first place are not predicted. As a result, conventional prognostication does not facilitate timely and therapeutic intervention, and therefore does not improve survival.

Management of severely septic surgical patients could be optimized by identifying and monitoring the onset and resolution of organ dysfunction, shock, and other systemic inflammatory conditions subclinically in individual patients. Considering the many inflammatory response mediators that have been associated with the development of shock, organ failure, and death, it seems logical that changes in circulating concentrations of such substances may be related prognostically to clinical events that lie biologically and temporally downstream from the original septic insult. We hypothesized that clinical manifestations of sepsis and its sequelae could be predicted through analyzing interactions between patient data and plasma inflammatory response mediators measured when sepsis was first diagnosed, and clinical events that occurred days later. This concept was tested retrospectively in a previous study [16], and has been developed methodologically as the SMART. Circulating eicosanoids and cytokines have been predicted also [17]. Building on the results of these pilot

projects, the objective of the present study was to further develop, and validate prospectively, SMART multivariate models that predict clinically important dysfunction of vital organs, in advance, in individual surgical patients with severe sepsis and septic shock.

Patients and methods

Data from 303 patients with severe sepsis and septic shock who were enrolled in the placebo arm of a phase III clinical trial [13] was tabulated. The clinical characteristics of these patients, including demographics, organ dysfunction, and types of infection, among other data, were described completely in the parent paper [18]. These patients then were assigned by a randomization program to a model-building training cohort ($n = 200$; 119 surgical) or a prospective validation, predictive cohort ($n = 103$; 55 surgical). Demographics, including sex, race, age, and comorbidities, were recorded at baseline for each patient. At baseline and on days 1 through 7, 14, 21, and 28, the physiologic parameters and hospital laboratory tests listed in Table 1 were recorded in all patients surviving at those observation points. In addition, at baseline and on days 1, 2, 3, and 4, plasma concentrations of IL-6, IL-8, and GCSF were measured by enzyme-linked immunosorbent assay (ELISA), using commercially available kits and standard ELISA laboratory methodology.

Using SAS software [19], data from the training cohort were analyzed by stepwise logistic regression. Multivariate models were developed that predicted the presence or absence of adult respiratory distress syndrome (ARDS), renal insufficiency, hepatobiliary dysfunction, and disseminated intravascular coagulation (DIC), all of which were defined according to established diagnostic criteria in the literature for these entities. These definitions are listed in Table 2. Also recorded were the number of lung quadrants on chest radiography that were affected by pulmonary edema (0–4), and 28-day survival.

Independent variables for each model were limited to 10 [20] and all-ways elimination was utilized in the model-building process [21]. Glasgow Coma Scale score less than 11 was chosen as the threshold for cerebral dysfunction because of the automatic absence of an appropriate verbal response for endotracheally intubated patients who otherwise might have intact cerebral function. The SMART multiple regression models derived for these dichotomous dependent variables then were validated prospectively by entering raw data from the 55 predictive cohort patients into the training cohort logistic regression formulae. Discrimination (the ability of the models to separate patients with and without the predicted dichotomous dependent variable) was assessed by calculating the AUC of ROC statistics [22]. Calibration (the degree of correspondence between predictions and observed results) was assessed using the Hosmer–Lemeshow goodness-of-fit test [23].

Table 1

Independent variables in trauma and surgical patients with severe sepsis

Age	Renal failure
Sex	Comorbidities (alcohol abuse, cirrhosis; HIV; dialysis; neutropenia; COPD; solid tumor; hematologic malignancy; chronic renal failure)
Race	Admitting service (surgery, medicine)
Albumin	A-aDO ₂
Alkaline phosphatase	Base deficit
Alanine aminotransferase	pH
Aspartate aminotransferase	PaO ₂
Blood urea nitrogen	SaO ₂
Calcium	FiO ₂
Cholesterol	Fluids in
Creatinine	Fluids out
γ-glutamyl transferase	PaO ₂ /FiO ₂ ratio
Glucose	Chloride
Hematocrit	Eosinophils
Hemoglobin	Lymphocytes
Mean corpuscular hemoglobin	Mononuclear cells
Mean corpuscular hemoglobin concentration	Metamyelocyte
Mean corpuscular volume	Segmental neutrophils
Phosphorous	Band neutrophils
Platelet count	Basophils
Potassium	Granulocytes (% granulocytes, % lymphocytes)
Total protein	Eosinophils
Prothrombin time	Lactic acid
Partial thromboplastin time	PCWP
Red blood cells	Cardiac index
Sodium	Systemic vascular resistance
Total bilirubin	PEEP
Triglycerides	Pressure support
Uric acid	Respiratory rate
White blood cells	Mechanical ventilation
IL-6	Trauma
IL-8	Systolic blood pressure
GCSF	Diastolic blood pressure
Electrocardiography (PR interval, QT interval)	Heart rate
DIC	Mean arterial pressure
Glasgow Coma Scale score	Temperature
Hepatobiliary baseline	Height
Shock	Weight
ARDS	

A-aDO₂, alveolar-arterial oxygen difference; COPD, chronic obstructive pulmonary disease; FiO₂, fractional inspired oxygen; PaO₂, arterial oxygen tension; PCWP, pulmonary capillary wedge pressure; PEEP, positive end-expiratory pressure; SaO₂, arterial oxygen saturation.

Stepwise multivariate logistic regression models that predicted dichotomous dependent variables 24 h after baseline used only baseline data. For predictions beyond 24 h, SMART modeling was carried out in two ways for each variable at each time point: from baseline data only; and from serial data, where baseline independent variable measures and/or subsequent determinations up to 24 h before the time being prognosticated were incorporated into the multiple regression and/or multivariate stepwise logistic regression modeling. For both baseline and baseline plus serial data approaches, a separate, unique predictive model was generated for each dependent variable, at each observation point (days 1–7, 14, 21, and 28).

Results

Prospectively validated SMART predictions from baseline data of the dichotomous dependent variables chest radio-

graphy score, ARDS, DIC, hepatobiliary failure, renal insufficiency, shock, and cerebral dysfunction are shown in Table 3. Ninety different predictive models were attempted; data were sufficient for building 65 successfully. Training cohort model failures are represented by blank spaces in the tables. SMART baseline models were validated at clinically useful levels of accuracy, with 30 out of 49 (61%) predicted versus observed ROC AUC determinations up to 7 days after baseline exceeding 0.700. For predictions of pulmonary edema score, ARDS, hepatobiliary failure, and renal insufficiency, SMART baseline models at 14, 21, and 28 days achieved prospective ROC AUCs exceeding 0.700 in 10 out of 12 (83%). Baseline modeling was not successful in generating predictions for mechanical ventilation or survival.

Prospectively validated SMART predictions of shock, organ failure, and 28-day survival from baseline plus serial

Table 2

Clinical definitions for organ failure and shock	
Clinical entity	Definition
Shock	Systolic blood pressure ≤ 90 mmHg or mean ≤ 70 mmHg, or need for vasopressors
Renal dysfunction	Serum creatine ≥ 2.0 mg/dl, the need for dialysis, or increase in serum creatine ≥ 20 mg/dl above pre-existing normal
Hepatic dysfunction	Two or more of the following: serum bilirubin ≥ 25 mg/dl; alanine aminotransferase or aspartate aminotransferase ≥ 2 times normal; and prothrombin time ≥ 1.5 times normal
DIC	Two or more of the following: prothrombin or partial thromboplastin time ≥ 1.2 times normal; platelet count $\leq 100\,000$ mm ³ ; and fibrin split products or D-dimer > 0.5 mg/l
ARDS	Lung injury score ≥ 7 [20]
Cerebral dysfunction	Glasgow Coma Scale score < 11

input of independent variables are listed in Table 4. Eighty-one models were attempted; 74 were successfully validated. ROC AUCs above 0.700 were achieved for 30 out of 54 (56%) of SMART models for days 2–7, and 15 out of 27 (56%) of predictions for days 14, 21, and 28. Survival was predicted from baseline plus serial and dependent variables with varying accuracy. From day 5 onward, the need for mechanical ventilation was predicted with five out of six ROC AUC determinations exceeding 0.800.

In 37 out of 63 (59%) predictive points at which both SMART baseline and baseline plus serial data models were validated prospectively, the ROC AUC of the

baseline plus serial prognostication was higher than baseline alone.

Hosmer–Lemeshow goodness-of-fit test results for SMART predictions of shock and organ failure from baseline data only are listed in Table 5. Goodness-of-fit statistics were not significant for the four-point dichotomous dependent variable chest radiography score. In the remainder of the baseline models for ARDS, DIC, hepatobiliary failure, renal failure, shock, and cerebral dysfunction, outcomes predicted by the models corresponded to the actual observed results with a probability of $P < 0.05$ in 60 out of 40 of these equations (67%). Correspondence was not uniform, but rather ranged from statistically significant results in nine out of 10 models predicting cerebral dysfunction and eight of 10 models predicting hepatobiliary failure, to only three out of 10 models that predicted DIC and four models that predicted shock.

Calibration analysis for SMART models developed from baseline plus serial data are shown in Table 6. Again, predictive versus observed results for chest radiography score did not correspond significantly. Similarly, Hosmer–Lemeshow goodness-of-fit statistics were significant for only two models that predicted presence or absence of mechanical ventilation (days 14 and 21), and for two for DIC (days 5 and 21). The Hosmer–Lemeshow test was significant in six out of 10 models which predicted hepatobiliary failure, renal failure, shock, and cerebral dysfunction. None of the models predicting survival status achieved significant goodness-of-fit statistics from baseline plus serial data.

IL-6, IL-8, and/or GCSF were weighted independent variables that contributed to predictive accuracy in 52 out of 70 (74%) of baseline SMART models, and in 51 out of 81 (63%) of baseline plus serial data predictions.

Table 3

Independent variable	Day									
	1	2	3	4	5	6	7	14	21	28
Chest radiography score*	0.874	0.802	0.798	0.761	0.820	0.625	0.686	0.598	0.786	1.000
ARDS	0.776	0.829	0.523	0.687	0.698	0.856	0.755	0.867	0.966	–
DIC	0.833	0.857	0.851	0.821	0.823	0.808	0.817	0.323	–	–
Hepatobiliary failure	0.788	0.796	0.631	0.954	0.832	0.856	0.778	0.733	0.724	0.921
Renal insufficiency	0.845	0.779	0.913	0.879	0.827	0.650	0.696	0.703	0.740	0.809
Shock	0.756	0.567	0.479	0.627	–	0.760	0.564	0.500	0.438	–
GCS score < 11	0.725	0.563	0.602	0.591	0.649	0.526	0.564	0.360	0.460	0.482

Values are AUC ROC determinations. *Chest radiography score = 0–4 quadrants pulmonary edema. Organ failure and shock definitions were from Fisher *et al* [18]. –, training cohort model failure; GCS, Glasgow Coma Scale.

Table 4

SMART: prediction of shock and organ failure and survival in severe sepsis from baseline plus serial data

Independent variable	Day									
	1	2	3	4	5	6	7	14	21	28
Chest radiography score*	–	0.847	0.902	0.807	0.859	0.629	0.709	0.467	0.794	1.000
Mechanical ventilation	–	–	–	–	0.857	1.000	0.905	0.697	0.827	0.913
ARDS	–	0.782	0.586	0.583	0.808	0.884	0.759	0.861	1.000	4.000
DIC	–	0.857	0.923	0.919	0.620	0.914	0.466	–	0.450	–
Hepatobiliary failure	–	0.778	0.679	0.940	0.888	0.890	0.843	0.856	0.947	0.938
Renal insufficiency	–	0.858	0.805	0.983	0.892	0.699	0.857	0.921	0.865	0.646
Shock	–	0.471	0.193	0.817	0.452	0.532	0.474	0.500	0.352	–
Survival Status	–	0.740	0.625	0.586	0.620	0.620	0.541	0.695	0.919	0.874
GCS score <11	–	0.667	0.760	0.576	0.617	0.604	0.708	0.500	0.873	–

Values are AUC ROC determinations. *Chest radiography score = 0–4 quadrants pulmonary edema. Organ failure and shock definitions were from Fisher *et al* [18]. –, training cohort model failure; GCS, Glasgow Coma Scale.

Table 5

SMART: Hosmer–Lemeshow goodness-of-fit test results for prediction of shock and organ failure and survival in severe sepsis from baseline data only

Independent variable	Day									
	1	2	3	4	5	6	7	14	21	28
Chest radiography score*	–	–	–	–	–	–	–	–	–	–
ARDS	24.702 (0.0009)	0.9667 (NS)	100.87 (0.0001)	61.879 (0.001)	35.996 (0.0001)	11.435 (0.2470)	102.25 (0.0001)	18.64 (0.0061)	4.524 (0.807)	3.3114 (NS)
DIC	7.4046 (0.3880)	4.581 (0.8692)	4.0234 (0.8550)	4.3578 (0.8235)	13.231 (0.0667)	9.4627 (0.221)	31.293 (0.0001)	617.34 (0.0001)	9.247 (NS)	–
Hepatobiliary failure	14.349 (0.0453)	5032 (0.0001)	15.346 (0.0318)	2.8502 (0.9434)	22.643 (0.0121)	18.508 (0.0177)	16.223 (0.0232)	242.39 (0.0001)	24.656 (0.0009)	7.3931 (0.4949)
Renal insufficiency	67.172 (0.0001)	200.8 (0.0001)	22.777 (0.0037)	18.597 (0.0172)	26.06 (0.0005)	131.14 (0.0001)	37.378 (0.0001)	12.759 (NS)	25.071 (0.0007)	15.32 (0.0532)
Shock	12.115 (0.1461)	10.962 (0.2783)	10.233 (0.2491)	20.961 (0.0073)	0.0443 (NS)	2.2016 (NS)	22.485 (0.0075)	119.08 (0.0001)	15.938 (0.0433)	0.0009 (NS)
GCS score <11	405.35 (0.0001)	67.866 (0.0001)	65.427 (0.0001)	22.233 (0.0001)	83.405 (0.0001)	46.674 (0.0001)	21.497 (0.0059)	21.66 (0.010)	137.24 (0.0001)	12.087 (0.0977)

Values are Hosmer–Lemeshow goodness-of-fit test statistics (*P* values). *Chest radiography score = 0–4 quadrants pulmonary edema. Organ failure and shock definitions were from Fisher *et al* [18]. GCS Glasgow Coma Scale; NS, not significant.

The equations developed to predict each dependent variable are available from the author for research and independent confirmation.

Discussion

The results of the present study demonstrate that, using data from a diverse population of septic patients, SMART predicts shock, organ dysfunction, mechanical ventilation, and 28-day survival in advance, in individual surgical

patients with severe sepsis and septic shock. Multiple logistic regression models, both from baseline independent variables and from baseline plus serial data, that predicted shock, chest radiography score, DIC, ARDS, cerebral dysfunction, and liver and renal failure in surgical sepsis were validated prospectively. In addition, baseline plus serial models predicted 28-day survival and the need for mechanical ventilation. The importance of measuring inflammatory response mediators to this kind of prognostication was

Table 6

SMART: Hosmer–Lemeshow goodness-of-fit test results for prediction of shock and organ failure and survival in severe sepsis from baseline plus serial data

Independent variable	Day									
	1	2	3	4	5	6	7	14	21	28
Chest radiography score*	–	–	–	–	–	–	–	–	–	–
Mechanical ventilation	–	–	–	3.5003 (NS)	11.159 (0.0836)	3.5366 (0.1706)	4.7716 (0.6878)	16.473 (0.0001)	14.65 (0.048)	3.8411 (0.8712)
ARDS	–	50.182 (0.0001)	5.1079 (NS)	229.18 (0.0001)	6.7868 (0.079)	7.4788 (0.2788)	162.02 (0.0001)	20.147 (0.0026)	13.294 (0.0099)	5.5629 (0.5916)
DIC	–	6.061 (0.7338)	2.513 (0.4261)	2.1413 (0.9764)	179.27 (0.0001)	4.8029 (0.6840)	2.4973 (NS)	8.0683 (NS)	245.38 (0.0001)	–
Hepatobiliary failure	–	4969 (0.0001)	42.872 (0.0001)	11.154 (0.1321)	19.672 (0.201)	25.623 (0.0003)	40.83 (0.0001)	23.322 (0.0015)	3.1979 (0.959)	9.207 (0.3251)
Renal insufficiency	–	0.8232 (NS)	3297.6 (0.0001)	1.4857 (0.9604)	20.514 (0.0086)	2172 (0.0001)	14.25 (0.0759)	20.981 (0.0072)	12.247 (0.0927)	8.1239 (0.0044)
Shock	–	14.611 (0.0413)	89.307 (0.0001)	11.089 (0.1348)	20.737 (0.0020)	74.51 (0.0001)	38.134 (0.0001)	62.76 (0.0001)	14.042 (0.0001)	–
Survival status	–	10.315 (0.2436)	15.226 (0.0849)	8.8376 (0.2645)	13.146 (0.0686)	13.146 (0.0686)	11.177 (0.2688)	13.851 (0.5835)	5.54 (0.5944)	7.3106 (0.5835)
GCS score <11	–	1064.3 (0.0001)	105.34 (0.0001)	729.5 (0.0001)	5418 (0.0001)	33.064 (0.0001)	16.397 (0.0217)	16.07 (0.0654)	4.2473 (0.9355)	0.8312 (NS)

Values are Hosmer–Lemeshow goodness-of-fit test statistics (*P* values). *Chest radiography score = 0–4 quadrants pulmonary edema. Organ failure and shock definitions were from Fisher *et al* [18]. GCS Glasgow Coma Scale; NS, not significant.

demonstrated by the inclusion of IL-6 and/or IL-8 and/or GCSF as weighed predictors in the development of 74% of baseline SMART models and in building 63% of baseline plus serial data formulae. The possible advantage of serial SMART models was suggested by higher serial ROC AUC determinations in 59% of baseline versus baseline plus serial comparisons.

A review of the literature indicates that such a prospectively validated method for predicting end-organ dysfunction in individual septic surgical patients, based on a mixed specialty septic database, has not previously been reported, and is a significant finding of the present study.

The SMART models in this study are applicable to individual septic surgical patients, and they frequently performed at clinically useful levels of accuracy. Other injury or illness severity scoring systems [1–11] have generally predicted only grouped risk of intensive care unit or hospital mortality. Some authors have correlated measurements of circulating inflammatory mediators with broad groupings of multiple organ failure [24]. Even when serial physiologic assessments [25] or Bayesian analysis [26] were included, prognostication was still limited to relative risk of mortality, and therefore was not applicable to individual patients. Similarly, attempts at predicting multiple organ dysfunction [24], ventilator dependence [27], or duration

of stay and hospital costs [28] have limited clinical value. In contrast, the prospectively validated SMART models presented here predicted important clinical end-points accurately in individual patients, with high levels of prognostic accuracy in most cases.

The SMART prognostic modeling approach differs from conventional surgical scoring systems in several ways. Although traditional outcomes research methods use clinically obvious information to predict mortality probability, SMART analyzes relationships between pathophysiology and standard laboratory tests, as well as circulating inflammatory response mediators and downstream pathophysiology, to predict clinical events that determine the course of each patient. Mortality risk assessments [1–7] group patients together who have similar mortality risks, but who also may vary widely in their mechanisms of disease; SMART predicts major clinical changes that may affect survival, on the basis of results from patients with similar illnesses. Conventional models assess relative risk of death, while assuming, statistically at least, that contributing factors remain constant and that outcome is not altered by treatment. SMART predicts conditions that may not yet be evident externally, and that therefore might improve outcomes by facilitating early intervention and timely modification of the host inflammatory response. It may be possible for SMART to facilitate improved outcomes in

the intensive care unit by directing clinicians toward effective early intervention in impending life-threatening conditions, and by identifying objectively those septic surgical patients who are about to manifest ARDS, DIC, liver or renal failure, or shock.

In this study, circulating levels of inflammatory response mediators were prominent as significant independent variables that contributed to the predictive power of SMART models. Plasma IL-6 and/or IL-8 and/or GCSF were involved in most of the models here. One might speculate from these results that SMART models for septic surgical patients could be optimized further by measuring additional inflammatory response mediators that have also been associated with sepsis, shock, and organ failure. Therefore, if such mediators (eg IL-1 β [29], IL-2 [30], tumor necrosis factor- α [29], intercellular adhesion molecule and other adhesion molecules [31], leukotrienes and prostaglandins [29], and activated complement [29]) were all included simultaneously as independent variables in SMART models, along with IL-6, IL-8, and GCSF as measured in the present study, then it might be that clinically useful predictive accuracy could be achieved consistently for many more dependent variables than was possible with the present database. Although immunoassays for these potential independent variables at present require separate ELISA determinations, as SMART databases grow the technology for programmable automated immunoanalyzers will become more readily available, and CD-based bedside analyzers are under development. Thus, the next level of SMART prognostication may become a practical reality.

Whether incorporating serial measurements of inflammatory response mediators and physiologic data into SMART models improves predictive accuracy in septic surgical patients is not clear from the present data. Predictions that are based on serial input were validated prospectively at higher ROC AUC values than were baseline prognostications in 59% of the points at which they could be compared directly. On the other hand, more baseline SMART models than baseline plus serial data equations were validated at ROC AUC determinations above 0.700. Adequate answers to the baseline versus baseline plus serial data SMART debate, then, will require detailed analysis of more expanded databases than were available here.

The requirement for mechanical ventilation in septic surgical patients proved difficult to predict in this study. From baseline data, SMART models were generated with only modest success for days 6 and 14. In contrast, models derived from baseline plus serial input were validated with respectable ROC AUC determinations for days 5 through 28. The question of whether more comprehensive measurements of inflammatory response mediators associated with ventilatory dysfunction could yield useful prognostications based on baseline data is not clear from the present

results. Also unanswered is the speculation that the excellent prediction by baseline plus serial models at 5 days and beyond may simply represent continued ventilator dependence of already compromised patients.

Clinical shock was also a difficult independent variable to predict consistently. Predicted versus observed ROC AUC determinations above 0.700 were achieved in only five out of 19 models for septic shock. One might speculate that the multiple factors that contribute to septic hypotension made it difficult to find strong prognostic interactions with either baseline data or baseline plus serial input. That many key inflammatory mediators associated with systemic vasodilatation in sepsis, such as prostacyclin [32] and nitric oxide [33], were not measured also may have influenced the suboptimal results.

It is important to note that the predictions for septic surgical patients presented in here were validated prospectively on SMART models built from a mixed medical/surgical/gynecologic sepsis training cohort. This contrasts with previous reports that suggested prognostic incompatibility of such mixed databases with surgical outcomes. For example, Cerra *et al* [34] reported that the APACHE II score did not accurately predict mortality risk in trauma patients. Determining whether prognostic models based strictly on surgical patients may be superior to the present results will require direct performance comparisons of SMART surgical and mixed sepsis databases.

Although predicting 28-day survival was not the primary focus of the present study, it was predicted with moderate success using baseline plus serial data SMART models. Not surprisingly, possibly reflecting a more defined mortality/survival population later during the course of severe sepsis/septic shock, predictions of survival were most accurate from serial data at 21 and 28 days. A similar phenomenon has been reported for the APACHE III [25].

Although the ability of the models presented here to discriminate between patients with or without ARDS, DIC, hepatobiliary failure, renal insufficiency, shock, or cerebral dysfunction was fairly consistent, Hosmer–Lemeshow statistics indicated that calibration (the degree of correspondence between the outcomes predicted by the present models and the actual outcomes) was somewhat more variable. Hosmer–Lemeshow goodness-of-fit statistics were statistically significant in the majority of models that predicted ARDS, hepatobiliary failure, renal insufficiency, and cerebral dysfunction, but were significant in less than half of the models that predicted DIC or shock. Calibration was not achieved at significant levels for chest radiography score. These findings are another indication of the need to further develop the SMART approach to prognostication in larger studied populations before this concept can become a worthwhile adjunct to clinical judgment.

The present study prospectively validated multiple logistic regression models from a population of septic surgical, medical, and gynecologic patients that predicted shock, pulmonary edema, mechanical ventilation, ARDS, DIC, and hepatobiliary, renal, and cerebral dysfunction in septic surgical patients, up to 28 days in advance. Models were confirmed prospectively also that predicted the occurrence of shock and organ system dysfunction. These prognostications are applicable to individual patients, and frequently at clinically useful levels of accuracy. Plasma cytokine concentrations contributed significantly as weighted independent variables to many quantitative models, and to most predictions of dichotomous dependent variables.

References

- Dominioni L, Dionigi R, Zanello M, Monico R, Cremaschi R, Dionigi R, Ballabio A, Massa M, Comelli M, Dal Ri P: **Sepsis score and acute-phase protein response as predictors of outcome in septic surgical patients.** *Arch Surg* 1987, **122**:141–146.
- Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ: **Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome.** *Crit Care Med* 1995, **23**:1638–1652.
- Vincent JL, de Mendonca A, Cantraine F, Moreno R, Takala J, Suter PM, Sprung CL, Colardyn F, Blecher S: **Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study.** *Crit Care Med* 1998, **26**:1793–1800.
- Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J: **Mortality probability models (MPM II) based on an international cohort of intensive care unit patients.** *JAMA* 1993, **270**:2478–2486.
- Le Gall J, Lemeshow S, Saulnier F: **A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study.** *JAMA* 1993, **270**:2957–2963.
- Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A: **The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults.** *Clin Invest Crit Care* 1991, **100**:1619–1636.
- Champion HR, Copes WS, Sacco WJ, Frey CF, Holcroft JW, Hoyt DB, Weigelt JA: **Improved predictions from A Severity Characterization of Trauma (ASCOT) over Trauma and Injury Severity Score (TRISS): results of an independent evaluation.** *J Trauma Injury Infect Crit Care* 1996, **40**:42–49.
- Rutledge R, Osler T, Emery S, Kromhout-Schiro S: **The end of Injury Severity Score (ISS) and the Trauma and Injury Severity Score (TRISS): ICISS, an International Classification of Diseases, Ninth Revision-based prediction tool, outperforms Both ISS and TRISS as predictors of trauma patient survival, hospital charges, and hospital length of stay.** *J Trauma Injury Infect Crit Care* 1998, **44**:41–49.
- Boyd CR, Tolson MA, Copes WS: **Evaluating trauma care: the TRISS method.** *J Trauma* 1987, **27**:370–378.
- Hannan EL, Mendeloff J, Farrel LS, Cavien CG, Murphy JG: **Validation of TRISS and ASCOT using a non-MTOS trauma registry.** *J Trauma Injury Infect Crit Care* 1995, **38**:83–88.
- Lett RR, Hanley JA, Smith JS: **The comparison of injury severity instrument performance using likelihood ratio and ROC curve analyses.** *J Trauma Injury Infect Crit Care* 1995, **38**:142–148.
- Talmor M, Hydo L, Barie PS: **Relationship of systemic inflammatory response syndrome to organ dysfunction, length of stay, and mortality in critical surgical illness.** *Arch Surg* 1999, **134**:81–87.
- Rixen D, Siegel JH, Friedman HP: **'Sepsis/SIRS,' physiologic classification, severity stratification, relation to cytokine elaboration and outcome prediction in post-trauma critical illness.** *J Trauma Injury Infect Crit Care* 1996, **41**:581–598.
- Meduri GU, Headley S, Kohler G, Stentz F, Tolley E, Umberger R, Leeper K: **Persistent elevation of inflammatory cytokines predicts a poor outcome in ARDS. Plasma IL-1 β and IL-6 levels are consistent and efficient predictors of outcome over time.** *Chest* 1995, **107**:1062–1073.
- Slotman GJ, Burchard KW, Yellin SA, Williams JJ: **Prostaglandin and complement interaction in clinical acute respiratory failure.** *Arch Surg* 1986, **121**:271–274.
- Slotman GJ, Quinn JV: **Multivariate regression modeling for the prediction of inflammation, systemic pressure, and end-organ function in severe sepsis.** *Shock* 1997, **8**:225–231.
- Slotman GJ, Quinn JV: **Multivariate regression modeling for the prediction of plasma eicosanoid and cytokine concentrations in patients with severe sepsis.** *J Surg Outcomes* 1998, **1**:24–30.
- Fisher CJ, Dhainaut JFA, Opal SM, Pribble JP, Balk RA, Slotman GJ, Iberti TJ, Rackow EC, Shapiro MJ, Greenman RL, Reines HD, Shelly MP, Thompson BW, LaBrecque JF, Catalano MA, Knaus WA, Sadoff JC, for the Phase III rIL-1ra Sepsis Syndrome Study Group: **Recombinant human interleukin-1 receptor antagonist in the treatment of patients with sepsis syndrome.** *JAMA* 1994, **271**:1836–1843.
- SAS Institute Inc: *SAS/STAT User's Guide, Version 6, Fourth Edition, Volumes 1 & 2.* Cary, NC: SAS Institute Inc; 1994.
- Afifi A, Clark C: *Computer-Aided Multivariate Analysis*, 2nd ed. New York: Van Nostrans and Reinhold Co; 1990:179.
- Kirkwood BR: *Essentials and Medical Statistics.* Oxford, UK: Blackwell Scientific Publications; 1988.
- Hanley JA, McNeil BJ: **The meaning and use of the area under a receiver operating characteristic (ROC) curve.** *Radiology* 1982, **143**:29–36.
- Harrell FE, Lee KL, Mark D: **Tutorial in biostatistics multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.** *Stat Med* 1996, **15**:361–387.
- Sauaia A, Moore FA, Moore EE, Haenel JB, Read RA, Lezotte DC: **Early predictors of post-injury multiple organ failure.** *Arch Surg* 1994, **129**:39–45.
- Wagner DP, Knaus WA, Harrell FE, Zimmerman JE, Watts C: **Daily prognostic estimates for critically ill adults in intensive care units. Results from a prospective, multicenter, inception cohort analysis.** *Crit Care Med* 1994, **22**:1359–1372.
- Millili JJ, Philiponis VS, Nusbaum M: **Predicting surgical outcome using Bayesian analysis.** *J Surg Res* 1998, **77**:45–49.
- Sellers BJ, Davis BL, Larkin PW, Morris SE, Saffle JR: **Early prediction of prolonged ventilator dependence in thermally injured patients.** *J Trauma Injury Infect Crit Care* 1997, **43**:899–903.
- Osler TM, Rogers FB, Glance LG, Cohen M, Rutledge R, Shackford SR: **Predicting survival, length of stay, and cost in the surgical intensive care unit: APACHE II versus ICISS.** *J Trauma Injury Infect Crit Care* 1998, **45**:234–237.
- Slotman GJ, Friedman B, Brathwaite C, Mure A, Quinn JV, Shapiro E: **Interleukin-1 mediates increased plasma levels of eicosanoids and cytokines in patients with sepsis syndrome.** *Shock* 1995, **4**:318–323.
- Pinsky MR, Vincent JL, Deviere J, Alegre M, Kahn RJ, Dupont E: **Serum cytokine levels in human septic shock. Relation to multiple-system organ failure and mortality.** *Chest* 1993, **103**:565–575.
- Boldt J, Wollbrück M, Kuhn D, Linke LC, Hempelmann G: **Do plasma levels of circulating soluble adhesion molecules differ between surviving and non-surviving critically ill patients.** *Chest* 1995, **107**:787–792.
- Slotman GJ, Burchard KW, Williams JJ, D'Arezzo A, Yellin SA: **Interaction of prostaglandins, activated complement, and granulocytes in clinical sepsis and hypotension.** *Surgery* 1986, **99**:744–750.
- Wray GM, Millar CG, Hinds CJ, Thiernemann C: **Selective inhibition of the activity of inducible nitric oxide synthase prevents the circulatory failure, but not the organ injury/dysfunction, caused by endotoxin.** *Shock* 1998, **9**:329–335.
- Cerra FB, Negro F, Abrams J: **APACHE II score does not predict multiple organ failure or mortality in postoperative surgical patients.** *Arch Surg* 1990, **125**:519–522.

Author's affiliation: Department of Surgery, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School at Camden, Cooper Hospital/University Medical Center, Camden, New Jersey, USA

Correspondence: Gus J Slotman, MD, Department of Surgery, Three Cooper Plaza, Suite 411, Camden, NJ 08103, USA