

Les scores de gravité généraux et de dysfonctions d'organes en réanimation pédiatrique : quoi de neuf en 2016 ?

Severity of Illness and Organ Dysfunction Scoring in Pediatric Intensive Care Unit: What News in 2016?

S. Leteurre · M.-E. Lampin · B. Grandbastien · M. Recher · A. Duhamel

Reçu le 29 avril 2016 ; accepté le 6 juillet 2016
© SRLF et Lavoisier SAS 2016

Résumé Les scores de gravité généraux et de dysfonctions d'organes (DO) sont de plus en plus utilisés dans les services de réanimation adultes et pédiatriques. Cette utilisation est justifiée par la nécessaire évaluation objective de la gravité et du niveau des DO des patients durant le séjour en réanimation. En réanimation pédiatrique, les scores de gravité généraux sont établis, indépendamment du diagnostic, en prenant en compte l'hétérogénéité des populations en termes d'âge notamment. Ces scores permettent d'évaluer la gravité dans les premières heures après l'admission en réanimation. Deux « systèmes » de scores de gravité généraux pédiatriques ont été proposés pour la population allant du nouveau-né (prématurés exclus) à l'adolescent : le système Pediatric Index of Mortality (PIM), le PIM3 étant la dernière version établie en 2013 et le système Pediatric Risk of Mortality (PRISM), le PRISM IV, dont la dernière version date de 2016. Les scores de DO, prenant en compte des paramètres physiologiques reflétant les principales DO, ont été développés et validés pour décrire les DO durant le séjour en réanimation. Le système de score de DO pédiatrique le plus utilisé est le Pediatric Logistic Organ Dysfunction (PELOD), dont la dernière version (PELOD-2) date de 2013. L'objectif de cette revue

est : 1) de donner un état des lieux actualisé des scores de gravité généraux et des DO en réanimation pédiatrique ; 2) d'explicitier les moyens de mesure de la performance et de l'adaptation des scores ; 3) de donner l'utilité, les limites et les perspectives des scores en réanimation pédiatrique.

Mots clés Score · Réanimation · Pédiatrie · Gravité · Dysfonction d'organe

Abstract Severity of illness and organ dysfunction scoring systems are increasingly used in adult and pediatric intensive care unit (PICU) to characterize disease severity and degree of organ dysfunction. In PICU, severity of illness scores are developed independently of diagnosis, taking into account the age of the different populations. Severity of illness scores were developed to better describe the severity of illness at baseline of groups of critically ill patients. Two systems are available: the pediatric index of mortality (PIM)—latest version PIM3 published in 2013, and the Pediatric RISK of Mortality (PRISM) system, which should be used in critically ill neonates, infants, children, or adolescents, (excluding premature infants)—latest version PRISM IV published in 2016. Organ dysfunction scores were developed to better describe the severity of illness during stay in the PICU. The Pediatric Logistic Organ Dysfunction score is the most commonly used pediatric organ dysfunction score—latest version PELOD-2 published in 2013. The purpose of this review is to provide an update about severity of illness and organ dysfunction scoring systems in PICU, to describe tools to evaluate performance and customization, and then discuss utility, limits and perspectives for scoring systems in PICU.

S. Leteurre (✉) · M.-E. Lampin · M. Recher
CHU Lille, service de réanimation pédiatrique,
F-59000 Lille, France
e-mail : stephane.leteurte@chru-lille.fr

S. Leteurre · M.-E. Lampin · B. Grandbastien · A. Duhamel
Univ. Lille, EA 2694,
Santé publique : épidémiologie et qualité des soins,
université de Lille, F-59000 Lille, France

B. Grandbastien
CHU Lille, Unité de lutte contre les infections nosocomiales,
service du risque infectieux, des vigilances et d'infectiologie,
F-59000 Lille, France

A. Duhamel
CHU Lille, Service de biostatistiques, F-59000 Lille, France

Keywords Scoring system · Pediatric · Severity of illness · Organ dysfunction

Rationnel

Les scores de gravité généraux et de dysfonctions d'organes (DO) sont de plus en plus utilisés dans les services de réanimation adultes et pédiatriques [1]. Cette utilisation est justifiée par la nécessaire évaluation objective de la gravité des patients, notamment pour les essais cliniques, mais également pour des aspects médicoéconomiques. Les scores disponibles sont modifiés dans le temps en réponse aux évolutions médicales. Compte tenu de l'utilité de ces scores pour les cliniciens, leur performance et les moyens de la mesurer doivent être parfaitement définis et explicités. Les limites d'utilisation des scores doivent aussi être connues tant sur le plan de l'impact individuel que sur le plan de l'analyse de la performance des services [2,3].

L'objectif de cette revue est :

- de donner un état des lieux actualisé des scores de gravité généraux et de DO en réanimation pédiatrique ;
- d'expliciter les moyens de mesure de la performance et de l'adaptation des scores ;
- de donner l'utilité, les limites et les perspectives de ces scores en réanimation pédiatrique.

Introduction

L'évaluation de la performance du système de soins en réanimation vise idéalement à comparer l'état de santé du patient immédiatement avant l'événement aigu justifiant l'admission en réanimation à l'état de santé du patient à distance de la sortie de réanimation. Si l'état de santé d'un enfant avant l'admission peut facilement être évalué lorsque ce dernier n'a aucun antécédent et que son admission est la conséquence d'un événement médical ou d'un accidentel brutal, le délai nécessaire à l'évaluation de l'état de santé à distance pour les enfants pose question. Ce délai peut être estimé à plusieurs années s'il s'agit, par exemple, de déterminer des capacités d'apprentissage et ne permet donc pas de répondre de façon pragmatique à cette évaluation de la performance du système de soins. Ainsi, depuis plusieurs décennies et afin de s'affranchir de ce délai, l'évaluation de la performance des soins en réanimation a cherché à comparer l'état de santé du patient à l'admission en réanimation à l'état de santé du patient à la sortie de réanimation [4].

Utilisation de la mortalité comme critère de jugement en réanimation

Dans les services de réanimation adulte et pédiatrique, la mortalité est respectivement d'environ 20 % [5] et 2,7 à 5 % [6] et représente un critère de jugement de type *gold*

standard. Ce critère de jugement objectif est établi soit à la sortie de réanimation [7,8], soit, plus rarement, à la sortie de l'hôpital [6]. Pour quantifier l'état de santé des patients à l'admission en réanimation, des scores de gravité à l'admission ont été développés et validés, prenant en compte des paramètres physiologiques recueillis durant les premières heures d'hospitalisation en réanimation. En réanimation pédiatrique, ces scores pronostiques (ou prédictifs de mortalité) sont établis, indépendamment du diagnostic, en prenant en compte l'hétérogénéité des populations en termes d'âge notamment [2,3]. Ils sont nommés scores de gravité généraux.

Parallèlement, durant le séjour en réanimation, la description et la quantification des DO sont apparues importantes dès les années 1990. En effet, la fréquence de survenue de ces DO est liée à la mortalité [9,10]. Ces DO peuvent exister dès l'admission ou apparaître durant le séjour en réanimation. Ainsi, des scores de DO, prenant en compte des paramètres physiologiques reflétant les principales DO, ont été développés et validés initialement pour décrire les DO durant le séjour en réanimation (et non comme des outils prédictifs de mortalité) [11,12]. Ainsi, le recueil quotidien du score Sequential Organ Failure Assessment (SOFA) chez l'adulte a permis d'établir que l'aggravation du score entre le jour j1 et le jour j2 était affectée d'une mortalité supérieure à 50 % quelle que soit l'évolution du score (aggravation, stagnation ou amélioration) entre j2 et j4 [13]. Chez l'enfant, le recueil quotidien du Pediatric Logistic Organ Dysfunction (PELOD) a permis d'établir que la mortalité des patients était supérieure à 50 % s'il existait une aggravation du score entre j1 et j2 et entre j2 et j5 [7]. Cependant, le recueil quotidien des scores de DO est très consommateur de ressources médicales. Des jours dits « ciblés » correspondant aux jours de recueil du score PELOD pour lesquels celui-ci est le plus lié à la mortalité durant le séjour en réanimation (Hasard ratios de mortalité significatifs pour chacun de ces jours ciblés) ont été déterminés : j1, j2, j5, j8, j12, j16 et j18 [7]. Ainsi, le critère de jugement (*gold standard*) pour le développement et la validation des scores de DO (scores SOFA, PELOD et PELOD-2) est également la mortalité. Cependant, il est établi qu'une fois construit, l'objectif de ces scores de DO est bien d'être utilisé comme un critère de jugement primaire ou secondaire, en analysant par exemple une variation du score entre deux jours successifs ou en quantifiant le nombre de DO (*new or progressive multiple organ dysfunction syndrome* — NPMODS) [14,15].

Quels sont les scores de gravité disponibles en réanimation pédiatrique ?

En réanimation pédiatrique, l'intérêt d'évaluer la gravité est renforcé par l'hétérogénéité des populations, du nouveau-né à l'adolescent, et par la variété des affections rencontrées.

Deux « systèmes » ont été proposés pour la population allant du nouveau-né (prématurés exclus) à l'adolescent.

Système du score PRISM

Le système du score PRISM peut être utilisé du nouveau-né à terme jusqu'à l'adolescent. La première version de 1984, nommée Physiologic Stability Index (PSI), incluait 24 variables [16]. En 1988, Pollack et al. publiaient une nouvelle version du score, nommée PRISM (Pediatric Risk of Mortality) [dénommée PRISM II par certains], qui incluait 14 variables [17]. En 1996, une nouvelle adaptation, le score PRISM III, incluait 17 variables (Annexe A) [18]. Les données du score PRISM III sont collectées dans les 12 ou 24 premières heures après l'admission en réanimation. La valeur la plus pathologique de chacune des variables est prise en compte sur la période (12 ou 24 heures). Les forces pour le score PRISM III sont :

- d'être validé à partir d'un échantillon de 11 165 patients issus de 32 services de réanimation pédiatrique aux États-Unis ;
- d'être adapté périodiquement à partir d'un site américain de collecte de données des services de réanimation pédiatrique [19]. De plus, il est possible de calculer la valeur du score PRISM III, indépendamment du calcul de la probabilité de décès.

Il existe deux principales limites au score PRISM III :

- la période relativement longue de recueil des données du score PRISM III (12 à 24 heures) n'est pas uniquement le reflet de la gravité initiale, mais aussi de la prise en charge durant cette période. Ainsi, un patient bénéficiant d'une prise en charge optimale dans un service pourra voir ses paramètres s'améliorer, alors qu'un patient identique dans un autre service et ayant un retard de prise en charge pourrait voir ses paramètres se dégrader. Le patient pris en charge de façon retardée aura un risque de décès supérieur au patient similaire bénéficiant d'une prise en charge optimale. La mortalité prédite dans le service à soins non optimaux objectivera un état de gravité des patients plus important, reflet non pas de la gravité initiale, mais de la prise en charge ;
- les coefficients de chaque variable nécessaires au calcul de la probabilité de décès ne sont pas dans le domaine public. L'utilisation de l'équation du score PRISM III nécessite le paiement de droits aux auteurs.

Début 2016, la version du PRISM IV vient d'être publiée [6]. Les données ont été obtenues sur une période entre 2011 et 2013, incluant une cohorte prospective de 10 078 admissions en réanimation (nouveau-nés à 18 ans), issues de sept services nord-américains. Le critère de jugement est le caractère vivant/décédé à la sortie de l'hôpital, après la première

admission en réanimation pédiatrique. La période de recueil est entre les deux heures avant l'admission en réanimation (prise en charge Smur) et les quatre premières heures après l'admission en réanimation. Les variables collectées et leurs catégorisations sont identiques au score PRISM III [20]. L'équation de calcul de la probabilité de décès est donnée et prend en compte les critères d'âge, de provenance, d'arrêt cardiaque dans les 24 heures précédentes, de cancer, de dysfonction principale à bas risque à l'admission et de scores selon les catégories des variables neurologiques et non neurologiques (Annexe B) [6].

Système du score PIM

Le système du score PIM peut être utilisé chez les nouveau-nés à terme jusqu'aux enfants âgés de 16 ans. La première version du score PIM (Pediatric Index of Mortality) de 1997 incluait huit variables collectées durant la première heure après l'admission en réanimation pédiatrique [21]. En 2003, le PIM2 était développé et validé sur 20 787 patients en Australie et au Royaume-Uni et incluait dix variables [22]. Le PIM3, développé en 2013 à partir d'un échantillon de 53 112 patients, inclut dix variables [8]. Les variables collectées sont identiques entre les versions des scores PIM2 et PIM3, mais les items des variables ont été réorganisés (diagnostics à bas risque et à haut risque dans le PIM2 redistribués en diagnostic à bas risque, haut risque et très haut risque dans le PIM3) (Annexe C). Les forces du score PIM3 sont la taille importante de la population de validation, l'évaluation quantitative dès la première heure après l'admission en réanimation et la publication des coefficients de chacune des variables et de l'équation de probabilité de décès. La principale limite du score PIM3 est d'avoir été construit uniquement à partir des variables du score PIM2 (sans tester de nouvelles variables potentielles). Une étude récente dans 17 services de réanimation pédiatrique italiens et incluant 11 109 patients montrait une bonne performance du PIM3 [23].

Ces deux « systèmes » présentent des caractéristiques différentes qui peuvent orienter le choix vers l'un ou l'autre selon les priorités choisies (Tableau 1). Cependant, certains auteurs mentionnent que les scores disponibles ne sont pas adaptés aux pays en voie de développement [24].

Quels sont les scores de dysfonctions d'organes actuellement disponibles ?

Le score PELOD a été développé sur une population de 594 patients dans trois services de réanimation en 1999 [25], puis validé sur une population externe de 1 806 patients dans sept services de réanimation (Canada, France, Suisse) [11,12]. Le score PELOD contenait six DO et 12 variables. La principale limite du score PELOD était de présenter des valeurs inobservables sur une échelle discrète de 0 à 71

Tableau 1 Principales caractéristiques comparatives entre les scores PRISM IV et PIM3		
	PRISM IV [6]	PIM3 [8]
Population de validation	Nord-américaine <i>n</i> = 10 078	Européenne/australienne <i>n</i> = 53 112
Nombre de variables	17	10
Durée d'observation nécessaire pour le recueil	6 heures	1 heure
Obtention d'une valeur du score	Oui	Non
Obtention d'une probabilité de décès	Oui	Oui
Équation de calcul de la probabilité de décès	Gratuite	Gratuite
Validation externe	Non	Oui [23]
PRISM : Pediatric Risk of Mortality ; PIM : Pediatric Index of Mortality.		

(valeurs possibles du score : 0, 1, 2, 3, 4, 10, 11, 12, 13, 14, 20, etc.). Cela posait des difficultés d'interprétation lors du calcul des moyennes ou médianes du score [26]. Le score PELOD-2, développé et validé en 2013 à partir d'un échantillon de 3 761 patients issus de 15 services (France et Belgique), comporte cinq DO et dix variables [27]. Concernant les DO et les variables, les principales différences entre les deux versions sont la suppression de la DO hépatique (non statistiquement significative) et le remplacement des variables pression artérielle systolique (statistiquement significative, mais redondante avec la pression artérielle moyenne) et fréquence cardiaque (non statistiquement significative) par respectivement pression artérielle moyenne et lactatémie dans le score PELOD-2 (Annexe D). De plus, pour le score PELOD-2, toutes les valeurs discrètes entre 0 et 33 points sont possibles (Annexe D). Le recueil du système PELOD est fondé sur un recueil journalier sur une période de 24 heures, à partir de l'horaire d'admission. Les jours de recueil les plus pertinents (dits « jours ciblés ») pour prédire la mortalité ont pu être déterminés (j1, j2, j5, j8, j12, j16 et j18) pour les scores PELOD et PELOD-2 [7,28]. Les équations de calcul des probabilités de décès du système PELOD ont été publiées. Compte tenu des évolutions de ce score, il est aujourd'hui préférable de privilégier l'utilisation du score PELOD-2.

Le score Pediatric Multiple Organ Dysfunction (P-MODS) a été développé et validé dans un seul service américain incluant 6 456 patients en 2005 [29]. Le score P-MODS inclut cinq DO (cardiovasculaire, respiratoire, rénale, hématologique, hépatique), mais pas la dysfonction neurologique. Chacune de ces cinq DO est caractérisée par une variable biologique. Le score P-MODS n'a jamais fait l'objet de validation

externe publiée. L'équation de calcul de probabilité de décès n'a pas été publiée [29].

Calcul des probabilités de décès

Le calcul de la probabilité de décès des scores en réanimation peut être réalisé de deux façons différentes selon les scores :

- les systèmes PRISM et PELOD permettent de calculer une valeur du score pour chaque patient. Cette valeur du score est transformée en une probabilité de décès grâce à une équation. Cette équation est librement disponible pour le système PELOD (Annexe D), les scores PRISM et PRISM IV, et est disponible après paiement d'une licence pour le score PRISM III ;
- le système PIM (scores PIM, PIM2 et PIM3) ne permet pas de calculer la valeur du score, mais permet de calculer directement la probabilité de décès. L'équation de calcul est directement intégrée à partir des variables et est disponible librement pour ces scores. Le score PIM3 est présenté dans l'Annexe C.

Quelle que soit la modalité d'obtention de la probabilité de décès, cette probabilité sera « confrontée » à l'état vivant/décédé à la sortie de réanimation (ou dans certains cas à la sortie de l'hôpital).

Qualité des scores en réanimation

Les qualités requises d'un score de gravité sont multiples, avant qu'il puisse être utilisé de façon courante. Les variables qui le composent doivent être pertinentes par rapport aux recommandations médicales, usuelles, objectives, de recueil facile, rapide et précoce après l'admission. Le score pronostique doit avoir une bonne reproductibilité intra- et interobservateur et une capacité à détecter des variations fines de gravité entre les patients (sensibilité au changement) ; il doit être validé après comparaison à d'autres scores ou indices pronostiques traditionnellement reconnus, être « acceptable » pour le patient, simple d'utilisation pour le médecin, de faible coût et « faisable » dans tout service susceptible de l'appliquer [24]. Tous ces critères de qualité justifient l'actualisation régulière des scores de gravité et de DO [2,15].

Performance et adaptation des scores

Performance des scores : qu'évalue-t-on et comment mesurer la performance ?

Les scores utilisés en réanimation doivent aider les cliniciens à répondre à certaines questions :

- « sont-ce bien les patients les plus graves qui décèdent dans mon service ? » ;

- « ce score est-il capable de prédire le nombre de décès dans mon service ? » ;
- « est-ce que ce nombre de décès prédits est cohérent avec le niveau de gravité mesuré par le score? ».

Les outils statistiques utilisés pour évaluer la performance des scores et permettre de répondre successivement à ces trois questions sont décrits ci-dessous.

La discrimination du score permet de répondre à la première question

Pour les patients qui vont vivre, les valeurs des scores devraient être plus basses que celles observées pour les patients qui vont décéder. La discrimination d'un score consiste à mesurer la capacité d'un score à « affecter » des valeurs du score (ou des probabilités de décès pour les scores de gravité généraux) les plus basses aux patients qui vont vivre et à « affecter » des valeurs (ou des probabilités de décès) du score les plus hautes aux patients qui vont décéder. La discrimination peut être évaluée soit à partir de la valeur du score, soit à partir de la probabilité de décès calculée à partir du score. En effet, la transformation permettant de passer du score à la probabilité de décès est une fonction monotone (fonction logarithmique), qui ne modifie pas l'ordre du classement entre valeur du score et probabilité de décès (Annexe E). De plus, la discrimination du score tient compte uniquement du classement des patients en fonction de la valeur du score ou de la probabilité de mortalité, cela indépendamment des valeurs des scores ou des probabilités de mortalité obtenues. Ainsi, il serait théoriquement possible que tous les patients d'un service soient parfaitement classés entre les vivants et les décédés sur une tranche de probabilité de décès entre 1 et 13 % par exemple. Dans cet exemple, si tous les patients survivants étaient classés entre 1 et 4 % et tous les patients décédés entre 5 et 13 %, le score aurait une discrimination parfaite, alors qu'aucun patient n'aurait une probabilité de décès supérieure à 13 % : on perçoit ainsi une limite de ce critère de discrimination, puisque la valeur de la probabilité elle-même (ou la valeur du score) obtenue n'est pas prise en compte (mais uniquement le classement).

La discrimination est évaluée par le calcul de l'aire sous la courbe ROC (*area under the receiver operating curve* — AUC — des Anglo-Saxons). La courbe ROC est obtenue en faisant varier successivement les seuils du score et en calculant pour chacun des seuils, les sensibilités et spécificités. Un exemple de calcul des sensibilités et spécificités en fonction des seuils est décrit dans l'Annexe E. La courbe ROC représente la variation de (1-spécificité) en fonction de la sensibilité du score (Fig. 1). L'aire sous la courbe s'interprète de la manière suivante : une aire sous la courbe ROC égale à 0,50 signifie que le score n'est pas plus discriminant que le hasard, une aire entre 0,70 et 0,79 est consi-

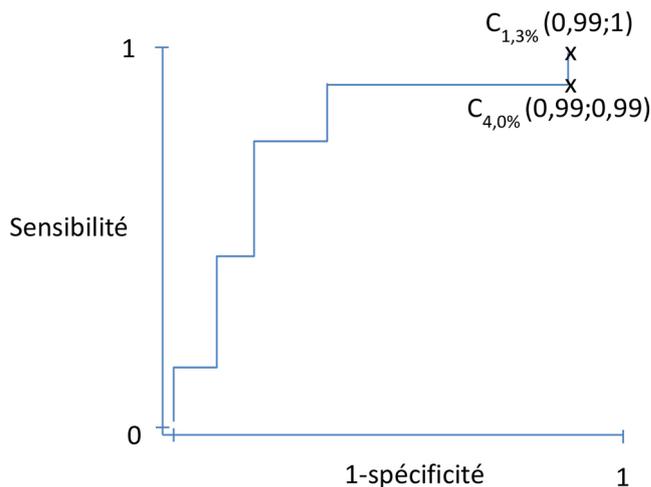


Fig. 1 Exemple de courbe ROC pour un score représentant l'exemple de l'Annexe E

dérée comme correcte, une aire entre 0,80 et 0,89 est considérée comme bonne et une aire supérieure à 0,90 excellente [30,31]. Un intervalle de confiance est calculé, dont la borne supérieure ne peut être supérieure à 1 [32].

Le calcul du ratio standardisé de mortalité a pour objectif de répondre à la deuxième question

Le ratio standardisé de mortalité (RSM ou Standardized Mortality Ratio — SMR — des Anglo-Saxons) est défini comme le rapport entre le nombre de décès observés divisé par le nombre de décès prédits. Le nombre de décès observés est le nombre de décès sur la population étudiée (69 dans l'exemple de l'Annexe E). Le nombre de décès prédits est obtenu en effectuant la somme de toutes les probabilités de décès des patients de la population (74,2 dans l'exemple de l'Annexe E). Lorsque le SMR est inférieur à 1, le nombre de décès observés est inférieur au nombre de décès prédits (exemple : Annexe E). Lorsque le SMR est supérieur à 1, le nombre de décès observés est supérieur au nombre de décès prédits. Un intervalle de confiance est déterminé pour le SMR : si l'intervalle de confiance du SMR inclut 1, la différence entre le nombre de décès observés et le nombre de décès prédits n'est pas significative (exemple : Annexe E). Si l'intervalle de confiance n'inclut pas 1, l'écart entre le nombre de décès observés et le nombre de décès prédits est significatif.

La calibration du score permet de répondre à la troisième question

La calibration est la mesure de l'adéquation entre la mortalité prédite et la mortalité observée par niveau de gravité. Les niveaux de gravité peuvent être définis de plusieurs façons.

Généralement, on considère dix groupes (ou classes) de niveau de gravité : 0–10 %, 10–20 %, etc. : les patients sont classés dans le sous-groupe de leur probabilité de décès [33]. Cette classification peut aboutir à un déséquilibre du nombre de patients par sous-groupe (plus de patients dans certains sous-groupes et très peu de patients dans d'autres sous-groupes). Ainsi, il est également possible de trier les patients par ordre croissant de leur probabilité de décès puis de considérer entre cinq et dix groupes comportant un même nombre de patients. Pour dix groupes, ces groupes correspondent aux déciles des probabilités prédites [34]. Dans chacun des groupes, on calcule deux effectifs prédits : le nombre de décès prédits (qui correspond à la somme des probabilités prédites de décès sur tous les individus du groupe) et le nombre de vivants prédits (qui est la somme, sur tous les individus du groupe, de 1 moins la probabilité prédite de décès du groupe). Si l'on considère les décès, il est attendu en général deux éléments :

- le nombre de décès observés et de décès prédits est moindre dans les sous-groupes ayant une probabilité de décès basse comparativement aux sous-groupes à haute probabilité de décès ;
- dans chacun des groupes, le nombre de décès observés est proche du nombre de décès prédits.

Le test statistique de Goodness of Fit de Hosmer-Lemeshow effectue une comparaison globale des effectifs observés (décès et vivants) aux effectifs théoriques à l'aide du test du χ^2 [33,34]. Une valeur de p est déduite, après avoir défini le nombre de degrés de liberté (ddl). Le nombre de ddl est égal au nombre de sous-groupes moins deux (huit dans notre exemple) pour le développement du score. Le nombre de ddl est égal au nombre de groupes pour la validation des scores [35]. Puisqu'il est attendu l'absence de différence entre le nombre de décès observés et le nombre de décès prédits, la calibration du score est bonne (ou adéquate) lorsque le test n'est pas significatif au niveau 5 %, et on attend une valeur du p supérieure à 0,05 [31]. La calibration est un test exigeant : si le nombre de décès observés est très différent du nombre de décès prédits dans un seul groupe, la calibration du score risque d'être mauvaise ($p < 0,05$) [36,37]. Lorsque la calibration d'un score est bonne sur une population, il est permis de conclure que le nombre de décès observés est proche du nombre de décès prédits et que cette adéquation est fonction du niveau de gravité des patients.

Adaptation des scores

Les scores de gravité généraux ont pour objectif d'établir une probabilité de décès à l'admission en réanimation (à h1 ou dans les premières 24 heures) ; ce sont des scores prédictifs (ou pronostiques). Ainsi, la discrimination et la calibra-

tion sont les tests habituellement utilisés pour la validation de ces scores. Par contre, les scores de DO ont pour objectif d'évaluer les DO durant le séjour en réanimation et ne sont pas des scores prédictifs de mortalité [1,15]. Ainsi pour les scores de DO, seul le critère de discrimination est nécessaire pour évaluer la performance, le critère de calibration n'étant pas justifié. Cependant, certains auteurs ont testé ou comparé les scores de gravité et les scores de DO comme des outils pronostiques et effectué le calcul de la calibration pour les deux types de scores. Les scores de DO étaient également fréquemment pertinents dans cet objectif [38].

Un score de gravité général ou de DO est développé puis validé sur une population générale de réanimation. Il est attendu que l'application de ce score sur une nouvelle population dans un lieu différent (validation externe géographique) puisse permettre de valider ce score sur cette nouvelle population. Cependant, l'expérience des validations externes a montré que les scores testés sur une nouvelle population avaient dans la grande majorité des cas une mauvaise calibration [39]. L'explication la plus reconnue est non pas une modification des performances d'une équipe par rapport à une autre, mais essentiellement un recrutement des services différents de par des particularités régionales ou nationales d'organisation des soins [40,41]. Ainsi, l'utilisation de l'équation de probabilité initiale ne permet pas de calculer une calibration fiable sur une nouvelle population externe. Il est reconnu comme nécessaire d'évaluer la calibration des scores en effectuant une adaptation (ou customisation des Anglo-Saxons) du score à la nouvelle population testée [5], même si cela compromet la comparabilité avec la population d'origine [42]. Il existe trois niveaux de customisation du score :

- la customisation de premier niveau consiste à affecter un coefficient correcteur global au score calculé afin de l'adapter à la nouvelle population (mais sans modifier les variables ou les coefficients affectés à chaque variable) [5]. Malheureusement, de nombreux auteurs s'affranchissent de cette étape d'adaptation et utilisent les scores de gravité et de DO comme des outils de prédiction et concluent de façon hâtive à une mauvaise calibration d'un score, sans avoir effectué ce premier niveau d'adaptation [26]. Cette customisation de premier niveau ne répond cependant pas à toutes les problématiques d'actualisation ;
- en effet, l'amélioration des soins permet de faire baisser la mortalité dans les services de réanimation au fil des années. Ainsi, les coefficients affectés à chaque variable dans les équations de calcul de la probabilité de mortalité perdent de leur exactitude. La customisation de second degré consiste à conserver chaque variable et à recalculer le coefficient affecté à chacune des variables en prenant en compte la mortalité de la nouvelle population testée. Ce type d'adaptation est régulièrement utilisé pour le score

PRISM III dont les coefficients des variables et l'équation ne sont pas publiques ;

- enfin, les scores sont établis à un moment donné, prenant en compte les variables cliniques et biologiques d'évaluation disponibles. Au fil des années, certains outils d'évaluation sont moins utilisés (PaO₂ sanglante par exemple), et d'autres plus pertinents apparaissent (PA moyenne, lactatémie...). La mise à jour complète du score consiste ainsi à actualiser les variables incluses dans les scores (et à calculer le coefficient de chaque variable du nouveau score). On change alors de version du score (PIM2 puis PIM3, PRISM III puis PRISM IV ou PELOD puis PELOD-2).

Pourquoi utiliser un score de gravité ou de dysfonctions d'organes en réanimation pédiatrique ?

Évaluation objective de la gravité et des dysfonctions d'organes d'un patient

L'évaluation de la gravité et des DO des patients est un objectif primordial dans les services de réanimation. Pour parvenir à cet objectif, le médecin en charge du patient prend en compte des données cliniques et paracliniques. La probabilité de décès calculée ne peut en aucun cas servir à une décision diagnostique ou thérapeutique individuelle pour la prise en charge du patient. En effet, sur un groupe de dix patients ayant chacun une probabilité de décès de 80 %, si une décision de limitation des thérapeutiques était prise pour chaque patient, les dix patients décèderaient. Or, il est « prévisible » que parmi ces dix patients ayant chacun une probabilité de mortalité de 80 %, deux (non identifiables par le calcul) devraient survivre. La probabilité de décès n'est donc interprétable qu'à l'échelle d'un groupe de patients [43].

Utilisation dans les études

Description du recrutement

Une première utilisation des scores de gravité et de DO est de permettre une description des patients inclus dans les études afin de caractériser la population étudiée [44].

Critère d'inclusion dans les études

Seule l'étude PROWESS, par une étude ancillaire, a montré un bénéfice de l'administration de drotrecogin alfa (activé) pour les patients ayant un score de gravité APACHE supérieur à 25 [45]. Cependant, l'utilisation d'un score de gravité comme critère d'inclusion dans les essais est très controversée. Les scores de gravité et de DO ne doivent pas être utilisés dans cet objectif [43]. Une stratification, en fonction de

la gravité évaluée par les scores, doit être préférée dans le plan d'analyse des résultats.

Outils de comparaison de gravité dans les essais randomisés

Les systèmes PRISM, PIM et PELOD permettent d'étudier la comparabilité des groupes dans les essais randomisés. Dans l'étude pédiatrique Transfusion Requirements in the Pediatric Intensive Care Unit (TRIPICU) recherchant le meilleur seuil transfusionnel de culots globulaires, le score PRISM était comparable après randomisation, entre le groupe « stratégie libérale : transfusion à un seuil de 9,5 g/dl » et le groupe « stratégie restrictive : transfusion à un seuil de 7 g/dl » [44]. Dans cette même étude pédiatrique, le critère de jugement principal était l'apparition ou l'aggravation d'une défaillance d'organe (*new or progressive MODS*). L'un des critères de jugement secondaires était le score PELOD. L'absence de différence significative entre les deux groupes pour les critères de jugement principal et secondaire et l'absence de différence pour la survenue de complications permettaient de recommander la stratégie restrictive de transfusion en culots globulaires chez les patients stables en réanimation pédiatrique [44]. Les scores de gravité et de DO peuvent également servir de critères d'ajustement dans les essais cliniques [46–48].

Évaluation du recrutement et de la performance d'un service

Le calcul d'un score de gravité et d'une probabilité de décès peut permettre d'apprécier l'évolution du recrutement et de déterminer le SMR d'un service. De même, il est envisageable d'effectuer cette approche dans plusieurs services et de comparer les services. Cette dernière utilisation a cependant des limites. En effet, si les scores de gravité généraux sont construits idéalement pour s'affranchir des diagnostics des patients et être appliqués sur toutes les populations de réanimation, il est apparu que le recrutement ou l'organisation des services (chirurgie cardiaque dans un centre, orientation néonatale dans un autre centre, politique d'admissibilité ou de sortie, existence d'une unité de surveillance continue d'aval, etc.), modifiaient la valeur du SMR et qu'une adaptation des scores pouvait être nécessaire afin de permettre une comparabilité [46–48].

Perspectives

Scores d'évaluation de la qualité de vie en réanimation

La réduction de la mortalité était le premier objectif de la création des services de réanimation pédiatrique. En Australie, la mortalité observée dans les services de réanimation

pédiatrique était de 4,7 % en 1996 ($n = 1\ 161$) [49]. Le registre américain des services de réanimation pédiatrique mentionnait un taux de mortalité de 3 % entre 2005 et 2008 ($n = 80\ 739$ patients) [50]. L'étude comparant des populations françaises et anglaises sur la période 2006–2007 objectivait respectivement un taux de mortalité de 7,4 % ($n = 5\ 602$ patients français) et de 4,9 % ($n = 20\ 693$ patients anglais) [47]. Ces variations internationales des taux de mortalité, établis sur des pays aux niveaux de développement comparables, s'expliquent probablement par des critères de recrutement différents et l'inclusion ou non des unités de surveillance continue (*intermediate care* des Anglo-Saxons). Dans tous les pays, il existe cependant comme point commun une réduction progressive de la mortalité. Ainsi, l'étude de Namachivayam et al. montrait que la mortalité observée dans une population australienne était de 4,7 % en 1996 ($n = 1\ 161$) et de 3,3 % en 2005–2006 ($n = 2\ 550$) [49]. En parallèle de cette diminution du taux de mortalité, l'étude de Pollack et al. objectivait que le taux des nouvelles morbidités était fréquent (4,6 %), survenant chez tous les types de patients [51]. Une conséquence récente de cette baisse de mortalité et de l'augmentation des morbidités a été de proposer une modification des critères de jugement établis à la sortie de réanimation. Ainsi, Pollack et al. ont développé et validé un outil prédictif établi à l'admission, prenant comme critère, un critère de jugement ternaire : vivant sans nouvelle morbidité, vivant avec une morbidité ou décédé [51]. Le statut de morbidité était quantifié en utilisant l'échelle développée par la même équipe en 2009 (échelle de performance fonctionnelle) qui prend en compte six domaines (conscience, sensorialité, communication, motricité, alimentation et respiration), avec pour chacun des domaines une quantification entre 1 (normal) et 5 (dysfonction très sévère) [52].

L'étape ultérieure est de quantifier la morbidité à moyen terme après la sortie de réanimation [53]. La revue d'Aspesberro et al. identifiait quatre échelles d'évaluation de la qualité de vie utilisables dans les essais en réanimation pédiatrique : l'échelle Pediatric Quality of life Inventory version 4.0 (Peds QL 4.0 Generix Core Scale) pour des enfants de 2 à 18 ans [54] bénéficiant d'une traduction française [55], l'échelle KIDSCREEN-27 [56] pour des enfants de 8 à 18 ans, KINDL pour des enfants de 6 à 18 ans et Child Health Questionnaire-Parent Form (CHQ-PF28) pour des enfants de 5 à 18 ans [53].

Évaluation des pratiques des services de réanimation pédiatrique

Les scores de gravité et de DO sont un des outils d'évaluation des pratiques dans les services de réanimation pédiatrique.

La mise à disposition des moyens de recueil des scores, le contrôle de qualité des données, l'adaptation (customisation) des scores et leur mise à jour régulière nécessitent un nombre important de patients regroupant plusieurs services de réanimation pédiatrique [5]. De plus, les résultats méritent d'être analysés collectivement, en connaissance des caractéristiques locales pour éviter des interprétations erronées. Ainsi, la nécessité d'un suivi national annuel des activités médicales et médicoéconomiques a conduit au développement de réseaux nationaux des services de réanimation pédiatrique dans de nombreux pays industrialisés : Amérique du Nord (Virtual PICU Performance System, « VPS » <https://portal.myvps.org/>) [19], Grande-Bretagne (Pediatric Intensive Care Audit Network, « PICAnet » <http://www.picanet.org.uk/>) [57] et Australie — Nouvelle-Zélande (Australian and New Zealand Pediatric Intensive Care Society, « ANZPICS » <http://www.anzics.com.au/pages/CORE/ANZPICR-registry.aspx>) [58]. À titre d'exemple, le réseau PICAnet augmente sa base de données de 16 000 patients par an. Le but de ces réseaux de collectifs des services de réanimation pédiatrique est de constituer une base de données dont les objectifs sont l'identification des meilleures pratiques, l'évaluation de l'offre et de la demande au niveau local, régional et national, une meilleure planification de la stratégie des soins de santé et une quantification des ressources nécessaires, le suivi de l'épidémiologie des maladies, du recrutement des services, des indicateurs de résultats (tels que la mortalité et la morbidité, et événements indésirables dont les infections nosocomiales) et la conduite d'un plus grand nombre d'études cliniques multicentriques.

Conclusion

Les scores en réanimation pédiatrique sont en perpétuelle évolution. L'utilisation d'un score de gravité général et d'un score de DO en réanimation pédiatrique est nécessaire pour objectiver respectivement la gravité à l'admission et durant le séjour en réanimation. La connaissance de leur actualisation est nécessaire, et les limites de l'interprétation de leurs résultats doivent être parfaitement connues, que ce soit pour le clinicien dans sa prise en charge individuelle (pronostic individuel, inclusion dans des protocoles...) ou sur le plan de l'analyse de la performance (nécessité des adaptations régulières avant toute conclusion [42]). Les perspectives d'un recueil automatisé des données permettant leur calcul, facilité par l'informatisation des services, sont une nécessité que les industriels doivent prendre en compte.

Liens d'intérêts : Les auteurs déclarent ne pas avoir de lien d'intérêt.

Annexe A. Variables des scores PRISM III et PRISM IV [6,18]

Pondérations	1	2	3	4	5	6	7	11	Total
Signes cardiovasculaires et neurologiques									
Pression Art. systolique (mmHg)									
Nouveau-né			40-55				< 40		
Nourrisson			45-65				< 45		
Enfant			55-75				< 55		
Adolescent			65-85				< 65		
Fréquence cardiaque (/min)									
Nouveau-né			215-225	> 225					
Nourrisson			215-225	> 225					
Enfant			185-205	> 205					
Adolescent			145-155	> 155					
Température (°C)									
			<33 ou >40						
Réaction pupillaire									
							1 réactive	2 aréactives	
État de conscience									
					Stupeur/coma (Glasgow < 8)		1 aréactive		
Acide base / gaz du sang									
Acidose : CO₂ total (mmol/l) ou pH									
		5-16,9 ou pH 7,0-7,28					<5 ou pH<7		
pH									
		7,48-7,55	>7,55						
pCO₂ (mmHg)									
	50-75		>75						
CO₂ total (mmol/l)									
				>34					
PaO₂ (mmHg)									
			42,0-49,9				<42		
Glycémie/ionogramme									
Glucose mg/dL (mmol/L)									
		>200 (>11)							
Potassium (mmol/L)									
			>6,9						
Créatinine mg/dL (µmol/L)									
Nouveau-né		>0,85 (>75)							
Nourrisson		>0,90 (>80)							
Enfant		>0,90 (>80)							
Adolescent		>1,30 (115)							
Urée g/L (mmol/L)									
Nouveau-né			>0,26 (>4,3)						
Autres âges			>0,33 (>5,4)						
Hématologie									
Globules blancs (/mm³)									
				< 3 000					
Plaquettes (/mm³)									
		100 000-200 000		50 000-99 999	< 50 000				
Temps de prothrombine (TP) ou Temps de céphaline activé (TCA)									
Nouveau-né			TP>22 ou TCA >85						
Autres âges			TP > 22 ou TCA > 57						
Autres facteurs									
<input type="checkbox"/> maladie cardiovasculaire non opératoire <input type="checkbox"/> anomalie chromosomique <input type="checkbox"/> cancer <input type="checkbox"/> précédente admission en réanimation pédiatrique <input type="checkbox"/> RCP avant l'admission en réanimation <input type="checkbox"/> postopératoire <input type="checkbox"/> diabète <input type="checkbox"/> transfert d'une autre unité (hormis les postopératoires)									

PRISM : Pediatric RISK of Mortality ; Nouveau-né : < 1 mois ; Nourrisson : 1-12 mois ;
 Enfant : 13 mois-12 ans ; Adolescent > 12 ans ; RCP : réanimation cardiopulmonaire.

Annexe B. Modalité de calcul de la probabilité de décès du score PRISM IV [6]

Variable	Coefficient ^c	Odds ratio de mortalité ^c (IC 95 %)
Constante	-5,776	
Âge		
0 à 14 jours	1,311	3,708 (2,251–6,107)
14 jours à 1 mois	0,968	2,632 (0,891–7,773)
1 à 12 mois	0,357	1,429 (0,956–2,135)
Provenance vs (bloc opératoire ou salle de réveil)		
Autre hôpital	1,012	2,750 (1,739–4,349)
Transfert interne de l'hôpital	1,626	5,085 (3,124–8,278)
Service d'urgence	0,693	1,999 (1,224–3,263)
Réanimation cardiopulmonaire dans les 24 h avant l'admission en réanimation (oui/non)	1,082	2,949 (1,580–5,507)
Cancer (aigu ou chronique)	0,766	2,152 (1,304–3,551)
Dysfonction primaire à bas risque^a	-1,697	0,183 (0,056–0,600)
Variable du score PRISM III^b		
Neurologique	0,197	1,218 (1,176–1,261)
Non neurologique	0,163	1,177 (1,147–1,207)

PRISM : pediatric RiSk of Mortality

^a Dysfonction d'organe primaire de type : endocrine, hématologique, musculosquelettique et rénal

^b Pour un point d'augmentation de chacune des variables du score. Variable neurologique : réactivité pupillaire et état de conscience. Variable non neurologique : fréquence cardiaque, pression artérielle systolique, température, PaO₂, pH, PCO₂, bicarbonates, glycémie, kaliémie, urée, créatinine, globules blancs, plaquettes, temps de céphaline activé (TCA), temps de prothrombine (TP).

^c Seule la colonne des coefficients doit être utilisée pour déterminer la probabilité de décès (cf. exemple ci-dessous). La colonne des odds ratio est donnée à titre indicatif du risque de mortalité lié à chaque variable

Logit PRISM IV = - 5,776

+ 1,311 si âge entre 0 et 14 jours ou + 0,968 si âge entre 14 jours et 1 mois ou + 0,357 si âge entre 1 et 12 mois ou + 0 si âge > 1 an

+ 1,012 si autre hôpital ou 1,626 si transfert interne ou 0,693 si transfert des urgences

+ 1,082 si réanimation cardiopulmonaire dans les 24h avant l'admission en réanimation

+ 0,766 si cancer

- 1,697 si dysfonction primaire de bas risque

+ scores des variables neurologiques X 0,197

+ scores des variables non neurologiques X 0,163

Probabilité de décès PRISM IV = 1 / [1 + e^{-logit PRISM IV}]

Annexe C. Score PIM3 [8]

Variable / Item	Coefficient ^a	Odds ratio de mortalité ^a (IC 95 %)
Constante	-1,7928	0,1665 (0,1027-0,27)
Admission programmée (élective) (Oui/Non)		
Oui	-0,5378	0,584 (0,4851-0,7031)
Admission postopératoire (Oui/Non)		
Oui :		
admission postopératoire sans chirurgie cardiaque	-1,5164	0,2195 (0,1653-0,2914)
admission postopératoire de chirurgie cardiaque sans CEC	-0,8762	0,4164 (0,2889-0,6001)
admission postopératoire de chirurgie cardiaque après CEC	-1,2246	0,2939 (0,225-0,3838)
Affection à TRÈS haut risque (Oui/Non)		
Oui :	1,6225	5,0657 (4,3517-5,8968)
-arrêt cardiaque intra- ou extrahospitalier		
-déficit immunitaire combinée sévère		
-leucémie ou lymphome après première cure d'induction		
-receveur de transplantation de moelle osseuse		
-insuffisance hépatique (principale cause d'admission)		
Affection à HAUT risque (Oui/Non)		
Oui :	1,0725	2,9228 (2,4771-3,4487)
-hémorragie cérébrale spontanée		
-cardiomyopathie ou myocardite		
-hypoplasie du cœur gauche		
-maladie neurodégénérative		
-entérocolite nécrosante		
Affection à BAS risque (Oui/Non)		
Oui :	-2,1766	0,1134 (0,0835-0,154)
-admission pour crise d'asthme		
-admission pour bronchiolite		
-admission pour laryngite		
-admission pour apnée obstructive du sommeil		
-admission pour acidocétose diabétique		
-admission pour convulsion		
Mydriase bilatérale aéroactive et fixe (Oui/Non)		
Oui	3,8233	45,7554 (31,7561-65,926)
Ventilation invasive (Oui/Non)		
Oui	0,9763	0,584 (0,4851-0,7031)
PAS minimale valeur (si inconnu : 120)		
PAS	-0,0431	0,9578 (0,949-0,9668)
PASXPAS/1000	0,1716	1,1872 (1,1329-1,244)
Base excess (mmol/L)		
Valeur absolue (si inconnu : 0)	0,0671	1,0694 (1,0593-1,0797)
[FiO2X100]/PaO2		
Valeur (si oxygène par « cloche » sans FiO2 connue : 0,23)	0,4214	1,5241 (1,3928-1,6678)

PIM : Paediatric Index of Mortality ; CEC : Circulation ExtraCorporelle ; PAS : Pression artérielle Systolique

^a Seule la colonne des coefficients doit être utilisée pour déterminer la probabilité de décès (cf. exemple ci-dessous). La colonne des odds ratio est donnée à titre indicatif du risque de mortalité lié à chaque variable

Logit PIM 3 = -1,7928
+ Admission programmée (0/1) x -0,5378
+ Admission postopératoire (0/1) x :
* admission postopératoire sans chirurgie cardiaque : -1,5164
* ou admission de chirurgie cardiaque sans CEC : -0,8762
* ou admission de chirurgie cardiaque après CEC : -1,2246
+ Affection à très haut risque (0/1) x 1,6225
+ Affection à haut risque (0/1) x 1,0725
+ Affection à bas risque (0/1) x -2,1766
+ Mydriase bilatérale aréactive et fixe (0/1) x 3,8233
+ Ventilation invasive (0/1) x 0,9763
+ PAS minimale (valeur) x -0,0431
+ [PAS² (valeur) / 1000] x 0,1716
+ Base excess (valeur absolue) x 0,0671
+ [(100 x FiO₂/PaO₂) (valeur)] x 0,4214
Probabilité de décès PIM3 = 1 / [1 + e^{-logit PIM3}]

Annexe D. Score Pediatric Logistic Organ Dysfunction (PELOD)-2 score [27]

Dyfonctions d'organe et variables ^a	Points						
	0	1	2	3	4	5	6
Neurologique^b							
Score de Glasgow	≥11	5-10			3-4		
Réaction pupillaire	Absence de mydriase bilatérale					Mydriase bilatérale	
Cardiovasculaire^c							
Lactatémie (mmol/L)	<5,0	5,0-10,9			≥11,0		
Pression artérielle moyenne (mmHg)							
0 à < 1 mois	≥46		31-45	17-30			≤16
1 à 11 mois	≥55		39-54	25-38			≤24
12 à 23 mois	≥60		44-59	31-43			≤30
24 à 59 mois	≥62		46-61	32-44			≤31
60 à 143 mois	≥65		49-64	36-48			≤35
≥144 mois	≥67		52-66	38-51			≤37
Rénal							
Créatinine (µmol/L)							
0 à < 1 mois	≤69		≥70				
1 à 11 mois	≤22		≥23				
12 à 23 mois	≤34		≥35				
24 à 59 mois	≤50		≥51				
60 à 143 mois	≤58		≥59				
≥144 mois	≤92		≥93				
Respiratoire^d							
PaO₂ (mmHg) /FiO₂	≥61		≤60				
PaCO₂ (mmHg)	≤58	59-94		≥95			
Ventilation invasive	non			oui			
Hématologique							
Globules blancs (×10⁹/L)	>2		≤2				
Plaquettes (×10⁹/L)	≥142	77-141		≤76			

^a Toutes les variables doivent être collectées. Cependant, la mesure n'est effectuée que si l'état du patient le justifie. Si une variable n'est pas mesurée, elle est considérée comme normale. Si une variable est mesurée plus d'une fois durant la période de 24 heures, la pire valeur de la période est retenue pour le calcul du score. FiO₂ : fraction inspirée d'oxygène.

^b Dysfonction neurologique : score de Glasgow : retenir la pire valeur. Si le patient est sédaté, retenir la valeur du score estimée avant la sédation. Le score de Glasgow est évalué uniquement chez les patients ayant une affection du système nerveux central aiguë connue ou suspectée. Réaction pupillaire : pupilles aréactives > 3 mm. Ne pas tenir compte d'une dilatation pupillaire iatrogène.

^c Dysfonction cardiovasculaire : pression artérielle moyenne : ne pas l'évaluer pendant les pleurs ou une agitation provoquée.

^d Dysfonction respiratoire : PaO₂ : utiliser une mesure artérielle uniquement. Rapport PaO₂/FiO₂ : ce rapport est considéré comme normal en cas de cardiopathie cyanogène.

PaCO₂ : peut être mesurée à partir d'un prélèvement artériel, capillaire ou veineux.

Ventilation invasive : l'utilisation d'un masque de ventilation n'est pas considérée comme une mesure de ventilation invasive.

Logit PELOD-2 = -6.61 + 0.47 × PELOD-2 score. Probabilité de mortalité PELOD-

$$2 = 1 / [1 + e^{-\text{logit PELOD-2}}]$$

Annexe E. Exemple de calcul de seuil de discrimination et du SMR sur une population de 500 patients pour un score

Patients	Valeur du score (ordre croissant)	Probabilité de décès	(%)	Statut sortie de réanimation (vivant 0 /décédé 1)
1	5	0,010	(1 %)	0
2	6	0,012	(1,2 %)	0
3	7	0,014	(1,4 %)	1
4	7	0,014	(1,4 %)	0
5	10	0,035	(3,5 %)	0
6	12	0,043	(4,3 %)	1
.....
499	30	0,957	(95,7 %)	1
500	31	0,968	(96,8 %)	1
Total		74,2		69

SMR = Décès observés/décès prédits = 69/74,2 = 0,93 (IC 95 % : 0,87-1,02)

Exemple de calcul des sensibilités et spécificités pour un seuil à 1,3 % :

	Décès	Vivant	
Proba DC > 1,3 %	69	429	498
Proba DC < 1,3 %	0	2	2
	69	431	500

Sensibilité (seuil 1,3 %) = 69/69 = 1 ; Spécificité (seuil 1,3 %) = 2/431 = 0,0046

Couple de point (1-spécificité ; sensibilité) pour le seuil 1,3 % : (0,99 ; 1)

Exemple de calcul des sensibilités et spécificités pour un seuil à 4,0 % :

	Décès	Vivant	
Proba DC > 4,0 %	68	427	495
Proba DC < 4,0 %	1	4	5
	69	431	500

Sensibilité (seuil 4,0 %) = 68/69 = 0,985 ; Spécificité (seuil 4,0 %) = 4/431 = 0,0092

Couple de point (1-spécificité ; sensibilité) pour le seuil 4,0 % : (0,99 ; 0,985)

SMR : Standardized Mortality Ratio ; Proba DC : probabilité de décès

Références

- Vincent JL, Bruzzi de Carvalho F (2010) Severity of illness. *Semin Respir Crit Care Med* 31:31–8
- Lacroix J, Cotting J, Pediatric Acute Lung Injury and Sepsis Investigators (PALISI) Network (2005) Severity of illness and organ dysfunction scoring in children. *Pediatr Crit Care Med* 6: S126–S34
- Marcin JP, Pollack MM (2007) Review of the acuity scoring systems for the pediatric intensive care unit and their use in quality improvement. *J Intensive Care Med* 22:131–40
- Martinot A, Leclerc F, Hue V, et al (1995) Les scores de gravité généraux en réanimation pédiatrique. *Réanimation Urgences* 4:335–63
- Bakhshi-Raiez F, Peek N, Bosman RJ, de Jonge E, de Keizer NF (2007) The impact of different prognostic models and their customization on institutional comparison of intensive care units. *Crit Care Med* 35:2553–60
- Pollack MM, Holubkov R, Funai T, et al (2016) The Pediatric Risk of Mortality Score: update 2015. *Pediatr Crit Care Med* 17:2–9
- Leteurtre S, Duhamel A, Grandbastien B, et al (2010) Daily estimation of the severity of multiple organ dysfunction syndrome in critically ill children. *CMAJ* 182:1181–7
- Stranely L, Clements A, Parslow RC, et al (2013) Pediatric index of mortality 3: an updated model for predicting mortality in pediatric intensive care. *Pediatr Crit Care Med* 14:673–81
- Wilkinson JD, Pollack MM, Ruttimann UE, Glass NL, Yeh TS (1986) Outcome of pediatric patients with multiple organ system failure. *Crit Care Med* 14:271–4
- Proulx F, Fayon M, Farrell CA, Lacroix J, Gauthier M (1996) Epidemiology of sepsis and multiple organ dysfunction syndrome in children. *Chest* 109:1033–7
- Leteurtre S, Martinot A, Duhamel A, et al (2003) Validation of the pediatric logistic organ dysfunction (PELOD) score: prospective, observational, multicentre study. *Lancet* 362:192–7
- Leteurtre S, Duhamel A, Grandbastien B, Lacroix J, Leclerc F (2006) Pediatric logistic organ dysfunction (PELOD) score. *Lancet* 367:897
- Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL (2001) Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA* 286:1754–8
- Villeneuve A, Joyal JS, Proulx F, Ducruet T, Poitras N, Lacroix J (2016) Multiple organ dysfunction syndrome in critically ill children: clinical value of two lists of diagnostic criteria. *Ann Intensive Care* 6:40
- Vincent JL, Moreno R (2010) Clinical review: scoring systems in the critically ill. *Crit Care* 14:207
- Yeh TS, Pollack MM, Ruttimann UE, Holbrook PR, Fields AI (1984) Validation of a physiologic stability index for use in critically ill infants and children. *Pediatr Res* 18:445–51
- Pollack MM, Ruttimann UE, Getson PR (1988) Pediatric risk of mortality (PRISM) score. *Crit Care Med* 16:1110–6
- Pollack MM, Patel KM, Ruttimann UE (1996) PRISM III: an updated Pediatric Risk of Mortality Score. *Crit Care Med* 24:743–52
- VPS PICU. In: <http://www.myvps.org/>. <https://www.myvps.org/>
- Pollack MM, Dean JM, Butler J, et al (2013) The ideal time interval for critical care severity-of-illness assessment. *Pediatr Crit Care Med* 14:448–53
- Shann F, Pearson G, Slater A, Wilkinson K (1997) Pediatric index of mortality (PIM): a mortality prediction model for children in intensive care. *Intensive Care Med* 23:201–7
- Slater A, Shann F, Pearson G, Pediatric Index of Mortality (PIM) Study Group (2003) PIM2: a revised version of the Pediatric Index of Mortality. *Intensive Care Med* 29:278–85
- Wolfler A, Osello R, Gualino J, et al (2016) The importance of mortality risk assessment: Validation of the Pediatric Index of Mortality 3 Score. *Pediatr Crit Care Med* 17:251–6
- Gulla KM, Sachdev A (2016) Illness severity and organ dysfunction scoring in pediatric intensive care unit. *Indian J Crit Care Med* 20:27–35
- Leteurtre S, Martinot A, Duhamel A, et al (1999) Development of a pediatric multiple organ dysfunction score: use of two strategies. *Med Decis Making* 19:399–410
- Garcia PCR, Eulmesekian P, Branco RG, et al (2010) External validation of the pediatric logistic organ dysfunction score. *Intensive Care Med* 36:116–22
- Leteurtre S, Duhamel A, Salleron J, et al (2013) PELOD-2: an update of the Pediatric Logistic Organ Dysfunction Score. *Crit Care Med* 41:1761–73
- Leteurtre S, Duhamel A, Deken V, Lacroix J, Leclerc F; Groupe Francophone de Réanimation et Urgences Pédiatriques (2015) Daily estimation of the severity of organ dysfunctions in critically ill children by using the PELOD-2 score. *Crit Care* 19:324
- Graciano AL, Balko JA, Rahn DS, Ahmad N, Giroir BP (2005) The Pediatric Multiple Organ Dysfunction Score (P-MODS): development and validation of an objective scale to measure the severity of multiple organ dysfunction in critically ill children. *Crit Care Med* 33:1484–91
- Shann F (2002) Are we doing a good job: PRISM, PIM and all that. *Intensive Care Med* 28:105–7
- Keegan MT, Gajic O, Afessa B (2011) Severity of illness scoring systems in the intensive care unit. *Crit Care Med* 39:163–9
- Deeks JJ, Altman DG (1999) Sensitivity and specificity and their confidence intervals cannot exceed 100%. *BMJ* 318:193–4
- Lemeshow S, Hosmer DW (1982) A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 115:92–106
- Peek N, Arts DG, Bosman RJ, van der Voort PH, de Keizer NF (2007) External validation of prognostic models for critically ill patients required substantial sample sizes. *J Clin Epidemiol* 60:491–501
- Murphy-Filkins R, Teres D, Lemeshow S, Hosmer DW (1996) Effect of changing patient mix on the performance of an intensive care unit severity-of-illness model: how to distinguish a general from a specialty intensive care unit. *Crit Care Med* 24:1968–73
- Pearson GA, Stickley J, Shann F (2001) Calibration of the pediatric index of mortality in UK pediatric intensive care units. *Arch Dis Child* 84:125–8
- Tibby SM, Murdoch IA (2002) Calibration of the pediatric index of mortality score for UK pediatric intensive care. *Arch Dis Child* 86:65; author reply 65–66
- Minne L, Abu-Hanna A, de Jonge E (2008) Evaluation of SOFA-based models for predicting mortality in the ICU: a systematic review. *Crit Care* 12:R161
- Metnitz PG, Lang T, Vesely H, Valentin A, Le Gall JR (2000) Ratios of observed to expected mortality are affected by differences in case mix and quality of care. *Intensive Care Med* 26:1466–72
- Metnitz B, Schaden E, Moreno R, et al (2009) Austrian validation and customization of the SAPS 3 Admission Score. *Intensive Care Med* 35:616–22
- Kramer AA, Higgins TL, Zimmerman JE (2014) Comparison of the mortality probability admission model III, national quality forum, and acute physiology and chronic health evaluation IV hospital mortality models: implications for national benchmarking. *Crit Care Med* 42:544–53
- Wolfler A, Silvani P, Musicco M, Salvo I; Italian Pediatric Sepsis Study (SISPe) Group (2007) Pediatric Index of Mortality 2 score in Italy: a multicenter, prospective, observational study. *Intensive Care Med* 33:1407–13

43. Vincent JL, Opal SM, Marshall JC (2010) Ten reasons why we should NOT use severity scores as entry criteria for clinical trials or in our treatment decisions. *Crit Care Med* 38:283–7
44. Lacroix J, Hébert PC, Hutchison JS, et al (2007) Transfusion strategies for patients in pediatric intensive care units. *N Engl J Med* 356:1609–19
45. Bernard GR, Vincent JL, Laterre PF, et al (2001) Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med* 344:699–709
46. Glance LG, Osler T, Shinozaki T (2000) Effect of varying the case mix on the standardized mortality ratio and W statistic: a simulation study. *Chest* 117:1112–7
47. Leteurtre S, Grandbastien B, Leclerc F, et al (2012) International comparison of the performance of the pediatric index of mortality (PIM) 2 score in two national data sets. *Intensive Care Med* 38:1372–80
48. Kramer AA, Higgins TL, Zimmerman JE (2015) Comparing observed and predicted mortality among ICUs using different prognostic systems: why do performance assessments differ? *Crit Care Med* 43:261–9
49. Namachivayam P, Shann F, Shekerdemian L, et al (2010) Three decades of pediatric intensive care: Who was admitted, what happened in intensive care, and what happened afterward. *Pediatr Crit Care Med* 11:549–55
50. Epstein D, Wong CF, Khemani RG, et al (2011) Race/ethnicity is not associated with mortality in the PICU. *Pediatrics* 127:e588–e97
51. Pollack MM, Holubkov R, Funai T, et al (2015) Simultaneous prediction of new morbidity, mortality, and survival without new morbidity from pediatric intensive care: a new paradigm for outcomes assessment. *Crit Care Med* 43:1699–709
52. Pollack MM, Holubkov R, Glass P, et al (2009) Functional Status Scale: new pediatric outcome measure. *Pediatrics* 124:e18–e28
53. Aspesberro F, Mangione-Smith R, Zimmerman JJ (2015) Health-related quality of life following pediatric critical illness. *Intensive Care Med* 41:1235–46
54. Varni JW, Seid M, Kurtin PS (2001) PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Med Care* 39:800–12
55. Tessier S, Vuillemin A, Lemelle JL, Briançon S (2009) Propriétés psychométriques du questionnaire générique français « Pediatric Quality of Life Inventory Version 4.0 » (PedsQLTM 4.0). *Eur Rev Appl Psychol* 59:291–300
56. Ravens-Sieberer U, Auquier P, Erhart M, et al (2007) The KIDSCREEN-27 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. *Qual Life Res* 16:1347–56
57. PICANet. <http://www.picanet.org.uk/>
58. Straney LD, Clements A, Alexander J, Slater A (2010) Measuring efficiency in Australian and New Zealand pediatric intensive care units. *Intensive Care Med* 36:1410–6