

Indices de gravité et applications en réanimation

Severity scores in Intensive Care Medicine

B. Guidet^{1,2,3}, Ph. Aegerter^{3,4}

¹AP-HP, Hôpital Saint Antoine, Service de Réanimation Médicale, Paris, F-75012 France;

² Université Pierre et Marie Curie-Paris6, UMR S 707, Paris, F-75012 France;

³INSERM, U707, Paris, F-75012 France;

⁴Département de Biostatistique Hôpital Ambroise Paré, 92100 Boulogne.

Correspondance: Bertrand Guidet

Tel : 33 1 49 28 23 18

Fax : 33 1 49 28 21 45

e. mail : bertrand.guidet@sat.aphp.fr

Points essentiels à retenir en pratique :

- Les scores de gravité sont établis à partir de vastes bases de données.
- La sélection et la pondération des variables constitutives du score doit reposer sur une méthode statistique de régression multivariée et non sur dire d'expert.
- La validation du score nécessite d'étudier la discrimination (capacité à prédire la survie ou le décès), habituellement évaluée par l'aire sous la courbe ROC (receiver operating characteristic), et la calibration (performance du score dans toute la gamme de gravité), habituellement évaluée par le test d'adéquation du modèle (goodness-of-fit), mais aussi sa robustesse à des variations de recrutement (uniformity of fit).
- La difficulté principale de l'application de ces scores est la variété des cas traités (case-mix). Effectivement, la pathologie joue un rôle majeur dans le risque de décès et ce facteur n'est que très imparfaitement pris en compte dans les scores.
- Afin d'adapter un score à un service ou à une sous-population de malades, il a été proposé de l'ajuster (customisation), soit en modifiant globalement la relation entre le score et le risque de décès, soit en donnant un poids différent à chacune des variables constitutives.
- Le score de gravité permet de prédire le risque de décès. Il est donc possible de calculer pour un service, le rapport de la mortalité observée par la mortalité prédite par le score. Si ce rapport O/P est inférieur à 1, le service peut-être considéré comme performant.
- Les imperfections des scores ne permettent pas de les utiliser comme aide à la décision individuelle.

Résumé :

Evaluer objectivement la gravité d'un patient par un score unique est certainement un objectif utile. Cependant, la méthodologie de construction et de validation du score doit être connue afin d'avertir l'utilisateur potentiel des bénéfices et des limites de l'utilisation des scores de gravité.

Cette revue générale aborde les différents aspects d'élaboration, de validation et d'utilisation des scores de gravité en réanimation.

L'utilisation en pratique quotidienne d'un score de gravité nécessite que la saisie des données élémentaires constitutives du score soit exhaustive et exacte. Le score doit être valide lorsqu'il est utilisé dans une population différente de la population incluse dans le set d'élaboration. Un score peut être utilisé pour décrire la population des patients hospitalisés dans un service donné, en vue de comparer des services cependant, le rapport de la mortalité observée à la mortalité prédite ne peut être utilisé seul comme marqueur de performance car de très nombreux facteurs, indépendants du service, interviennent dans ce rapport. Enfin, le résultat d'un score ne peut être utilisé au niveau d'un individu comme outil d'aide à la décision d'admission en réanimation ni même d'aide à la décision thérapeutique individuelle.

Mots clés : Scores, réanimation, base de données, performance des services,

Summary :

Severity scores are commonly used in ICU since more than 25 years. In order to understand usefulness but also limitation of these scores, ICU physicians need to know how they are developed, validated and updated.

The construction of such scores requires large data base and modern statistical method in order to identify and weight the included variables.

Validation of a score is assessed through three sets of tests:

- Discrimination representing its the ability to predict death or survival. This is measured with the Area under the ROC curves. The closest ROCa it is to one, the better is the discrimination.
- Calibration is the assessment of the quality of the score according to different strates of severities. A score might be very accurate for very severe patients but with huge discrepancies between predicted and observed mortality for less severe patients. This is analysed with C test developed by Hosmer and Lemeshow
- Goodness-of-fit that tests the performance of a score with a different case-mix. A score might have good discrimination and calibration for certain type of patients but poor results in a different ICU with different patients.

Scores might be improved through customisation techniques.

The most recent scores identify different parts : i.e. the SAPS 3 scores identify three boxes : box1 patients characteristics before ICU admission; box2 Circumstances of ICU admission; box3 degree of physiologic derangement.

There is a general international consensus to consider that individual score should not be used to help in deciding whether a patient should be admitted or not in ICU. Scores are useful tools to assess ICU performance and to characterise patients included in trials.

Key words : scores, severity, intensive care unit, intensive care medicine, benchmarking; performance.

Introduction

La littérature médicale foisonne de scores dont les modalités d'élaboration, de validation et d'utilisation méritent d'être rappelées.

Les scores diagnostiques et les scores de charge en soins (TISS, NEMS, PRN, SIIPS,...) ne seront pas traités dans cette revue générale afin de pouvoir aborder en détail les scores de gravité généralistes et certains scores spécifiques. Pour une revue détaillée, le lecteur est invité à se référer au texte de la deuxième conférence de consensus européenne en réanimation traitant des "facteurs pronostiques chez les malades de réanimation" [1].

Quel peut être l'intérêt de disposer de scores de gravité ? Ils permettent au niveau collectif de décrire la population soignée et d'évaluer la performance des unités et ils pourraient, au niveau individuel, apporter une aide à la décision (politique d'admission, décision d'intervention, ou décision d'arrêt de soins actifs). Ces objectifs se heurtent à des difficultés méthodologiques, mais également éthiques, qui justifient de connaître les limites de l'utilisation de tous ces scores pronostiques. Quelles sont les variables retenues dans les scores permettant de prédire la mortalité dans le service de réanimation, la mortalité hospitalière, à distance (3 mois, 6 mois,...) ; la qualité de vie (quand ? avec quelle échelle de mesure?), la durée de séjour ou les coûts générés. Nous ne parlerons ici que des scores de gravité permettant de prédire la mortalité hospitalière.

I – Elaboration des scores

I – 1 – Pré-requis

Exhaustivité, exactitude et reproductibilité des informations collectées pour le calcul d'un score sont un préalable à l'élaboration comme à l'exploitation de ce score. Cet impératif

impose une “ retenue ” dans la multiplication des champs à saisir et des indicateurs à suivre en tenant compte de contraintes de “ faisabilité ”. Les procédures d’assurance-qualité des données peuvent être mises en oeuvre à la saisie (champs obligatoires, bornes incompatibles, cohérence entre items), à l’étape de validation de chaque fiche ou après export dans une base de données. Cependant, seul un audit externe permet de façon objective d’évaluer le respect des règles de codage et les erreurs de saisie. Il faut donc vérifier qu’une procédure de double codage de dossiers tirés au sort a été réalisée dans l’échantillon d’élaboration et de validation du score afin d’évaluer la reproductibilité (concordance) des informations. Habituellement, 5% des dossiers sont ainsi contrôlés.

A titre d’exemple, un tel travail a été réalisé avec les services adhérents à la base de données des services de Réanimation d’Ile-de-France (CUB-REA) [2]. Les dossiers tirés au sort ont été codés à nouveau par un médecin réanimateur d’un des autres services de la base CUB-REA. Les résultats font apparaître que la concordance est faible pour le statut fonctionnel et pour l’indice de Mac Cabe (classification en trois classes selon la présence d’une maladie sous jacente non mortelle, ou réputée mortelle en moins de 5 ans ou moins d’un an) [3]. L’analyse du score IGS I [4] ou IGS II [5] montre une très bonne concordance si les malades avec une donnée manquante sont exclus. Si les valeurs manquantes sont considérées comme normales, la concordance diminue mais l’écart constaté en terme de prédiction de mortalité n’est que de 3%.

Le développement des systèmes d’information hospitalier (SIH) devrait permettre d’alimenter automatiquement les bases de données et ainsi de renseigner plusieurs champs nécessaires au calcul des scores de gravité (données administratives, examens validés par le plateau technique ...). Les bases de données administratives ont l’avantage d’être obligatoires, disponibles et de concerner des populations importantes, elles peuvent permettre dans une certaine mesure d’appréhender la qualité [6] mais elles ont l’inconvénient d’être souvent

dissociées du résultat médical ou économique, de ne pas comporter d'évaluation des procédures ni d'évaluation de la qualité des données saisies.

En résumé, les données retenues pour figurer dans un score doivent être faciles à saisir et faciles à contrôler. Comme pour tout processus de mesure faisant l'objet d'un protocole, les définitions doivent être non ambiguës et connues des codeurs. Les conditions de recueil des informations élémentaires doivent être données avec, en particulier, la précision du délai entre l'admission et le recueil de ces informations. L'utilisation du score nécessite d'avoir le moins possible de données manquantes, sinon les données manquantes sont réputées normales ce qui tend à diminuer la performance du score.

I – 2 – Choix de l'échantillon

I- 2- 1 Choix des services pour l'élaboration du score

Le choix des services participant à l'élaboration du score n'est pas anodin. Soit, il s'agit d'un échantillon représentatif de l'ensemble des services d'une région, d'un pays, voire d'un continent et dans ce cas, il n'y a pas de pré requis pour pouvoir participer à l'échantillon d'élaboration du score (taille de l'unité, vérification de critère de qualité de soins) cependant l'échantillonnage tiendra compte de facteurs de structure (taille de l'unité ou de l'établissement, spécialisation, caractère universitaire) ; soit il s'agit de services qui satisfont à des critères minimum - ou maximum de qualité ; soit enfin la base de données repose pour partie sur le volontariat des services, ce qui est généralement le cas, en raison de la charge de travail supplémentaire induite. L'utilisation ultérieure du score, en particulier comme marqueur de performance, est bien sûr influencée par la qualité des services qui ont permis de l'élaborer. Ainsi, la référence peut être la moyenne obtenue par des services hautement sélectionnés sur des critères d'excellence et guère représentatifs de la moyenne des services. Cette démarche où le comparateur est un service idéal s'apparente à une approche de

type “ benchmarking”. Au contraire, s’il n’y a pas eu de sélection des services participants à l’échantillon d’élaboration du score, son application à l’ensemble des services poursuit un objectif différent, de classement relatif et de description épidémiologique.

En résumé, il faut caractériser les unités ou services ayant permis d’élaborer et de valider le score (type d’hôpital, nombre de lits, nombre de patients par unité et exigence de critères de qualité de soins) afin d’en estimer l’applicabilité dans un contexte différent (réanimation, hôpital).

I-2-2 Choix des patients

La plupart des scores excluent certains patients, considérés comme atypiques ou relevant d’une autre structure de prise en charge, ainsi l’étude d’élaboration du score IGS II [5] a écarté les patients hospitalisés pour une durée inférieure à 8 heures, les enfants mineurs, les sujets hospitalisés dans la période post-opératoire d’un pontage aorto-coronarien ou les patients brûlés qui sont donc également exclus de l’utilisation du score. Le pourcentage de patients exclus varie d’un service à un autre. Dans l’étude EURICUS I [7] regroupant 89 services de réanimation de 12 pays européens, certains services avaient plus de 30 % de leurs patients répondant aux critères d’exclusion, ce qui bien évidemment diminue la validité du score dans ces services.

I-3 : Construction du score.

Selon l’ancienneté de conception des scores, la méthode de construction est empirique ou statistique.

I-3-1. Scores anciens :

Pour le premier score Apache (Acute Physiologic and Chronic Health Status)[8] en 1981 et l’IGS 1 (Indice de Gravité Simplifié) en 1984 [4], les items retenus relevaient d’un choix d’experts.

Pour le score APACHE II, publié en 1985 [9], étaient retenues 12 variables physiologiques qui constituent l'Acute Physiology Score (APS) auxquels s'ajoutent l'âge et un certain nombre de maladies préexistantes. Chaque variable physiologique est évaluée pendant les 24 premières heures d'hospitalisation en réanimation et affectée d'une valeur allant de 0 (zone de normalité) à 4 (valeur la plus anormale). Cette pondération de chaque variable est arbitraire et ne repose pas sur une analyse de sa contribution à la mortalité observée à l'hôpital. Le score APACHE II tient compte du motif d'admission en réanimation à partir d'une liste prédéterminée de 50 diagnostics.

I-3-2. Scores de deuxième génération (tableau 1):

Il s'agit de trois scores généralistes : APACHE III, MPM II (mortality prediction model), IGS II. Pour chacun de ces scores, le choix des items constitutifs du score ainsi que les poids relatifs ont été établis par une méthode de régression logistique. Tous ces scores permettent de calculer une probabilité de décès hospitalier. Leur performance est meilleure que celle des scores plus anciens [10].

- Score APACHE III [11] :

Ce score a été élaboré à partir d'une base comportant les données de 40 services de réanimation nord américains représentatifs du pays. Le score va de 0 à 299 points dont 252 points pour les 18 variables physiologiques, 24 points pour l'âge et 23 points pour les maladies chroniques. Les valeurs les plus anormales des 24 premières heures sont considérées dans le calcul du score. La conversion du score en probabilité de décès utilise des équations de régression logistique pour chacun des 78 diagnostics et tenant compte des 9 origines possible. Ces équations ne sont pas du domaine public et sont commercialisées par l'APACHE Medical System, Washington, DC. En dehors du coût d'achat, l'application de ce score se heurte à la difficulté de choisir un seul diagnostic [13].

- Score IGS II

Ce score a été construit et validé sur une base de données nord américaine et européenne. Il comporte 12 variables physiologiques dont la profondeur du coma évaluée par le score de Glasgow [14], l'âge, le type d'admission (médicale, chirurgicale programmée ou non programmée), et 3 maladies sous-jacentes (SIDA, cancer métastaté et maladie hématologique). Comme d'autres, son utilisation en routine nécessite de suivre les conditions de sa validation, ainsi il faut considérer dans le calcul les valeurs les plus anormales observées pendant les 24 premières heures de l'hospitalisation et respecter les bornes pour chaque classe. Le mode de calcul de chaque item doit donc être précisé [15]. Les erreurs de date ou de borne exposent à des erreurs dans le calcul de ce score [16]. Cependant les conditions de recueil des variables physiologiques (continue ou discontinue) ne semblent avoir qu'une influence modérée sur le calcul de score de gravité [17]. Les erreurs les plus communes concernent l'évaluation du score de Glasgow [18]. Le score IGS II va de 0 à 163 points dont 116 points pour les 12 variables physiologiques y compris 26 points au maximum pour le score de Glasgow, 17 points pour l'âge et 30 points au maximum pour les maladies chroniques.

- MPM II [12].

Cet indicateur a été construit en 1993 sur une base de données nord américaine et ajusté sur une base internationale afin de prédire la mortalité hospitalière. Il peut intégrer les données présentes à l'admission (MPM₀) ou après 24 heures d'hospitalisation (MPM₂₄). Le score MPM₀ comporte 15 variables : âge, 3 variables physiologiques (coma, FC et PA systolique), 3 maladies chroniques (insuffisance rénale, cirrhose et cancer métastatique), 5 diagnostics (insuffisance rénale aiguë, troubles du rythme cardiaque, accident vasculaire cérébral, hémorragie digestive et effet de masse intracérébral), type d'admission, ventilation mécanique et arrêt cardiaque avant l'admission.

Toutes ces variables sont collectées dans l'heure qui précède l'admission en réanimation. Le MPM₂₄ utilise 13 variables : âge, 6 variables physiologiques (coma, créatinine, infection confirmée, hypoxémie, TP, diurèse), 3 variables connues à l'admission (cirrhose, effet de masse intracérébral et cancer métastatique), type d'admission, ventilation mécanique et drogues vasoactives.

I-3-3. Scores de troisième génération (tableau 2):

Ces nouveaux scores sont récents, intègrent de nouvelles variables et répondent des variables anciennes en fonction de la prise en charge actuelle des patients de réanimation.

- IGS III [19,20]

Ce score est construit à partir de trois types d'information :

- Groupe 1 : âge, comorbidités, durée de séjour et localisation dans l'hôpital avant transfert en réanimation, utilisation de drogues vasoactives au moment du transfert.
- Groupe 2 : Motif d'admission en précisant s'il s'agit d'une admission programmée ou non, si c'est en rapport avec une intervention chirurgicale en donnant le site opéré, et s'il existe une infection.
- Groupe 3 : il s'agit de variables physiologiques classiques mais dont les valeurs sont recueillies pendant l'heure qui précède ou suit l'admission et non pendant les 24 heures suivant l'admission comme dans les scores plus anciens.

- APACHE IV [21]

Ce score purement américain est construit à partir des données recueillies dans les 24 premières heures. Il est également très performant avec une bonne discrimination et une très bonne calibration (voir infra). Son intérêt principal réside dans la description détaillée des diagnostics avec analyse de la performance du test pour prédire la mortalité en fonction de chaque classe de diagnostic.

I-4 : Transformation d'un score de gravité en probabilité de décès.

Un score est établi à partir des différentes variables qui le constituent, chacune ayant un poids spécifique. Le choix des variables, leur transformation éventuelle (logarithme ou découpage en classes) sont le résultat du processus d'élaboration d'un modèle statistique multivarié [22]. Ainsi, la variable à expliquer, ici le décès, exprimé sous la forme logit faisant intervenir la probabilité de décès ($\text{logit}(\text{proba décès}) = \log [\text{proba décès}/(1 - \text{proba décès})]$), est liée aux différentes variables par une équation du type logistique :

$$\text{Logit} = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$$

où b_0 est l'intercept du modèle, x_1 à x_k représentent les variables explicatives et b_1 à b_k les coefficients de régression associés à chaque variable et estimés par l'ajustement du modèle de régression logistique.

La transformation suivante permet d'obtenir la probabilité de décès

$$\text{Pr} = \frac{e^{\text{Logit}}}{1 + e^{\text{Logit}}}$$

où Pr représente la probabilité de décès et Logit est comme décrit ci-dessus. La transformation logistique permet de modéliser une relation sigmoïde (de type dose-effet) existant entre deux variables et s'avère adaptée aux réponses binaires (oui/non, vivant/décédé) ; enfin les coefficients obtenus peuvent se traduire en terme d'odds-ratio, mesure d'association utilisée en épidémiologie. La figure 1 représente la relation entre le score IGS II et la probabilité de décès hospitalier. Il est intéressant de noter que quand l'IGS II passe de 20 à 30, la probabilité de décès passe de 3 à 10 % et que quand l'IGS II passe de 80 à 100 la probabilité de décès passe de 92 à 98 %, alors que pour les

valeurs intermédiaires, une faible variation de l'IGS II modifie de façon majeure la probabilité de décès. Ainsi quand l'IGS II passe de 45 à 50, la probabilité de décès passe de 34 à 46%.

II – Validation des scores

La validation du score nécessite d'avoir un échantillon de validation du score distinct de l'échantillon de construction, en effet la validation du score sur le même échantillon entraîne un biais d'optimisme.

Le processus de validation peut être réalisé de trois manières différentes non exclusives et de valeur croissante : validation croisée, validation temporelle, validation externe. Pour la validation croisée, il est classique de séparer la population en deux groupes : par exemple les 2/3 de la population totale pour construire le score et 1/3 pour le valider, cependant cette méthode réduit la puissance statistique et ne met pas à l'abri d'idiosyncrasies et de surparamétrisations néfastes à la robustesse de l'équation. Aussi préfère-t-on désormais avoir recours à des méthodes de rééchantillonnage (bootstrap) lors de cette phase qui tend de plus en plus à s'intégrer au processus d'élaboration [22]. La validation temporelle consiste à évaluer la performance des équations sur une population issue ultérieurement des mêmes services tandis que la validation externe consiste à évaluer la performance des équations sur une population indépendante (autres services). Cette étape est importante car plusieurs facteurs, au-delà des fluctuations d'échantillonnage et des erreurs dues à une mauvaise modélisation, concourent à la diminution de la pertinence du score: variabilité dans l'éventail des cas traités ("case-mix" des anglo-saxons)(tableau 2), politiques locales d'admission ou de sortie, qualité des soins, qualité de la collecte des informations servant à calculer les scores. Il faut donc vérifier, sur une population indépendante de la population de construction, la capacité d'un score de gravité à prédire correctement le décès. Ceci suppose de répondre à trois ordres de

questions : calibration, discrimination et robustesse (identification de groupes de patients mal évalués par le score).

II.1 – Calibration - discrimination

Par construction, lorsqu'il est appliqué à son échantillon de construction, le modèle donne un nombre total de décès estimés P égal au nombre total de décès observés O , le rapport O/P , appelé rapport standardisé de mortalité (standardized mortality ratio ou SMR) est donc égal à 1. La bonne calibration d'un score signifie que le modèle est capable de prédire correctement le taux de décès dans toute la gamme de gravité. Il s'agit de comparer la probabilité estimée de mortalité et le taux de mortalité observé obtenus en répartissant les patients par strates de gravité croissante par un test du χ^2 résumant les écarts entre les proportions prédites et observées. Ce test d'adéquation du modèle, s'il est significatif, traduit une mauvaise calibration. Le test C de Hosmer-Lemeshow [22] utilise 10 strates de taille équivalente déterminées sur les déciles de mortalité prédite. Cette méthode de classification est préférable à l'utilisation de valeurs seuil fixées *a priori*, en particulier pour tenir compte du faible nombre de patients pour les valeurs extrêmes. L'échantillon global doit être de taille suffisante sinon il faut réduire le nombre de strates. Les courbes de calibration représentent le taux de mortalité observé en fonction du taux de mortalité prédit et le comparent à la droite d'identité (diagonale) correspondant à un SMR de 1. Dans l'exemple donné avec le score MPM-II (figure 2), testé sur la population des patients de l'étude Euricus -I [7], il est noté que le score est bien calibré jusqu'à une probabilité de décès à 30% mais que pour des malades plus graves, le score surestime le risque de mortalité (40% de mortalité observée pour 60% de mortalité prédite). Le score APACHE III a fait l'objet d'une validation externe à partir d'une base de données nord-américaine constituée entre 1993 et 1996, avec 285 services de réanimation [23]. La mortalité hospitalière observée est de 12,35% et la mortalité prédite de 12,27% ce qui témoigne d'un bon ajustement global du modèle. En revanche, la calibration est médiocre pour certaines situations

diagnostiques (infarctus aigu du myocarde, tentative de suicide, traumatisme crânien sans chirurgie, polytraumatisme sans chirurgie, cf infra).

La discrimination consiste, pour le modèle, à faire correctement la distinction entre les observations ayant une issue favorable (survie) ou celles à issue défavorable (décès) sur la base du risque de décès prédit. En faisant varier le seuil de risque à partir duquel une observation est classée dans un groupe ou l'autre, on construit une courbe ROC (receiver operating characteristic), l'aire sous la courbe étant un indice de discrimination (ROCa. : aire sous la courbe ROC). Un modèle qui ne permet pas une meilleure discrimination que le tirage au sort à pile ou face aura une ROCa égale à 0,5 alors qu'un modèle parfait aura une aire de 1. Pour les trois scores généralistes récents, l'aire sous la courbe ROC est supérieure à 0,8 ce qui témoigne d'une bonne discrimination. La figure 3 représente les deux courbes ROC correspondant aux scores MPM II et IGS II chez les malades de l'étude Euricus-I [7]. Le score IGS II a une aire discrètement supérieure à celle du MPM-II mais celle-ci reste inférieure à 0,9 ce qui le rend inutilisable à l'échelon individuel, par exemple comme aide à la décision de poursuivre un traitement actif de réanimation. La figure 4, indique que dans une étude multicentrique portugaise (R Moreno, étude non publiée), la mauvaise corrélation entre le score IGS II et Apache II dans la zone de gravité intermédiaire correspondant à une probabilité de décès entre 20 et 60 pour cent, ce qui souligne également la prudence avec laquelle un score doit être utilisé chez un patient donné.

Pour les nouveaux scores de gravité, la discrimination est très bonne avec une aire sous la courbe ROC de 0,848 pour l'IGS III [20] et de 0,88 pour l'APACHE IV[21] et une très bonne calibration.

La population des patients pris en charge par un service est loin d'être homogène. Il faut donc vérifier que la performance d'un score (discrimination et calibration) reste bonne lorsqu'il est appliqué à des situations particulières [24]. Les principaux éléments à prendre en considération sont résumés par le terme anglais de case-mix que l'on peut traduire par éventail des cas traités (tableau 3). L'importance de ces différents facteurs apparaît clairement dans le tableau 4 qui a été établi à partir des données de l'année 2005 de la base de données francilienne, CUB-REA, regroupant 34 services de réanimation [2]. Globalement, sur l'ensemble des patients, le rapport de la mortalité hospitalière observée à la mortalité hospitalière prédite par le score IGS II est inférieur à 1 ce qui suggère que les réanimations franciliennes sont en 2005 plus performantes que les réanimations ayant permis de construire le score IGS II en 1991. En fait, ce rapport doit être interprété et ne peut être utilisé seul comme marqueur de la performance. Effectivement, un service de réanimation qui n'accueillerait que des patients souffrant d'un coma toxique ou d'un asthme aigu grave serait considéré comme très performant au vu d'un rapport O/P très bas. On voit donc que l'ajustement sur la gravité, estimée par l'IGS II, ne permet pas de s'affranchir de l'effet des pathologies traitées. On doit donc envisager d'améliorer le score par des procédés d'ajustement tenant compte de ces facteurs confondants.

II.3. Ajustement (“ customisation ”)

Afin d'étendre l'utilisation du score à des populations particulières, il est possible de l'ajuster en adaptant les équations qui permettent de transformer un score en probabilité de décès. Ce processus d'ajustement encore appelé “ customisation ” peut être réalisé de deux manières [25]:

- ajustement de premier niveau qui implique une modification globale de la relation entre le résultat du score et le résultat clinique (survie ou décès).
- ajustement de deuxième niveau qui consiste à recalculer le poids de chacune des variables constitutives.

Ainsi, l'ajustement de premier niveau n'est rien d'autre qu'une translation mathématique des logit originaux de manière à obtenir une probabilité de mortalité différente pour un même score. Ceci permet d'améliorer la calibration du modèle mais non sa discrimination. Cette technique doit donc être utilisée si le modèle a déjà une bonne discrimination. S'il s'agit d'améliorer la discrimination du modèle, la seule technique possible est l'adjonction de nouvelles variables ou la repondération des variables existantes. Ces techniques nécessitent des effectifs très importants (comme lors de l'élaboration initiale) et ne peuvent en pratique être réalisées qu'à partir de très larges bases de données multicentriques.

Un exemple de l'intérêt de l'ajustement de premier ordre est fourni par l'étude de la prédiction de décès chez les patients ayant un sepsis sévère précoce [26].

Dans une population de 1130 patients ayant un sepsis sévère précoce, la performance de l'IGS II ou du MPM24 est mauvaise. Sur l'échantillon de construction, l'ajustement a permis d'améliorer la calibration de l'IGS II et du MPM24 ($p = 0,92$ pour le test d'adéquation du modèle) et la discrimination avec une aire sous la courbe ROC de 0,79 pour les deux scores ajustés. La performance des deux scores ajustés restait bonne dans l'échantillon de validation ($p = 0,52$ pour le test d'adéquation du modèle et ROC_a de 0,75)

L'ajustement de deuxième niveau est plus fin et permet en principe d'adapter au mieux le modèle mais la différence avec la méthode de premier niveau n'est pas toujours sensible et il nécessite d'avoir connaissance de la valeur de chaque variable constitutive du score.

II.4. Adaptation globale de l'IGS II [27,28]

Nous avons réalisé une étude multicentrique prospective, regroupant 26 services de Réanimation d'Ile-de-France, participant à la base Cub-Réa, soit plus de 30 000 séjours sur 2 ans (1999-2000).

Une modélisation par régression logistique, intégrant l'IGS II, le type d'admission et l'information diagnostique permet d'estimer le taux de mortalité en réanimation et à l'hôpital. Différents niveaux

d'information diagnostique ont été utilisés : diagnostic principal ou groupes définis au moyen d'une méthode de classification par arbre. Tous les modèles ont bénéficié d'une validation interne et externe. Tous les nouveaux modèles affichent une bonne calibration et une discrimination améliorée par rapport à l'IGS II originel (ROCa de 0,93). La stabilité du modèle est acquise par l'intégration d'information diagnostique. Donc, l'IGS II peut être enrichi par l'intégration d'informations diagnostiques, afin d'améliorer la stabilité du modèle face à une grande variété de patients traités.

III- Scores spécifiques

De nombreux scores ont été établis pour des situations spécifiques. Citons par exemple, le score de Ranson pour les pancréatites aiguës [29], le score de Fine pour les pneumopathies communautaires [30], le score de Child pour les cirrhoses [31]. La méthode de validation et l'objectif de ces scores varient. Ainsi, le score de Fine constitue une aide à la décision d'admission [30], le score de Child permet d'évaluer le risque d'infection du liquide d'ascite justifiant une antibiothérapie. Nous détaillerons uniquement le score de Ranson en le comparant à un score généraliste.

Le Score de Ranson [29] intègre 5 paramètres recueillis à l'admission (âge > 55 ans, globules blancs > 16 000/mm³, glycémie > 11 mmol/l, LDH > 350 UI/l, SGOT > 250 UI) et 6 paramètres recueillis à la 48^{ème} heure d'hospitalisation (calcium sérique < 2 mmol/l, chute de l'hématocrite > 10%, PaO₂ < 60 mmHg, déficit en bases > 4 mmol/l, augmentation de l'urée > 5 mg/dl (1,8 mmol/l), séquestration liquidienne > 6 litres). Un score strictement inférieur à 3 identifie une population où la mortalité est de 3%. Celle-ci passe à 15% pour un score entre 3 et 5 et dépasse 50% si le score est supérieur à 6. De nombreux travaux ont étudié la valeur de ce score avec une sensibilité à identifier les formes sévères variant de 36,7% à 87%, et une spécificité de 60% à 96%. Les disparités sont en partie dues à la qualité inégale de ces travaux, aux différentes définitions de gravité utilisées et aussi

probablement à la grande variabilité du moment de la mesure car si le score de Ranson est mesuré à l'admission et à la 48^{ème} heure, le temps séparant le début de la symptomatologie et le calcul du score est rarement précisé. Les deux travaux méthodologiquement corrects retrouvent néanmoins des valeurs de sensibilité, spécificité, valeur prédictive positive et négative très proches avec respectivement 76.9% et 75%, 70.3% et 68%, 39.6% et 37%, 92.4% et 91%, permettant de classer correctement 71.3% et 69% des patients. Plusieurs travaux ont évalué la valeur pronostique du score APACHE II au cours de la pancréatite aiguë. Dans un travail prospectif incluant 290 pancréatites aiguës (280 évaluable dont 59 sévères : complications systémiques ou locales), un score APACHE II > 9 quel que soit le moment où il est mesuré, prédit 76% des défaillances viscérales graves et 73% des collections péri pancréatiques avec une spécificité de 84% pour les deux complications. Ce travail montre la supériorité du score APACHE II comparativement au score spécifique de Ranson. Le score APACHE II s'élève de 2 points entre l'admission et la 24^{ème} heure et de 3 points entre la 24^{ème} et 48^{ème} heure en cas de forme sévère alors qu'en cas de forme modérée, il ne se modifie pas dans les premières 24 heures et qu'il diminue d'un point au 2^{ème} jour. Le seuil à partir duquel on parle de forme grave n'est pas établi bien qu'un score > 7 ait été proposé lors de la conférence d'Atlanta. Ce score manque de sensibilité, il est néanmoins calculable chaque jour et ce dès l'admission. Sa bonne valeur prédictive négative (93%) permet d'éliminer une forme grave quand il ne dépasse pas 10 à la 24^{ème} heure et de postuler l'absence de décès si le score ne dépasse pas 10 au cours des 3 premiers jours. Ceci plaide pour une utilisation d'un score généraliste, type APACHE II plutôt que du score de Ranson, au cours des premières 48 heures d'une pancréatite.

IV – Utilisation des scores

Les scores peuvent avoir de multiples utilisations que les procédures d'élaboration et de validation doivent considérer.

IV.1. Epidémiologie descriptive globale.

La récente réforme de l'organisation du système de santé français justifie que les médecins élaborent des outils d'évaluation de leur activité. Dans une période de restriction budgétaire et de réallocation des budgets entre et à l'intérieur des régions, l'activité de la Réanimation qui représente 5% des séjours, mais environ 15 % des coûts hospitaliers, apparaît comme un domaine important à analyser. L'interprétation des tableaux de bord de réanimation [32] n'est possible que si la population des patients admis est correctement décrite, intégrant en particulier l'évaluation de la gravité. Le résultat des scores peut être comparé à une référence historique du service (courbe de tendance), à une moyenne de services équivalents, enfin à un service idéal.

IV.2. Mesure de la performance. [33].

La définition même de la qualité est variable selon le point de vue du patient, du service ou du payeur. La mesure de la performance des services de Réanimation bute sur plusieurs difficultés concernant les indicateurs de procédure, de résultats, le contrôle de la variété des cas, l'espace de référence, et enfin l'impact des facteurs organisationnels.

Le critère de résultat le plus pertinent pour des patients hospitalisés dans un service de Réanimation est la mortalité. Or, cet indicateur est influencé par de multiples facteurs. Certains sont d'une certaine mesure indépendants de la qualité intrinsèque de la réanimation : variété des cas traités (case-mix), environnement de la réanimation (qualité des services d'amont et d'aval, taille de l'hôpital, bassin de population), alors que d'autres sont dépendants de la qualité du service comme l'organisation interne, la qualité des équipes, le respect des recommandations de bonne pratique, l'hygiène,... [28].

Les pathologies traitées conditionnent largement les résultats en terme d'évaluation de l'efficacité (cf supra). Elles doivent donc être prises en compte de façon différenciée afin de

dépasser les limites d'un indice unique global ne fournissant aucune identification des situations à prise en charge déficiente.

Il est régulièrement avancé que le rapport mortalité observée sur mortalité prédite par un score comme l'IGS II pourrait être un indicateur de performance d'un service de réanimation. Ainsi la figure 5 positionne les services de réanimation inclus dans le protocole européen Euricus I [7] en fonction du rapport de la mortalité observée à la mortalité prédite (SMR). Une lecture rapide de ce graphique suggère que certaines réanimations ont une bonne performance car leur SMR est largement inférieur à 1 alors que d'autres ont un SMR supérieur à 1 indiquant une mortalité observée supérieure à la mortalité prédite par l'IGS II. Cependant, plusieurs études montrent que l'IGS II doit être recalibré lorsqu'il est appliqué à d'autres populations, ou à des sous-groupes particuliers. On sait aussi que la capacité de prédire le décès hospitalier par l'IGS II diminue avec la prolongation du séjour [34]. Enfin, cet indice est construit pour la mortalité hospitalière, méconnaissant ainsi la mortalité en Réanimation. Or la différence entre le taux de mortalité en réanimation et le taux de mortalité à l'hôpital est variable d'une institution à une autre et d'une pathologie à une autre. En fait, l'écart entre la prédiction et l'observation peut avoir plusieurs causes [35]: "case-mix", organisation du service et environnement hospitalier, qualité des soins, enfin erreurs aléatoires du modèle. Par exemple, la part de la variable expliquée, ici le décès, dans l'étude IGS III n'est que de 25% pour les variables du groupe 3 (BOX 3) qui regroupe les désordres physiologiques. Ce n'est que sur ces variables que l'apport de la prise en charge dans un service de réanimation peut être décisive. (Tableau 5).

Il est intéressant de noter que dans notre étude, l'utilisation d'un score IGS II recalibré ne change pas fondamentalement l'ordre des services les mieux classés mais qu'en revanche au-delà de la sixième place l'ordre change considérablement (tableau 6). Ceci est illustré de

manière graphique sur la figure 6. La figure 7 permet de montrer que l'analyse de la mortalité en réanimation et de la mortalité hospitalière est complémentaire.

Malgré les critiques générales que l'on peut formuler vis-à-vis du SMR, il faut bien reconnaître que les nouveaux scores, en particulier l'IGS III, permettent d'identifier des groupes de pays avec des performances des services de réanimation assez variables et conformes à la connaissance que nous avons des divers systèmes de soins [20].

IV.3. Prise en compte dans un essai thérapeutique

Dans le cadre d'un essai thérapeutique comparatif multicentrique, la connaissance du score de gravité permet de pratiquer une randomisation stratifiée sur les niveaux de gravité ou d'ajuster l'évaluation du traitement sur le score, ce qui s'avère primordial si le critère de jugement principal est le décès.

IV.4. Aide à la décision individuelle.

Il existe un consensus [1,35,36,37] pour ne pas recommander l'utilisation de scores de gravité comme critère d'admission en réanimation. La sensibilité et la spécificité trop médiocres de ces scores n'autorisent pas leur utilisation aveugle et automatique qui serait de toute façon contraire aux principes éthiques qui gouvernent notre pratique.

Conclusion

Un certain nombre de précautions doivent être apportées avant d'utiliser un score. Il faut tout d'abord disposer de données correctes, exhaustives, vérifiées. Les scores sont habituellement élaborés sur des échantillons de services ou de malades différents de la population locale, ce qui justifie une revalidation des scores au minimum au niveau national pour tenir compte

de la variation des cas traités. Certains scores sont déjà anciens et méritent d'être actualisés. Il faut actuellement considérer les scores comme marqueur global de performance d'un service ou groupe de services de réanimation et non comme simple outil d'aide à la décision au niveau individuel.

Bibliographie :

1. Deuxième conférence de consensus européenne en réanimation. Facteurs pronostiques chez les malades de réanimation. Réa Urgences 1994, 3, N°2 bis.
2. Aegerter Ph, Auvert B, Buonamico G, Sznajder M, Beauchet B, Guidet B, LeGall JR et CUB-REA. Mise en oeuvre et évaluation d'une base de données commune aux services de réanimation d'Ile-de-France. Rev Epidémio Santé Publ 1998, 46 : 226-237.
3. Mac Cabe W. R., Jackson G. G. Gram negative bacteremia : Etiology and Ecology. Arch Intern Med 1962 ; 110 : 847-852.
4. Le Gall JR, Loirat P, Alperovitch A et coll. A simplified acute physiologic score for ICU patients. Crit Care Med 1984 ; 12 : 975-977.
5. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA. 1993, 270 : 2957-63.
6. Iezzoni LI. Assessing quality using administrative data. Ann Int Med 1997, 127 : 666-674.
7. Reis Miranda D, Ryan DW, Schaufeli WB, Fidler V. Organisation and management of intensive care. A prospective study in 12 European countries. Update in Intensive care and emergency medicine, ed JL Vincent. Springer Vol 29, 1997.
8. Knaus WA., Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE – Acute physiology and chronic health evaluation : a physiologically based classification system. Crit Care Med 1981 ; 9 : 591-597.
9. Knaus WA., Draper EA., Wagner DP., Zimmerman JE. APACHE II : A severity of disease classification system. Crit Care Med 1985 ; 13 : 818-819.
10. Castella X, Artigas A, Bion J, Kari A. The European/North American Severity Group. A comparison of severity of illness scoring systems for intensive care unit patients : results of a multicenter, multinational study. Crit Care Med 1995 ; 23 : 1327-35.
11. Knaus WA. Wagner DP. Draper EA. Zimmerman JE. Bergner M. Bastos PG. Sirio CA. Murphy DJ. Lotring T. Damiano A. et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest. 1991, 100 : 1619-36.
12. 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 : 2478-86.
13. Teres D. Lemeshow S. Why severity models should be used with caution? Critical Care Clinics. 1994, 10 : 93-110.

14. Teasdale G, Jenett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* 1974 ; 2 : 81-84.
15. Recommandations de codage d'un malade de réanimation. Guidet B, Misset B, Maner Ph. In *PMSI et Réanimation* eds B Guidet et B Misset ; Masson, Paris 2000, pp93-104.
16. Féry-Lemonnier E, Landais P, Loirat P, Kleinknecht D, Brivet F. Evaluation of severity scoring systems in ICUs. Translation, conversion and definition ambiguities as a source of inter-observer variability in Apache II, SAPS and OSF. *Intensive Care Med* 1995, 21 : 356-60.
17. Pollack MM, Patel KM, Ruttimann U, Cuerdon T. Frequency of variable measurement in 16 pediatric intensive care units: influence on accuracy and potential for bias in severity of illness assessment. *Crit Care Med*. 1996, 24 : 74-7.
18. Bastos PG, Sun X, Wagner DP, Wu AW, Knaus WA. Glasgow Coma Scale score in the evaluation of outcome in the intensive care unit: findings from the Acute Physiology and Chronic Health Evaluation III study. *Critical Care Medicine*. 1993, 21 : 1459-65
19. Metnitz PG, Moreno RP, Almeida E et al. SAPS 3 – From evaluation of the patient to evaluation of the intensive care unit. Part 1: objectives, methods and cohort description. *Intensive care med* 2005, 31:1336-1344
20. Moreno RP, Metnitz PG, Almeida et al. SAPS 3 – From evaluation of the patient to evaluation of the intensive care unit. Part 2; Development of a prognostic model for hospital mortality at ICU admission. *Intensive care med* 2005, 31:1345-55
21. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006, 34:1297-1310
22. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*. 1996, 15 : 361-87.
23. Zimmerman JE, Wagner DP, Draper EA, Wright L, Alzola C, Knaus WA. Evaluation of acute physiology and chronic health evaluation III predictions of hospital mortality in an independent database. *Crit Care Med* 1998, 26 :1317-26.
24. Moreno R, Apolone G, Miranda DR. Evaluation of the uniformity of fit of general outcome prediction models. *Intensive Care Medicine*. 1998, 24 : 40-7.
25. Moreno R, Apolone G. Impact of different customization strategies in the performance of a general severity score. *Crit Care Med*. 1997, 25 : 2001-8.

26. Le Gall JR, Lemeshow S, Leleu G, Klar J, Huillard J, Rue M, Teres D, Artigas A. Customized probability models for early severe sepsis in adult intensive care patients. *Intensive Care Unit Scoring*. JAMA. 1995, 273 : 644-50.
27. Aegerter P, Boumendil A, Retbi A, Minvielle E, Dervaux B, Guidet B. Peut-on mesurer la performance des services grâce à l'IGS II ? *REANIMATION* 2003, 12, suppl 2 : 31s-41s
28. Aegerter P, Boumendil A, Retbi A, Minvielle E, Dervaux B, Guidet B. SAPS II revisited. *Intensive Care Med* 2005, 31 : 416-423
29. Ranson JHC., Rikfind KM, Turner JW. Prognostic signs and non operative peritoneal lavage in acute pancreatitis. *Surg Gynecol Obstet* 1976 ; 143 : 209-219.
30. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, Coley CM, Marrie TJ, Kapoor WN. A prediction rule to identify low-risk patients with community acquired pneumonia. *N Engl J Med* 1997, 336 : 243-50.
31. Child CG, Turcotte JG. Surgery and portal hypertension. In: Child CG, editor. *The liver and portal hypertension*. Philadelphia: WB Saunders; 1964. p. 50-72.
32. Guidet B, Fosse J Ph, Gervais C. Description de l'activité en Réanimation. Tableau de bord. dans *Management en réanimation, évaluation, organisation et éthique*. Collection Réanimation. Elsevier Eds F Saulnier et J Bion ;Paris 2000, pp 133-152.
33. Chouaid C, Hejblum G, Guidet B, Valleron AJ. De l'évaluation de la qualité des soins à la performance des soins. *Rev Mal Resp*. 2006, 23, Sup4 : 87-98.
34. Sicignano A, Carozzi C, Giudici D et al. The influence of length of stay in the ICU on power of discrimination of a multipurpose severity score (SAPS). *Intensive Care Med* 1996; 22: 1048-51.
35. Teres D, Lemeshow S. Why severity scores should be used with caution?. *Crit Care Clin* 1994 ; 10 : 93-110.
36. Lemeshow S, Klar J, Teres D. Outcome prediction for individual intensive care patients: useful, misused, or abused?. *Intensive Care Med*. 1995, 21 : 770-6.
37. Gunning K, Rowan K. ABC of intensive care: outcome data and scoring systems *BMJ*. 1999, 319 : 241-4.

Tableau 1 : Trois principaux scores de gravité généralistes de deuxième génération

Caractéristiques	APACHE III [11]	IGS II [4]	MPM II [12]
Année de publication	1991	1993	1993
N pays	1	12	12
N réanimations	40	137	140
N patients	17440	12997	19124
Variables			
Age	Oui	Oui	Oui
Origine	Oui	Non	Non
Statut chirurgical	Oui	Oui	Oui
Maladies chroniques	Oui	Oui	Oui
Physiologie	Oui	Oui	Oui
Diagnostic	Oui*	Non	Oui
N Variables	26	17	15\$
Prédiction de la mortalité hospitalière	Oui	Oui	Oui

* Choisis à partir d'une liste de 78 diagnostics médicaux ou chirurgicaux

\$ Le modèle MPM24 n'utilise que 13 variables

Tableau 2 : Nouveaux scores de gravité

Caractéristiques	APACHE IV [21]	SAPS III [19,20]
Année de publication	2006	2005
N pays	1	35
N réanimations	104	307
N patients	110 558	19 577
N Variables	26	20
Discrimination (aROC)	0,88	0,85
Calibration	Bonne	Bonne

Tableau 3 : Principaux éléments de l'éventail des cas traités " case mix "

Localisation dans l'hôpital avant l'admission en réanimation et durée de ce séjour préalable (" lead time bias ")
Statut chirurgical (chirurgie programmée, urgente ou malade médical)
Sévérité de la maladie
Réserve physiologique (âge, maladies sous jacentes)
Pathologies ayant motivées l'admission

Tableau 4 : Rapports de la mortalité hospitalière observée sur la mortalité hospitalière prédite par l'IGS II (O/P) pour les séjours de l'année 2005 des services participants à la base de données CUB-Réa.

Situations	n	IGS II	Mortalité observée	Mortalité O/P
Totalité de la base	16 647	41,5	23%	0,71
Types de séjour				
Médical	14 186	42,1	24%	0,71
Chirurgical	2 461	37,9	20%	0,71
Mode d'entrée				
Direct ou urgence	8965	41,6	21%	0,63
Transfert interne	6728	41,3	26%	0,82
Transfert externe	954	41,2	23%	0,72
Pathologies				
Coma toxique	924	40,8	3%	0,09
Asthme aiguë grave	284	27,2	8%	0,57
Etat de choc	3704	58,1	46%	0,83
SDRA	1164	60,4	56%	0,96

Tableau 5 : variables de l'IGS III

	IGS III	Contribution à l'estimation de la mortalité hospitalière
Groupe 1: Caractéristiques avant admission	Age Co-morbidités Drogues vasoactives Lieu d' hospitalisation Durée de séjour antérieure à l'admission en réanimation	50%
Groupe 2: Circonstances de l'admission en réa	Motif d'admission Admission programmée Statut chirurgical Site de l'intervention Infection Communautaire Nosocomiale	22,5%
Groupe 3: Altérations physiologiques	Score de Glasgow Fréquence cardiaque Pression artérielle Bilirubine Température Créatinine Leucocytes Plaquettes pH Support ventilatoire	27,5%

Tableau 6 : Performance des services (SMR et rang) pour la mortalité en réanimation en fonction de l'IGS2 recalibré utilisé [27]

Service	Modèle A			Modèle B			Modèle C			Modèle D		
	smr	IC à 95%	rang	smr	IC à 95%	rang	smr	IC à 95%	rang	smr	IC à 95%	rang
1	0,57	(0.4-0.78)	1	0,56	(0.4-0.77)	1	0,58	(0.42- 0.8)	1	0,66	(0.47-0.91)	2
2	0,72	(0.51- 1)	2	0,67	(0.47-0.92)	2	0,66	(0.47-0.91)	2	0,61	(0.43-0.85)	1
3	0,82	(0.67-0.99)	3	0,84	(0.69-1.01)	4	0,88	(0.72-1.05)	6	0,89	(0.73-1.08)	6
4	0,84	(0.69-1.01)	4	0,84	(0.7-1.01)	6	0,82	(0.68-0.99)	5	0,83	(0.69- 1)	5
5	0,85	(0.68-1.06)	5	0,83	(0.66-1.03)	3	0,81	(0.64- 1)	3	0,80	(0.64- 1)	3
6	0,90	(0.73-1.11)	6	0,84	(0.68-1.03)	5	0,81	(0.66- 1)	4	0,81	(0.65-0.99)	4
7	0,95	(0.77-1.16)	7	0,91	(0.74- 1.1)	7	0,92	(0.75-1.12)	7	0,93	(0.75-1.13)	7
8	0,99	(0.81-1.19)	8	1,01	(0.83-1.22)	10	1,00	(0.82- 1.2)	10	1,01	(0.83-1.22)	13
9	0,99	(0.82-1.19)	9	1,02	(0.84-1.22)	11	1,03	(0.86-1.24)	13	1,05	(0.87-1.26)	15
10	1,01	(0.86-1.18)	10	1,05	(0.89-1.22)	14	1,00	(0.85-1.16)	12	0,97	(0.83-1.12)	10
11	1,02	(0.85- 1.2)	11	1,13	(0.94-1.34)	17	1,15	(0.96-1.37)	19	1,09	(0.91-1.29)	18
12	1,02	(0.81-1.26)	12	0,99	(0.79-1.22)	8	0,98	(0.78-1.21)	8	1,01	(0.8-1.25)	12
13	1,04	(0.86-1.24)	13	1,00	(0.82-1.19)	9	0,98	(0.81-1.17)	9	0,99	(0.82-1.18)	11
14	1,07	(0.92-1.24)	14	1,04	(0.89- 1.2)	13	1,00	(0.86-1.16)	11	0,93	(0.8-1.08)	8
15	1,08	(0.91-1.28)	15	1,07	(0.9-1.26)	15	1,10	(0.93- 1.3)	16	1,07	(0.9-1.26)	16
16	1,11	(0.95-1.29)	16	1,16	(0.99-1.35)	19	1,29	(1.1- 1.5)	24	1,23	(1.05-1.43)	23
17	1,11	(0.86-1.42)	17	1,03	(0.8-1.31)	12	1,21	(0.94-1.54)	21	0,96	(0.74-1.22)	9
18	1,12	(0.93-1.34)	18	1,11	(0.92-1.34)	16	1,08	(0.9- 1.3)	14	1,04	(0.86-1.25)	14
19	1,14	(0.94-1.37)	19	1,20	(1-1.44)	23	1,14	(0.94-1.36)	18	1,16	(0.96-1.39)	21
20	1,17	(1-1.35)	20	1,20	(1.03-1.38)	21	1,08	(0.93-1.25)	15	1,08	(0.93-1.25)	17
21	1,17	(0.99-1.37)	21	1,19	(1.01- 1.4)	20	1,19	(1.01- 1.4)	20	1,19	(1.01-1.39)	22
22	1,18	(0.94-1.46)	22	1,20	(0.96-1.49)	22	1,13	(0.9-1.41)	17	1,10	(0.87-1.36)	19
23	1,26	(1.01-1.57)	23	1,14	(0.91-1.41)	18	1,22	(0.98-1.51)	22	1,30	(1.04-1.61)	24
24	1,30	(0.95-1.75)	24	1,28	(0.93-1.72)	24	1,24	(0.9-1.66)	23	1,11	(0.81-1.49)	20
25	1,72	(1.3-2.23)	25	1,64	(1.24-2.12)	25	1,53	(1.15-1.98)	25	1,53	(1.16-1.99)	25

Légende des figures :

Fig 1 : Relation entre le score d'indice de gravité simplifié (IGS) II et la probabilité de décès hospitalier.

Fig 2 : Etude de la calibration du score MPM II dans la population des patients inclus dans l'étude Euricus I [7]. La courbe indique une surestimation du risque de décès à partir d'une probabilité de décès de 30%.

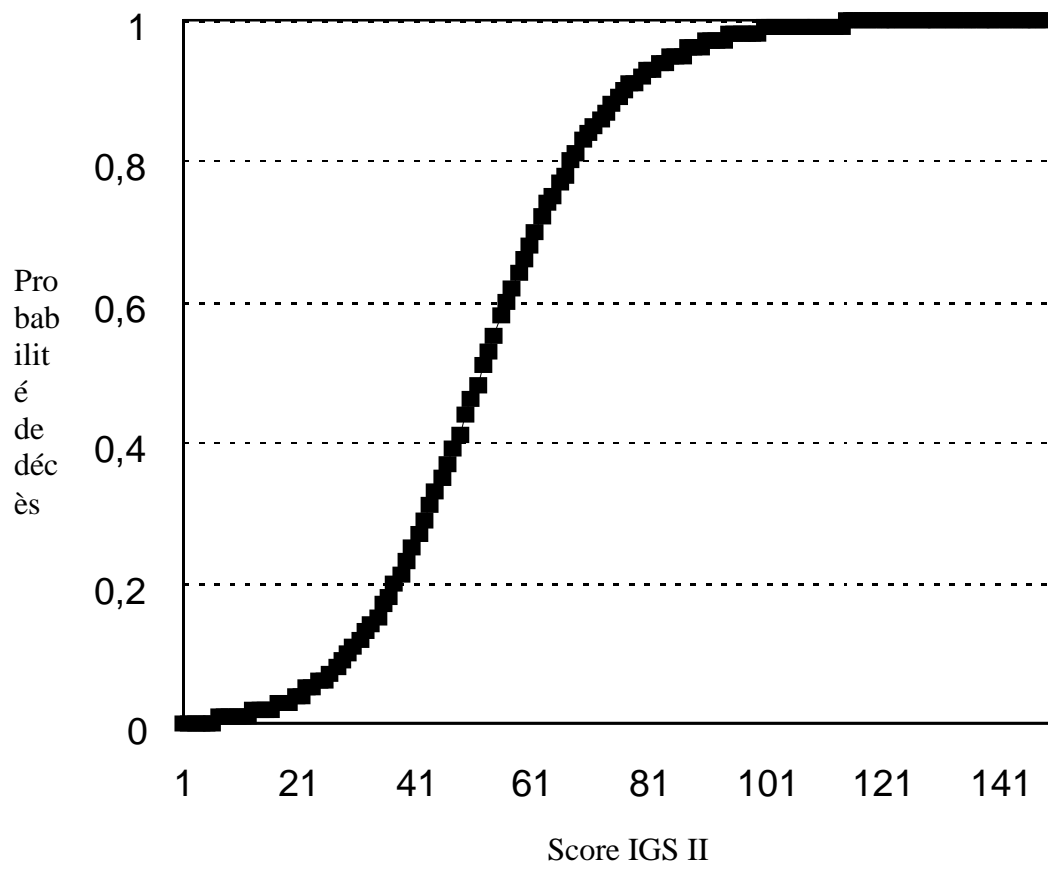
Fig 3 : Etude de la discrimination du score MPM II (*) et IGS II (+) dans la population des patients inclus dans l'étude Euricus I [7]. L'aire sous la courbe ROC est de 0,822 pour l'IGS II et de 0,785 pour le MPM II.

Fig 4 : Corrélation entre le score IGS II et APACHE II. La corrélation est mauvaise dans la zone de probabilité intermédiaire de décès. Cette faible corrélation dans la zone la plus pertinente pour le clinicien doit rendre prudente l'utilisation individuelle d'un score pour prédire le décès.

Fig 5 : Classement des 89 services de réanimation européens inclus dans l'étude Euricus I [7] selon le rapport de la mortalité observée à la mortalité prédite (SMR : standardised mortality ratio, avec intervalle de confiance à 95%). Certains services ont un SMR largement inférieur à 1 suggérant une bonne performance alors que d'autres ont un SMR supérieur à 1 suggérant une mauvaise performance. Cet indicateur doit être interprété en intégrant d'autres informations (voir texte). Chaque barre verticale représente un service.

Fig 6 : SMR des services en fonction du modèle de mortalité en réanimation [28]

Fig 7 : SMR des services en fonction du modèle de mortalité en réanimation ou hospitalière [28]



Figure

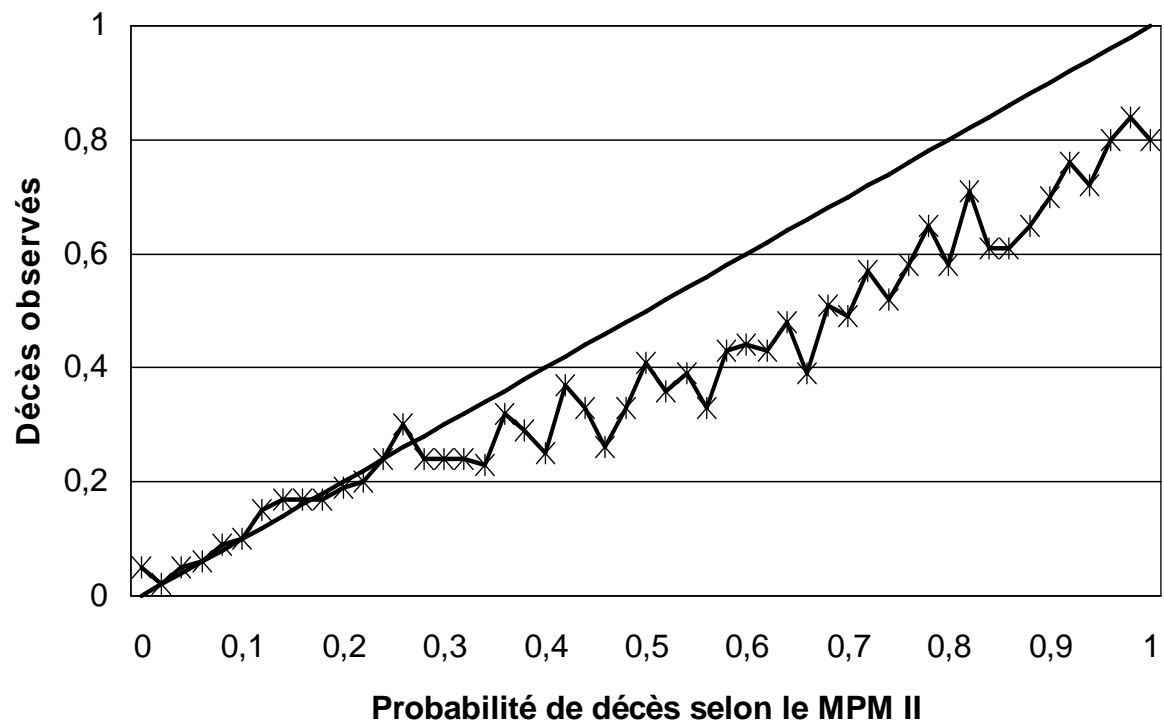


Figure 2

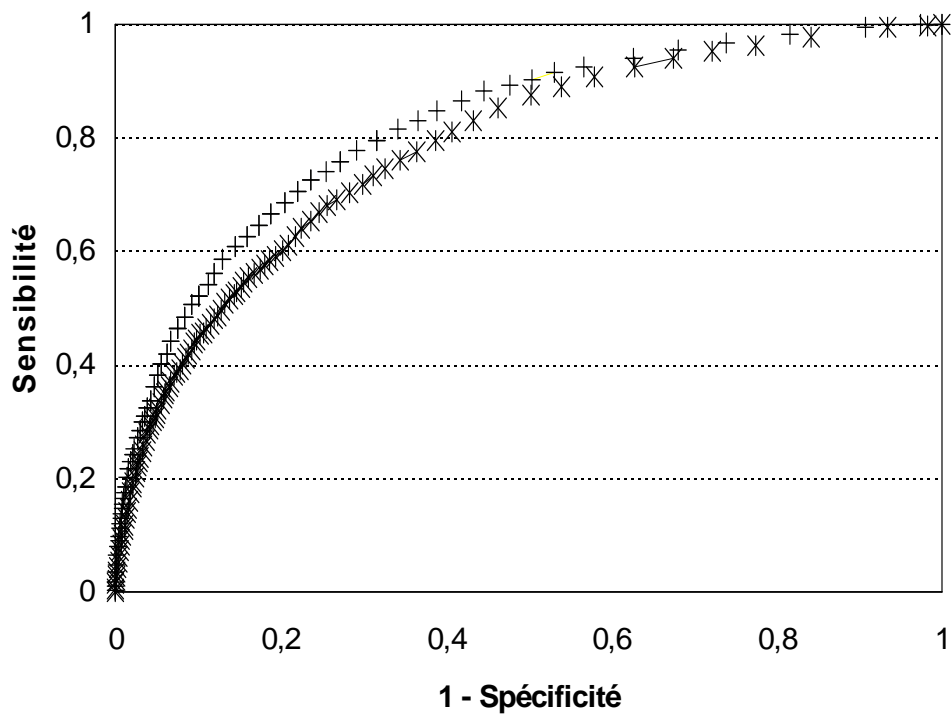


Figure 3

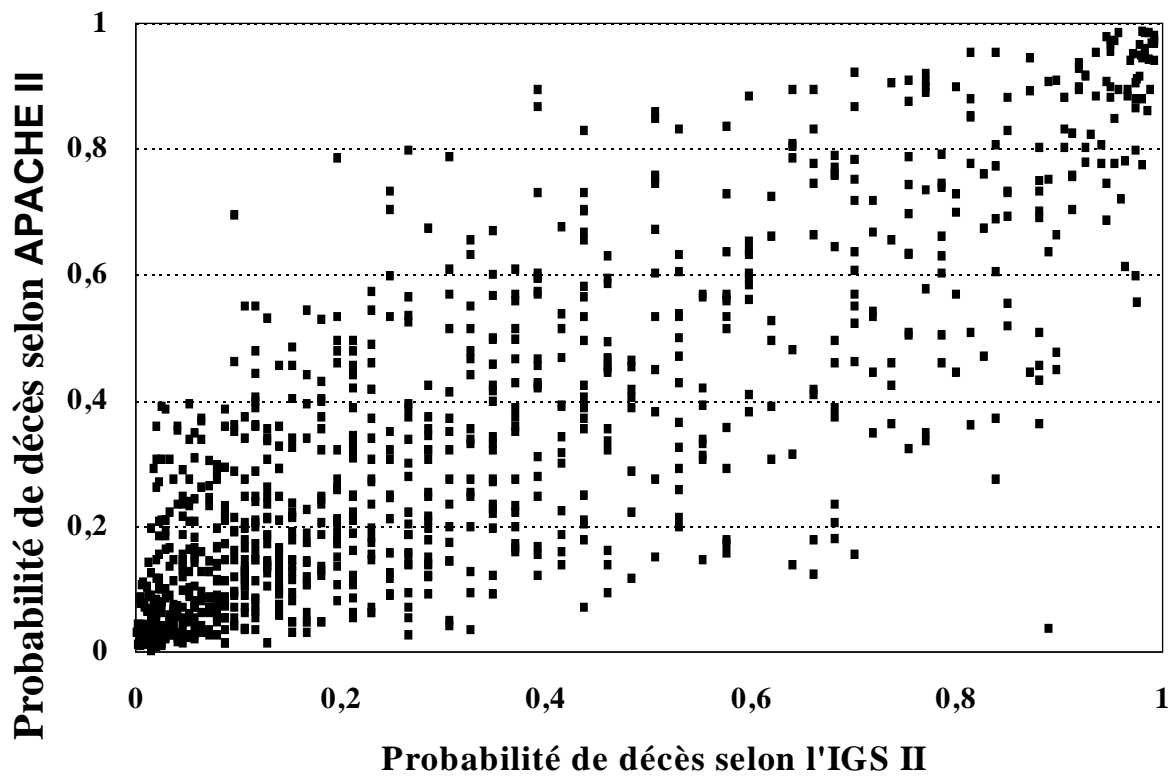


Figure 4

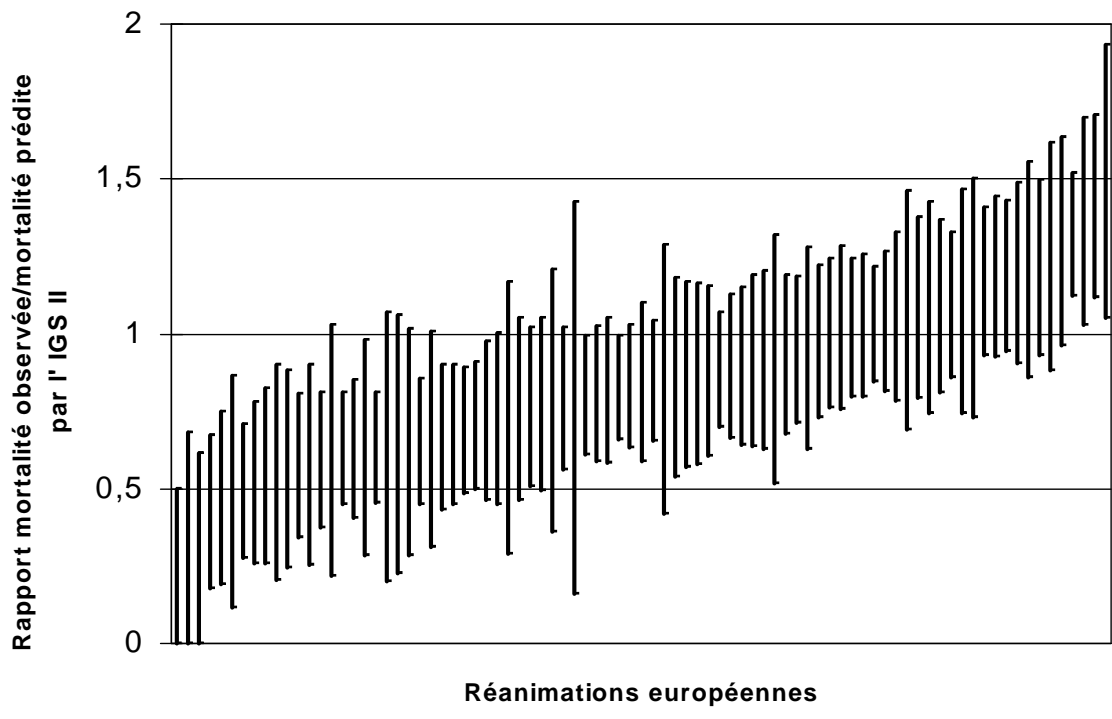


Figure 5

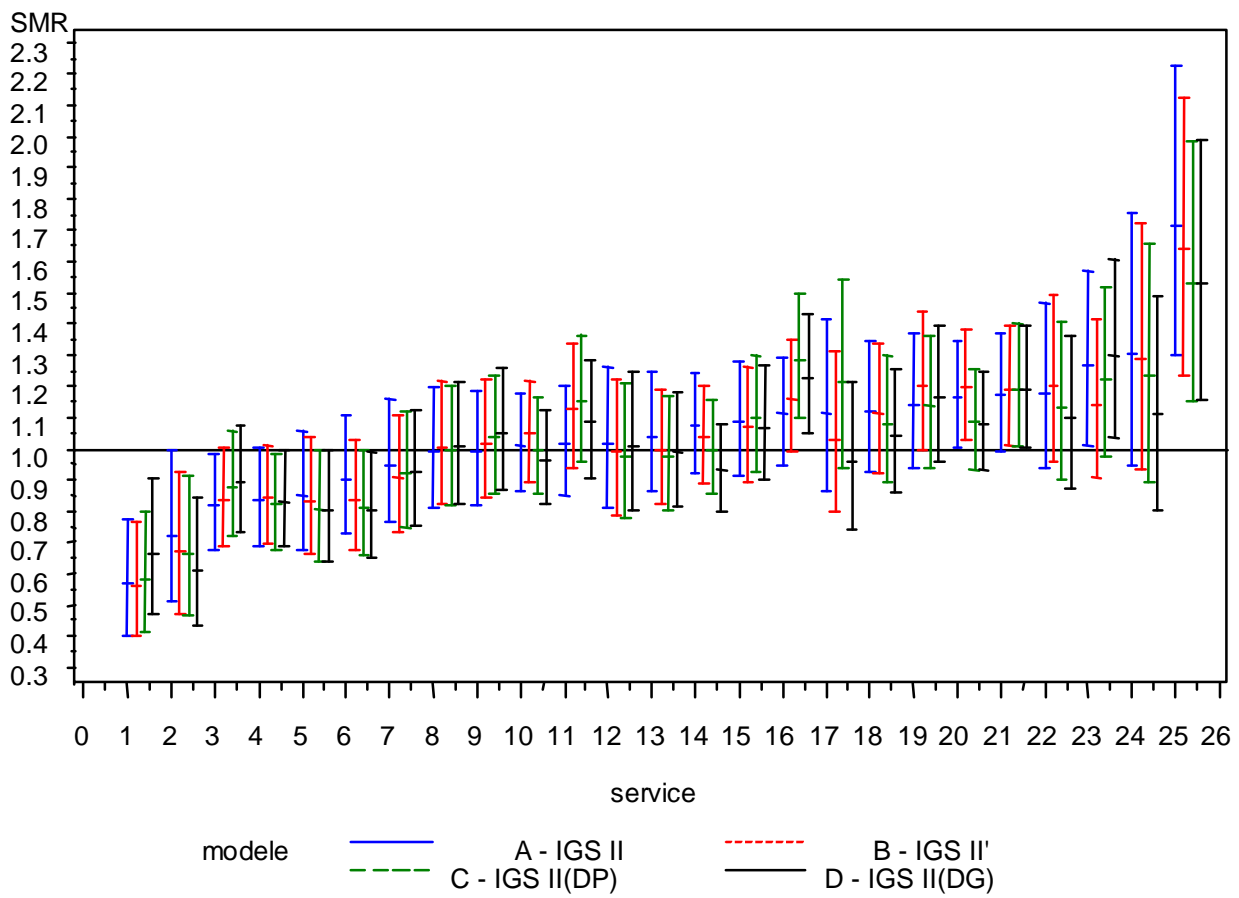


Fig 6

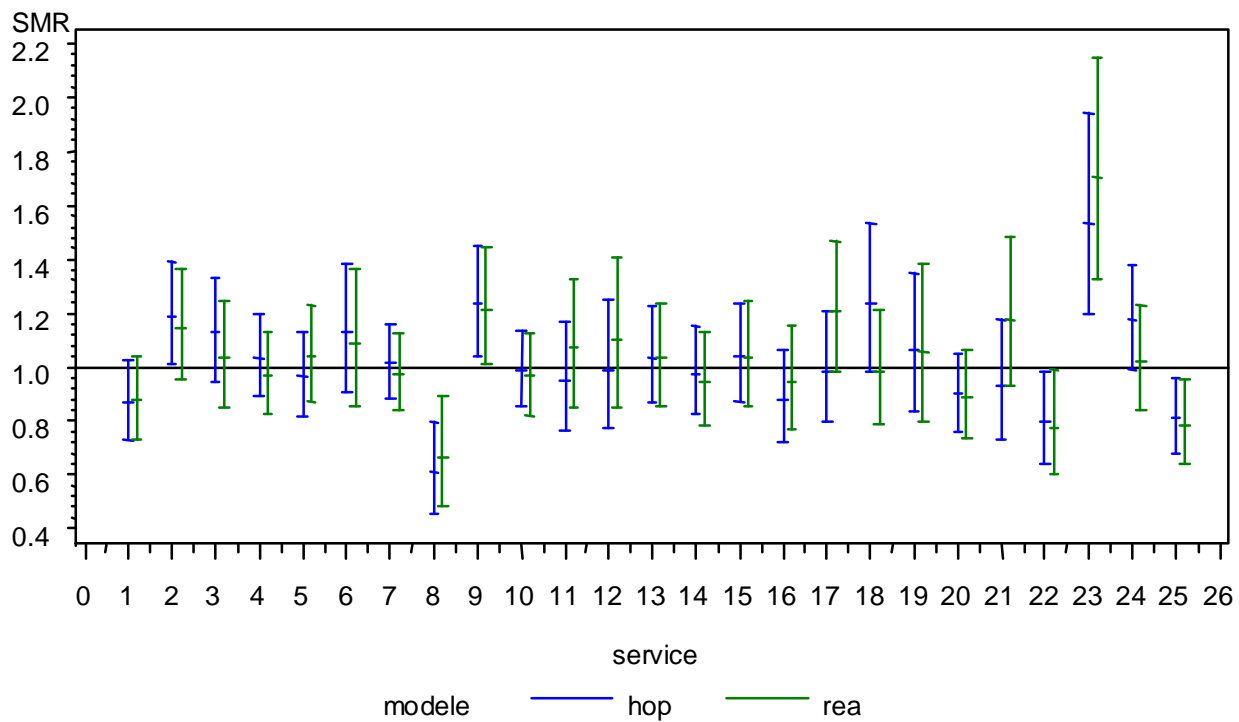


Fig 7