Critical Care Medicine
Issue: Volume 23(8), August 1995, pp 1327-1335

Publication Type: [Clinical Investigation]
ISSN: 0090-3493
Accession: 00003246-199508000-00005

[Clinical Investigation]

A comparison of severity of illness scoring systems for intensive care unit patients: Results of a multicenter, multinational study

Castella, Xavier MD; Artigas, Antoni MD; Bion, Julian MBBS MRCP, FRCA, MD; Kari, Aarno MD

Author Information

The European/North American Severity Study Group*; From the Critical Care Department (Dr. Castella), Manresa General Hospital, Barcelona, Spain; the Intensive Care Medicine Service (Dr. Artigas), Hospital de Sabadell, Barcelona, Spain; the Department of Anesthesia and Intensive Care (Dr. Bion), the University of Birmingham, Queen Elizabeth Hospital, Birmingham, UK; and the Department of Intensive Care (Dr. Kari), Kuopio University Hospital, Kuopio, Finland.
This study was supported, in part, by departmental funding from the participating institutions.
*Participants in the European/North American Severity Study are listed in Appendix 1.

Abstract

Objective: To compare the performance of three severity of illness scoring systems used commonly for intensive care unit (ICU) patients in a large international data set. The systems analyzed were versions II and III of the Acute Physiology and Chronic Health Evaluation (APACHE) system, versions I and II of the Simplified Acute Physiology Score (SAPS), and versions I and II of the Mortality Probability Model (MPM), computed at admission and after 24 hrs in the ICU.

Design: A multicenter, multinational cohort study.

Setting: One hundred thirty-seven ICUs in 12 European and North American countries.

Patients: During a 3-month period, 14,745 patients were consecutively admitted to 137 ICUs enrolled in the study.

Interventions: Collection of information necessary to compute the APACHE II and APACHE III scores, SAPS I and SAPS II, and MPM I and MPM II scores. Patients were followed until hospital discharge. Statistical comparison, including indices of calibration (goodness-of-fit) and discrimination (area under the receiver operating characteristic curve).

Measurements and Main Results: Despite having acceptable receiver operating characteristic areas, the older versions of the systems analyzed (APACHE II, SAPS, and MPM I computed at admission- MPM I computed after 24 hrs in the ICU) demonstrated poor calibration for the whole database. The new versions of the systems (SAPS II and MPM II) were superior to their older counterparts. This superiority is reflected by larger receiver operating characteristic areas and better fit. The APACHE III system improved its receiver operating characteristic area compared with the APACHE II system, which showed the best fit of the old systems analyzed.

Conclusions: The new versions of the severity systems analyzed (APACHE III, SAPS II, MPM II) perform

better than their older counterparts (APACHE II, SAPS I, and MPM I). APACHE II, SAPS II, and MPM II show good discrimination and calibration in this international database.

(Crit Care Med 1995; 23:1327-1335)

KEY WORDS: severity of illness index; patient outcome assessment; mortality rate; critical care; intensive care; Acute Physiology and Chronic Health Evaluation; receiver operating characteristic curve

The need for quantitative methodology in evaluating medical activities is becoming increasingly appreciated. Such methods have been accepted widely for use in critical care patients. Systems to analyze severity of disease have been used since 1981 when the Acute Physiology and Chronic Health Evaluation (APACHE) scoring system was introduced [1]. This system was designed using a subjective method to weight several variables determined by a panel of experts to be the most important. Although validation of the system was never well documented and was cumbersome to use, its introduction was a landmark since it represented the first time that a quantitative evaluation of severity of disease was introduced into the intensive care environment. In 1984, Le Gall et al. [2] showed that an abbreviated version of the original APACHE score performed with similar effectiveness. This system, known as the Simplified Acute Physiology Score (SAPS), has been broadly used, especially in France and many European countries. As is true with the APACHE system, the values of the variables chosen to compute the score reflect the worst level in the first 24 hrs in the intensive care unit (ICU). In 1985, a new version of APACHE, known as APACHE II, was published [3]. Although the APACHE II system also used an abbreviated list of physiologic variables, it required that a single primary diagnosis be selected from a list. The APACHE II system used a logistic regression equation to compute the probability of death for each patient. Since then, APACHE II has been the most broadly used system in the world.

Using a completely different approach, based on the selection of the most influential variables by logistic regression, Teres et al. [4] developed a model in 1982 for estimating the probability of mortality for ICU patients. This model eventually evolved into the Mortality Probability Model (MPM) in 1985 [5]. These models were validated and refined in 1988 [6] and are now being used routinely in many hospitals. The system was built upon three unique models--one used at the time of admission to the ICU, another at 24 hrs, and a third at 48 hrs of ICU stay. An over-time model was also developed [5]. As was true for SAPS, the MPM system did not require the selection of a primary diagnosis.

Recently, new versions of the three systems have been released. APACHE III is similar in concept to APACHE II [7,8]. Logistic regression was used to select the ranges of the variables and to allocate point assignments to the various ranges. Additional information was added to the system with a lengthening of the list of primary diagnoses and location of the patient before ICU admission [9]. At the time of this study, the APACHE III model could be used only in conjunction with a computerized system that must be purchased through APACHE Medical Systems, Washington, DC. The new version of SAPS, SAPS II [10], also used logistic regression to select the variables, its ranges, its point assignments, and the algorithm to compute the probability of death. Using the same methodology, the new version of MPM, MPM II, was published in 1993 [11] and completed with additional equations for 48 and 72 hrs [12].

Comparison of the models in the same population, using state of the art statistical methods, has been considered imperative [13]. However, due to the complexity of carrying out such a study, direct comparison of the performance of the systems has been uncommon [14-16]. The first European and North American study on severity of disease of intensive care patients was designed with the main objective of setting up a large database for the purpose of comparing the different systems. A large portion of the database assembled in this study was also used as the developmental sample for the SAPS II and MPM II models [10,11].

This article presents the results of the first comparison of the systems currently in use in the ICU environment in a large population of patients from 12 countries during the same study period. We include an analysis of the comparison of the performance of the older versions of the systems (APACHE II, SAPS, MPM), a comparison of the performance of the old vs. the new versions of each of the systems (APACHE II vs. APACHE III; SAPS I vs. SAPS II; MPM I vs. MPM II), and a comparison of the new systems among each other.

MATERIALS AND METHODS

Organization of the International Study.

One hundred thirty-seven ICUs in 12 countries volunteered to participate in the study Appendix 1 (Table 5, Table 6, and Table 7). A coordinator was appointed in every country to centralize the information collected, check its accuracy, and mail the assembled data every 2 wks to the University of Massachusetts. Data were then checked for accuracy and completeness and compiled into a large database. Data were entered onto paper forms and into a specially designed computer program that checked for out-of-range values, and was available in both IBM and Macintosh formats. Data collection included all the variables necessary to complete the APACHE II, APACHE III score, MPM I or SAPS I models, along with information for developing the new versions of SAPS and MPM. A comprehensive operations manual, describing study procedures, data collection requirements, and variable definitions, was distributed to each participating institution. Continuous support was provided during the study by telephone, fax, and electronic mail.

Country/Country Coordinator	Participating Hospital	City	Data Collectors
Austria			
Prof. Dr. med. H. Burchardi			
	University Hospital	Innsbruck	H. Benzer, C. Huber
Belgium			
Dr. J.P. Alexander			
	A. Z. Middelheim	Antwerpen	J.P. Alexander, M. Delande
	C.H.U. de Liège	Liège	D. Ledoux, J.L.Canivet, P. Damas
	A.Z. Stuivenberg	Antwerpen	I. Demeyer, K. Vissers
	Clin. St. Pierre	Ottignies	Th. Dugernier
	A. Z. VUB	Brussel	L. Huyghens, M. Diltour, N. De Wit
	A. Z. Middelheim	Antwerpen	J. Nagler, F. Cools
	O.L.V. Ziekenhuis	Aalst	G. Nollet, J. Verbeke
	U.Z. Gent	Gent	J. Poelaert, f. Collardyn
	Clin. Univ. St. Luc	Bruxelles	P.F. Laterre, A. Dougnac, M. Reynaert
	St. Vincentius	Antwerpen	R. Rutsaert, L. Colemont
	U.Z. Gasthuisberg	Leuven	M. Schetz, P. Lauwers
Canada			
Dr. D. Teres			
	University of Alberta Hospitals	Edmonton	S. Hamilton, C. Norris
	Royal Alexandra Hospital	Edmonton	A. Shustack, R. Johnston, E. Konopad
Finland			
Dr. A. Kari			
	Central Hospital of North Karelia	Joensuu	P. Hannonen
	Central Hospital of Central Finland	Jyväskylä	K. Hersio
	Vaasa Central Hospital	Vaasa	P. Kairi
	Kuopio University Hospital	Kuopio	A. Kari, M. Niskanen
	Turku University Central Hospital	Turku	J. Klossner
	Oulu University Central Hospital	Oulu	E. Saarela
	Central Hospital of Southern Saimaa	Lappeenranta	M. Vähämurto
France			
Dr. F. Saulnier			
	Centre Hospitalier Regional de Nimes	Nimes	C. Arice
	Centre Hospitalier d'Annonay	Annonay	B. Bedocq
	Hôpital General	Dijon	B. Blettery
	Hôpital Saint Joseph	Paris	B. Misset, J. Carlet
	Hôpital Louis Mourier	Colombes	L. Mier, D. Dreyfuss
	Hôpital Avicenne	Bobigny	J.P. Fosse
	Centre Hospitalier Morvan	Brest	B. Garo
	Centre Hospitalier de Bourg en Bresse	Bourg en Bresse	G. Demingeon, L. Holzhapfel
	Pavillon Pasteur - CHU	Strasbourg	J. Kopferschmitt
	CHU St. Etienne Nord	St. Priest en Jarez	P. Mahul
	Institut Gustave-Goussy	Villejuif	G. Nitemberg
	Centre Hospitalier d'Agen	Agen	F. Plouvier
	Hôpital Albert Calmette	Lille	F. Saulnier
	CHU - Hôtel Dieu	Nantes	D. Villers
Germany			
Prof. Dr. med. H. Burchardi			
	University Hospital	Ulm	H. Wiedeck
	University Hospital	Göttingen	H. Burchardi, H. Klingler
	University Hospital	Mainz	W. Dick, F. Brost
	Zentralklinikum	Augsburg	J. Eckart, P. Wengert
	University Hospital	Freiburg/Br.	K. Gelger, K. Armbruster
	Zentralkrankenhaus St. Jürgenstrasse	Bremen	H.-D. Kamp, M. Rothe
	Städtisches Krankenhaus München-Bogenhausen	München	B. Landauer, T.-O. Schmid
	University Hospital Grosshadern	München	K. Peter, H Forst
	Klinikum der Landeshaupstadt	Wiesbaden	C. Piper
	University Hospital Steglitz	Berlin	K. Reinhart, T. Rudolph

Table 5. Appendix 1. The European/North American Severity Study Group: Participating hospitals, country coordinators and data collectors

Country/Country Coordinator	Participating Hospital	City	Data Collectors
Germany (cont'd)			
Prof. Dr. med. H. Burchardi			
	University Nürnberg - Erlangen	Erlangen	E. Rügheimer, E. Pscheidl
	Klinikum der Landeshaupstadt	Wiesbaden	J. Schmitz
	Städtisches Krankenhaus	Hildesheim	H.-P. Schuster, K.F. Bodmann
	Klinikum Friedrichshain	Berlin	D. Stober, C. Dressler
Italy			
Dr. G. Iapichino			
	IRCCS Policlinico S. Matteo	Pavia	F. Albertario, L. Carnevale
	Ospedale Civile	Vimercate	F. Bassi, B.M. Graziani
	IRCCS Policlinico S. Matteo	Pavia	F. Bobbio Pallavicini, C. Cassini
	Ospedale Policlinico	Bari	G. Cinella, A. Brienza
	Policlinico Umberto I	Roma	G. Conti, A. de Blasi
	Ospedale Maggiore	Bologna	M.T. Fiandri, D. Cosco
	Nuovo Ospedale S. Gerardo	Monza	R. Fumagalli, L. Avalli
	Ospedale S. Raffaele	Milano	D. Giudici, G. Gallioli
	Arcispedale S. Anna	Ferrara	M. Capuzzo, R. Ragazzi, C.A. Volta
	Ospedale Civile	Melegnano	A. Guarino, G. Merli
	Ospedale di Cattinara	Trieste	L. Serra, S. Fasiolo
	Ospedale Maggiore	Milano	G. Iapichino, S. Rotelli
	Ospedale S. Bortolo	Vicenza	L. Lacquaniti, T. Moretti
	Policlinico S. Orsola	Bologna	R. Melotti, G. Negro
	Ospedale Maggiore	Parma	M. Mergoni, A. Saccani
	Ospedale Niguarda Ca'Granda	Milano	A. Ravizza, G. Casella
	Ospedale S. Carlo Borromeo	Milano	D. Ripamonti, A. Favero
	Ospedale S. M. Battuti	Treviso	G. Simini, A. Manuali
	Ospedale Regionale	Aosta	S. Vernero, A. Viale
	Ospedale S. Paolo	Milano	S. Vesconi, A. Sicignano
The Netherlands			
Dr. D. Reis Miranda			
	Academisch Ziekenhuis	Rotterdam	H.A. Bruining
	Ziekenhuis Leyenburg	The Hague	J. de Haas
	Academisch Ziekenhuis Vrije Universiteit	Amsterdam	D. de Jong
	Medisch Centrum Alkmaar	Alkmaar	M. de Jong
	Scheperziekenhuis	Emmen	W.P. Haanstra
	Academisch Ziekenhuis	Utrecht	P.F. Hulstaert
	Zuiderziekenhuis	Rotterdam	R. Jairam
	Academisch Ziekenhuis	Nijmegen	R. van Dalen
	Academisch Ziekenhuis	Maastricht	S. van der Geest
	Twenteborg Ziekenhuis	Almelo	A.J.J. Woittiez
	Academisch Ziekenhuis	Groningen	J.H. Zwaveling
Spain			
Dr. A. Artigas			
	H. Son Dureta	Palma Mallorca	R. Abizanda, B. Balerdí, Ll. Socías
	H. de Sabadell, Parc Taulí	Sabadell	A. Artigas, X. Castella, J. Mestre
	H. Josep Trueta	Girona	A. Bonet, A Alvarez
	H. de la Vall d'Hebró	Barcelona	J.L. Bóveda, I. Manzanares, I. Salgado
	H. de Barcelona	Barcelona	L. Cabré, G. Carrasco
	H. Comarcal d'Igualada	Igualada	M. Casanovas, E. Faraidun, J.M. Bausili
	H. de la Creu Roja d'Hospitalet	Hospitalet de Llobregat	M. Cerdà, M. Ibars, C. Gimeno
	C. S. de Bellvitge	Hospitalet de Llobregat	A. Díaz Prieto, H. Torrado
	H. de Tarragona "Joan XXIII"	Tarragona	J.J. Guardiola, C Boqué
	C. H. Unitat Coronària de Manresa	Manresa	M. Guirado-Alaíz
	H. de Lleida Arnau de Vilanova	Lleida	F. Iturbe, C. Barberà, C. Rabasso
	H. Mútua de Terrassa	Terrassa	J. M. Nava Caballero
	H. de la Sta. Creu i St. Pau	Barcelona	A. Roglán, A. Net
	H. de l'Alianza	Barcelona	J. Ruíz, L. Garcia
	H. de Badalona, G. Trias i Pujol	Badalona	X. Sarmiento, J.M. Toboso
	H. del Mar	Barcelona	J. Solsona, A. Vazquez
	H. General de Catalunya	Sant Cugat del Vallés (Barcelona)	M. Nolla

Table 6. Appendix 1 continued

Country/Country Coordinator	Participating Hospital	City	Data Collectors
Switzerland			
Dr. A. de Torrenté			
	University Hospital	Zürich	P.C. Baumann, H.-M. Vonwiller
	Regional Hospital "La Carita"	Locarno	G. Domenighetti, D. Erba
	University Hospital	Genève	J.C. Chevrolet, Ph. Jolliet
	University Hospital	Lausanne	R. Chioléro, A. Messikommer
	City Hospital "Les Cadolles"	Neuchâtel	J.F. Enrico, R. Kehtari
	Community Hospital	La Chaux-de-Fonds	A. de Torrenté, A. Kocher
	"Bürger" Hospital	Solothurn	G. Lupi
	Canton Hospital	Chur	A. Frutiger, M. Reigner
	University Hospital	Basel	R. Ritz, S. Durrer
	University Hospital	Lausanne	C.H. Perret, M.D. Schaller
	University Hospital	Genève	P. Suter, B. Ricou
United Kingdom			
Dr. J. Bion			
	Queen Elizabeth Hospital	Birmingham	J. Bion, M. Bowden
	John Radcliffe Hospital	Oxford	C. Garrard
	Southampton General Hospital	Southampton	B. Randalls
	Royal Devon and Exeter Hospital	Exeter	I. Wilson
United States			
Dr. D. Teres			
	St. Michael's Medical Center	Newark, NJ	M. Adelman, R.A. Miller, B. Quinones
	Oregon Health Sciences University	Portland, OR	C.L. Baer, J. Schwamacher, L. Renner
	Maine Medical Center	Portland, ME	P.M. Cox, S. Prato
	Albany Medical Center	Albany, NY	I.A. Fein, A. Veeder
	Dartmouth-Hitchcock Medical Center	Hanover, NH	A. Gettinger, K. Holmes
	St. Vincent's Hospital & Medical Center	New York, NY	M.E. Astiz, J. Saxon, G. DeGent
	East Pasco Medical Center	Zephyrhills, FL	L. Grossbard, R. Ruchti
	Hermann Hospital	Houston, TX	G. Gutierrez, C. Clark, J. Witherspoon
	South Shore Hospital	South Weymouth, MA	F. Harris, M. Higgins
	SUNY Health Science Center	Syracuse, NY	M.S. Jastremski, A. Milewski, K. Bunch
	Mercy Hospital	Springfield, MA	G. Karras, C. Barghoud, N. Richard
	Buffalo General Hospital	Buffalo, NY	F.V.McL. Booth, R. Kerins, J. Booth
	St. Vincent Hospital	Worcester, MA	S.A. Nasraway, F.D. Sottile, P. Sigel
	The Genesee Hospital	Rochester, NY	C.R. Ortiz, J. Cromiller
	St. Elizabeth's Hospital	Boston, MA	K.A. Porter
	St. Francis Medical Center	Pittsburgh, PA	H. Rafkin, S. Ermakov
	Medical College of Virginia	Richmond, VA	S. Retchin, H.D. Reines, M. Casado
	Riverside Methodist Hospital	Columbia, OH	H. Rogove, S. Morrow, K. Chupka, E. Foster
	New England Medical Center	Boston, MA	S.D. Schwaitzberg, J. Hayes, J. Scaramuzzi
	Geisinger Medical Center	Danville, PA	J.L. Smith, R. Burns, D. Hammaker
	Baystate Medical Center	Springfield, MA	D. Teres, C. Desrosiers, A. Moineau
	Mercy Hospital	Pittsburgh, PA	D. Thompson, M.E. Sipperly
	Akron General Medical Center	Akron, OH	D. Heiselman, T. Hofer, R. Vidovich
	Lahey Clinic Medical Center	Burlington, MA	J.M. O'Donnell, A. Gray, F.G. Davis
	Hermann Hospital	Houston, TX	A.S. Tonnesen, L.S. Cronin, C. Jennings

Table 7. Appendix 1 continued

From September 30, 1991 to December 27, 1991, all consecutive admissions >or=to18 yrs of age were enrolled in the study. Burn, coronary care, and cardiac surgery patients were excluded from statistical analyses. Patients were followed to hospital discharge, and their survival status was registered. Any patient still in the hospital on February 28, 1992 was dropped from the study.

Interobserver quality control was assessed by having each country coordinator complete a second set of forms for a 5% random sample of every ICU's patients. The original and quality control forms were compared and discrepancy was evaluated using the kappa statistic [17].

Data Analysis.

Probability of death for every patient in the database was calculated, using published coefficients and equations along with the description of each ICU model [3,5,10,11]. Performance of every system was analyzed, using indices of calibration and discrimination according to state of the art recommendations [13].

Calibration refers to the ability of a model to describe the mortality pattern in the data and is assessed using formal goodness-of-fit testing [18]. However, it is applicable only to methods producing probabilities of mortality. When the mortality predicted by a model differs significantly from the observed pattern, this model does not calibrate well and the goodness-of-fit statistics are highly significant. These parameters (C and H) result from two different strategies for grouping the estimated probabilities of mortality [18].

Discrimination refers to the ability of the model to separate those patients predicted to live from those patients predicted to die and is measured using the area under the receiving operating characteristic curve [19]. Tossing a coin to classify patients as dead or alive would produce an area under the receiver operating characteristic curve of 0.50. A model is considered to discriminate well when this area is >0.8; as a general rule, the bigger the area the better the discriminatory capability of the model. This method is available for scores and probabilities but is only meaningful once the model has been shown to calibrate well. Methods to compare areas under the receiver operating characteristic curve correcting for the degree of correlation among observations have been described [20]. We decided not to use statistics based on 2 times 2 tables as measures of performance of the model because they convert probabilities to dichotomous values. We believe that with this approach, a considerable amount of information is lost and resulting conclusions can be misleading.

Using random numbers, the whole database was split into a developmental and a validation data set. In total, 65% of cases were assigned to the developmental sample, and 35% to the validation sample. SAPS II and MPM II were developed and validated, using these data sets according to standard statistical criteria [21]. We were able to use the whole database when we compared the old severity systems. However, any analysis including any of the new systems was restricted to the validation subsample of this European/North American data-base. The methodologic approach used in the fragmentation and analysis of different subsamples is depicted in Figure 1.

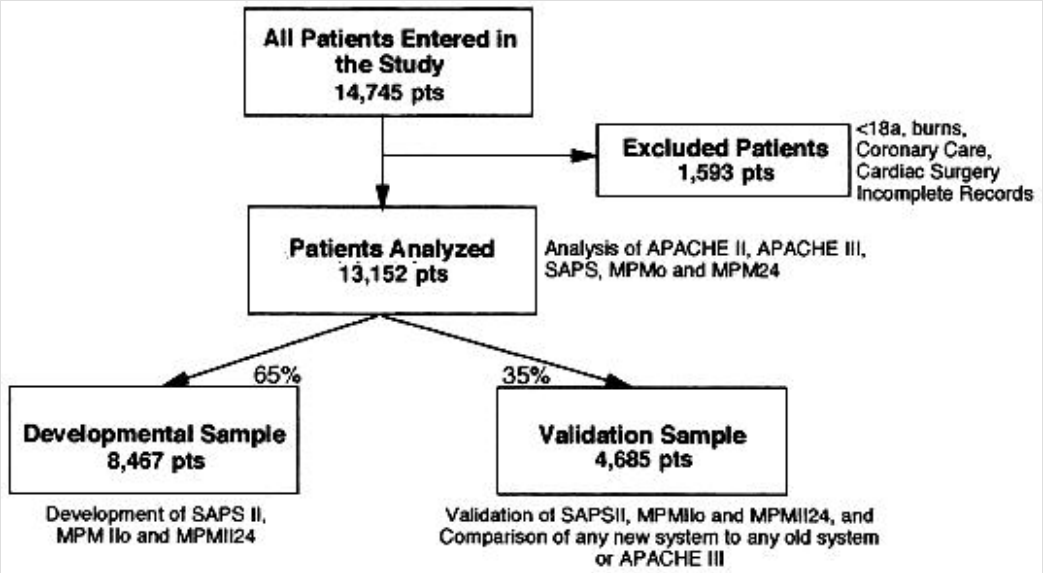

Figure 1. Study design. pts, patients; APACHE II-III, Acute Physiology and Chronic Health Evaluation; SAPS-SAPS II, Simplified Acute Physiology Score; MPMo-MPMIIo, Mortality Probability Model at intensive care unit (ICU) admission; MPM24-MPMII24, Mortality Probability Model at 24 hrs after ICU admission.

Although a general description of the APACHE III system has been published [7,8] and the distribution of points for the acute physiology part of the system is available, the coefficients attributed to each diagnosis were proprietary at the time of the study. As a consequence, we were able to compute APACHE III scores but not probability of death, using this system.

RESULTS

Data were collected on 14,745 patients in 137 ICUs in 12 countries. Excluded from the study were 1,593 patients because of incomplete records or the presence of one or more exclusion criteria. This exclusion left 13,152 valid cases for data analysis. General characteristics of these patients by country are presented in Table 1. The developmental sample included 8,467 patients, leaving 4,685 patients for validating and comparing the systems Figure 1. Neither subsample differed significantly from the other in any of the variables used in the models or diagnostic grouping.

	Hospital Mortality			Sex (%)		Type of Admission (%)			Length of Stay (Days)	
Country	No. of Patients	Rate (%)	Age (yr)	Male	Female	A	B	C	In ICU	In Hospital[a]
Belgium	1,091	21.7	57.5 ± 18.2	62.1	37.9	15.2	33.5	42.8	6.2 ± 9.8	21.5 ± 21.7
Finland	720	17.6	56.3 ± 17.4	59.0	41.0	25.8	28.2	46.0	4.1 ± 5.5	14.0 ± 13.7
France	1,393	28.9	56.9 ± 19.1	61.5	38.5	12.0	8.8	79.3	9.7 ± 12.4	18.9 ± 20.2
Germany/Austria	1,807	15.7	56.6 ± 18.2	58.6	41.4	28.1	50.9	21.1	6.0 ± 8.7	21.0 ± 18.5
Italy	1,297	31.3	58.1 ± 18.3	62.6	37.4	23.2	33.2	43.6	7.2 ± 10.2	20.5 ± 19.3
Spain	1,270	27.1	54.9 ± 18.3	62.9	37.0	18.3	26.1	55.2	9.5 ± 12.2	22.8 ± 21.2
Switzerland	756	13.8	54.9 ± 18.7	62.2	37.8	15.5	21.7	62.8	4.9 ± 5.6	17.6 ± 17.1
The Netherlands	950	20.0	60.3 ± 17.1	63.3	36.7	25.4	45.8	28.7	5.5 ± 8.3	19.3 ± 15.9
United Kingdom	136	32.4	57.4 ± 18.7	61.8	38.2	27.9	24.3	47.8	5.7 ± 7.1	14.8 ± 15.9
USA/Canada	3,732	19.7	57.9 ± 19.0	55.1	44.9	16.9	29.5	53.5	5.9 ± 8.4	17.1 ± 18.2
Total	13,152	21.8	57.2 ± 18.5	59.6	40.4	19.6	31.2	48.4	6.6 ± 9.5	19.1 ± 18.9

A, Emergency surgical admission; B, scheduled surgical admission; C, medical; ICU, intensive care unit; USA, United States of America.
[a]Denotes days in hospital from beginning of ICU stay.

Table 1. General characteristics of patients in the study (mean +/- SD)

We analyzed the old versions of SAPS, MPM computed at admission, MPM computed after 24 hrs in the ICU, and the APACHE II, using a common sample for which complete information for the four systems was available for the same individual subjects (12,802 patients in the whole database and 4,101 in the validation sample) Table 2.

Model	Sample	ROC Area	GOF (C)	GOF (H)
APACHE II probability	Population	0.852	<0.0001	<0.0001
	Validation	0.857	0.0245	0.0074
MPM 0	Population	0.773	<0.0001	<0.0001
	Validation	0.778	<0.0001	<0.0001
MPM 24	Population	0.825	<0.0001	<0.0001
	Validation	0.816	<0.0001	<0.0001
SAPS score	Population	0.798	NA	NA
	Validation	0.799	NA	NA

ROC, receiver operating characteristic; GOF, goodness-of-fit; for definition of C and H, see reference 18; APACHE II, Acute Physiology and Chronic Health Evaluation II; MPM, Mortality Probability Model; MPM 0, Mortality Prediction Model at admission; MPM 24, Mortality Prediction Model at 24 hrs after admission; SAPS, Simplified Acute Physiology Score; NA, not applicable.

Table 2. Comparison of the old models (population, n = 12,802; Validation, n = 4,101)

It was possible to evaluate the performance of the APACHE II score, the APACHE II probability of death, and the discrimination of the APACHE III score, using those patients in the whole database who had all three measures available for the same individual subjects (12,899 patients) Table 3. When we tested the improvement in the receiver operating characteristic curve area of APACHE III compared with APACHE II, it was highly significant ($p < .0001$). APACHE II and APACHE III scores were highly correlated ($r^2 = .827$).

	Sample	ROC Area	GOF (C)	GOF (H)
APACHE System (n = 12,899)				
APACHE II probability	Whole population	0.853	<0.0001	<0.0001
APACHE II score	Whole population	0.848	NAp	NAp
APACHE III score	Whole population	0.866	NAv	NAv
MPM 0 System (n = 4,605)				
MPM 0	Validation	0.766	<0.0001	<0.0001
MPM II 0	Validation	0.805	0.0694	0.0143
MPM 24 System (n = 4,101)				
MPM 24	Validation	0.815	<0.0001	<0.0001
MPM II 24	Validation	0.833	0.1531	0.0247
SAPS System (n = 4,605)				
SAPS score	Validation	0.784	NAp	NAp
SAPS II score	Validation	0.855	NAp	NAp
SAPS II probability	Validation	0.847	0.0244	0.1019

ROC, receiver operating characteristic; GOF, goodness-of-fit; for definition of C and H, see reference 18; APACHE, Acute Physiology and Chronic Health Evaluation; MPM, Mortality Probability Model; MPM 0, Mortality Prediction Model at admission; MPM 24, Mortality Prediction Model 24 hrs after admission; SAPS, Simplified Acute Physiology Score; NAp, not applicable; NAv, not available.

Table 3. Comparison of the old vs. the new systems

The performance of the old vs. the new MPM admission models was analyzed, using a common data set in the validation sample only, with complete information for both systems in the same subjects (4,605 patients) Table 3. The calibration of the new MPM admission model was superior to that observed for the older version. Area under the receiver operating characteristic curve for MPM II computed at admission was also higher (81%) than was the case for the old MPM I computed at admission (77%) ($p < .0001$). The correlation coefficient between MPM I computed at admission and MPM II was $r^2 = .583$.

The performance of MPM I after 24 hrs in the ICU vs. MPM II after 24 hrs in the ICU was compared using only validation set patients who had both of these probabilities evaluated (4,101 patients) Table 3. The area under the receiver operating characteristic curve for MPM II after 24 hrs in the ICU (83%) was higher than for MPM I after 24 hrs in the ICU (82%) ($p = .0001$) and the correlation coefficient between MPM I after 24 hrs in the ICU and MPM II after 24 hrs in the ICU was $r^2 = .795$.

Since probabilities of mortality were unavailable for the original SAPS system, calibration could only be assessed for the SAPS II probabilities. Discrimination could be assessed both for the original SAPS and the SAPS II probability. Using only validation sample cases with information for the two systems (4,605 patients), the SAPS II model demonstrated acceptable fit ($p = .02$) and showed an increase in its discriminatory capability (85%) when compared with the old SAPS (78%) ($p < .0001$) Table 3. Correlation of SAPS and SAPS II scores was $r^2 = .636$.

The performance of the new systems, APACHE III, SAPS II, MPM II at admission, and MPM II after 24 hrs in the ICU probabilities of death, were compared using that subset of the patients in the validation sample on whom information for all four systems was available for the same subjects Table 4. Each of the new systems for which we could calculate probabilities calibrated acceptably well. The discriminatory performance of each of these systems is good, as reflected by the areas under the receiver operating characteristic curves.

	Sample	ROC Area	GOF (C)	GOF (H)
MPM II 0	Validation	0.805	0.0720	0.0148
MPM II 24	Validation	0.833	0.0932	0.0255
SAPS II probability	Validation	0.847	0.0244	0.1019
APACHE III score	Validation	0.861	NAv	NAv

ROC, receiver operating characteristic; GOF, goodness-of-fit; for definition of C and H, see reference 18; MPM, Mortality Probability Model; MPM 0, Mortality Prediction Model at admission; MPM 24, Mortality Prediction Model 24 hrs after admission; SAPS, Simplified Acute Physiology Score; APACHE, Acute Physiology and Chronic Health Evaluation; NAv, not available.

Table 4. Comparison of the new models (n = 4,099)

DISCUSSION

Mortality prediction models are routinely used in many ICUs all over the world and they have been used to compare critical care in different countries [22-25] and even to study ICU outcomes between ethnic groups [26]. However, since patient characteristics and medical practice differ from country to country and over time, exportation of one system developed in one time and one place to other settings should always be preceded by formal goodness-of-fit testing. Such testing ensures that a given system fits the mortality pattern observed in the country or hospital analyzed. Once calibration is appropriate, discrimination of the model can be analyzed.

The study on which the results presented in this paper were based included 12 developed countries in which societies and medical practices could be considered similar enough to analyze them as a whole. Defining reference models based on large international databases such as the one used in this report is extremely important since they represent the standard to which any local analysis should refer to for the purpose of quality comparisons.

The health status of a society, socioeconomic variables, and medical practices change with time. As a result, it is not surprising that any system estimating probability of mortality based on data and technology that were appropriate years ago would need updating. Despite showing acceptable discrimination, none of the old systems (APACHE II, MPM I at admission, or MPM after 24 hrs in the ICU) calibrated well in the whole international database and only APACHE II provided acceptable performance in the validation sample. Given the huge number of patients studied in the whole population, it is not surprising that the p values for the goodness-of-fit tests were so significant. The clinical significance of these large discrepancies is probably less impressive than their statistical implications. Nevertheless, any improvement in the fit of the predictive ability of a system to the mortality observed in the data set has to be considered indisputable.

When comparing the new models to the old ones, improvement in performance was consistently observed. To the best of our knowledge, this is the first research paper that formally proves that validity of severity models change over time and that severity systems can be improved, using statistical techniques and updated databases.

We were particularly careful to compare systems only in the validation data set since none of the patients entered in these analyses was used in model development. The new models developed using this international database (SAPS II, MPM II) validate well in these 12 countries taken as a unit. However, when we move from group summary to individual patient prediction, every country (or perhaps every hospital or ICU) must be assured that the model they intend to use fits their own data, since only a perfectly calibrated model would produce meaningful and accurate estimated probabilities. Even with all this information available, one should be extremely cautious in its use for individual prediction [27,28].

Our study shows that the new systems represent real improvement in severity model performance. However, none of them stand out as being clearly superior to the others, and all of them can be used with considerable reliability. Whatever the model chosen, it is essential for every individual user to know the capabilities and limitations of that system, its goodness-of-fit in their area of application (country/hospital /ICU), as well as its discriminatory capability.

REFERENCES

1. Knaus WA, Zimmerman JE, Wagner DP, et al: APACHE--Acute Physiology and Chronic Health Evaluation: A physiologically based classification system. Crit Care Med 1981; 9:591-597 SFX | Bibliographic Links | [Context Link]

2. Le Gall J, Loirat P, Alperovith A, et al: A simplified acute physiology score for ICU patients. Crit Care Med 1984; 12:975-977 SFX | Bibliographic Links | [Context Link]

3. Knaus WA, Draper EA, Wagner DP, et al: APACHE II: A severity of disease classification system. Crit Care Med 1985; 13:818-829 SFX | Bibliographic Links | [Context Link]

4. Teres D, Brown RB, Lemeshow S: Predicting mortality of intensive care unit patients. The importance of coma. Crit Care Med 1982; 10:86-95 SFX | Bibliographic Links | [Context Link]

5. Lemeshow S, Teres D, Pastides H, et al: A method for predicting survival and mortality of ICU patients using objectively derived weights. Crit Care Med 1985; 13:519-525 SFX | Bibliographic Links | [Context

Link]

6. Lemeshow S, Teres D, Avrunin J, et al: Refining intensive care unit outcome prediction by using changing probabilities of mortality. Crit Care Med 1988; 16:470-477 SFX | Bibliographic Links | [Context Link]

7. Knaus W, Wagner D, Draper E, et al: The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized patients. Chest 1991; 100:1619-1636 SFX | Bibliographic Links | [Context Link]

8. Knaus W, Wagner D, Lynn J: Short-term mortality predictions for critically ill hospitalized adults: Science and ethics. Science 1991; 254:389-394 SFX | Bibliographic Links | [Context Link]

9. Zimmerman J, Wagner D, Draper E, et al: Improving intensive care unit discharge decisions: Supplementing physician judgment with predictions of next day risk for life support. Crit Care Med 1994; 22:1373-1384 SFX | Bibliographic Links | [Context Link]

10. 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 [Context Link]

11. Lemeshow S, Teres D, Klar J, et al: Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. JAMA 1993; 270:2478-2486 SFX | Full Text | Bibliographic Links | [Context Link]

12. Lemeshow S, Klar J, Teres D, et al: Mortality probability models for patients in the intensive care unit for 48 or 72 hours: A prospective, multicenter study. Crit Care Med 1994; 22:1351-1358 SFX | Bibliographic Links | [Context Link]

13. Hadorn D, Keeler E, Rogers W, et al: Assessing the Performance of Mortality Prediction Models. Santa Monica, CA, RAND/UCLA/Harvard Center for Health Care Financing Policy Research, 1993 [Context Link]

14. Lemeshow S, Teres D, Avrunin JS, et al: A comparison of methods to predict mortality of intensive care unit patients. Crit Care Med 1987; 15:715-722 SFX | Bibliographic Links | [Context Link]

15. Castella X, Gilabert J, Torner F, et al: Mortality prediction models in intensive care: Acute Physiology and Chronic Health Evaluation II and Mortality Prediction Model compared. Crit Care Med 1991; 19:191-197 SFX | Bibliographic Links | [Context Link]

16. Rowan K, Kerr J, Major E, et al: Intensive Care Society's Acute Physiology and Chronic Health Evaluation (APACHE II) study in Britain and Ireland: A prospective, multicenter, cohort study comparing two methods for predicting outcome for adult intensive care patients. Crit Care Med 1994; 22:1392-1401 SFX | Bibliographic Links | [Context Link]

17. Cohen J: A coefficient of agreement for nominal scales. Educ Psychol Measur 1960; 20:37-46 [Context Link]

18. Lemeshow S, Hosmer DJ: A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 1982; 115:92-106 [Context Link]

19. Hanley J, McNeil B: The meaning and use of the area under a Receiver Operating Characteristic (ROC) Curve. Radiology 1982; 143:29-36 SFX | Bibliographic Links | [Context Link]

20. Hanley J, McNeil B: A method of comparing the areas under Receiver Operating Characteristic Curves derived from the same cases. Radiology 1983; 148:839-843 SFX | Bibliographic Links | [Context Link]

21. Hosmer D, Lemeshow S: Model-building strategies and methods for logistic regression. In: Applied Logistic Regression. Hosmer D, Lemeshow S (Eds). New York, John Wiley & Sons, 1989, pp 82-134 [Context Link]

22. Teik O, Hutchinson R, Short S, et al: Verification of the Acute Physiology and Chronic Health Evaluation scoring system in a Hong Kong intensive care unit. Crit Care Med 1993; 21:698-705 [Context Link]

23. Sirio C, Tajimi K, Tase C, et al: An initial comparison of intensive care in Japan and the United States. Crit Care Med 1992; 20:1207-1215 SFX | Bibliographic Links | [Context Link]

24. Zimmerman JE, Knaus WA, Judson JA, et al: Patient selection for intensive care: A comparison of New Zealand and United States hospitals. Crit Care Med 1988; 16:318-326 SFX | Bibliographic Links | [Context Link]

25. Knaus WA, Le GJ, Wagner DP, et al: A comparison of intensive care in the USA and France. Lancet 1982; ii:642-646 [Context Link]

26. Williams JF, Zimmerman JE, Wagner DP, et al: African-American and white patients admitted to the intensive care unit: Is there a difference in therapy and outcome? Crit Care Med 1995; 23:626-636 Ovid Full Text | SFX | Bibliographic Links | [Context Link]

27. Wagner D, Knaus W, Harrell F, et al: 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 SFX | Bibliographic Links | [Context Link]

28. Rogers J, Fuller H: Use of daily Acute Physiology and Chronic Health Evaluation (APACHE) II scores to predict individual patient survival rate. Crit Care Med 1994; 22:1402-1405 SFX | Bibliographic Links | [Context Link]

IMAGE GALLERY

Select All

Export Selected to PowerPoint

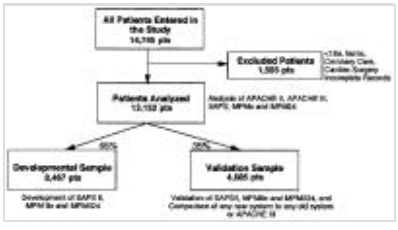

Figure 1

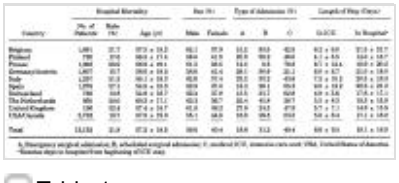

Table 1

Model	Sample	ROC Area	GOF (C)	GOF (H)
APACHE II probability	Population	0.852	<0.0001	<0.0001
	Validation	0.857	0.0245	0.0074
MPM 0	Population	0.773	<0.0001	<0.0001
	Validation	0.778	<0.0001	<0.0001
MPM 24	Population	0.825	<0.0001	<0.0001
	Validation	0.816	<0.0001	<0.0001
SAPS score	Population	0.708	NA	NA
	Validation	0.709	NA	NA

ROC, receiver operating characteristic; GOF, goodness-of-fit; for definition of C and H, see reference 18; APACHE II, Acute Physiology and Chronic Health Evaluation II; MPM, Mortality Probability Model; MPM 0, Mortality Prediction Model at admission; MPM 24, Mortality Prediction Model at 24 hrs after admission; SAPS, Simplified Acute Physiology Score; NA, not applicable.

Table 2

	Sample	ROC Area	GOF (C)	GOF (H)
APACHE System (n = 12,899)				
APACHE II probability	Whole population	0.853	<0.0001	<0.0001
APACHE II score	Whole population	0.848	NAp	NAp
APACHE III score	Whole population	0.866	NAv	NAv
MPM 0 System (n = 4,805)				
MPM 0	Validation	0.766	<0.0001	<0.0001
MPM II 0	Validation	0.805	0.0894	0.0143
MPM 24 System (n = 4,101)				
MPM 24	Validation	0.815	<0.0001	<0.0001
MPM II 24	Validation	0.833	0.1531	0.0247
SAPS System (n = 4,605)				
SAPS score	Validation	0.784	NAp	NAp
SAPS II score	Validation	0.855	NAp	NAp
SAPS II probability	Validation	0.847	0.0244	0.1019

ROC, receiver operating characteristic; GOF, goodness-of-fit; for definition of C and H, see reference 18; APACHE, Acute Physiology and Chronic Health Evaluation; MPM, Mortality Probability Model; MPM 0, Mortality Prediction Model at admission; MPM 24, Mortality Prediction Model 24 hrs after admission; SAPS, Simplified Acute Physiology Score; NAp, not applicable; NAv, not available.

Table 3

	Sample	ROC Area	GOF (C)	GOF (H)
MPM II 0	Validation	0.805	0.0720	0.0148
MPM II 24	Validation	0.833	0.0932	0.0255
SAPS II probability	Validation	0.847	0.0244	0.1019
APACHE III score	Validation	0.861	NAv	NAv

ROC, receiver operating characteristic; GOF, goodness-of-fit; for definition of C and H, see reference 18; MPM, Mortality Probability Model; MPM 0, Mortality Prediction Model at admission; MPM 24, Mortality Prediction Model 24 hrs after admission; SAPS, Simplified Acute Physiology Score; APACHE, Acute Physiology and Chronic Health Evaluation; NAv, not available.

Table 4

Table 5

Table 6

Table 7

Back to Top