ORIGINAL

Philipp G. H. Metnitz
Rui P. Moreno
Eduardo Almeida
Barbara Jordan
Peter Bauer
Ricardo Abizanda Campos
Gaetano Iapichino
David Edbrooke
Maurizia Capuzzo
Jean-Roger Le Gall
on behalf of the SAPS 3
Investigators

SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description

P. G. H. Metnitz (✉)
Dept. of Anesthesiology
and General Intensive Care,
University Hospital of Vienna,
Vienna, Austria
e-mail: philipp@metnitz.biz
Fax: +43-1-9522843

R. P. Moreno
Unidade de Cuidados Intensivos
Polivalente,
Hospital de St. António dos Capuchos,
Centro Hospitalar de Lisboa (Zona Central),
Lisboa, Portugal

E. Almeida
Unidade de Cuidados Intensivos,
Hospital Garcia de Orta,
Pragal, Portugal

B. Jordan
Dept. of Medical Statistics,
University of Vienna, Vienna, Austria

R. A. Campos
Dept. of Intensive Care,
Hospital Universitario Asociado General
de Castelló, Castello, Spain

G. Iapichino
Dept. of Anesthesia
& Intensive Care Medicine,
Hospital San Paolo, Università degli Sudi,
Milano, Italy

D. Edbrooke
Critical Care Directorate,
Royal Hallamshire Hospital,
Sheffield, United Kingdom

M. Capuzzo
Dept. of Anesthesia
& Intensive Care Medicine,
Hospital of Ferrara,
Ferrara, Italy

J.-R. Le Gall
Dept. Réanimation Médicale,
Hôpital St. Louis, Université Paris VII,
Paris, France

Abstract *Objective:* Risk adjustment systems now in use were developed more than a decade ago and lack prognostic performance. Objective of the SAPS 3 study was to collect data about risk factors and outcomes in a heterogeneous cohort of intensive care unit (ICU) patients, in order to develop a new, improved model for risk adjustment. *Design:* Prospective multicentre, multinational cohort study. *Patients and setting:* A total of 19,577 patients consecutively admitted to 307 ICUs from 14 October to 15 December 2002. *Measurements and results:* Data were collected at ICU admission, on days 1, 2 and 3, and the last day of the ICU stay. Data included sociodemographics, chronic conditions, diagnostic information, physiological derangement at ICU admission, number and severity of organ dysfunctions, length of ICU and hospital stay, and vital status at ICU and hospital discharge. Data reliability was tested with use of kappa statistics and intraclass-correlation coefficients, which were >0.85 for the majority of variables. Completeness of the data was also satisfactory, with 1 [0–3] SAPS II parameter missing per patient. Prognostic performance of the SAPS II was poor, with significant differences between observed and expected mortality rates for the overall cohort and four (of seven) defined regions, and poor calibration for most tested subgroups. *Conclusions:* The SAPS 3 study was able to provide a high-quality multinational database, reflecting heterogeneity of current ICU case-mix and typology. The poor performance of SAPS II in this cohort underscores the need for development of a new risk adjustment system for critically ill patients.

Keywords Intensive care unit · Severity of illness · ICU mortality · Hospital mortality · Risk adjustment

Introduction

Following the publication in the early 1980s of the Acute Physiology and Chronic Health Evaluation Score (APACHE [1]), Simplified Acute Physiology Score (SAPS [2]), and—some years later—APACHE II [3] systems, outcome prediction became an important topic among European intensivists. Ten years later, a new generation of these instruments was published: APACHE III [4], SAPS II [5], and Mortality Probability Model (MPM) II [6]. All of these newer systems were developed by using sophisticated statistical techniques in large multinational databases, and were found to perform better than their predecessors [7, 8].

The availability of such sophisticated methods for risk adjustment facilitated outcome research in critically ill patients, which became increasingly important over time. Risk adjustment systems now have a fixed place in critical care research for various purposes. At the patient level, the reporting of severity of illness and the use of risk-adjusted mortality rates to draw inferences from their results are a prerequisite for any study to be published. At the intensive care unit (ICU) level, observed-to-expected mortality ratios (or the use of direct standardisation techniques based on severity scores) have become standard for assessing the impact of ICU-related factors on outcome, such as the effects of organisation and management [9, 10].

However, a series of studies assessing the performance of risk adjustment systems unveiled a lack of prognostic performance of these systems: In most cases, lack of calibration was evident over several subgroups of patients, often accompanied by an underestimation of mortality in low-risk patients and an overestimation in high-risk patients. This pattern was observed for all published outcome prediction models in several countries [11, 12, 13, 14, 15, 16, 17, 18] and seemed to be worsening over time [19].

For this reason, several researchers tried to improve the prognostic performance of various systems through recalibration, using one of two possible approaches. A level 1 customization requires calculation of a new equation for the prediction of hospital mortality (without changing the weights of the constituent variables). A level 2 customization involves a reweighting of each variable contained in the model. Although recalibration was able to improve prognostic accuracy in some cases [13, 14], it generally did not solve the various problems inherent in the models.

These problems can be classified as either user-, patient-, or model-dependent. User-dependent problems include differences in the definitions and application criteria [20, 21]. Patient-dependent problems are mainly shifts in the baseline characteristics of the populations over time [22]: age distribution, distribution of illnesses, and the development of new treatments, all of which affect prognosis. Model-dependent problems have many different causes, such as the lack of important prognostic variables (e.g., diagnostic information [4, 23]) or the presence, location and aetiology of infection [24, 25, 26]. Confounding variables and statistically wrong assumptions [9, 27] also distort performance results.

If recalibration is not sufficient to improve the performance of the prognostic model, the only alternative is to develop a new model that takes into account the results of studies done since the original model was developed. This means incorporating missing variables that have been shown to affect outcome, minimizing problems with the application of the model, and reducing the possibility of other confounders.

The objective of the SAPS 3 project was to cope with the above-stated problems by developing a new model for improved risk adjustment in critically ill patients. Another important goal was to make the new model available free of charge for use in the scientific community.

In the SAPS 3 study (which took place at the end of 2002), data about risk factors and outcomes in an international multicentric cohort of critically ill patients were prospectively collected so that a high-quality database would be available for further analysis of the associations between risks and outcomes in our patients.

Materials and methods

Project Organization

The SAPS 3 project was conducted by the SAPS 3 Outcomes Research Group. The project was endorsed by the European Society for Intensive Care Medicine (ESICM, http://www.esicm.org) and conducted in cooperation with the Section on Health Services Research and Outcome of the ESICM. The SAPS 3 Outcomes Research Group consists of a project coordinator and a steering group. The steering group was responsible for the scientific conduct and consistency of the project. An additional advisory board integrated further scientists with special expertise who were asked for comments on the scientific content and for help in conducting the project. The complete board lists can be found in Appendix D of the Electronic Supplementary Material (ESM).

During the data collection phase, a coordination and communications centre (CCC) was installed. The CCC was responsible for the management and control of the project. This included the administration of all project tasks and implementation of actions and activities as necessary; communication between project partners (e.g., centres, researchers and institutions) through sampling and distribution of necessary information; and pooling and administration of the data provided by project participants. In addition, the CCC provided almost around-the-clock service to answer urgent questions and resolve problems during the phase of data collection.

In each country, a country coordinator was responsible for operational management and direct communication with the participating ICUs in that country, including giving specific help when necessary. The country coordinator was responsible for ensuring completion of the various tasks required of the participating ICUs. The list of country coordinators can be found in Appendix E of the ESM.

At the ICU level, an ICU coordinator was responsible for local activities, such as obtaining approval from the local ethics or data-

protection committees where applicable. In addition, the ICU co-ordinator was responsible for supervising the daily data collection, problem management, controlling the completeness of the data, data quality control, training medical and nonmedical staff for data collection, management of the data, and transmission of the data to the CCC or country coordinator. The list of ICU coordinators can be found in Appendix F of the ESM.

Data collection

Patient data were recorded by using either online data collection software (provided by *i*MD*soft*, Tel Aviv, Israel) or the SAPS 3 stand-alone database system (provided by the CCC). The latter software used a Microsoft Access database (Microsoft Corporation, Redmond, WA, USA) for data storage and needed no Internet connection for data entry. Both systems maintained a variety of plausibility controls to ensure the quality of the recorded data. Each variable was precisely defined before the start of data collection (see Appendix C of the ESM). Detailed definitions of the variables were available to participants in both paper and electronic form. To facilitate plausibility checking, each variable was assigned a probability range, encompassing the range of probable values for that variable. In addition, a range of possible values (storage range) for that variable was defined (e.g., for FiO_2, no values <21% or >100% could be accepted). Thus, formal plausibility controls in the software systems were used wherever possible and ensured the maximum of data quality checking during data collection.

Participants who could not use one of the two software options were allowed to record the data on paper forms and submit them to the CCC (*n*=26 ICUs). Patient data were then entered into the SAPS 3 stand-alone software system and thus checked for plausibility. In cases of uncertainty, ICU coordinators were contacted for clarification.

In addition, each ICU received a questionnaire with detailed questions about ICU structures and about resources available in other areas of the hospital.

Data were collected at ICU admission, on days 1, 2 and 3, and on the last day of the ICU stay. Data from the day of admission (aside from sociodemographic data such as age and sex) were categorized into different levels: (i) data about the condition of the patient before ICU admission, such as chronic conditions and medical diseases; (ii) data about the patient's condition at ICU admission, such as the reason for admission, infection at admission, and surgical status; and (iii) data about the patient's physiologic derangement at ICU admission. These data were collected within an hour before or after ICU admission.

On the following days of the ICU stay, further information was collected: severity of illness, as measured by the SAPS II [5]; number and severity of organ dysfunction, as measured by the Sequential Organ Failure Assessment (SOFA) [28]; length of ICU and hospital stay; and outcome data, including vital status at ICU and hospital discharge. All patients were subjected to mandatory follow-up until hospital discharge, but not longer than 90 days after ICU admission. Patients still remaining in the hospital at 90 days were at that time classified as being "still in the hospital".

To record diagnoses, a three-level system was used. (i) An a*cute medical disease* was recorded for all patients, independent of surgical status, i.e., the acute (or acute on chronic) disease that best explained the ICU admission. If the reason for ICU admission was infectious disease, then this was recorded. (ii) *Surgical status* at admission and the anatomic site of surgery were recorded for all patients undergoing surgery during the hospital stay before ICU admission. (iii) A concrete *reason for admission* had to be selected. At least one reason for admission was required, but several selections were possible (one within each organ system). If no other reason was present, at least "basic and observational care" had to be selected.

All participants received detailed documentation of patient- and ICU-based data items as well as a detailed description of the data collection process. Moreover, specific forms to check the completeness of the patient-based documentation were provided. Additionally, a training session for ICU coordinators was organised at the 15th Annual Congress of the ESICM in Barcelona, Spain, before the start of data collection. Throughout the project, the project website (http://www.saps3.org) provided all necessary information. In addition, the CCC was available to answer questions by email, fax and phone. Data were to be collected from all consecutively admitted patients between 14 October and 15 December 2002. ICUs with a high number of beds (and thus also admissions) could stop patient enrolment after contributing 100 patients.

Database

Data were collected and pooled by the CCC. The final database file was then imported into the SAS system, Version 8e (SAS Institute Inc., Cary, USA). Data cleaning was accomplished through the application of a variety of additional plausibility controls and cross-checking of information between redundant data fields.

A total of 22,791 admissions were recorded in the 309 participating ICUs during the study period. For patients who were admitted more than once (*n*=1,455), only the first admission was included, giving 21,336 admitted patients. Patients who were <16 years of age (*n*=628), those without ICU admission or discharge data (*n*=1,074), and those with records that lacked an entry in the field "ICU outcome" (*n*=57) were excluded. The *Basic SAPS 3 Cohort* thus comprises 19,577 patients from 307 ICUs.

For the development of a predictive model for hospital mortality as outcome, patients with a missing entry in the field of "vital status at hospital discharge" (*n*=2,540) or an entry of "still in the hospital" at the end of the follow-up period (*n*=253) were further excluded. The *SAPS 3 Hospital Outcome Cohort* thus comprises 16,784 patients from 303 ICUs.

Because the study was an observational study and no additional interventions were performed, the need for informed consent was waived by the institutional review board. Each ICU, however, was made responsible for obtaining local permissions as necessary.

Data quality

Recorded data were evaluated for completeness of the documentation and reliability. Interrater quality control was performed through rescoring of the data from day 0 (the day of ICU admission) for three randomly selected patients in each ICU. From the rescored data, kappa coefficients and intra-class correlation coefficients were calculated, as appropriate. Availability of the variables necessary to calculate the SAPS II was used as an indicator for the completeness of the data.

Statistical analysis

Statistical analysis was performed using the SAS system, version 8e (SAS Institute Inc., Cary, NC, USA). A *P* value of <0.05 was considered significant. Unless otherwise specified, results are expressed as median and interquartile ranges (quartile). Observed-to-expected (O/E) mortality ratios were calculated by dividing the number of observed deaths per group by the number of expected deaths per group (as predicted by the SAPS II). To test for statistical significance, we calculated 95% confidence intervals (CI) according to the method described by Hosmer and Lemeshow [29]. The Hosmer-Lemeshow goodness-of-fit Ĥ-statistic and Ĉ-statistic [30] were used to evaluate the calibration of the SAPS II. Discrimination was tested by measuring the area under the receiver

operating characteristic (aROC) curve, as described by Hanley and McNeil [31]. Reliability of data collection was analysed using K-statistics or intra-class correlation coefficients, as appropriate. Statistical methods used for the development of the predictive model are described in Part 2 of this report.

Results

Data quality

Four hundred eighty-three rescored patients could be identified and linked to their original counterparts (2.5% of admitted patients). Data quality was found to be excellent, with the majority of coefficients being >0.85. Only two of the more than 50 tested variables had coefficients <0.80 (body weight, 0.79; positive end-expiratory pressure, 0.72), and only one was <0.70 (leukocytes [maximum], 0.57). For a detailed list of coefficients see Table E1 in the ESM. Data completeness was also found to be satisfactory, with 1 [0–3] SAPS II parameter missing per patient.

Description of ICUs

The Basic SAPS 3 cohort includes 307 ICUs from 35 countries. On average each ICU contributed 50 (27–78) patients to the cohort. To assess heterogeneity of results between different geographic regions, seven regions were defined: Australasia, Central and South America, Central and Western Europe, Eastern Europe, North America, Northern Europe, and Southern Europe and Mediterranean countries. The allocation of countries to these regions can be seen from Table E10 of the ESM.

Seventy percent of the participating ICUs identified themselves as mixed medical-surgical (Table E2, ESM). Roughly half of the ICUs (46%) were located in university-affiliated or teaching hospitals. Eighty-four percent of ICUs (*n*=258) reported having a full-time medical director, and 272 (88.6%) reported having a full-time nursing director. On weekdays, 76.6% of ICUs reported having an intensivist available on the ICU 24 hours per day, whereas 6.2% had an intensivist available in the hospital. In 12.1% of ICUs, the intensivist was at home, on-call, during the daytime. During weekends, this proportion did not change much (74.3%, 5.5%, and 15.0% on the ICU, in the hospital, and on-call, respectively). None of the participating ICUs reported having no intensivists available during night or weekend shifts.

Description of patients

The Basic SAPS 3 Cohort comprises 19,577 patients admitted to participating ICUs during the study period. More than 70% of patients were admitted from the same

Table 1 ICU admission data for the two cohorts (*Basic cohort:* SAPS 3 basic cohort; *HO cohort:* SAPS 3 Hospital Outcome Cohort; *n:* number of patients)

	Basic cohort		HO cohort	
	n	*%*	*n*	*%*
Number of patients	19,577		16,784	100.0
Gender				
Female	7,678	39.2	6,610	39.4
Male	11,881	60.7	10,161	60.5
Missing	18	0.1	13	0.1
Age, years (median, quartiles)	63	49-74	64	49-74
Origin				
Home	2,810	14.4	2,343	14.0
Same hospital	13,926	71.1	12,063	71.9
Chronic care facility	74	0.4	64	0.4
Public place	519	2.7	432	2.6
Other hospital	2,125	10.9	1,791	10.7
Other	80	0.4	59	0.4
Missing	43	0.2	32	0.2
Intra-hospital location before ICU admission				
Emergency room	5,419	27.7	4,630	27.6
Intermediate care unit/ High dependency unit	562	2.9	475	2.8
Operating room	7,537	38.5	6,449	38.4
Other	552	2.8	413	2.5
Other ICU	698	3.6	611	3.6
Recovery room	482	2.5	400	2.4
Ward	3,411	17.4	3,036	18.1
Missing	916	4.7	770	4.6
ICU admission status				
Planned admission	6,750	34.5	5,598	33.4
Unplanned admission	12,338	63.0	10,801	64.4
Missing	489	2.5	385	2.3
Acute Infection at ICU admission				
No infection	15,254	77.9	12,968	77.3
Clinically improbable/ colonization	342	1.7	298	1.8
Clinically probable/ documented	2,761	14.1	2,422	14.4
Microbiologically documented	1206	6.2	1,083	6.5
Missing	13	0.1	13	0.1
Surgical status				
No surgical procedure	8,437	43.1	7,305	43.5
Scheduled surgery	6,800	34.7	5,700	34.0
Emergency surgery	3,321	17.0	2,930	17.5
Missing	1,019	5.2	849	5.1

hospital as the ICU, with operating rooms, emergency departments and normal wards contributing most of the patients (Table 1). Almost two thirds of the admissions were classified as unplanned. The mean age of patients was 60.0±17.7 years (Fig. 1), and 39.2% were female. Comorbidities were recorded in 65% of admitted patients, with arterial hypertension, chronic obstructive pulmonary disease, and chronic heart failure being the most frequent (Table E3, ESM).

Cardiovascular, respiratory and neurologic diseases were the most frequent organ-specific reasons for admission (Table E4, ESM). The acute medical diseases

Fig. 1 Age distribution and associated mortality. The figure shows the age distribution of the Basic SAPS 3 Cohort (*n*=19,577) and the corresponding ICU mortality rates for each stratum. *Columns:* Number of patients as percentages of the whole cohort; *squares:* ICU mortality rates for the corresponding stratum
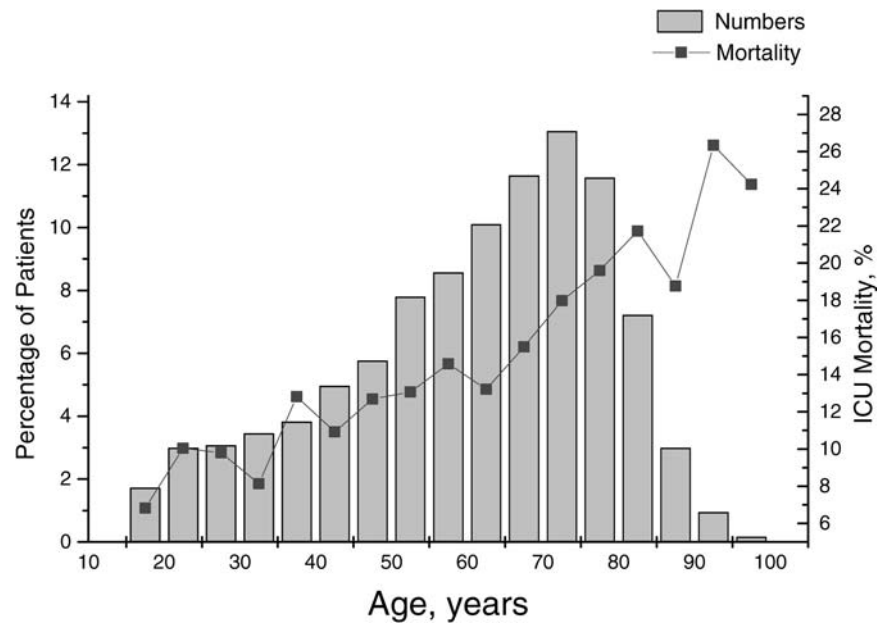

Table 2 ICU discharge and outcome data for the two cohorts (*Basic cohort:* SAPS 3 basic cohort; *HO cohort:* SAPS 3 Hospital Outcome Cohort; *n:* number of patients; *ICU LOS:* ICU length of stay; *IMCU/HDU:* intermediate care unit/high dependency unit; *Q1, Q3:* lower and upper interquartile range, respectively)

	Basic cohort		HO cohort	
	n	%	n	%
Number of patients	19,577	100.0	16,784	100.0
ICU LOS, days (median, quartiles)	2	1–6	2	1–6
ICU discharge–destination				
Home	438	2.2	361	2.2
Same hospital	14,946	76.3	12,477	74.3
Other hospital	1,029	5.3	852	5.1
Missing	3,164	16.2	3,094	18.4
Intrahospital discharge				
Emergency room	58	0.3	50	0.3
IMCU/HDU	2,222	11.4	1,873	11.2
Other	303	1.5	257	1.5
Other ICU	583	3.0	479	2.9
Recovery room	306	1.6	218	1.3
Ward	12,250	62.6	10,291	61.3
Missing	3,855	19.7	3,616	21.5
ICU discharge—status				
Planned discharge	14,872	76.0	12,262	73.1
Unplanned discharge	1,595	8.1	1,467	8.7
Missing	3,110	15.9	3,055	18.2
Risk adjustment				
SAPS II score (median, Q1–Q3)	30	20–42	31	21–43
SOFA score (median, Q1–Q3)	9	6–11	9	6–11
Outcome				
ICU mortality (%)		15.2		17.7

necessitating ICU admission included a broad spectrum of diagnoses (Table E5, ESM). Approximately one half of the patients underwent surgery before ICU admission, with abdominal, cardiac and vascular surgery being the most frequent procedures (Table E6, ESM).

Regarding discharge details (Table 2), it is notable that a high percentage of patients were discharged unplanned (8.15%), i.e., without at least a 12-hour planning window. 15.2% of patients from the SAPS 3 Basic cohort died within the ICU. As can be seen from Table 3, patient cohorts differed significantly between regions. Both, ICU and hospital mortality rates exhibited a broad spectrum between ICUs: hospital mortality was on average 28% (17–42%) in the SAPS 3 Hospital outcome cohort.

Performance of the SAPS II

The performance of the original SAPS II model [5] (using data from the first 24 hours) was tested in the SAPS 3 Hospital Outcome Cohort (*n*=16,784). Discrimination was good with an aROC of 0.83 (95% CI: 0.824–0.838). SAPS II showed underestimation of hospital mortality: The O/E ratio of the overall cohort was 1.08 (1.06–1.10). O/E ratios significantly differed between regions: from 0.86 (0.81–0.91) for Central and Western Europe to 1.32 (1.25–1.38) for Central and South America, with four out of the seven defined regions exhibited O/E ratios significantly different from 1 (Table E7, ESM). Calibration, as assessed by the Hosmer-Lemeshow $\hat{H} + \hat{C}$ statistics, was poor for the overall cohort: $\hat{H}$ 227.21, $\hat{C}$ 184.70; both $p<0.0001$; This lack of calibration was present for all tested subgroups except for the region of North America (see Table E7, ESM).

Table 3 ICU admission and discharge data for the seven defined geographic regions (SAPS 3 Basic Cohort; n=19,577)

	Australasia		Central & South America		Eastern Europe		Central and Western Europe		Northern Europe		Southern Europe and Mediterranean countries		North America	
Number of patients	2,235		2,540		1,084		4,712		355		7,854		797	
Females, %	38.0		44.6		42.2		40.1		42.5		36.9		38.0	
Age, years (median, quartiles)	59	45–71	31	46–74	62	48–72	65	52–75	66	51–76	64	59–74	64	49–75
SAPS II score (median, Q1-Q3)	28	19–40	30	20–42	27	18–43	29	21–40	35	25–46	32	22–44	29	20–39
SOFA score (median, Q1-Q3)	8	6–10	8	6–11	9	7–11	8	5–11	9	7–11	9	7–11	9	6–11
ICU mortality, %	12.7		17.4		16.9		10.8		20.6		18.1		8.5	

Discussion

To the best of our knowledge, the SAPS 3 study is the largest prospective epidemiologic multicentre, multinational study conducted in health services and outcomes research in intensive care medicine to date.

The project was first intended to focus on Europe because it was believed such a strategy would produce a more homogeneous cohort of patients, which would in turn provide a more stable reference line for further comparisons. This idea was discussed during several investigator meetings and finally abandoned—first, because interest from outside Europe was enormous: 39% of ICUs that registered for the project were located outside Europe. The SAPS 3 board members thus agreed that such a high level of interest should not be ignored. Second, some investigators questioned whether a concentration on European ICUs would be successful in reducing heterogeneity anyway. Provision of intensive care in Europe is extremely variable, with enormous differences in severity of illness, provision of treatments and mortality from north to south and from west to east [32, 33].

For these reasons ICUs from regions outside Europe were invited to participate. Our results prove that we were right in our assumptions: First, one can easily see that the four European regions (as defined in our study) are hardly comparable: severity of illness as measured by the SAPS II varied from 27 to 35 points, and ICU mortality ranged from 10.8 to 20.6%—almost a doubling of mortality figures (Table 3). Second, almost a third of the patient cohort (28.5%) was contributed from regions outside Europe.

Although the decision to accept ICUs worldwide probably increased the heterogeneity of our sample, it also allowed the SAPS 3 database to better reflect important differences in patients' and health care systems' baseline characteristics that are known to affect outcome. These include, for example, different genetic makeups, different styles of living or a heterogeneous distribution of major diseases within different regions, as well as issues such as access to the health care system in general and to intensive care in particular, or differences in availability and use of major diagnostic and therapeutic measures within the ICUs [32, 34]. Although the integration of ICUs outside Europe and the U.S. surely increased it's representativeness, it must be acknowledged, that the extent to which the SAPS 3 database reflects case-mix on ICUs worldwide cannot be determined yet.

It should additionally be noted that allocation of countries to regions does not always follow geographic borders (Table 3; see also Table E10 in the ESM). Partitioning of the sample was done to adjust for some of the above-stated differences between different populations and to develop a system that uses several different reference lines to compare ICUs on a similar level. Thus,

patients are not necessarily representative of their respective regions.

To minimize possible seasonal influences, we chose late fall in the Northern Hemisphere for data collection. Thus, participants in both late fall/winter (Northern Hemisphere) and spring/summer (Southern Hemisphere) are represented in our cohort. A recent study [35] showed, moreover, that differences in seasonal mortality rates, at least in a sample of ICUs in the United Kingdom, were related to variations in case mix rather than to a specific impact of season on outcome.

Performance of the SAPS II was, not surprisingly, found to be similar to that in previous studies: acceptable discrimination but lack of calibration. Possible reasons for this have already been alluded to in the Introduction. In contrast to previous studies, however, we found an underestimation of hospital mortality, which contradicts the rationale that the shifting in calibration is due only to the development of new and possibly better therapies and thus to better ICU performance [19].

Analyzing the various geographic regions provides evidence that the underestimation of hospital mortality by the SAPS II might be partially attributable to the composition of the cohort: SAPS 3 is the first large international study on severity of illness systems to include patients from all continents. South America, for example, where provision of intensive care is much more limited than it is in Europe or North America, contributed extensively to the patient cohort. High O/E ratios have already been reported for this continent [36] and are probably linked to the limited availability of resources.

Data quality was one of our major concerns. Completeness of the documentation was found to be satisfactory: The amount of missing data is in fact smaller than reported from previous cohort studies on severity of illness systems [11, 12, 16]. With respect to reliability, intraclass-correlation coefficients and kappa coefficients were generally similar to or even better than those found in previous studies, showing a high degree of interrater agreement (see Table E1 in ESM) [37, 38].

We did, however, experience problems with the cohort of rescored patients: First, we had to exclude all rescored patients for whom the original counterpart was also excluded due to the application of any of the exclusion criteria. Second, in some cases the original patient identification was either missing or documented in such a way that a unique allocation was not possible. Both of these exclusions reduced the number of rescored patients available for analysis.

Two strategies to build up a cohort are available: first, to recruit only patients who meet well-documented inclusion criteria (such as documented vital status at hospital discharge) or, second, to document all patients and then exclude patients based on a predefined set of exclusion criteria. For the SAPS 3 study we chose the second option—to form two different cohorts—because we needed to provide a basic cohort for all further analyses of the SAPS 3 database. Since some studies will focus on different outcomes (e.g., ICU outcome rather than hospital outcome), we decided to use missing ICU outcome (and not hospital outcome) as an exclusion criterion for the basic cohort.

A possible limitation of the SAPS 3 database is that vital status at hospital discharge was not available for all admitted patients. Despite several efforts from the CCC and sufficient time to allow for a close follow-up, we did not succeed to receive all hospital outcomes documented. Recording of hospital outcome (or later outcomes) still poses major problems for ICUs in European and non-European hospitals, either because of technical problems or possibly because of data security algorithms in the hospitals. Exclusion of these patients did, however, not affect major criteria, such as geographic representation, ICU admission or discharge data, co-morbidities, or the distribution of the reasons for admission (Tables 1 and 2).

We conclude that the SAPS 3 database is within the above discussed limits of high quality and reflects the heterogeneity of current intensive care provision. As such, it provides an excellent basis for the development of a new risk adjustment system.

References

1. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. (1981) APACHE—acute physiology and chronic health evaluation: a physiologically based classification system. Crit Care Med 9:591–597

2. Le Gall J-R, Loirat P, Alperovitch A. (1983) Simplified acute physiological score for intensive care patients. Lancet ii:741

3. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. (1985) APACHE II: a severity of disease classification system. Crit Care Med 13:818–829

4. Knaus WA, Wagner DP, Draper EA et al. (1991) The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest 100:1619–1636

5. Le Gall JR, Lemeshow S, Saulnier F. (1993) A new simplified acute physiology score (SAPS II) based on a European / North American multicenter study. JAMA 270:2957–2963

6. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. (1993) Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. JAMA 270:2478–2486

7. Castella X, Artigas A, Bion J, Kari A, The European / North American Severity Study Group. (1995) A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. Crit Care Med 23:1327–1335

8. Bertolini G, D'Amico R, Apolone G et al. Predicting outcome in the intensive care unit using scoring systems: is new better? (1998) A comparison of SAPS and SAPS II in a cohort of 1.393 patients. Med Care 36:1371–1382

9. Organization and management of Intensive Care: a prospective study in 12 European countries. Berlin Heidelberg: Springer-Verlag, 1997 (Reis Miranda D, Ryan DW, Schaufeli WB, Fidler V, eds. vol 29)

10. Moreno R, Miranda DR, Matos R, Fevereiro T. (2001) Mortality after discharge from intensive care: the impact of organ system failure and nursing workload use at discharge. Intensive Care Med 27:999–1004

11. Apolone G, D'Amico R, Bertolini G et al. (1996) The performance of SAPS II in a cohort of patients admitted in 99 Italian ICUs: results from the GiViTI. Intensive Care Med 22:1368–1378

12. Moreno R, Morais P. (1997) Outcome prediction in intensive care: results of a prospective, multicentre, Portuguese study. Intensive Care Med 23:177–186

13. Moreno R, Reis Miranda D, Fidler V, Van Schilfgaarde R. (1983) Evaluation of two outcome predictors on an independent database. Crit Care Med 1998;26:50–61

14. Metnitz PG, Valentin A, Vesely H et al. (1999) Prognostic performance and customization of the SAPS II: results of a multicenter Austrian study. Intensive Care Med 25:192–197

15. K Rowan KM, Kerr JH, Major E, McPherson K, Short A, Vessey MP. (1993) Intensive Care Society's APACHE II study in Britain and Ireland—II: Outcome comparisons of intensive care units after adjustment for case mix by the American APACHE II method. Br Med J 307:977–981

16. Bastos PG, Sun X, Wagner DP, Knaus WA, Zimmerman JE, The Brazil APACHE III Study Group. (1996) Application of the APACHE III prognostic system in Brazilian intensive care units: a prospective multicenter study. Intensive Care Med 22:564–570

17. Rivera-Fernandez R, Vazquez-Mata G, Bravo M et al. (1998) The Apache III prognostic system: customized mortality predictions for Spanish ICU patients. Intensive Care Med 24:574–581

18. Zimmerman JE, Wagner DP, Draper EA, Wright L, Alzola C, Knaus WA. (1998) Evaluation of acute physiology and chronic health evaluation III predictions of hospital mortality in an independent database. Crit Care Med 26:1317–1326

19. Popovich MJ. (2002) If most intensive care units are graduating with honors, is it genuine quality or grade inflation? Crit Care Med 30:2145–2146

20. Rowan K. The reliability of case mix measurements in intensive care. (1996) Curr Opin Crit Care 2:209–213

21. Fery-Lemmonier E, Landais P, Kleinknecht D, Brivet F. (1995) Evaluation of severity scoring systems in the ICUs: translation, conversion and definitions ambiguities as a source of inter-observer variability in APACHE II, SAPS, and OSF. Intensive Care Med 21:356–360

22. Zhu B-P, Lemeshow S, Hosmer DW, Klarm J, Avrunin J, Teres D. (1996) Factors affecting the performance of the models in the mortality probability model and strategies of customization: a simulation study. Crit Care Med 24:57–63

23. Knaus WA, Sun X, Nystrom PO, Wagner DP. (1992) Evaluation of definitions for sepsis. Chest 101:656–662

24. Alberti C, Brun-Buisson C, Burchardi H, Martin C, Goodman S, Artigas A, Sicignano A, Palazzo M, Moreno R, Boulme R, Lepage E, Le Gall R. (2002) Epidemiology of sepsis and infection in ICU patients from an international multicentre cohort study. Intensive Care Med; 28:108–121

25. Azoulay E, Alberti C, Legendre I, Buisson CB, Le Gall JR. (2005) Post-ICU mortality in critically ill infected patients: an international study. Intensive Care Med; 31 (1):56–63

26. Fagon J-Y, Chastre J, Novara A, Medioni P, Gilbert C. (1993) Characterization of intensive care unit patients using a model based on the presence or absence of organ dysfunctions and/or infection: the ODIN model. Intensive Care Med 19:137–44

27. Pollack MM, Alexander SR, Clarke N et al. (1990) Improved outcomes from tertiary center pediatric intensive care: a statewide comparison of tertiary and nontertiary care facilities. Crit Care Med 19:150–159

28. Vincent JL, de Mendonca A, Cantraine F et al. (1983) Use of the SOFA score to assess the incidence of organ dysfunction/ failure in intensive care units: results of a multicentric, prospective study. Crit Care Med 1998 26:1793–1800

29. Hosmer DW, Lemeshow S. (1995) Confidence interval estimates of an index of quality performance based on logistic regression models. Statistics in Medicine 14:2161–2172

30. Lemeshow S, Hosmer DW (1982) A review of Goodness-of-fit Statistics for Use in the development of logistic regression models. American-J-Epidemiology 115:92–106

31. Hanley JA, McNeil BJ (1981) The Meaning and Use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36

32. Miranda DR, Ryan DW, Schaufeli WB, Fidler V (Eds.) Organization and Management of Intensive Care. In: Vincent JL (Ed.) Update in Intensive Care Medicine, Vol. 29. Springer-Verlag Berlin-Heidelberg 1998

33. Vincent JL. (1999) Forgoing life support in western European intensive care units: the results of an ethical questionnaire. Crit Care Med. 27:1626–33

34. Metnitz PhGH, Kopp A, Jordan B, Lang Th. (2004) More interventions do not necessarily improve outcome in critically ill patients. Intensive Care Med 30:1586–1593

35. Harrison DA, Lertsithichai P, Brady AR, Carpenter JR, Rowan K. (2004) Winter excess mortality in intensive care in the UK: an analysis of outcome adjusted for patient case mix and unit workload. Intensive Care Med 30:1900–1907

36. Bastos PG, Knaus WA, Zimmerman JE, Magalhaes A Jr, Sun X, Wagner DP. (1996) The importance of technology for achieving superior outcomes from intensive care. Brazil APACHE III Study Group. Intensive Care Med 22:664–669

37. Damiano AM, Bergner M, Draper EA, Knaus WA, Wagner DP. (1992) Reliability of a measure of severity of illness: acute physiology and chronic health evaluation. II. J Clin Epidemiol 45:93–101

38. Chen LM, Martin CM, Morrison TL, Sibbald WJ. (1999) Interobserver variability in data collection of the APACHE II score in teaching and community hospitals. Crit Care Med. 27:1999–2004