

# Scoring Systems for Organ Dysfunction and Multiple Organ Dysfunction: The PODIUM Consensus Conference

Luregn J. Schlapbach, MD, PhD, FCICM,<sup>a,b,\*</sup> Scott L. Weiss, MD, MSCE,<sup>c,\*</sup> Melania M. Bembea, MD, PhD,<sup>d</sup> Joseph A. Carcillo, MD,<sup>e</sup> Francis Leclerc, MD,<sup>f,g</sup> Stephane Leteurre, MD,<sup>f,g</sup> Pierre Tissieres, MD, DSc,<sup>h</sup> James L. Wynn, MD,<sup>i</sup> Jerry Zimmerman, MD PhD,<sup>j</sup> Jacques Lacroix, MD, FRCPC,<sup>k</sup> on behalf of the Pediatric Organ Dysfunction Information Update Mandate (PODIUM) Collaborative

abstract

**CONTEXT:** Multiple scores exist to characterize organ dysfunction in children.

**OBJECTIVE:** To review the literature on multiple organ dysfunction (MOD) scoring systems to estimate severity of illness and to characterize the performance characteristics of currently used scoring tools and clinical assessments for organ dysfunction in critically ill children.

**DATA SOURCES:** Electronic searches of PubMed and Embase were conducted from January 1992 to January 2020.

**STUDY SELECTION:** Studies were included if they evaluated critically ill children with MOD, evaluated the performance characteristics of scoring tools for MOD, and assessed outcomes related to mortality, functional status, organ-specific outcomes, or other patient-centered outcomes.

**DATA EXTRACTION:** Data were abstracted into a standard data extraction form by a task force member.

**RESULTS:** Of 1152 unique abstracts screened, 156 full text studies were assessed including a total of 54 eligible studies. The most commonly reported scores were the Pediatric Logistic Organ Dysfunction Score (PELOD), pediatric Sequential Organ Failure Assessment score (pSOFA), Pediatric Index of Mortality (PIM), PRISM, and counts of organ dysfunction using the International Pediatric Sepsis Definition Consensus Conference. Cut-offs for specific organ dysfunction criteria, diagnostic elements included, and use of counts versus weighting varied substantially.

**LIMITATIONS:** While scores demonstrated an increase in mortality associated with the severity and number of organ dysfunctions, the performance ranged widely.

**CONCLUSIONS:** The multitude of scores on organ dysfunction to assess severity of illness indicates a need for unified and data-driven organ dysfunction criteria, derived and validated in large, heterogenous international databases of critically ill children.



<sup>a</sup>Children's Research Center and Department of Intensive Care and Neonatology, University Children's Hospital Zurich, Zurich, Switzerland; <sup>b</sup>Child Health Research Centre, The University of Queensland, and Queensland Children's Hospital, Brisbane, Australia; <sup>c</sup>Department of Anesthesiology and Critical Care, Children's Hospital of Philadelphia, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania; <sup>d</sup>Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland; <sup>e</sup>Department of Critical Care Medicine, Children's Hospital of Pittsburgh, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania; <sup>f</sup>ULR 2694 - METRICS: Évaluation des technologies de santé et des pratiques médicales, Université de Lille, CHU Lille, Lille, France; <sup>g</sup>EA 2694 Santé publique, épidémiologie et qualité des soins, Université de Lille, Lille, France; <sup>h</sup>Pediatric Intensive Care, Université Paris-Saclay, Le Kremlin-Bicêtre, France; <sup>i</sup>Department of Pediatrics and Department of Pathology, Immunology, and Laboratory Medicine, University of Florida, Gainesville, Florida; <sup>j</sup>Division of Pediatric Critical Care Medicine, Department of Pediatrics, Seattle Children's Hospital, Seattle Children's Research Institute, University of Washington School of Medicine, Seattle, Washington; and <sup>k</sup>Division of Pediatric Critical Care Medicine, Centre Hospitalier Universitaire de Sainte-Justine, Université de Montreal, QC, Canada

\* Contributed equally as co-first authors.

Drs Schlapbach and Weiss conceptualized and designed this review, drafted the initial manuscript, and reviewed and revised the manuscript; Drs Bembea, Lacroix, and Zimmerman designed, led, and supervised the Pediatric Organ Dysfunction Information Update Mandate project and contributed to sections of the manuscript; Drs Carcillo, Leclerc, Leteurre, Tissieres, and Wynn contributed to sections of the manuscript; and all authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Multiple organ dysfunction syndrome (MODS) in critically ill children remains associated with a high morbidity and persistently high mortality.<sup>1</sup> A recent study utilizing the Virtual Pediatric Systems database, including nearly 200 000 PICU admissions, revealed a mortality of 10.3% among children with MODS compared with 0.7% in children without MODS.<sup>2</sup> In MODS survivors, the risk of survival with poor functional status as assessed by the Pediatric Overall Performance Category/Pediatric Cerebral Performance Category was increased severalfold. Recent research into the pathophysiology of critical illness illustrates that different MODS phenotypes may reflect patient populations more likely to respond to distinct, targeted therapies. Reliable identification of patients with MODS is therefore required to: (1) accurately characterize epidemiology, (2) assist in prognostication, (3) select patient groups where risk/benefit of specific treatments may vary, and (4) efficiently enroll selected patients into targeted trials.

However, to date, diagnostic criteria of MODS remain a matter of debate and there is no agreement on a gold standard for MODS, which organs to include, and thresholds to define dysfunction for individual organ systems. A unified approach to MODS is further hampered by patient heterogeneity of previous studies. Some studies have focused primarily on the prediction of mortality, whereas others report on scores as a description of illness severity. Most scores for MODS have been in use for many years, but a comprehensive review of the performance of different scores is lacking.

As part of the Pediatric Organ Dysfunction Information Update Mandate (PODIUM) project, we

aimed to review the literature on MODS scoring systems to characterize the performance characteristics of currently used scoring tools and clinical assessments for organ dysfunction in critically ill children.

## METHODS

The PODIUM taskforce sought to develop evidence-based criteria for organ dysfunctions in children. As part of this process, a subgroup on MODS (SW, PL, EJ, CC, JLW, LJS) reviewed the literature on MODS scoring systems. The present article reports on the systematic review on organ dysfunction scoring systems performed as part of PODIUM and provides a critical evaluation of the available literature with recommendations for future research. Details on data sources, study selection, data extraction, data synthesis, and risk of bias assessment utilized by the PODIUM collaborative are presented in the PODIUM Executive Summary.<sup>3</sup>

## RESULTS

### Overview of Commonly Used Scores

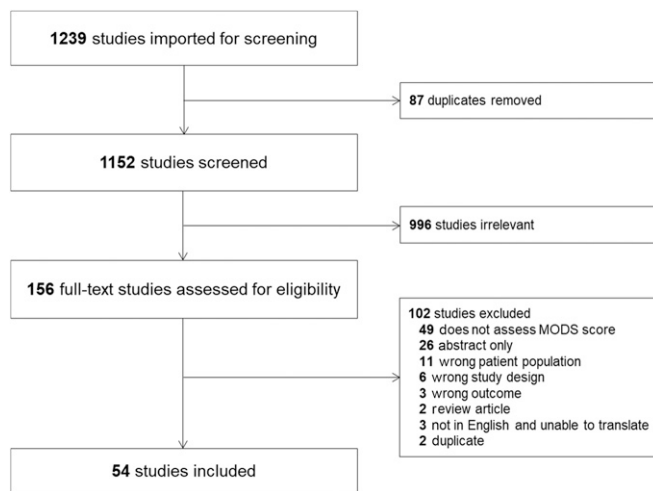
Out of 1152 unique abstracts, 159 full texts were reviewed, of which 54 provided data on scores for the purpose of this review, as shown in the PRISMA flowchart (Fig 1), data tables (Supplemental Information, Supplemental Tables 1, 2, and 3), and risk of bias assessment summary (Supplemental Information, Supplemental Fig 1). Many scores have been developed and reported in critically ill children (Table 1). Scores show substantial differences in their scope (predictive, descriptive, diagnostic, Fig 2), number and type of variables assessed (Table 2 and Fig 3), suitability to measure organ dysfunctions, time frame, and applicability to different clinical settings.

In terms of predictive scores, the Pediatric Index of Mortality-3

(PIM-3)<sup>4</sup> and the Pediatric Risk of Mortality-IV (PRISM-IV)<sup>5</sup> scores (and their predecessors) are the most commonly used. However, because these scores are intended for PICU patients with and without organ dysfunctions, they may have limited applicability for assessment specific to patients with MODS.<sup>4,5</sup> Although PIM-3 contains information on cardiovascular (systolic blood pressure), respiratory (need for mechanical ventilation) and neurologic dysfunction (dilated pupils), it does not lend itself to assessment of individual organ dysfunctions or MODS. The PRISM-IV physiologic score contains information on cardiac (heart rate, systolic blood pressure, temperature), neurologic (pupillary reactivity, mental status), respiratory (arterial PO<sub>2</sub>, pH, PCO<sub>2</sub>, total bicarbonate), hematologic (white blood cell count, platelet count, prothrombin, and partial thromboplastin time) and chemical score components (glucose, potassium, blood urea nitrogen, creatinine).

Commonly used descriptive scores include the Pediatric Logistic Organ Dysfunction Score-2 (PELOD-2)<sup>6</sup> and, more recently, the pediatric Sequential Organ Failure Assessment (pSOFA).<sup>7</sup> PELOD-2 assesses 5 (neurologic, cardiovascular, renal, respiratory, and hematologic) organ dysfunctions, and pSOFA includes 6 (including hepatic) organ dysfunctions. PELOD-2 was derived from a multicenter European PICU cohort. In contrast, pSOFA was constructed as a modification from the adult Sequential Organ Failure Assessment score with application of age-specific thresholds based on PELOD-2.

Diagnostic scores are designed to characterize presence of (multi)organ dysfunction for the purpose of correct classification and/or selection for clinical studies.



**FIGURE 1** Study flow diagram according to the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols recommendations.

Although not a score in the strict sense, the 2005 International Pediatric Sepsis Definition Consensus Conference (IPSDCC)<sup>8,9</sup> statement defined criteria for 6 organ dysfunctions which have been widely used, both in patients with and without sepsis.

In addition to these more commonly used scores, the literature search identified a number of articles proposing other approaches to assess organ dysfunctions in both broad and specific patient populations (Table 1 and Supplemental Tables 1 and 2).

### Predictive Scores

Predictive scores such as PIM-3<sup>4</sup> or PRISM-IV<sup>5</sup> describe the severity of illness at a defined baseline time point, which is often a time window around PICU admission, or time of randomization in clinical trials. The premise of predictive scores is founded on predicting the outcome with minimal influence by therapies provided to treat the condition (ie, is the observed severity of illness attributable to the disease that brings the patient to the PICU or to treatment given after PICU admission?), and on a temporal

separation between the prediction and the outcome (ie, is the score predicting, rather than describing, death?). The reliability of a predictive score is better if the data are collected before any care is given or if the data are unresponsive to care. The discriminative value of a test is estimated by measuring its area under the receiver operating characteristic curve and the Hosmer-Lemeshow goodness of fit, with death used most commonly as the outcome. Good calibration refers to the agreement between predicted and observed rates of death across the spectrum of the score and may be measured by Cox calibration regression or other techniques. Reproducibility across different sites and health care settings is desirable to enable comparison of baseline risk of death for benchmarking. Predicted scores are not intended to be used in individual patients to guide treatment or to inform end-of-life decisions because they were validated in whole PICU populations, not in single patients. These scores need to be updated regularly because the population of PICU patients changes over time and because the risk of mortality changes over time for many specific diseases.

There are no predictive scores specific to patients with MODS at PICU admission or at randomization. Although PIM and PRISM represent the most frequently used predictive scores, organ dysfunction scores such as PELOD or pSOFA, obtained in a time window around PICU admission (such as day 1), also have predictive value for mortality. In addition, PELOD-2 on day of admission and maximum and cumulative PELOD-2 scores were associated with health-related quality of life 3 months postdischarge in a recent pediatric sepsis cohort.<sup>10</sup>

### Descriptive Scores

Descriptive organ dysfunction scores estimate the severity of cases at defined time points or time intervals. Descriptive scores focus on the differentiation between patients with mild versus severe illness. Descriptive scores should reliably capture (un)responsiveness to care, as well as disease progression or resolution, and may thereby provide additional information not reflected in baseline prediction.<sup>11</sup> Although simplicity is desirable to facilitate clinical application, descriptive scores aim to characterize the number and severity of organ dysfunctions. For example, the final PELOD-2 score utilizes 10 out of 17 criteria assessed in the derivation<sup>6,12</sup> because these 10 were sufficient to explain the statistical variability related to the risk of death observed in the index population. The discriminative value of descriptive scores is estimated by measuring its area under the receiver operating characteristic to differentiate death and/or severe adverse outcomes. The calibration of a descriptive score to predict the risk of adverse outcomes should be excellent in the index population used to create and validate the score. On the other hand, calibration in other

**TABLE 1** Comparison of the Performance of Different Multiple Organ Dysfunction Scoring Tools

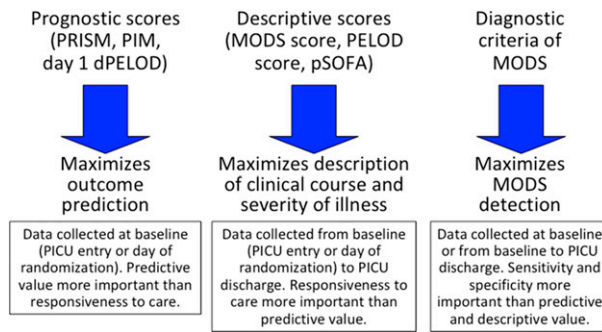
Criteria	SCORES														
	PIM-2/3	PRISM-III	PELOD-2 (dPELOD-2)	qPELOD	SOFA	pSOFA/mSOFA	LODS	MOSF/MODS (count)	PeRF	SICK	PEDIA	BEP	TISS	Arzeno 2015	Meyer 2005
Number of studies	5	4	8	1	4	2	1	4	1	1	1	1	1	1	1
Validation	Death PICU, Mening	Death PICU, sepsis, RRT	Death PICU	Death Sepsis	Death PICU, RRT, GI	Death C ICU, PICU	Death Hosp fever	Death PICU, ARDS, Sepsis	Death ARDS	Death Hosp fever, PICU	Death Hosp fever	Death Mening	Death PICU	Death PICU brain	Death PICU oncology
Reference standard	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	No	No	No	Yes	Yes
Case-mix applied to	No	Yes	No	No	Yes	Yes	No	Yes	Yes	No	No	No	No	Yes	Yes
Validity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Construct (score reflects MODS?)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Content (score includes all organs?)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Criterion (↑score → ↑ death)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Discrimination	Mod-good	Mod-good	Good	Good	Mod-good	Good	Good	Good	Mod-good	Mod-good	Good	Mod	Mod	Mod	NR
AUROC	Good	Good	Poor to good	Good	Mod to good	Good	Good	Good	Poor to good	Poor to good	Good	Unknown	Unknown	Good	Unknown
Calibration	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
Reliability	Mod	Mod	Mod	Good	Good	Good	—	Good	—	—	—	Good	—	—	—
Ease of Use	Mod	Mod	Good	Good	Good	Good	—	Good	—	—	—	Good	—	—	—
Ease of interpretation	Mod	Mod	Good	Good	Good	Good	—	Good	—	—	—	Good	—	—	—
External validity/generalizability	Unknown	Unknown	Unknown	Unknown	Mod	Unknown	Unknown	Good	Unknown	Mod	Unknown	Unknown	Unknown	Unknown	Unknown

dPELOD-2, Pediatric Logistic Organ Dysfunction on day 1; SOFA, Sequential Organ Failure Assessment; mSOFA, modified sequential organ failure assessment; LODS, logistic organ dysfunction score; MOSF, multiple organ system failure; PeRF, pediatric respiratory failure score; SICK, signs of inflammation in children that can kill score; PEDIA, pediatric early death index for Africa score; BEP, base excess and platelet count at presentation score; TISS, therapeutic intervention scoring system; RRT, renal replacement therapy; C ICU, cardiac ICU; Hosp fever, hospitalized patients with fever; ARDS, acute respiratory distress syndrome; Mening, meningococcal disease; AUROC, area under the receiver operating characteristic; Mod, moderate; NR, not reported; —, not applicable.

populations is less important because comorbidities and medical practice can differ significantly in different PICUs and in different countries. Updating descriptive scores over time is somewhat less important compared with predictive scores. Descriptive scores, as predictive scores, have been validated in large populations, not in individual patients and all subpopulations; thus, they should not be used to guide treatment or inform end-of-life decisions at the bedside. For example, the PELOD-2 score can be used in critically ill children with respiratory problems,<sup>13</sup> as well as children with suspected infection,<sup>14</sup> but we do not know how reliable the score is in other subpopulations of PICU patients, such as trauma patients.

**Diagnostic Criteria for MODS**

MODS represents a syndrome, not a specific disease entity, because MODS reflects a group of symptoms and signs that consistently occur together, the combination of which is associated with predictable outcomes. Diagnostic criteria are important to enable correct classification for (1) selecting specific monitoring, interventions and clinical pathways, (2) prognostication, and (3) reliably characterizing epidemiology. Contrary to a syndrome such as trisomy 21, where a consistent list of symptoms and signs relates to one common, genetic finding that defines the “gold standard,” the diagnosis of many conditions often depends on “the subjective interaction of an observer, and its defining boundaries are both arbitrary and a little fuzzy.”<sup>15</sup> Presently, there is no reference standard for MODS, and diagnosis is based on different approaches to physiologic data, such as blood pressure, interventions (such as ventilation), and laboratory parameters (such as creatinine



**FIGURE 2**  
Purpose of different scoring systems

concentrations). Since formal criteria for pediatric organ dysfunction were first proposed in 1987 by Wilkinson,<sup>16</sup> subsequent iterations, such as criteria proposed in 1996 by Proulx<sup>17</sup> and in 2005 by Goldstein,<sup>8</sup> were largely independent rather than the result of a consistent, iterative revision process. Importantly, these initial criteria for MODS were not data-driven but defined by expert consensus opinion. Although there is ample evidence for the association of increasing MODS severity with risk of death in critically ill children,<sup>1,18</sup> the diagnostic performance of these (multi)organ dysfunction criteria in terms of sensitivity and specificity has been understudied. When it was studied, results indicated substantially worse performance compared with descriptive scores.<sup>19,20</sup>

### Limitations of Current (Multi)Organ Dysfunction Scores

Clinicians and researchers base management and diagnostic decisions at least partially on objective physiologic parameters such as blood pressure, heart rate, or neurologic state. Although there is ample observational evidence to support the relevance of individual organs in relation to outcomes, a closer look reveals substantial differences in thresholds applied. For example, an adolescent with a

creatinine of 100 micromol/L will score 2 points for renal dysfunction in PELOD-2 and pSOFA but not be counted as kidney dysfunction by IPSCC. To complicate the matter further, ICUs internationally care for an increasing proportion of children with complex chronic health care conditions, and only some scores incorporate changes from baseline.<sup>21</sup> In addition, thresholds may vary with concomitantly administered therapy; for example, Glasgow Coma Scale in presence of sedation and/or neuromuscular blockade.

Furthermore, the comparison of scores (Table 2, Fig 3) reveals inconsistencies in terms of which organs are included; for example, lactate is measured in PELOD-2 and IPSCC only, whereas hepatic dysfunction is not included in PELOD-2. Although some of these differences stem from score design methodology (a priori versus derivation), they may simply relate to whether some of the criteria were available in the databases used for derivation/validation. The issue is further accentuated as scores variably consider the level of support provided to an organ; only pSOFA and IPSCC consider vasoactive-inotrope support, for example, and none consider renal replacement therapies or extracorporeal membrane oxygenation. A child may thus

exhibit severe MODS requiring extracorporeal membrane oxygenation and renal replacement therapies resulting in normal blood pressure, blood gases, and creatinine, yet only be scored for the mechanical ventilation component by some tools. In addition, scores often do not fully account for the evolution of critical care. For example, pSOFA includes vasoactive-inotropic support, but milrinone, vasopressin, and angiotensin-II analogs are not included.<sup>22</sup> Importantly, the different approaches used to classify organ dysfunction severity (ie, binary in IPSCC, weighted score in PELOD-2, discrete score in pSOFA) hinder direct comparisons of absolute score levels in relation to MODS. For example, a score as high as 4 in both PELOD-2 or pSOFA may either reflect severe dysfunction in a single organ system or mild dysfunction across several organ systems.

It is also important to consider that, despite the merits of the procedures applied for derivation, validation, and calibration, the patient cohorts used should be considered historic and were almost exclusively biased toward PICUs in the United States, Canada, Western Europe, Australia, and New Zealand. Considering the expansion of PICU services around the world over the past 2 decades, it is imperative to ensure scores, or adapted versions, are applicable to different health care settings, some of which may have different resource levels. Finally, the focus of organ dysfunction scores has inevitably been on children admitted to PICUs that have the capacity to collect data on organ dysfunctions and severity. Yet, patient care represents a continuum, including emergency department to PICU, operating theater to PICU, or interhospital transfer to PICU journeys; hence, it



**TABLE 2** Comparison of Characteristics of Main Organ Dysfunction Assessment Tools

Scores	PIM-3	PRISM-III and IV	PELOD-2 (dPELOD-2)	qPELOD	pSOFA	2005 IPSDCC (Goldstein)
Organ systems	Not specific to organs	Not specific to organs	Cardiovascular system, plus metabolic (lactate), respiratory, hematologic, renal, CNS	Cardiovascular system, CNS	Cardiovascular system, respiratory, hematologic, hepatic, renal, CNS	Cardiovascular system, respiratory, hematologic, hepatic, renal, CNS
Main purpose (at time of design of the score)	Prediction of mortality, ICU benchmarking	Prediction of mortality, ICU benchmarking	Description of severity	Prediction of mortality in sepsis	Description of severity	Diagnosis of organ dysfunction
Number of items	11	17	10	3	8 (12 if counting Sp <sub>o</sub> <sub>2</sub> and individual inotropes)	18
Number of laboratory items (number of laboratory items available as POC, ie, blood gas components, glucose, and lactate)	2 (2)	12 (5)	6 (3)	0	3 (1)	9 (3)
Development methods	Derivation cohort Australia, New Zealand, Ireland, and the United Kingdom <i>n</i> = 53 112 in 2010–2011	Derivation cohort United States <i>n</i> = 10 078 in 2011–2013	Derivation cohort French/Belgium <i>n</i> = 3671 in 2006–2007	A priori (aligned with qSOFA and PELOD-2)	A priori (aligned with SOFA and PELOD-2)	A priori (expert statement)
Validation/calibration	Multiple validations	Multiple validations	Multiple validations	NA	Multiple validations	NA
Time frame	Within 60 min of admission, including first contact outside PICU by PICU team	First 4 h of PICU admission, minus 2 h to 4 h for laboratory variables	Daily every 24 h of PICU admission	First 24 h of PICU admission	Daily every 24 h of PICU admission	Not specified
Patient information	Yes	Yes in PRISM-IV	No	No	No	No
Treatment information	Yes (ventilation)	No	Yes (ventilation)	No	Yes (ventilation, vasoactives)	Yes (ventilation, vasoactives)
Applicability outside PICU	Poor	Poor	Moderate	Very good	Moderate	Good
Applicability in resource-limited setting	Good	Poor	Poor	Very good	Moderate	Moderate

dPELOD-2, Pediatric Logistic Organ Dysfunction on day 1; qPELOD, Quick Pediatric Logistic Organ Dysfunction; CNS, central nervous system; Sp<sub>o</sub><sub>2</sub>, pulse oxygen saturation; POC, point of care; NA, not applicable.

would be desirable if scores could be readily applied outside the PICU environment.

### Future Developments

An ideal MODS reference score should be specific to MODS and fulfill the following quality criteria:

1. highly sensitive and specific performance for clinically relevant outcomes such as mortality;
2. operator-independence;
3. criteria should be met as soon as possible while MODS is developing;
4. good reproducibility;
5. readily available in diverse PICU and non-PICU settings; and
6. good performance, as well, for nonmortality outcomes such as prolonged dependency on ICU support

Organ	Variable	Score					
		PiM-3	PRISM-IV	PELOD-2	qPELOD	PSOFA	IPSOFA
Cardiovascular	Systolic Blood Pressure						
	Mean Blood Pressure						
	Inotrope use						
	Oliguria: urine output 0.5 ml/kg/hr						
	Prolonged capillary refill						
	Core to peripheral temperature gap						
	Heart rate/tachycardia						
	Lactate						
Respiratory	Mechanical ventilation						
	Invasive ventilation						
	PaO2						
	PaO2 and FIO2						
	SpO2/FIO2						
	PCO2						
Neurology	Glasgow Coma Scale						
	Pupillary response						
Renal	Creatinine						
	Blood Urea Nitrogen						
Hepatic	Bilirubin						
	ALAT						
Haematologic	Platelet count						
	White cell count						
	Partial thromboplastin time						
	International normalized ratio/prothrombin time						
Metabolic	Base excess						
	pH						
Endocrine	Glucose						
History	Admission source characteristics						
	Age						
	Risk related to specific diagnostic groups						
Other	Temperature						
	Potassium						

**FIGURE 3**  
Comparison of variables used to calculate commonly used organ dysfunction scores.

and mid- to long-term quality of life and functional status.

In the era of electronic health records (EHRs), the availability of multisite, multinational, and preferably, multisetting granular data from initial presentation throughout intensive care stay to discharge or death is promising and will enable better data-driven, rather than purely expert-based, approaches.<sup>23</sup> Acknowledging that

developing new scores from EHR data will inherently result in a bias toward high-resource settings such as selected PICUs in the United States, the international shift toward EHR in many countries, and the creation of high-quality databases in resource-limited settings, opens new opportunities for validation and adaptation to meet the requirements of different settings. Data-driven approaches should rigorously derive and validate criteria in large and

independent databases, assess the intrarater and interrater reproducibility of the new list of diagnostic criteria, and compare the discriminative capacity of any new diagnostic criteria for MODS versus existing scoring systems.

Recent developments in clinician-driven approaches from the more traditional Delphi study are worth considering because they have the potential to be pragmatic and combine both the clinician's perception of the gestalt of a condition and the data collected during the course of a disease. For example, a "temporary (presumptive) diagnosis" of MODS put forward by one or more independent clinicians can be compared with a "confirmatory (post hoc) diagnosis" of MODS, the latter diagnosis being made by an adjudicating committee.<sup>24,25</sup> The analysis can then assess the sensitivity and specificity of items used for temporary and confirmatory diagnosis. Using a Bayesian strategy and likelihood ratio to ascertain the diagnosis can further improve the reliability of the diagnosis of MODS made by the members of the adjudicating committee.<sup>26</sup>

The challenge of developing 1 universal MODS gold standard can be partially overcome with appropriate methodologies. For example, latent class analysis may serve to identify surrogate gold standards. Supervised and unsupervised learning algorithms can identify clusters of patients with similar features and outcomes, which may serve to characterize phenotypes more likely to respond to certain interventions.<sup>27</sup> Importantly, computational approaches carry enormous promise to overcome the limitations of static (1 assessment in a time window) score measures because dynamic measures of change over time may

be more informative on a patient's pattern of disease, response to disease, and response to treatment.

It is important to note that even future large-scale, data-derived criteria and scores for MODS are likely to fall short from a number of perspectives. First, recent developments such as cytokine and gene expression profiles, proteomics, genomics, or highly granular analysis of EHR data, such as heart rate variability, may lead to improved tools. However, their clinical usefulness remains to be determined, and applicability to different settings will pose major challenges because of the resources and expertise involved. Second, although it is feasible to derive best cutoffs for individual organ dysfunction based on optimal performance in terms of sensitivity and specificity, we are currently unable to delineate when organ dysfunction begins (for example, is a slightly elevated creatinine in a child with gastroenteritis receiving enteral rehydration equivalent to the same creatinine concentration in a child heading toward sepsis-

related MODS?). Third, some alterations in physiology may reflect adaptive hypo- (eg, hibernation) or hyperfunction (tachycardia to meet increased cardiac output requirements), but current approaches struggle to discriminate these from dysfunction associated with worse outcomes.

### CONCLUSIONS

After more than 3 decades of MODS research in critically ill children,<sup>16</sup> and a large body of observational data demonstrating worse short- and long-term outcomes in children with MODS, present approaches remain hampered by lack of validation, standardization, and applicability, indicating an urgent need for revised MODS criteria. The creation of large international research networks contributing high-resolution data and the advances in computational science are expected to lead to a paradigm shift in the development and application of organ dysfunction scores. It is highly desirable to combine efforts aiming to yield data-driven criteria for organ and multiorgan dysfunction. In addition,

these efforts should aim to derive parsimonious scores for easier application at an early stage, even in settings where resources are limited, to pave the way toward interventions more likely to improve outcomes for children with MODS globally.

### ABBREVIATIONS

EHR: electronic health record  
IPSDCC: International Pediatric Sepsis Definition Consensus Conference  
MODS: multiple organ dysfunction syndrome  
PELOD: Pediatric Logistic Organ Dysfunction score  
PIM: Pediatric Index of Mortality  
PODIUM: Pediatric Organ Dysfunction Information Update Mandate  
PRISM: Pediatric Risk of Mortality  
pSOFA: Sequential Organ Failure Assessment

The guidelines/recommendations in this article are not American Academy of Pediatrics policy, and publication herein does not imply endorsement.

DOI: <https://doi.org/10.1542/peds.2021-052888D>

Accepted for publication Sep 24, 2021

Address correspondence to Luregn Schlapbach, MD, PhD, FCICM, Department of Intensive Care and Neonatology University Children's Hospital Zurich - Eleonore Foundation, Steinwiesstrasse 75, CH-8032 Zurich, Switzerland. E-mail: [luregn.schlapbach@kispi.uzh.ch](mailto:luregn.schlapbach@kispi.uzh.ch)

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2021 by the American Academy of Pediatrics

**FINANCIAL DISCLOSURE:** The authors have indicated they have no financial relationships relevant to this article to disclose.

**FUNDING:** Dr Schlapbach was supported by a National Health and Medical Research Council practitioner fellowship and by the Children's Hospital Foundation, Brisbane, Australia. This work was supported by National Institutes of Health, National Institute of Neurological Disorders and Stroke, grant R01 NS106292 to Dr Bembea. The funders had no role in the design and conduct of the study. Funded by the National Institutes of Health (NIH).

**CONFLICT OF INTEREST:** The authors have indicated they have no potential conflicts of interest to disclose.

### REFERENCES

1. Watson RS, Crow SS, Hartman ME, Lacroix J, Odetola FO. Epidemiology and outcomes of pediatric multiple organ dysfunction syndrome. *Pediatr Crit Care Med*. 2017;18(Suppl 1):S4-S16
2. Typpo K, Watson RS, Bennett TD, Farris RWD, Spaeder MC, Petersen NJ; Pediatric Existing Data Analysis (PEDAL) Investigators and Pediatric Acute Lung Injury and Sepsis Investigators (PALISI) Network. Outcomes of day 1 multiple organ dysfunction syndrome in the picu. *Pediatr Crit Care Med*. 2019;20(10):914-922
3. Bembea MM, Agus M, Acan-Arikan A, et al. Pediatric organ dysfunction information update mandate (PODIUM) contemporary organ dysfunction criteria:



- executive summary. *Pediatrics*. 2022;149(suppl 1):e2021052888B
4. Straney L, Clements A, Parslow RC, et al. ANZICS Paediatric Study Group and the Paediatric Intensive Care Audit Network. Paediatric index of mortality 3: an updated model for predicting mortality in pediatric intensive care\*. *Pediatr Crit Care Med*. 2013;14(7):673–681
  5. Pollack MM, Holubkov R, Funai T, et al. Eunice Kennedy Shriver National Institute of Child Health and Human Development Collaborative Pediatric Critical Care Research Network. The pediatric risk of mortality score: update 2015. *Pediatr Crit Care Med*. 2016;17(1):2–9
  6. Leteurtre S, Duhamel A, Salleron J, Grandbastien B, Lacroix J, Leclerc F; Groupe Francophone de Réanimation et d'Urgences Pédiatriques (GFRUP). PELOD-2: an update of the pediatric logistic organ dysfunction score. *Crit Care Med*. 2013;41(7):1761–1773
  7. Matics TJ, Sanchez-Pinto LN. Adaptation and validation of a pediatric sequential organ failure assessment score and evaluation of the sepsis-3 definitions in critically ill children. *JAMA Pediatr*. 2017;171(10):e172352
  8. Goldstein B, Giroir B, Randolph A. International Consensus Conference on Pediatric Sepsis. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med*. 2005;6(1):2–9
  9. Gebara BM. Values for systolic blood pressure. *Pediatr Crit Care Med*. 2005;6(4):500, author reply 500–501
  10. Zimmerman JJ, Banks R, Berg RA, et al. Life After Pediatric Sepsis Evaluation (LAPSE) Investigators. Critical illness factors associated with long-term mortality and health-related quality of life morbidity following community-acquired pediatric septic shock. *Crit Care Med*. 2020;48(3):319–328
  11. Leteurtre S, Duhamel A, Grandbastien B, et al. Daily estimation of the severity of multiple organ dysfunction syndrome in critically ill children. *CMAJ*. 2010;182(11):1181–1187
  12. Leteurtre S, Duhamel A, Deken V, Lacroix J, Leclerc F; Groupe Francophone de Réanimation et Urgences Pédiatriques. Daily estimation of the severity of organ dysfunctions in critically ill children by using the PELOD-2 score. *Crit Care*. 2015;19(324):324
  13. Leclerc F, Duhamel A, Deken V, Lacroix J, Leteurtre S. Groupe Francophone de Réanimation et d'Urgences Pédiatriques. Nonrespiratory pediatric logistic organ dysfunction-2 score is a good predictor of mortality in children with acute respiratory failure. *Pediatr Crit Care Med*. 2014;15:590–593
  14. Leclerc F, Duhamel A, Deken V, Grandbastien B, Leteurtre S; Groupe Francophone de Réanimation et Urgences Pédiatriques (GFRUP). Can the pediatric logistic organ dysfunction-2 score on day 1 be used in clinical criteria for sepsis in children? *Pediatr Crit Care Med*. 2017;18(8):758–763
  15. Diamond GA, Forrester JS. Metadiagnosis. An epistemologic model of clinical judgment. *Am J Med*. 1983;75(1):129–137
  16. Wilkinson JD, Pollack MM, Glass NL, Kanter RK, Katz RW, Steinhart CM. Mortality associated with multiple organ system failure and sepsis in pediatric intensive care unit. *J Pediatr*. 1987;111(3):324–328
  17. Proulx F, Fayon M, Farrell CA, Lacroix J, Gauthier M. Epidemiology of sepsis and multiple organ dysfunction syndrome in children. *Chest*. 1996;109(4):1033–1037
  18. Typpo K, Watson RS, Bennett TD, et al. Outcomes on day 1 multiple organ dysfunction syndrome in the PICU\*. *Pediatr Crit Care Med*. 2019;20(10):914–922
  19. Schlapbach LJ, Straney L, Bellomo R, MacLaren G, Pilcher D. Prognostic accuracy of age-adapted SOFA, SIRS, PELOD-2, and qSOFA for in-hospital mortality among children with suspected infection admitted to the intensive care unit. *Intensive Care Med*. 2018;44(2):179–188
  20. Villeneuve A, Joyal JS, Proulx F, Ducruet T, Poitras N, Lacroix J. Multiple organ dysfunction syndrome in critically ill children: clinical value of two lists of diagnostic criteria. *Ann Intensive Care*. 2016;6(1):40
  21. Moynihan KM, Alexander PMA, Schlapbach LJ, et al. Australian and New Zealand Intensive Care Society Pediatric Study Group (ANZICS PSG) and the ANZICS Centre for Outcome and Resource Evaluation (ANZICS CORE). Epidemiology of childhood death in Australian and New Zealand intensive care units. *Intensive Care Med*. 2019;45(9):1262–1271
  22. Gaies MG, Gurney JG, Yen AH, et al. Vasoactive-inotropic score as a predictor of morbidity and mortality in infants after cardiopulmonary bypass. *Pediatr Crit Care Med*. 2010;11(2):234–238
  23. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ*. 1995;311(7001):376–380
  24. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess*. 2007;11(50):iii, ix-51.
  25. Bertens LC, Broekhuizen BD, Naaktgebooren CA, et al. Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. *PLoS Med*. 2013;10(10):e1001531
  26. Eddy DM, Clanton CH. The art of diagnosis: solving the clinicopathological exercise. *N Engl J Med*. 1982;306(21):1263–1268
  27. Seymour CW, Kennedy JN, Wang S, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA*. 2019; 321(20):2003–2017