Université de Montréal

Étapes préliminaires à l'élaboration de systèmes d'aide au diagnostic automatisé de l'hypoxémie aigüe pédiatrique

Par

Michaël Sébastien SAUTHIER

Cycles supérieurs en sciences biomédicales, Faculté de Médecine

Thèse présentée en vue de l'obtention du grade de Philosophiæ doctor (Ph.D.)

en sciences biomédicales, option médecine expérimentale

Aout 2021

Université de Montréal

Cycles supérieurs en sciences biomédicales, Faculté de médecine

---

Cette thèse intitulée

Étapes préliminaires à l'élaboration de systèmes d'aide au diagnostic automatisé de l'hypoxémie aigüe pédiatrique

Présentée par

Michaël Sébastien SAUTHIER

A été évaluée par un jury composé des personnes suivantes

Emmanuel CHARBONNEY
Président-rapporteur

Philippe JOUVET
Directeur de recherche

Guillaume EMERIAUD
Codirecteur

Guy WOLF
Membre du jury

François LELLOUCHE
Examinateur externe

Yoan LAMARCHE
Représentant du doyen

Résumé

L'insuffisance respiratoire hypoxémique aigüe (IRHA) est une des causes les plus fréquentes d'admission aux soins intensifs pédiatriques. Elle est liée à plusieurs mécanismes dont le plus grave est l'œdème pulmonaire lésionnel conduisant au syndrome de détresse respiratoire aigüe (SDRA) pédiatrique qui représente 5-10 % des patients admis aux soins intensifs. Actuellement, les recommandations internationales de prise en charge de l'IRHA et du SDRA sont sous-appliquées du fait d'un défaut de diagnostic ou d'un diagnostic tardif. Ceci est probablement en partie responsable d'une ventilation mécanique prolongée dans le SDRA pédiatrique. Afin d'améliorer les critères d'évaluation de l'IRHA chez les enfants et éventuellement leur devenir, les 3 objectifs de cette thèse sont d'améliorer le diagnostic précoce d'IRHA chez l'enfant, informatiser un score de gravité de défaillance d'organes (score PELOD-2) utilisable comme critère de jugement principal en recherche en remplacement de la mortalité qui est faible dans cette population et prédire la ventilation prolongée chez la population la plus fragile, les nouveau-nés.

Pour réaliser ces objectifs, nous avons : 1) optimisé une base de données haute résolution temporelle unique au monde, 2) validé un indice continu d'oxygénation utilisable en temps réel et robuste à toutes les valeurs de saturations pulsées en oxygène, 3) validé une version informatisée du score PELOD-2 utilisable comme critère de jugement principal en recherche, 4) développé un modèle prédictif d'IRHA persistante dû à l'influenza et 5) proposé une définition de la ventilation prolongée en pédiatrie applicable quel que soit l'âge et le terme de l'enfant et 6) étudié le devenir des nouveau-nés ayant une ventilation prolongée et proposé un modèle prédictif du sous-groupe le plus grave. Les méthodes utilisées à travers ces différentes études ont associé la science des données massives pour le regroupement, la synchronisation et la normalisation des données continues. Nous avons également utilisé les statistiques descriptives, la régression linéaire et logistique, les forêts aléatoires et leurs dérivés, l'apprentissage profond et l'optimisation empirique d'équations mathématiques pour développer et valider des modèles prédictifs. L'interprétation des modèles et l'importance de chaque variable ont été quantifiées soit par l'analyse de leurs coefficients (statistiques conventionnelles) soit par permutation ou masquage des variables dans le cas de modèles d'apprentissage automatique. En conclusion, l'ensemble de ce travail, soit la

reconnaissance et la pronostication automatique de l'IRHA chez l'enfant vont me permettre de développer, de valider et d'implanter un système d'aide à la décision en temps réel pour l'IRHA en pédiatrie.

Mots-clés : enfant, insuffisance respiratoire, hypoxémie, données massives, données continues, base de données à haute résolution temporelle, système d'aide à la décision clinique.

Abstract

Acute hypoxemic respiratory failure (AHRF) is one of the most frequent causes of admission to pediatric intensive care units. It is related to several mechanisms, the most serious of which is lesional pulmonary edema leading to pediatric acute respiratory distress syndrome (ARDS), which accounts for 5–10% of patients admitted to intensive care. Currently, international guidelines for the management of ARDS are under-implemented due to failure to diagnose or late diagnosis. This is probably partly responsible for prolonged mechanical ventilation in pediatric ARDS. In order to improve the criteria for assessing AHRF in children and possibly their outcome, we aimed to improve the early diagnosis of ARDS in children, to automate an organ failure severity score (PELOD-2 score) that can be used as a primary endpoint in research to replace mortality, which is low in this population, and to predict prolonged ventilation in the most fragile population, neonates.

To achieve these objectives, we have: 1) optimized a unique high temporal resolution database, 2) validated a continuous oxygenation index usable in real time and robust to all values of pulsed oxygen saturation, 3) validated a computerized version of the PELOD-2 score usable as a primary outcome in research, 4) developed a predictive model of persistent AHRF due to influenza and 5) proposed a definition of prolonged ventilation in pediatrics applicable regardless of the age and term of the child and 6) studied the outcome of newborns with prolonged ventilation and proposed a predictive model of the most severe subgroup. The methods used across these different studies combined big data science for clustering, synchronization, and normalization of continuous data. We also used descriptive statistics, linear and logistic regression, random forests and their derivatives, deep learning, and empirical optimization of mathematical equations to develop and validate predictive models. The interpretation of the models and the importance of each variable were quantified either by analyzing their coefficients (conventional statistics) or by permuting or masking the variables in the case of machine learning models. In conclusion, all this work, i.e. the recognition and automatic prognosis of AHRF in children will allow me to develop, validate and implement a real-time decision support system for AHRF in pediatrics.

Keywords: children, respiratory failure, hypoxemia, big data, continuous data, high temporal resolution database, computer decision support.

Table des matières

Liste des tableaux

Liste des figures

Liste des sigles et abréviations

AGPL	Licence publique générale Affero (*Affero General Public License*)
aPELOD-2	Version automatique du score PELOD-2
AUROC	Aire sous la courbe ROC (*Receiver Operating Characteristics*)
CHU	Centre hospitalier universitaire
CPAP	Mode ventilatoire à pression expiratoire positive continue (*Continuous Positive Airway Pressure)*
ECMO	Oxygénation extracorporelle (*ExtraCorporeal Membrane Oxygenation*)
$ETCO_2$	Pression partielle en gaz carbonique en fin d'expiration (*end-tidal carbon dioxide*)
EWS.O2	Indice d'alerte précoce (*Early Warning Score*) d'hypoxémie
FiO_2	Fraction inspirée en oxygène
Go	Gigaoctet, soit 10^9 octets
HL7	Langage de type *Health Level 7*
OI	Indice d'oxygénation
ORI	Indice de réserve d'oxygène (*Oxygen Reserve Index*)
OSI	Indice de saturation en oxygène
PALICC	Conférence de consensus sur l'insuffisance respiratoire aigüe (*Pediatric Acute Lung Injury Consensus Conference*)
PaO_2	Pression partielle en oxygène dans le sang artériel
PCO_2	Pression partielle en gaz carbonique dans le sang
PELOD-2	2^e version du score pédiatrique de la défaillance d'organes (*PEdiatric Logistic Organ Dysfunction*)
PEP	Pression expiratoire positive
PF	Rapport PaO_2/FiO_2
P_{moy}	Pression moyenne dans les voies aériennes
ROX	Indice respiratoire d'oxygénation
SADC	Système d'aide à la décision clinique
SaO_2	Saturation artérielle en oxygène

SDRA	Syndrome de détresse respiratoire aigüe
SF	Rapport SpO_2/FiO_2
SpO_2	Saturation pulsée en oxygène
SQL	Langage de requête structurée (*Structured Query Language*)
To	Téraoctet, soit 10^{12} octets
VPN	Réseau virtuel privé (*Virtual Private Network*)
XGBoost	*eXtreme Gradient Boosting*, méthode d'apprentissage automatique

À Aurélie, Noémie, Estelle et Camille

Remerciements

- À Philippe JOUVET, qui est un mentor exceptionnel et qui a su me guider et me soutenir depuis le tout début de ce parcours il y a bientôt 10 ans ;
- À Guillaume EMERIAUD, pour être un modèle de rôle par sa curiosité, sa rigueur et son excellence dans absolument tous les domaines ;
- To Adrienne RANDOLPH, who mentored me and support me during my training in Boston;
- Aux membres du jury, Emmanuel CHARBONNEY, Philippe JOUVET, Guillaume EMERIAUD, Guy WOLF, François LELLOUCHE et Yoan LAMARCHE, pour leur temps passé à réviser cette thèse, leur bienveillance et leurs précieux conseils ;
- À mes collègues et amis qui m'ont formé et soutenu tout au long de ma formation : Laurence DUCHARME-CREVIER, Geneviève DU PONT-THIBODEAU, Catherine FARRELL, Karen HARRINGTON, Jean-Sébastien JOYAL, Jacques LACROIX, Géraldine PETTERSEN, François PROULX, Nadia ROUMELIOTIS, Céline THIBAULT, Baruch TOLEDANO et Marisa TUCCI ;
- À mes collaborateurs impliqués depuis longtemps dans ces projets : Sally AL-OMAR, David BROSSIER et Atsushi KAWAGUCHI ;
- À Jean-Gilles GUIMOND, qui m'a fait découvrir les soins intensifs ;
- À Philippe SAUTHIER et Magali BREGUET, qui ont eu tellement raison d'avoir eu tort trop tôt ;
- À Nicolas et Marie-Amandine SAUTHIER, pour leur support constant et sans faille depuis le tout début et pour être l'exemple du frère et de la sœur parfaits ;
- À Gabrielle, bien sûr, qui m'a permis d'aller jouer dans le nuage tout en gardant les pieds sur terre, et qui donne un sens à toute cette quête.

Introduction

L'insuffisance respiratoire hypoxémique aigüe (IRHA) représente une capacité diminuée de l'appareil respiratoire à oxygéner adéquatement le corps (1–3). Cette condition requiert un traitement urgent, car elle peut entrainer des lésions pouvant s'accompagner de séquelles à long terme ou du décès. Le diagnostic de l'IRHA repose sur une quantification de l'oxygène dans le sang. Plus cette quantité est faible, plus il sera nécessaire d'enrichir l'air inspiré par le patient avec de l'oxygène (1), afin que la quantité mesurée dans le sang reste à un niveau sécuritaire pour les organes. D'autres traitements, comme la ventilation mécanique, vont s'ajouter dans les cas plus graves. Le diagnostic et l'évaluation de la gravité de l'insuffisance respiratoire semblent être une tâche simple au premier abord étant donné le peu de paramètres impliqués. Cependant, les données physiologiques évoluent constamment et sont en interaction avec les autres systèmes du corps humain, ce qui les rend complexes à modéliser. En outre, les paramètres physiologiques doivent être relativisés et personnalisés aux antécédents et particularités du patient pour établir la gravité de l'atteinte et le plan de traitement.

L'IRHA est une entité appartenant à plusieurs syndromes, dont le plus grave est l'œdème pulmonaire lésionnel conduisant au syndrome de détresse respiratoire aigüe (SDRA). Le SDRA pédiatrique représente environ 5 % des admissions en soins intensifs pédiatriques et a une mortalité allant jusqu'à 40 % (4). Il est associé à plusieurs complications, en particulier le besoin de ventilation mécanique prolongée (4,5) et à des séquelles chez le patient et des conséquences sur la qualité de vie familiale (6,7). L'IRHA est un élément central des critères diagnostiques du SDRA pédiatrique (**Tableau 1**) (1,4).

Critère	**Description**	
Âge	Exclusion des patients avec atteinte pulmonaire périnatale.	
Délai d'apparition	Dans les 7 jours suivant la lésion clinique	
Origine de l'œdème pulmonaire	L'insuffisance respiratoire ne doit pas être pleinement expliquée par une insuffisance cardiaque ou par une surcharge liquidienne	
Imagerie thoracique	L'imagerie retrouve un ou des infiltrat(s) suggestif(s) d'une maladie pulmonaire parenchymateuse aigüe	
IRHA	*Ventilation invasive :* Légère : $4 \leq OI < 8$ ($5 \leq OSI^{*} < 7.5$) Modérée : $8 \leq OI < 16$ ($7.5 \leq OSI^{*} < 12.3$) Grave : $OI \geq 16$ ($OSI^{*} \geq 12.3$)	*Ventilation non invasive :* Masque facial avec 2 niveaux de pression ou CPAP≥5 cmH_2O et $PF \leq 300$ ou $SF^{*} \leq 264$

IRHA : Insuffisance respiratoire hypoxémique aigüe, OI : indice d'oxygénation (voir tableau suivant), OSI : indice de saturation (voir tableau suivant), CPAP : *Continous Positive Airway Pressure*, PF : rapport PaO_2/FiO_2, SF : rapport $SpO_2/FiO2$. Les mesures artérielles ont priorité sur les mesures de SpO_2 lorsque celles-ci sont disponibles. * $SpO_2 \leq 97$ %.

Tableau 1. – Critères diagnostics du syndrome de détresse respiratoire aigüe pédiatrique pour la population générale (2015)

Le diagnostic de l'insuffisance respiratoire hypoxémique aigüe

En 1994, les premiers critères diagnostics adultes de l'IRHA dans le cadre du SDRA ont été établis par le consensus américain et européen (8). Le critère d'IRHA du SDRA reposait sur le rapport de la pression partielle en oxygène dans le sang artériel (PaO_2) divisé par la fraction inspirée d'oxygène (FiO_2), appelé rapport PF (**Tableau 2**). L'utilisation de ce marqueur pour définir l'hypoxémie était surtout justifiée par l'association entre sa valeur et la mortalité hospitalière (8). Les définitions successives du SDRA chez l'adulte ont modifié certains critères, comme l'ajout d'une valeur de pression expiratoire positive (PEP) minimale (9), mais le rapport PF est resté central à la définition de l'hypoxémie (9).

Spécimen artériel	Ventilation invasive	Ventilation non invasive
Oui	$OI = \frac{P_{moy} \cdot FiO_2 \cdot 100}{PaO_2}$	$PF = \frac{PaO_2}{FiO_2}$
Non	$OSI = \frac{P_{moy} \cdot FiO_2 \cdot 100}{SpO_2}$	$SF = \frac{SpO_2}{FiO_2}$

OI : indice d'oxygénation, P_{moy} : pression moyenne dans les voies aériennes, FiO_2 : fraction inspirée en oxygène, PaO_2 : pression partielle en oxygène dans le sang artériel, OSI : indice de saturation. SpO_2 : saturation pulsée en oxygène (≤ 97 %).

Tableau 2. – Équations quantifiant l'hypoxémie en pédiatrie.

En pédiatrie, les premiers critères diagnostics du SDRA (**Tableau 1**) ont été établis en 2015 lors de la conférence de consensus sur l'insuffisance respiratoire aigüe (PALICC, *Pediatric Acute Lung Injury Consensus Conference*). Du point de vue de l'hypoxémie, ces critères différaient de ceux utilisés pour les patients adultes par l'utilisation de l'indice d'oxygénation (OI, *oxygenation index*) lors de la ventilation invasive (**Tableau 2**). Ce dernier intègre, en plus de la PaO_2 et de la FiO_2, la pression moyenne des voies aériennes (P_{moy}). Cette pression appliquée par les appareils de ventilation mécanique permet d'améliorer le recrutement alvéolaire et l'oxygénation, indépendamment de la FiO_2 (10,11). Cependant, elle induit un stress sur les alvéoles et peut également causer des lésions pulmonaires secondaires. L'OI étant plus fortement associé à la mortalité que le rapport PF en pédiatrie (5), il a été retenu comme la référence dans la classification de la gravité de l'IRHA pédiatrique (**Tableau 3**). Cependant, l'OI nécessite une mesure fiable de P_{moy}, ce qui n'est possible qu'en ventilation invasive et exclut par le fait même la ventilation non invasive et les lunettes nasales à haut débit. Dans ces cas, les lignes de conduite préconisent l'utilisation du rapport PF sans catégorisation de la gravité (5). Les critères diagnostics du SDRA pédiatrique sont en cours de révision (PALICC 2) et seront publiés vers au début 2022.

Gravité de l'hypoxémie	Indice d'oxygénation (OI)	Indice de saturation (OSI) [a]
Aucune ou minimale	< 4	< 5
Légère	4 – 7,9	5 – 7,4
Modérée	8 – 15,9	7,5 – 12,2
Grave	≥ 16	≥ 12,3

[a] $SpO_2 \leq 97$ %

Tableau 3. – Seuils de gravité de l'hypoxémie en pédiatrie (ventilation invasive)

Toutefois, on constate que le SDRA est largement sous-diagnostiqué avec une reconnaissance entre 25-60 % des cas par les cliniciens (12–14). De plus, même lorsqu'il est diagnostiqué par les soignants, il l'est plus souvent tardivement même lorsqu'il est grave (13,15,16). Ce sous-diagnostic a des implications importantes sur l'adhésion aux recommandations de prise en charge qui, lorsqu'elles sont suivies, sont associées à un meilleur devenir (4,17). En particulier, il été démontré en pédiatrie que l'utilisation d'un volume courant plus élevé que celui recommandé ou d'une pression expiratoire positive plus faible que celle recommandée était tous les deux indépendamment associée à une augmentation de la mortalité et de la durée de ventilation mécanique en pédiatrie (17). Les auteurs de cette étude constatent aussi que la supplémentation d'oxygène ne respecte pas les recommandations dans plus de 70 % du temps. En outre, plus le SDRA était grave, plus l'adhésion aux recommandations diminuait, notamment les cibles de volume courant, et plus les conséquences sur la mortalité étaient significatives (17). Les autres facteurs de risque d'adhésion insuffisante aux recommandations du volume courant étaient la gravité de l'IRHA (surtout) et le statut pondéral ou nutritionnel (17).

La reconnaissance et la catégorisation de l'IRHA sont donc capitales pour le SDRA et l'application adéquate des recommandations permettant un meilleur devenir. Dans ce contexte, notre hypothèse est qu'un système d'aide à la décision clinique (SADC) permettra de mieux diagnostiquer l'IRHA et plus rapidement, et donc d'optimiser l'adhésion aux lignes de conduite afin d'améliorer leur devenir.

Les systèmes d'aide à la décision clinique en temps réel

Comme leur nom l'indique, les systèmes d'aide à la décision clinique (SADC) sont des systèmes informatiques visant à compléter le clinicien en lui offrant une forme d'assistance dans son processus décisionnel pour en améliorer le résultat (18,19). Il est également espéré que les SADC puissent accélérer le transfert des connaissances, réputé sinon pour prendre jusqu'à 20 ans autrement (20). Ils ont été montrés efficaces, particulièrement en pédiatrie, dans les domaines de la prescription, de la demande de tests diagnostiques, de la documentation et de la vaccination (21). Les SADC en temps réel sont plus complexes, tant au niveau du noyau décisionnel que dans la couche d'interface. Au niveau du noyau, l'utilisation d'un flux de données continu apporte des défis non rencontrés avec des données ponctuelles, par exemple la gestion des interruptions et des valeurs corrompues, la gestion du délai de rétention et les fonctions d'agglomération (18). Sur le plan de l'interface, le circuit du message requiert un choix fiable du destinataire, la confirmation de la réception des notifications et la rétroaction sur les actions du soignant (18).

Dans le domaine respiratoire, on retrouve principalement 2 catégories : les SADC dédiés à améliorer le diagnostic et ceux ciblant l'amélioration des conduites thérapeutiques (22). Dans la première catégorie, plusieurs études adultes ont montré que des règles explicites amélioraient significativement l'identification des patients atteints de SDRA en passant de 25-50 % à 75-95 % de reconnaissance (12–15,23). D'autres modèles utilisant l'apprentissage automatique, notamment le traitement du langage naturel dans les notes médicales et les rapports de radiographie, arrivent aux mêmes conclusions (24–29). Toutefois, la plupart de ces études rétrospectives sont difficilement applicables dans un contexte de temps réel ou de détection précoce, car les notes médicales sont en retard sur le processus analytique du clinicien.

Dans la seconde catégorie, soit les SADC voués à améliorer le respect des lignes de conduite, plusieurs ont montré un certain bénéfice sur le volume courant (30–35) ou sur d'autres critères variés reconnus dans la littérature, comme la supplémentation d'oxygène ou la pression expiratoire positive (35–39). D'une façon similaire aux SADC, des appareils autonomes (boucle fermée) ou semi-autonomes et à différents stades de développements ont été validé dans le domaine du sevrage

de la ventilation mécanique (40–44) ou dans l'ajustement automatique de l'oxygénothérapie, selon une cible de SpO_2 prédéterminée (45). Toutefois, ces appareils ont une portée limitée en pédiatrie (nécessitent des patients > 2 ans) et les algorithmes sont souvent propriétaires et indissociables des appareils biomédicaux. Ce dernier point rend l'interopérabilité plus limitée, notamment avec le dossier médical afin de personnaliser automatiquement le traitement.

Dans ce contexte, notre hypothèse est qu'un SADC en temps réel représente une excellente solution pour non seulement reconnaitre l'IRHA précocement, mais également classifier automatiquement la gravité et améliorer l'adhésion aux recommandations. Toutefois, plusieurs barrières limitent la réalisation d'un tel système.

Problématique de travail

Le développement de SADC est une solution qui semble s'imposer pour améliorer la reconnaissance précoce de l'IRHA, mais plusieurs éléments limitent leur développement.

L'absence de prise en compte de la temporalité dans les marqueurs d'hypoxémie

La temporalité n'est pas définie pour le calcul des marqueurs d'hypoxémie (rapport PF ou OI). Il est sous-entendu que les composants (FiO_2, P_{moy} et la PaO_2) sont mesurés au même instant, mais aucune directive n'a été publiée sur la tolérance maximale de délai de recueil acceptable entre ces paramètres, la PaO_2 étant obtenue à partir d'un prélèvement sanguin et la FiO_2 et P_{moy} à partir d'un instrument médical ou d'une note de soignant. La littérature est divisée entre des études prospectives basées sur une mesure unique au chevet, donc simultanée, et d'autres rétrospectives utilisant des données extraites du dossier médical s'étalant entre 0 et 60 minutes avant la mesure de la PaO_2 (3,46–49).

De plus, le moment de la mesure de la PaO_2 est un choix, dont l'objectif est le plus souvent d'être représentatif de l'état du patient. Dans l'étude de validation des lignes de conduite PALICC (étude PARDIE), il était spécifié que les paramètres enregistrés devaient être pris à un moment de stabilité (4). En contexte clinique, les évènements pouvant mener à des perturbations de l'oxygénation sont fréquents : aspirations endotrachéales, examens diagnostics, artéfacts de mouvement, modification sur le respirateur, thérapies de recrutement, etc. De plus, le temps pour une normalisation de l'ensemble des paramètres n'est pas spécifié. Ce moment « moyen » est donc basé sur un ensemble de facteurs subjectifs définissant ce concept. En dehors d'un cadre de recherche, cette période de stabilité n'est pas définie, ce qui complique l'automatisation du processus.

Par ailleurs, la mesure continue de l'oxygénation (SpO_2) est très sujette à des variations artéfactuelles. En conséquence, le développement de systèmes automatisés nécessite l'établissement de critères de qualité des signaux qui soient robustes à l'instabilité. Les lignes de conduite actuelles pour l'utilisation continue de l'OSI et de rapport SF nécessitent donc que le signal soit stable, sans quoi leur application stricte fera en sorte qu'un patient pourra changer rapidement entre plusieurs catégories ou osciller autour d'un seuil. Afin de diminuer cet effet, certaines études dans la prévention de l'hyperoxie se sont intéressées à l'effet cumulatif plutôt qu'instantané (50). D'autres ont intégré une analyse continue pondérée dans des appareils d'ajustement automatique de la FiO_2 ou en sevrage de la ventilation (boucle fermée) (43,45,51). Dans les deux cas, ils se sont affranchis d'une catégorisation et pallient les données corrompues.

L'imprécision de la SpO_2

La PaO_2 est l'étalon d'or de l'hypoxémie, mais elle n'est mesurée que sur une minorité de patients en pédiatrie, en particulier à la phase précoce de la maladie (4). Son prélèvement nécessite soit une ponction artérielle, soit la pose d'un cathéter dans une artère, geste techniquement difficile chez les plus petits patients. Ces procédures sont associées à des complications, comme la douleur ou les thromboses, en particulier chez les plus jeunes patients (52). Par conséquent, cela crée un biais dans les études, car les mesures artérielles sont plus fréquentes chez les patients ayant l'atteinte la plus grave. Dans la population de SDRA pédiatrique, moins de 45 % des patients avaient des mesures artérielles, biaisant les conclusions potentiellement (4). En outre, la mesure de la PaO_2 n'est faite que ponctuellement, ce qui nuit au diagnostic rapide ou pour le suivi de l'IRHA. Les cathéters intraartériels permettant une mesure continue existent, mais ils sont réservés à des enfants plus âgés (>2 ans et via l'artère fémorale) (53). Bien qu'ils fournissent des résultats relativement précis et non biaisés (54), ils sont limités à quelques jours d'utilisation (55) et sont associés à des complications médicales et connus pour plusieurs problèmes techniques (56). Par conséquent, ils ne font pas partie de la pratique courante en pédiatrie.

L'alternative la plus courante est la SpO_2. Connue depuis les années 1970, cette technologie facilement accessible, peu couteuse, indolore et disponible en continu, fait partie de la surveillance

standard des patients en anesthésie ou en soins intensifs (2). À l'aide d'un capteur placé sur une extrémité du corps, la SpO_2 fournit une estimation de la saturation artérielle en oxygène (SaO_2). Cette estimation est généralement fiable (± 3 % absolu) (49). Toutefois, des différences significatives existent selon la pigmentation de la peau (SpO_2 mesurée sur peau blanche adulte est généralement plus élevée que sur une peau noire adulte (49)), selon le site anatomique de la mesure (57), selon la présence de certaines hémoglobinopathies (58,59) et possiblement aussi selon la marque du saturomètre (60–62). Bien que la relation SaO_2 – PaO_2 soit connue depuis plus d'un siècle (63), la SaO_2 était initialement décrite en fonction de la PaO_2. L'inverse est plus complexe en raison de leur relation asymptotique (**Figure 1**).

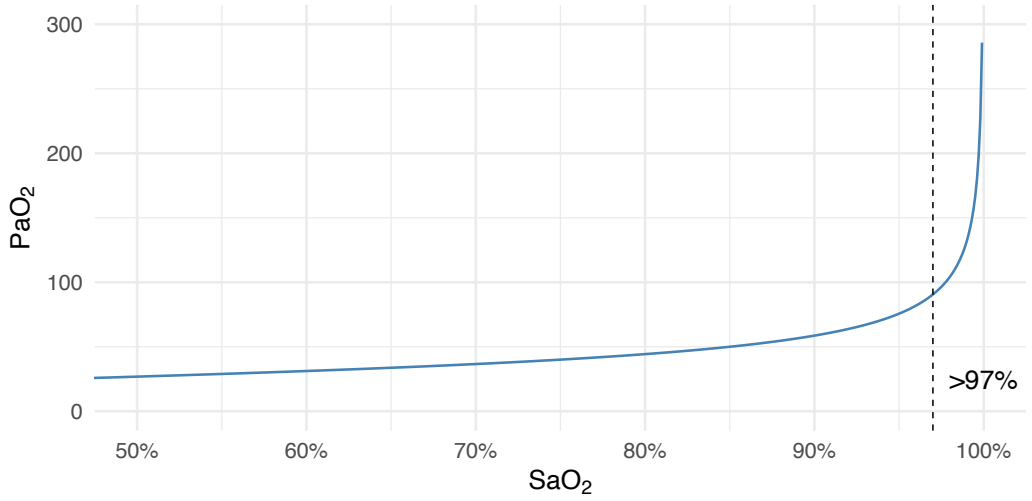


Figure 1. – Courbe d'oxyhémoglobine de Severinghaus avec la PaO_2 en variable dépendante

C'est en raison de ce risque d'imprécision que seules les valeurs de $SpO_2 \leq 97$ % ont été validées pour le diagnostic et la classification de l'hypoxémie à l'aide de l'indice de saturation en oxygène (OSI) et du rapport SF (SpO_2/FiO_2) (4,5,64–68). Pour les valeurs de $SpO_2 > 97$ %, il est plutôt recommandé de diminuer la FiO_2 jusqu'à ce que la SpO_2 soit dans une plage plus fiable. Les études de validation en pédiatrie utilisaient majoritairement les sondes Masimo et Nellcor (65). Malgré ces recommandations, on retrouve plus de 60 % des valeurs de SpO_2 qui sont au-dessus de ce seuil (69,70), rendant les équations actuelles inutilisables en dépistage automatisé de l'IRHA.

Les données corrompues

Un SADC présuppose un accès à des données fiables. Johnson et ses collègues ont décrit les 3 défis principaux dans l'utilisation des données aux soins intensifs pour un SADC : compartimentation, corruption et complexité (18). La compartimentation représente la difficulté d'obtenir les données en provenance de plusieurs sources. Dans le cas de l'hypoxémie, même un calcul simple comme l'OI ou l'OSI nécessite déjà plusieurs sources de données, notamment les laboratoires (PaO_2), les moniteurs de surveillance (SpO_2) et les ventilateurs (P_{moy} et FiO_2). Afin qu'elles soient exploitables, il faut s'assurer que les données soient synchrones, impliquant une gestion des horloges des différents appareils biomédicaux et une latence minimale dans le transit des données (71). Également, il faut prévoir devoir harmoniser des fréquences de mesure différentes entre les données. Par exemple, la FiO_2 ou la P_{moy} peuvent être mesurées à chaque seconde, mais la PaO_2 ne sera mesurée que toutes les quelques heures. De plus, dans les soins intensifs informatisés, il existe plusieurs sources de données pour une même variable. Par exemple, la FiO_2 peut être fournie automatiquement par le respirateur (FiO_2 mesurée et FiO_2 réglée), ou par l'appareil de diffusion du monoxyde d'azote, ou consignée par différents professionnels dans le dossier médical. Un travail de hiérarchisation et de validation est nécessaire à cette étape.

Le second défi est la corruption des données qui est détaillée en 3 sous-catégories : les données erronées, manquantes et imprécises. Comme mentionné plus haut, la PaO_2 est précise, mais n'est disponible qu'épisodiquement, ce qui la rend manquante le reste du temps. Elle peut occasionnellement être erronée si un problème survient avec l'analyse (bulle d'air) ou si le prélèvement est mal étiqueté (en particulier, spécimen veineux au lieu d'artériel). La SpO_2 est également très touchée par le problème de précision et devient manquante, selon les lignes de conduite, lorsque sa valeur est supérieure à 97 %. Elle peut également être erronée lorsque le signal est de mauvaise qualité, par exemple en cas de mauvaise perfusion dans les états de choc ou lorsque l'enfant bouge ou enlève le capteur. La pression moyenne de ventilation peut également être concernée par le problème de corruption. La lecture de celle-ci n'est fiable qu'en cas de ventilation invasive, la rendant manquante ou imprécise dans les autres circonstances (ventilation non invasive, lunettes nasales à haut débit). Elle peut également être perturbée par la respiration

spontanée du patient ou par les interventions des soignants, notamment lors de l'aspiration endotrachéale.

Le troisième défi est la complexité des données, en particulier dans le cas de données continues (temporelles) multimodales et asynchrones (**Figure 2**). L'évaluation de la tendance tout en excluant les données aberrantes est une tâche relativement simple pour un être humain. En revanche, pour une machine, cette tâche est encore hautement complexe.

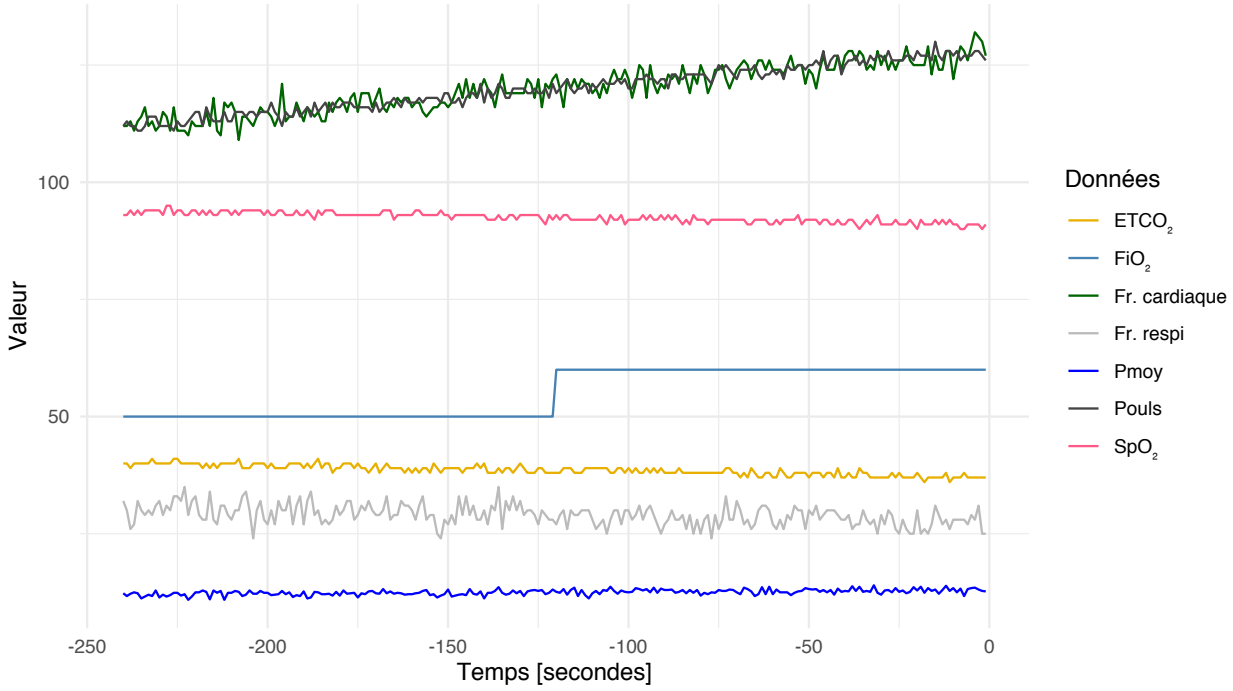


Figure 2. – Données multimodales sérielles (synthétiques) illustrant la complexité de l'évolution en temps réel (0 représente l'instant présent) des données chez un patient de soins intensifs

Conduite de la thèse

Objectifs

L'objectif principal de cette thèse est de développer les outils informatiques permettant de détecter et de prédire automatiquement le devenir des enfants en IRHA.

Les objectifs secondaires étaient :

1) Optimiser une base de données à haute résolution temporelle nécessaire au développement des outils informatiques ;
2) Mettre au point un marqueur non invasif continu d'hypoxémie à partir de 2 bases de données haute résolution temporelle ;
3) Valider l'informatisation d'un score de gravité permettant d'obtenir automatiquement le PELOD-2 non respiratoire comme critère de jugement possible ;
4) Modéliser le devenir des patients avec IHRA à la suite d'une infection grave à influenza ;
5) Créer une définition pédiatrique de ventilation prolongée adaptée à tous les âges et termes d'enfants ;
6) Modéliser le devenir des nouveau-nés avec ventilation prolongée.

Démarche scientifique

Cette thèse repose sur l'optimisation et le déploiement d'une base de données de recherche à haute résolution temporelle et 5 articles scientifiques réalisés entre 2016 et 2021 au CHU Sainte-Justine (Montréal, Canada) et au Boston Childrens's Hospital (Boston, États-Unis). Ils sont centrés sur la reconnaissance automatique de l'IRHA et de sa gravité dans différentes populations pédiatriques.

Optimisation d'une base de données recherche à haute résolution temporelle

Cette section décrit la réalisation et le déploiement d'un système automatisé et continu de collecte des données à haute résolution temporelle des appareils de surveillance et de maintien des fonctions vitales aux soins intensifs pédiatriques du CHU Sainte-Justine. Les données incluent les valeurs numériques enregistrées chaque seconde et le signal brut pour l'ensemble des patients admis. Les contraintes techniques et leurs solutions y sont abordées.

Article 1

Estimated PaO_2: A Continuous and Noninvasive Method to Estimate PaO_2 and Oxygenation Index. **Sauthier** M, Tuli G, Jouvet PA, Brownstein JS, Randolph AG. *Crit Care Explor.* 2021 Oct; *3*(10):e0546.

Cet article avait pour objectif de diagnostiquer et classifier en continu de façon automatique la gravité de l'hypoxémie. Ce travail compare 3 méthodes d'apprentissage profond multimodales et une méthode d'optimisation empirique, afin d'estimer la PaO_2 en continu en fonction des données continues. On y démontre la supériorité de cette méthode pour estimer l'indice d'oxygénation comparé à la référence actuelle (OSI) ainsi que sa validité pour les SpO_2 plus élevée que 97 %, ce qui permet d'être utilisable en tout temps pour reconnaitre l'IRHA.

Article 2

Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. **Sauthier** M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9.

Cette étude a démontré qu'un score de gravité validé initialement pour être calculé manuellement pouvait être automatisé engendrant probablement moins d'erreurs et un calcul incomparablement plus rapide. L'automatisation a fait appel au traitement des données à l'aide d'algorithmes par requêtes structurées. Ce développement était nécessaire, afin que les SADC soient en mesure de reconnaitre objectivement l'amélioration ou la dégradation des patients en IRHA.

Article 3

Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. **Sauthier** MS, Jouvet PA, Newhams, MM, Randolph AG. *Crit Care Explor*. 2020;2(8):e0175.

Cette étude sur une cohorte observationnelle prospective multicentrique d'enfants atteints d'influenza grave montre que la persistance de l'IRHA et leur besoin d'hospitalisation en soins intensifs à 7 jours est prévisible dès le premier 24 h d'hospitalisation. Cette étude décrit également l'importance d'autres marqueurs que l'hypoxémie dans la prédiction de l'évolution de l'état respiratoire. Enfin, cette étude démontre la fréquence des données manquantes ou inutilisables avec une application stricte des lignes de conduite et présente des solutions.

Article 4

Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. **Sauthier** M, Rose L, Jouvet P. *Respir Care*. 2017 Jan 1;62(1):49–53.

Cette revue de la littérature (*scoping review*) a permis de proposer la première définition opérationnelle de la ventilation prolongée en pédiatrie, afin d'en faire un critère de jugement alternatif à la mortalité. Le développement de nouveaux critères est nécessaire pour les futures études de validation des SADC dans une population où la mortalité diminue, mais la morbidité augmente. Celle-ci tient compte des particularités modernes retrouvées en soins intensifs pédiatriques, comme les populations hétérogènes en termes d'âges gestationnels (prématurité) ou de malformations congénitales et les modes ventilations alternatifs, comme les lunettes nasales à haut débit.

Article 5

Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. **Sauthier** M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et coll. *Front Pediatr*. 2021;9(July):1–10.

Cette revue rétrospective sur une population néonatale hétérogène visait à appliquer la définition de ventilation prolongée. Cette étude décrit cette population et son devenir à 18 mois d'âge corrigé. Elle montre également un modèle prédictif permettant d'identifier précocement le sous-groupe d'enfant ventilé le plus longtemps, afin de pouvoir les cibler lors de futures études interventionnelles.

Optimisation et déploiement d'une base de données à haute résolution temporelle

Le point de départ de tout SADC est l'accès aux données (72). En particulier dans le cas de SADC en temps réel, il est essentiel que l'ensemble des données soient immédiatement et continuellement disponibles. Lorsque nous avons entrepris ce développement, aucune solution informatique regroupant l'ensemble de la surveillance de patients de soins intensifs à haute résolution (moniteur de surveillance des fonctions vitales, appareils de maintien des fonctions vitales, pompes et pousse-seringues et autres appareils biomédicaux) associés au dossier médical électronique (démographie, diagnostic et thérapeutiques) du patient n'existait (73). La grande majorité des bases de données médicales électronique étaient celles du dossier médical enregistrant les données selon la pratique standard aux soins intensifs, soit un point par heure (73). La notion de « haute résolution temporelle » implique un nombre de points suffisamment élevés pour recréer l'évolution détaillée d'un patient ; un concept que nous avions nommé « patient perpétuel » (73). Dans un cadre de soins intensifs, les données à haute résolution se divisent en deux catégories : numérique et signal brut. Les données numériques sont le résultat calculé par le moniteur (par exemple la valeur de la fréquence cardiaque) à partir du signal brut (le tracé électrocardiographique dans ce cas-ci). L'un et l'autre sont complémentaires, mais ont des considérations techniques très différentes quant à leur capture, leur stockage et leur utilisation. Nous avons donc entrepris de développer une base de données à haute résolution temporelle qui enregistrerait toutes les données en continu (numérique et signal brut) pour l'ensemble des patients admis aux soins intensifs.

Le développement du prototype (73), auquel j'ai activement participé, a fait l'objet d'une validation subjective par les utilisateurs (74) et objective de la qualité des données (71). De telles validations sont extrêmement rares dans la littérature et notre travail était pionnier dans ce cas. Toutefois, cette preuve de concept souffrait de plusieurs limitations et j'ai entrepris alors un développement significatif de cette base de données qui sont décrites dans cette thèse, afin d'en faire un système autonome de collecte de données à haute résolution temporelle.

Réalisation du prototype

Le prototype a été développé entre 2014 et 2018 après avoir obtenu l'approbation du comité d'éthique de la recherche du CHU Sainte-Justine pour la constitution de la banque de données (73). L'approbation portait sur la collecte et l'enregistrement de toutes les données cliniques mesurées aux soins intensifs. Toutefois, l'exploitation de ces données à des fins de recherche nécessite une autorisation distincte du comité d'éthique et propre à chaque projet de recherche. De plus, la constitution du registre est indiquée aux familles sur le dépliant remis aux familles lors de l'admission et précise aussi la possibilité d'être exclue de la base de données.

La première version consistait à « écouter » les messages HL7 v2 (format de données *Health Level 7*, version 2) envoyés par les moniteurs de surveillance à chaque 5 secondes et à les enregistrer dans une base de données en utilisant un format de requêtes structurées (SQL, *Structured Query Language*), de type Transact-SQL (SQL Server 2008, Microsoft, Redmond, États-Unis). Dans les messages HL7 v2, l'ordre des données est variable (déterminé par l'émetteur) et le type d'information est précisé avant sa valeur. Dans ce prototype, les données étaient décodées du message et enregistrées dans l'ordre dans lequel le message était reçu dans des colonnes non spécifiques : valeur 1, valeur 2, valeur 3, etc. L'information de chaque valeur était séparée en 2 colonnes : nom de la variable, et valeur de la variable (**Tableau 4**).

Identifiant	**Horodate**	**Nom1**	**Valeur1**	**Nom2**	**Valeur2**	**Nom3**	**Valeur3**
464664	2019-03-12 14:55:03	Pouls	112	SpO_2	98	FR	33
464664	2019-03-12 14:55:08	SpO_2	98	Pouls	108	FR	30
464664	2019-03-12 14:55:13	SpO_2	98	FR	35	Pouls	110
…	…	…	…	…	…	…	…

Tableau 4. – Exemple d'enregistrement structuré des messages HL7 envoyés par moniteurs de surveillance pour un patient.

Bien que fonctionnelle, cette structure s'est révélée limitante par plusieurs aspects. Premièrement, le nombre fixe de colonnes dans la base de données étaient incompatibles avec une structure dynamique des messages HL7. Ainsi, les variables supplémentaires dépassant le nombre de colonnes prévues n'étaient pas enregistrées dans la base de données (71). Nous avons augmenté le nombre de colonnes jusqu'à 20 pour atténuer l'impact, mais la limitation persistait occasionnellement avec des valeurs non sauvegardées. Deuxièmement, l'extraction d'informations était complexe en raison de la répartition aléatoire des variables entre les colonnes. Par exemple, la fréquence cardiaque pouvait être dans la colonne 3 pendant un moment, puis dans la colonne 8 sans autre logique que l'ordre d'envoi par le moniteur de surveillance. Ainsi, à chaque interrogation de la base, il fallait systématiquement déconstruire l'ensemble des colonnes (virtuellement) pour sélectionner les variables d'intérêts, ce qui était problématique d'un point de vue de la performance. Troisièmement, les colonnes multiples contenaient divers types de données, principalement du texte, des nombres entiers et des nombres décimaux à différentes précisions. Dans le langage SQL, cela forçait l'utilisation d'un type générique de données (appelé « VARCHAR » en Transact-SQL), mais impactait négativement les performances des index et la gestion de l'espace de stockage. Quatrièmement, l'insertion continue et massive des données (>1 milliard de lignes) dans une base de données structurées a fini par rendre la gestion des index impossible. Ces derniers sont un répertoire organisé des données pour faciliter leur identification. Chaque ajout de ligne dans la base de données nécessite leur mise à jour. Lorsque la base atteint un volume significatif, l'insertion de nouvelles données est fortement ralentie par la mise à jour de ces index. Cinquièmement, ce prototype exigeait une intervention manuelle pour déclencher l'enregistrement de certaines données, en particulier le signal brut, causant des données manquantes et des doublons. Sixièmement, les données étaient décentralisées dans plus de 11 tableaux, compliquant les interrogations. Septièmement, les horloges des appareils biomédicaux n'étaient pas toutes synchronisées sur le temps universel contrôlé et ne comprenaient pas le fuseau horaire permettant de marquer, entre autres, l'heure avancée en été. De plus, l'étude de validation objective a révélé un délai dans le transit des données allant jusqu'à plusieurs dizaines de secondes occasionnellement. Par conséquent, les horodates inscrits dans la base de données n'étaient pas complètement fiables et comparables entre eux. Dernièrement, ce prototype reposait partiellement sur des technologies propriétaires, comme Windows Server (Microsoft, Redmond, États-Unis) ou

SQL Server (Microsoft, Redmond, États-Unis), rendant les couts d'installation élevés et limitant sa diffusion.

Contributions à la base de données

Afin d'améliorer les limitations de cette base de données, j'ai entrepris une réécriture complète. J'ai également réalisé l'importation et le nettoyage des données de la première version de la base de données pour les intégrer dans la nouvelle. Il en résulte un programme complet de collecte et de stockage des données reposant entièrement sur des technologies à licences ouvertes (*open source*). Ce programme gère également les accès aux plus de 30 chercheurs utilisant cette base de données aujourd'hui. Les améliorations se répartissent en plusieurs domaines décrits ci-dessous.

Structure et schéma de données

Le système de gestion de la base de données choisie est PostgreSQL v12.4 (PostgreSQL Global Development Group) pour son utilisation multiplateforme, sa gestion avancée des index, son intégration facilitée avec des bases non structurées (NoSQL) et sa licence ouverte et gratuite. Comme d'autres systèmes SQL, il assure des transactions de type ACID (*atomic*, *consistent*, *isolated*, *durable*) garantissant leur fiabilité. Ce principe est essentiel dans le cas d'une utilisation simultanée par plusieurs utilisateurs de la base de données.

La structure choisie est un schéma en étoile, dont le centre est l'identification du patient par le dossier médical électronique (**Figure 3**). Chaque tableau spécifique à une catégorie de données (numérique ou signal brut) a ensuite une structure dite « clé-valeur » ou EAV (*entity-attribute-value*) (75). Cette structure repose sur 4 colonnes principalement : l'identification du patient, l'horodate, le nom de la variable et sa valeur (**Tableau 5**). Les avantages retenus sont la structure très polyvalente et dynamique ; l'ajout d'une nouvelle variable ne pose aucun problème contrairement à un modèle avec des colonnes spécifiques. En revanche, ces tableaux nécessitent parfois d'être traités avant d'être analysé (pivotage) et occupent plus d'espace de stockage, car l'horodate et l'identification sont répétés pour chaque valeur (**Tableau 6**). De plus, des ajustements

sont nécessaires aux structures clés-valeurs pour maintenir les performances, car la longueur importante de ces tableaux complique la gestion des index (75).

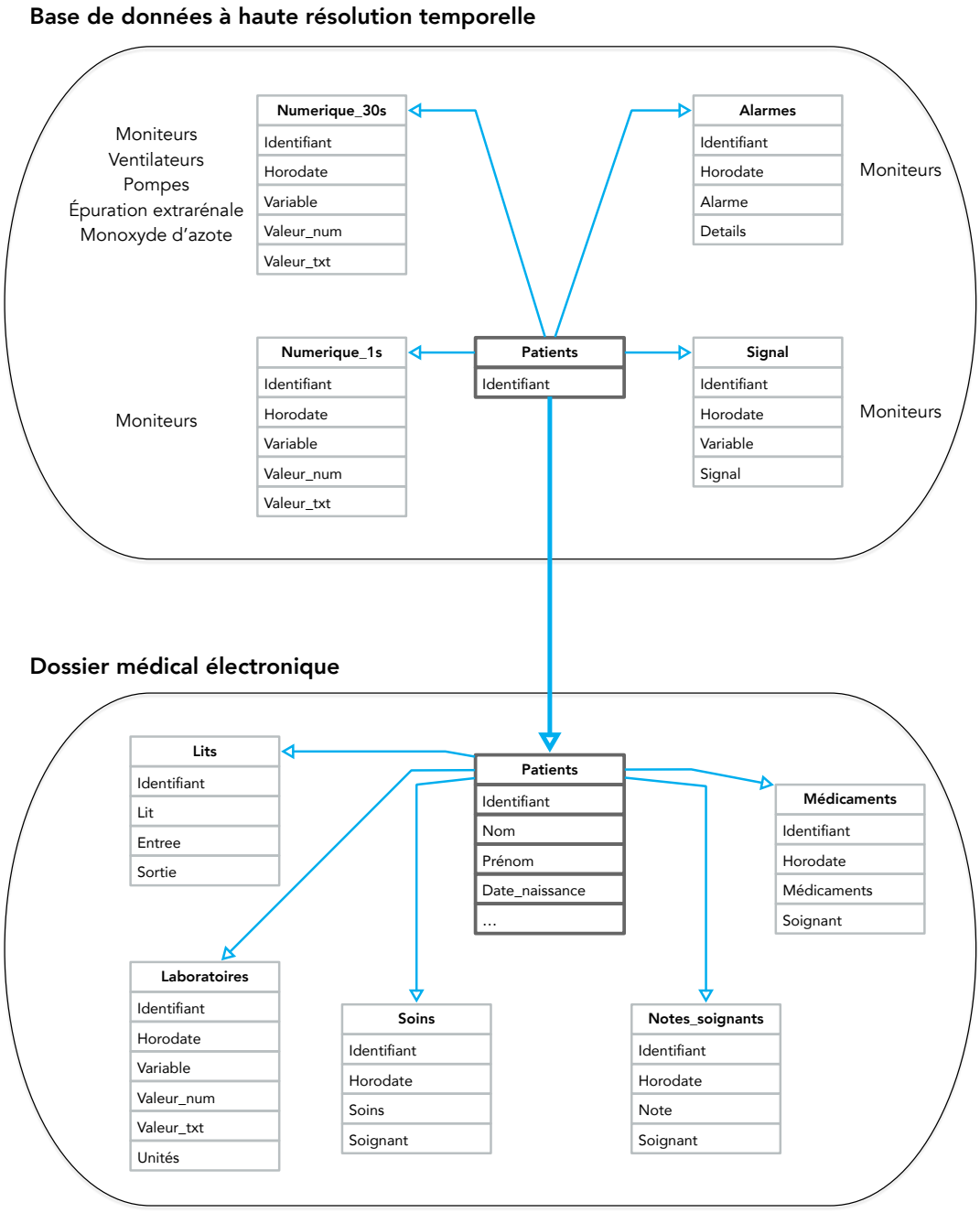


Figure 3. – Schéma simplifié de la structure de la base de données à haute résolution temporelle associée au dossier médical électronique

Identifiant	Horodate	Variable	Valeur_num	Valeur_txt
464664	2019-03-12 14:55:03	Pouls	112	
464664	2019-03-12 14:55:03	SpO_2	98	
464664	2019-03-12 14:55:03	FCardiaque	110	
464664	2019-03-12 14:55:03	FRespi	18	
464664	2019-03-12 14:55:03	Rythme	*NULL*	sinusal
464664	2019-03-12 14:55:05	Pouls	116	
464664	2019-03-12 14:55:05	SpO_2	98	
…	…	…	…	

Tableau 5. – Exemple d'une structure clé-valeur pour des données numériques et textuelles

Identifiant	Horodate	Pouls	SpO_2	FCardiaque	FRespi	Rythme
464664	2019-03-12 14:55:03	112	98	110	18	sinusal
464664	2019-03-12 14:55:05	116	98	*NULL*	*NULL*	
…	…	…	…	…	…	…

Tableau 6. – Exemple d'une structure clé-valeur après pivotage

Automatisation

Afin de développer des SADC en temps réel, il était capital que la base de données soit continuellement mise à jour. J'ai donc développé un programme qui automatise complètement les tâches de collecte, normalisation et insertion dans la base de données en temps quasi réel. Le langage du programme principal est écrit en Python v3.9 (Python Software Foundation) et utilise plusieurs modules accessibles librement (*numpy*, *python-dateutil*, *pyodbc*, *psycopg2*, *psycopg2-binary*, *setproctitle* et *simple-crypt*). Il est exécuté sur un serveur Linux de la distribution CentOS 8 au sein du réseau interne du CHU Sainte-Justine.

Contrairement au prototype qui captait les messages HL7, ce programme puise les données dans des sources déjà structurées, soit par le dossier médical électronique soit par le système des moniteurs de surveillance (Patient Information Center vB.01, Philips, Pays-Bas). Comme ces systèmes ne conservent l'information que pendant une courte période (24 à 48 h), il était essentiel de les stocker de façon permanente. Le programme se charge de déterminer quelles données sont nouvelles, de les collecter, de les normaliser et de les stocker.

Réidentification des données

L'identification des patients dans les données sources n'est pas complètement fiable, car elle dépend d'une intervention manuelle directement sur les moniteurs de surveillance. Dans notre validation interne, ce processus survenait en retard après l'admission du patient ou était erroné pendant un certain temps avant qu'un autre soignant vienne le corriger. Pour corriger cela, nous avons fait en sorte que le programme de collecte des données ré identifie systématiquement toutes les données entrantes à partir du lit et de l'heure de l'information en comparant avec les informations contenues dans le dossier médical. Ce processus a lieu automatiquement à l'insertion et est revalidé chaque 24 h en cas de retard avec les informations du dossier médical. De plus, le numéro identifiant le patient est à l'échelle du séjour aux soins intensifs et non à échelle de l'identité du patient, comme le numéro de dossier. Cette distinction permet de distinguer facilement les séjours répétés ou les réadmissions.

Ajout de sources de données

Moniteur de surveillance et signal brut

Les données numériques des moniteurs sont échantillonnées et enregistrées chaque seconde, pour tous les patients. Chaque horodate, synchronisée par l'horloge du serveur sur le temps universel coordonné, est précise à la milliseconde et inclut le fuseau horaire permettant de garder les données comparables lors des changements d'heures d'été et d'hiver.

Le signal brut est également enregistré et son stockage est plus complexe. La fréquence d'enregistrement dépend de la variable mesurée, allant jusqu'à 500 Hz (soit 500 points par seconde) pour le tracé électrocardiographique. Le signal est enregistré par bloc de 5 secondes contenant les coordonnées en ordonnées (*y*) seulement puisque les coordonnées en abscisse (le temps) sont déduites à partir de l'ordre de la valeur dans la séquence. Dans le système *Patient Information Center* (Philips, Pays-Bas), les données de signal brut sont stockées en format hexadécimal par bloc de 2 octets écrit en petit-boutiste (*little endian*). Ainsi, le décodage de la coordonnée *y* nécessite un échange d'octet (*swap*) de type de BA-DC (à partir d'une séquence ABCD), avant d'être converti du format hexadécimal (base 16) en format numérique standard (base 10). Ce processus (**Figure 4**) est fait en temps réel par le système que j'ai développé. De plus, chaque bloc de 5 secondes de signal doit être mis à l'échelle pour obtenir les unités réelles de mesures (par exemple, les millivolts ou les mmHg).

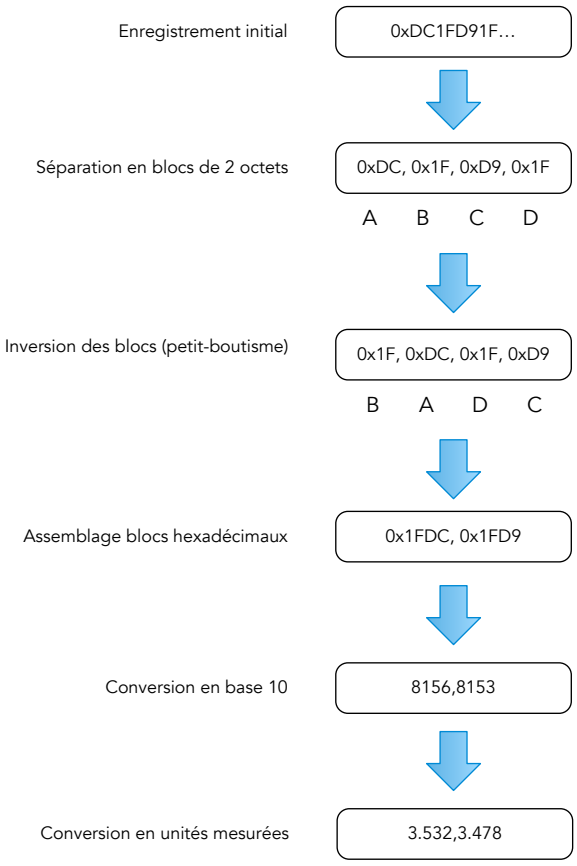


Figure 4. – Processus de décodage du signal brut capté par les moniteurs de surveillance

Alarmes

Les alarmes des moniteurs sont déclenchées selon des seuils manuellement indiqués par les soignants au chevet et dépendent de plusieurs facteurs : âge, pathologie, comorbidités, etc. Dorénavant, ces alarmes et l'action d'inhibition potentielle des soignants sont enregistrées et peuvent être corrélées au signal brut ou aux valeurs numériques au même moment.

Volumes totaux des pousse-seringues

Dans le prototype, seule la valeur instantanée du débit de la pompe ou du pousse-seringue était enregistrée. Bien qu'en théorie cette information permette de recalculer le volume total administré, en pratique les déconnexions des pompes dans le cas d'un déplacement (changement de chambre ou examen complémentaire) empêchaient de le faire. Dorénavant, le volume total est également sauvegardé.

Intégration des données du dossier médical

Un miroir des données du dossier médical a été créé au sein de la base de données à haute résolution temporelle facilitant une analyse intégrée des données démographiques, des résultats de laboratoires, des traitements administrés, des diagnostics, etc.

Enregistrement des données en provenance d'autres unités cliniques

Nous avons vérifié que le programme fonctionnait également pour d'autres unités de soins dans l'hôpital, notamment l'unité des soins intensifs néonataux.

Sécurité et protection des données

Il y a plusieurs niveaux de sécurité mis en place dans ce programme et certains détails ne sont volontairement pas divulgués dans cette thèse. Le premier niveau est que le système n'est accessible que depuis le réseau interne du CHU Sainte-Justine, que l'on soit physiquement dans l'enceinte de l'hôpital ou par l'entremise du réseau virtuel privé (VPN, *virtual private network*).

Cette première couche de sécurité est gérée par le CHU et le ministère de la Santé et des Services sociaux. Le second niveau consiste à n'ouvrir un accès qu'en lecture seule à un schéma ne contenant que des vues sur les données elles-mêmes. Ce système permet aux utilisateurs spécifiquement désignés de n'accéder qu'à un « miroir » des vraies données, avec 2 niveaux de protection empêchant une altération des données originales. Le troisième niveau est un enfermement des données sur le serveur. Les données ne peuvent être consultées et travaillées que directement sur le serveur à l'aide d'une plateforme web accessible uniquement depuis le réseau interne de l'hôpital (RStudio, RStudio Inc, États-Unis). La copie des données vers l'extérieur n'est possible que sur un domaine autorisé par le ministère de la Santé et des Services sociaux du Québec. Toutes les transactions SQL sont également journalisées. Dernièrement, le programme inclut un système automatique, périodique et incrémental de sauvegarde des données dans un répertoire accessible uniquement par les administrateurs du serveur. Le serveur lui-même est également sauvegardé avec des copies locales et distantes par le service informatique de l'hôpital.

Performance

Le programme de collecte automatique est dit « multifils » (*multithreads*), c'est-à-dire qu'il exécute simultanément plusieurs fonctions (programmation parallèle). D'une part, cela permet de mieux tirer profit des processeurs à plusieurs cœurs et d'accélérer l'exécution du programme. D'autre part, un processus lent n'entraine pas de ralentissement pour les autres. Dans la version actuelle, l'ensemble du travail est divisé en 15 tâches indépendantes.

De plus, comme mentionnée précédemment, la principale limite des structures clé-valeurs est la longueur des tables rendant les index de plus en plus lents à maintenir à jour. Pour contrecarrer cette limitation, j'ai utilisé des tableaux partitionnés. Ces derniers sont des sous-tableaux contenant une portion de l'ensemble des données et chacun possède son propre index, facilitant grandement leur mise à jour. La portion a été définie en blocs de 2 semaines de données par tableau partitionné, soit environ 150 Go de données. Le seuil a été déterminé empiriquement. L'autre avantage de ce système est qu'il accélère la recherche d'information, car le système de gestion de la base de données peut paralléliser le travail et consulter plusieurs sous-tableaux simultanément.

Déploiement

Une première version de test a été déployée en aout 2018 et la version définitive en mai 2019. En aout 2021, plus de 45'000 lignes de données sont enregistrées chaque minute, soit plus de 20 milliards de lignes par année et environ 6 To de données par année au rythme actuel. Aujourd'hui, plus de 6500 admissions en soins intensifs pédiatriques sont disponibles en haute résolution temporelle dans cette base de données et plus d'une trentaine de chercheurs l'ont utilisée pour leurs recherches.

Article 1 – Estimated PaO_2: A Continuous and Noninvasive Method to Estimate PaO_2 and Oxygenation Index

Préface

Ce travail a permis de développer un outil d'évaluation continue de l'IRHA applicable sans limites supérieures de SpO_2. Dans cet article, nous démontrons comment améliorer la précision de l'estimation en utilisant des données continues multimodales ainsi qu'une comparaison de plusieurs modèles d'apprentissage automatique et d'optimisation mathématique empirique afin d'estimer la PaO_2 et l'indice d'oxygénation en continu. Dans ce travail réalisé avec 4 autres collaborateurs, j'ai réalisé plus de 90 % du travail, soit la création de la base de données à haute résolution temporelle, la collecte et le formatage des données, le développement et la validation des modèles prédictifs, l'analyse des données et la rédaction du manuscrit. Il a été publié dans la revue *Critical Care Explorations* en 2021. Ce travail m'a valu d'être invité à l'élaboration des prochaines lignes directrices sur le SDRA pédiatrique au sein de la conférence de consensus international PALICC 2.

PREDICTIVE MODELING REPORT

OPEN

Estimated Pao_2: A Continuous and Noninvasive Method to Estimate Pao_2 and Oxygenation Index

Michaël Sauthier, MD, MBI[1–3]
Gaurav Tuli, PhD[4]
Philippe A. Jouvet, MD, PhD, MBA[2]
John S. Brownstein, PhD[4]
Adrienne G. Randolph, MD, MSc[1,3]

BACKGROUND: Pao_2 is the gold standard to assess acute hypoxic respiratory failure, but it is only routinely available by intermittent spot checks, precluding any automatic continuous analysis for bedside tools.

OBJECTIVE: To validate a continuous and noninvasive method to estimate hypoxemia severity for all Spo_2 values.

DERIVATION COHORT: All patients who had an arterial blood gas and simultaneous continuous noninvasive monitoring from 2011 to 2019 at Boston Children's Hospital (Boston, MA) PICU.

VALIDATION COHORT: External cohort at Sainte-Justine Hospital PICU (Montreal, QC, Canada) from 2017 to 2020.

PREDICTION MODEL: We estimated the Pao_2 using three kinds of neural networks and an empirically optimized mathematical model derived from known physiologic equations.

RESULTS: We included 52,879 Pao_2 (3,252 patients) in the derivation dataset and 12,047 Pao_2 (926 patients) in the validation dataset. The mean function on the last minute before the arterial blood gas had the lowest bias (bias −0.1% validation cohort). A difference greater than or equal to 3% between pulse rate and electrical heart rate decreased the intraclass correlation coefficients (0.75 vs 0.44; $p < 0.001$) implying measurement noise. Our estimated Pao_2 equation had the highest intraclass correlation coefficient (0.38; 95% CI, 0.36–0.39; validation cohort) and outperformed neural networks and existing equations. Using the estimated Pao_2 to estimate the oxygenation index showed a significantly better hypoxemia classification (kappa) than oxygenation saturation index for both Spo_2 less than or equal to 97% (0.79 vs 0.60; $p < 0.001$) and Spo_2 greater than 97% (0.58 vs 0.52; $p < 0.001$).

CONCLUSION: The estimated Pao_2 using pulse rate and electrical heart rate Spo_2 validation allows a continuous and noninvasive estimation of the oxygenation index that is valid for Spo_2 less than or equal to 97% and for Spo_2 greater than 97%. Display of continuous analysis of estimated Pao_2 and estimated oxygenation index may provide decision support to assist with hypoxemia diagnosis and oxygen titration in critically ill patients.

KEY WORDS: automatic data processing; clinical decision support systems; critical care; machine learning; oximetry

Acute hypoxic respiratory failure is a common reason of admission in PICUs and is associated with a high mortality and morbidity (1–3). The gold standard to assess hypoxemia severity in pediatric requires an arterial blood gas (ABG) to measure the oxygen partial pressure (Pao_2) and to calculate the oxygenation index (OI = [Fio_2 × 100 × mean airway pressure]/Pao_2) (4). Although Pao_2 is the reference, it is an invasive method available only for selected patients. Furthermore, only intermittent spot checks are routinely possible, precluding accurate continuous evaluation. Oxygen is most a rapidly and widely changing analyte. Clinically, this may lead to periods of undetected hypoxemia (5).

DOI: 10.1097/CCE.0000000000000546

Article 1 - *Estimated PaO_2: A Continuous and Noninvasive Method to Estimate PaO_2 and Oxygenation Index*

BACKGROUND: Arterial oxygen partial pressure (PaO_2) is the gold standard to assess acute hypoxic respiratory failure, but it is only routinely available by intermittent spot checks, precluding any automatic continuous analysis for bedside tools.

OBJECTIVE: To validate a continuous and non-invasive method to estimate hypoxemia severity for all SpO_2 values.

DERIVATION COHORT: All patients who had an arterial blood gas (ABG) and simultaneous continuous non-invasive monitoring from 2011-2019 at Boston Children's Hospital (BCH, United States) pediatric intensive care unit (PICU).

VALIDATION COHORT: External cohort at Sainte-Justine Hospital PICU (SJH, Montreal, Canada) from 2017-2020.

PREDICTION MODEL: We estimated the PaO_2 using three kinds of neural networks and an empirically optimized mathematical model derived from known physiological equations.

RESULTS: We included 52,879 PaO_2 (3,252 patients) in the derivation dataset and 12,047 PaO_2 (926 patients) in the validation dataset. The mean function on the last minute before the ABG had the lowest bias (bias -0.1% validation cohort). A difference ≥ 3% between pulse rate and electrical heart rate decreased the intraclass correlation coefficients (ICC, 0.75 vs 0.44, P<0.001) implying measurement noise. Our $ePaO_2$ equation had the highest ICC (0.38, 95% CI 0.36 - 0.39, validation cohort) and outperformed neural networks and existing equations (Hill 1910, Severinghaus 1979, Brockway 1998 and Gadrey 2019). Using the $ePaO_2$ to estimate the Oxygenation Index (OI) showed a significantly better hypoxemia classification (Kappa) than Oxygenation Saturation Index for both $SpO_2 \leq 97\%$ (0.79 vs 0.60, P<0.001) and $SpO_2 > 97\%$ (0.58 vs 0.52, P<0.001).

CONCLUSION: The $ePaO_2$ allows a continuous and non-invasive estimation of the oxygenation index that is valid for $SpO_2 \leq 97\%$ and for $SpO_2 > 97\%$. Display of continuous analysis of $ePaO_2$ and estimated OI may provide decision support to assist with hypoxemia diagnosis and oxygen titration in critically ill patients.

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

INTRODUCTION

Acute hypoxic respiratory failure is a common reason of admission in pediatric intensive care units (ICU) and is associated with a high mortality and morbidity (1–3). The gold standard to assess hypoxemia severity in pediatric requires an arterial blood gas (ABG) to measure the oxygen partial pressure (PaO_2) and to calculate the oxygenation index [OI = (FiO_2×100×Mean Airway Pressure) / PaO_2] (4). Although PaO_2 is the reference, it is an invasive method available only for selected patients. Furthermore, only intermittent spot checks are routinely possible, precluding accurate continuous evaluation. Oxygen is most rapidly and widely changing analyte. Clinically, this may lead to periods of undetected hypoxemia (5). Pulse oximetry (SpO_2) is the most common continuous surrogate, but the relationship between SpO_2, the actual arterial oxygen saturation (SaO_2) and PaO_2 is complex, especially when SpO_2 is >97% (6, 7). Therefore, pediatric and adult acute respiratory distress syndrome (ARDS) guidelines recommends to use only $SpO_2 \leq 97\%$ to assess hypoxemia severity (4, 8). This is an important limitation for a continuous estimation in the ICU as a majority of patients have SpO_2 above 95% (3, 9–13).

First equations describing oxyhemoglobin dissociation curve had important limitations (14–16). They cannot estimate the PaO_2 when SaO_2 is close to 100% (asymptotic relationship) and are significantly biased for high SaO_2 values (17). Gadrey et al. modified Hill's equation to address that limitation and showed improved performance in hospitalized non-ICU patients (17). Furthermore, all existing equations used a value of P_{50} (PaO_2 value when SaO_2=50%) that was measured on a few healthy adults (15), but this parameter is known to be different in a critically ill population (18). Other equations estimating the PaO_2/FiO_2 ratio from the SpO_2/FiO_2 ratio have been developed (19–23). Although PaO_2 can be mathematically solved, fit was generally poor (17, 19, 24).

Only few studies have used automatic continuously collected streams of data (6). This rich data source is helpful to precisely pair ABGs with other continuous monitoring data. Because no human is involved in the process, they require accurate automatic detection mechanisms to exclude potentially erroneous data (25). Different time windows and aggregation functions have been

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

described in the literature from -60 minutes to the near-exact time (6, 17, 22, 26) but no comparative methods have been published in the literature to our knowledge.

This study aimed to validate a method to continuously estimate the hypoxemia severity, valid for all SpO_2 values, by estimating the PaO_2 using noninvasive monitoring information. Such a method would facilitate early detection and monitoring of hypoxemia.

MATERIAL AND METHODS

We collected two independent datasets at Boston Children's Hospital (BCH, Boston, United States) ICU (excluding the cardiac unit) and at Sainte-Justine Hospital (SJH, Montreal, Canada) pediatric ICU. Both are academic referral centers that treat a broad range of highly specialized and common pediatric critical conditions, and they are approximately at the same altitude (<50 m above sea level). We used the BCH cohort to develop new models and the SJH cohort to validate them independently. We included any patients admitted in these pediatric ICUs who had an ABG with concomitant continuous SpO_2 measures. ABG were excluded if more than 20% of the SpO_2 were missing before the ABG. This study was approved by both BCH (reference IRB-P00021911) and SJH (reference 2018-1587) institutional review boards and waived the need for informed consent. We followed the 2020 standards for prediction models in critical care (27) and the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines for development and validation of predictive models (28).

Data Collection

Data were extracted from secured servers located in each hospital. At BCH, data were retrieved at a 5-second frequency from the T3 database (Etiometry Inc, Boston, United States) from October 2011 to December 2019. At SJH, data were collected from January 2017 to January 2020 and retrieved at 1-second rate from a software developed in-house (29). We collected SpO_2,

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

pulse rate and heart rate from -20 minutes to +10 minutes around the ABG. SJH data were resampled (closest 5-second time) to have a simulated 5-second frequency. In both centers, ABGs were analyzed using Radiometer ABL90 FLEX (Radiometer Medical ApS, Denmark). In both centers, SpO_2 were measured using Masimo (Masimo Corporation, Irvine, United States) probes (RD SET™ Neo and RD SET™ Inf). Body locations of the probe were not available as a continuous data. Once extracted, data were stored and managed on a secured local PostgreSQL 12.1 server and were analyzed using Python 3.7.7 and R 3.6.3 scripts. Neural networks models were built with Keras 2.3.0 on TensorFlow 2.1.0. In the models, we only used SpO_2, heart rate and pulse rate.

Statistical Analyses and Aggregation Function

Categorical data were described as count and percentage and continuous data as median and interquartile range (IQR). Confidence intervals and P values were calculated using 10,000 bootstrap repetitions (30). We used the intraclass correlation coefficient (ICC) as a global metrics for agreement, accuracy and precision (31). We used the two-way mixed effects ICC, also known as ICC(3,1) in the Shrout and Fleiss classification (32). The accuracy was estimated using the bias defined as the mean of the differences (mean biased error or fixed bias) (33). Precision was estimated with the mean absolute error (MAE) and the limits of agreements (95% confidence interval on the differences) (33). Agreement was shown with Bland-Altman plots grouped by range of SpO_2. We also calculated the proportional bias, defined by the slope on the Bland-Altman plot. Discrimination was estimated using the area under the receiving operating characteristic curve (AUROC) for mild ($OI \geq 4$) and severe ($OI \geq 16$) hypoxemia.

We used the direct relationship between SpO_2 and SaO_2 to find the best SpO_2 aggregation. We compared mean and median function and different time windows from an exact time match to a 20-minute period before the PaO_2. We chose the aggregation method that minimized the bias.

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

SpO_2 Quality Assessment

In ICU, heart rate is continuously monitored by an electrocardiogram monitor and pulse oximeters. As both measurements should be identical in most clinical situations, the similarity is a way to validate the quality of the SpO_2 measures as already used by our team (34). We investigated if a higher difference in percentage between the pulse rate and the electrical heart rate is associate with a greater MAE and lower ICC.

Empirical Optimization Model

As Hill's general equation is complex to optimize by traditional mathematical techniques (**Figure 1, equation 1**), we used an empirical method that calculated nearly all the combinations in the plausible ranges for SpO_2 (80% to 100%), P_{50} (20.0 mmHg to 40.0 mmHg), m (0.950 to 0.999) and Hill's number n_H (2.50 to 3.50). Then, we identified the optimal combination that minimized the mean squared error (MSE) in the derivation dataset. This approach, also called exhaustive search, brute-force search or grid search, is often used in machine learning hyperparameter optimization to solve a multidimensional problem (35). However, this optimization method requires a significant amount of time and computing resources.

$$PaO_2 = \left(\frac{(P_{50})^{n_H}}{\frac{1}{SpO_2} - m} \right)^{\frac{1}{n_H}} \quad \text{(Equation 1)}$$

$$_{Hill}PaO_2 = \left(\frac{26^{2.7}}{\frac{1}{SaO_2} - 1} \right)^{\frac{1}{2.7}} \quad \text{(Equation 2 [reference 14])}$$

$$_{Severinghaus}SaO_2 = \left(\left(\frac{28.603^3}{(PaO_2)^3 + (150 \cdot PaO_2)} \right) + 1 \right)^{-1} \quad \text{(Equation 3 [reference 15])}$$

$$_{Gadrey}PaO_2 = \left(\frac{28.603^3}{\frac{1}{SpO_2} - 0.99} \right)^{\frac{1}{3}} \quad \text{(Equation 4 [reference 17])}$$

$$_{Sauthier}PaO_2 = \left(\frac{27.8^{2.8}}{\frac{1}{SpO_2} - 0.99} \right)^{\frac{1}{2.8}} \quad \text{(Equation 5)}$$

Figure 1: Known equations to estimate PaO_2 using SaO_2 or SpO_2.

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

Neural Network Models

In order to train the neural network model, we randomly split at the patient level the derivation cohort (BCH) into a train (85%) and a validation (15%) dataset. Once finalized, models were tested on the external distinct cohort (SJH). We used the last 20 minutes before PaO_2 to build a standardized multidimensional matrix of SpO_2, pulse rate and electrical heart rate (241×3, chronologically ordered). Data were normalized relatively to the minimum and maximum value of the training cohort. Missing values were imputed with the last value available. We compared three common types of networks able to assess serial or repeated measurements: multilayer perceptron (MLP), convolutional neural networks (CNN) and long short-term memory networks (LSTM) (36–38). Architecture, optimizer and hyperparameters were adjusted to minimize the MSE.

Hypoxemia Severity Validation

Following the Pediatric Acute Lung Injury Consensus Conference (PALICC) (4), hypoxemia severity was defined using the OI: mild (values between 4 and 7.9), moderate (8 - 15.9) and severe (OI≥16). If no arterial sample is available, PALICC guidelines recommend using the Oxygen Saturation Index (OSI) where PaO_2 is replaced by SpO_2. OSI have different thresholds (mild 5-7.4, moderate 7.5-12.2, severe≥12.3) extrapolated from equations that estimate the OI (7). Mean airway pressures and FiO_2 were extracted at the exact same time that the ABG was drawn and were the same for all models. We selected the best PaO_2 model to estimate the OI and compared its performance with the OSI on severity classification (Kappa) and hypoxemia (mild and severe) discrimination (AUROC). Results were separated into two SpO_2 categories, below or equal to 97% and above 97%, even if OSI has not been validated in that upper range.

RESULTS

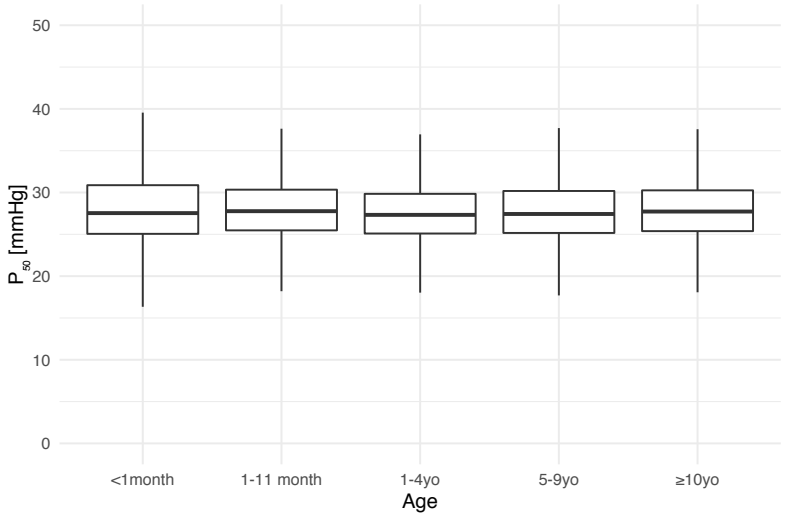
We included 64,926 PaO_2 (4,178 patients) with 46 million concomitant SpO_2 values. Patients and ABG characteristics in each cohort are described in **Table 1**. In Supplemental Digital Content (**SDC 1**), we showed that P_{50} varied from 20 to 40 mmHg but was similar between age groups.

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

Variables	**Boston Children's Hospital Derivation cohort** N (%) or median [IQR]	**Sainte-Justine Hospital Validation cohort** N (%) or median [IQR]
Arterial blood gases, n	52,879	12,047
Patients (total), n	3,252	926
Adult patients (≥18), n	308 (9.5)	15 (1.6)
Arterial blood gases per patient	5 [2 – 17]	6 [2 – 13]
Age [years]	2.4 [0.3 – 11.2]	1.4 [0.1 – 7.7]
Gender, female	23,545 (44.5 %)	5,420 (45 %)
pH	7.4 [7.3 – 7.4]	7.4 [7.4 – 7.5]
PaO_2 [mmHg]	97 [75.4 – 128]	94.6 [72.6 – 129]
SpO_2 [%]	98 [96 – 100]	98.1 [95.5 – 100]
Lactate [mmol/L]	1.2 [0.8 – 1.8]	1.2 [0.8 – 2]
Gases under invasive ventilation	14,378 (27.2 %)	7,547 (62.6 %)
Mean airway pressure [cmH_2O]	11.4 [9 – 14.4]	10.5 [8.5 – 13]
Inspired oxygen fraction	0.4 [0.3 – 0.6]	0.4 [0.3 – 0.6]
Oxygenation index	4.5 [2.8 – 8.1]	4.3 [2.6 – 7.9]

IQR: interquartile range.

Table 1. Patients and arterial blood gases characteristics.

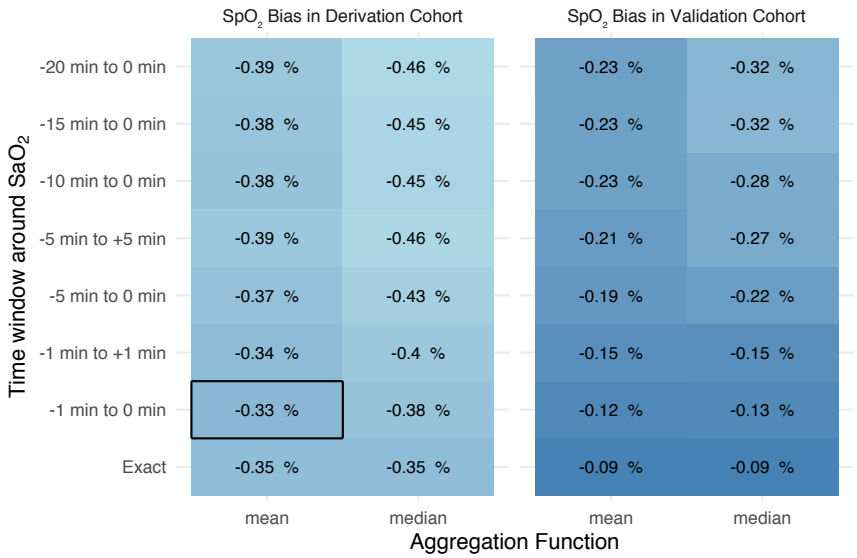


Supplemental Digital Content 1: P_{50} relationship with age

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

Aggregation Methods and Data Quality

We compared different time windows around the SaO_2 measurement (**SDC 2**) from -20 minutes to the exact time of the ABG. Although the differences were small, closer results to the exact time had a lower bias in both derivation and validation cohort. Median function provided very slightly more biased results than the mean in both datasets.

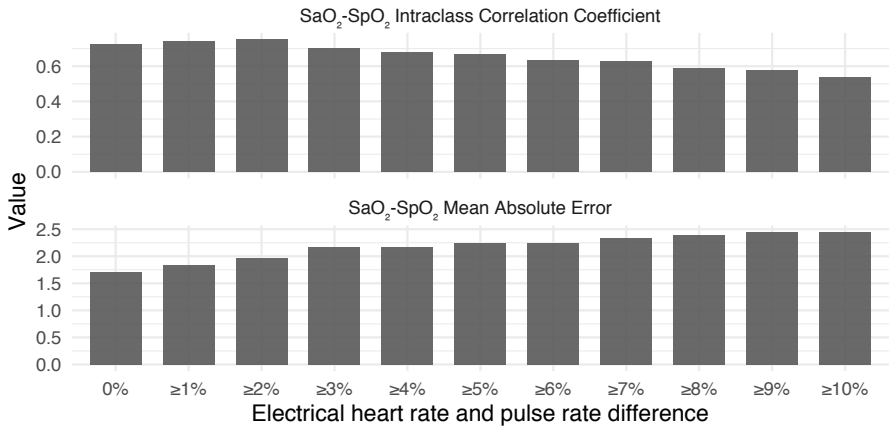


Supplemental Digital Content 2: Different aggregations of SpO_2 and the impact on the bias (mean of differences) compared to SaO_2.

Our results also showed that any difference between the pulse rate read by the pulse oximeter and the electrical heart rate was associated with a lower correlation and higher bias in the derivation cohort (**SDC 3**). The impact was more significant when the difference was ≥3%. We confirmed the results in the validation cohort when the difference was <3% (ICC 0.75, 95% CI 0.74-0.75), ≥3% (ICC 0.44, 95% CI: 0.38-0.49) or unable to read a pulse (ICC 0.49, 95% CI 0.39-0.57). MAE increased from 2.2 (difference<3%) to 3.4 (difference ≥ 3%) and 4.2 (no pulse read). ICC improved from 0.73 (95% CI 0.72-0.73) to 0.75

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

(95% CI 0.74-0.75, P<0.01) after excluding those values, even if these results represented a minority of the measurements (51 777, 866 and 236, respectively in the derivation cohort and 11 271, 582 and 194 in the validation cohort).



Supplemental Digital Content 3: Difference between pulse rate assessed by oximetry and electrical heart rate and the impact on the SpO_2 mean intraclass correlation coefficient and mean absolute error.

Empirical Optimization Model

Over the 1.015 million combinations of parameters using Hill's equation structure, we found that results that minimized the MSE in the derivation cohort was P_{50}=27.8 mmHg, m=0.99 and n_H=2.8 (**Figure 1, equation 5**). Performance in the derivation cohort is presented in **SDC 4**. Results for the validation cohort compared to the other models are shown in **SDC 5** and illustrated in **Figure 2** and **3**. The $_{Sauthier}ePaO_2$ equation outperformed all existing models for ICC (0.38 vs. 0.35 and lower, P=0.004) and slopes (1.01 vs 1.14 and further from one, P<0.001).

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

Models	ICC (95% CI)	Fixed bias (95% CI)	Slope (95% CI)	MAE (95% CI)
Mathematical equations				
Sauthier ℮PaO_2	0.42 (0.41 — 0.44)	5.4 (5.0 — 5.8)	0.98 (0.96 — 0.99)	26.3 (26.0 — 26.7)
Gadrey 2019	0.39 (0.38 — 0.40)	11. 8 (11.4 — 12.2)	1.11 (1.09 — 1.13)	26.1 (25.7 — 26.4)
Brockway 1998	0.30 (0.29 — 0.31)	17.8 (17.3 — 18.2)	1.38 (1.36 — 1.4)	28.1 (27.7 — 28.5)
Severinghaus 1979	0.37 (0.36 — 0.38)	-10.3 (-10.8 — -9.7)	0.34 (0.33 — 0.36)	30.2 (29.7 — 30.6)
Hill 1908	0.35 (0.34 — 0.36)	-17.1 (-17.7 — -16.5)	0.28 (0.27 — 0.29)	34.7 (34.2 — 35.2)
Neural networks				
Multilayer perceptron	0.42 (0.41 — 0.43)	2.3 (1.9 — 2.7)	1.04 (1.03 — 1.06)	27.3 (27.0 — 27.7)
Convolutional neural network	0.40 (0.39 — 0.41)	1.6 (1.2 — 2.0)	1.11 (1.09 — 1.12)	27.7 (27.3 — 28.0)
Long short-term memory	0.38 (0.37 — 0.39)	3.6 (3.2 — 4.0)	1.16 (1.14 — 1.18)	27.4 (27.0 — 27.7)

ICC: Intraclass Correlation Coefficient, CI: Confidence interval, MAE : Mean Absolute Error

Supplemental Digital Content 4: Performance of the models in the derivation cohort (Boston Children's Hospital).

Models	ICC (95% CI)	Fixed bias (95% CI)	Slope (95% CI)	MAE (95% CI)
Mathematical equations				
Sauthier ℮PaO_2	0.38 (0.36 — 0.39)	8.1 (7.1 — 9.1)	1.01 (0.97 — 1.04)	32.5 (31.7 — 33.3)
Gadrey 2019	0.35 (0.33 — 0.36)	14.3 (13.3 — 15.4)	1.14 (1.1 — 1.18)	32.3 (31.4 — 33.2)
Brockway 1998	0.28 (0.27 — 0.29)	20.7 (19.6 — 21.7)	1.36 (1.31 — 1.4)	33.5 (32.6 — 34.5)
Severinghaus 1979	0.20 (0.18 — 0.22)	-26.8 (-29.1 ; -24.5)	0.12 (0.11 — 0.14)	51.8 (49.7 — 54.0)
Hill 1908	0.16 (0.14 — 0.18)	-39.6 (-42.7 ; -36.6)	0.09 (0.08 — 0.1)	62.1 (59.3 — 65.0)
Neural networks				
Multilayer perceptron	0.35 (0.34 — 0.37)	4.5 (3.5 — 5.6)	1.05 (1.01 — 1.09)	33.5 (32.6 — 34.3)
Convolutional neural network	0.33 (0.32 — 0.35)	3.7 (2.7 — 4.8)	1.10 (1.06 — 1.14)	33.7 (32.9 — 34.6)
Long short-term memory	0.32 (0.31 — 0.34)	6.0 (5.0 — 7.0)	1.18 (1.14 — 1.22)	33.3 (32.5 — 34.2)

ICC: Intraclass Correlation Coefficient, CI: Confidence interval, MAE : Mean Absolute Error

Supplemental Digital Content 5: Performance of the models in the validation cohort (Sainte-Justine Hospital).

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

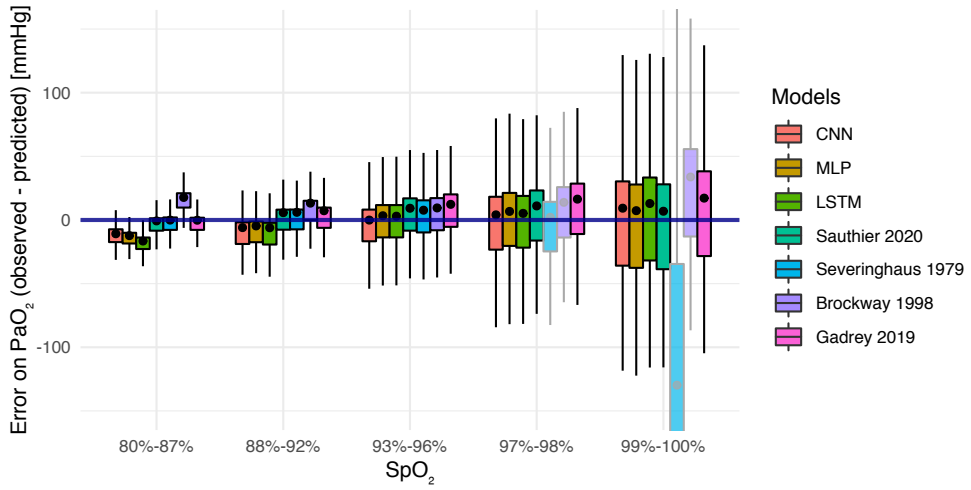


Figure 2: Grouped Bland-Altman plots showing models bias (mean of the errors, black dots) and limits of agreement (boxplot extremes) for different SpO_2 categories. Severinghaus and Brockway were not validated for SpO_2>97%.

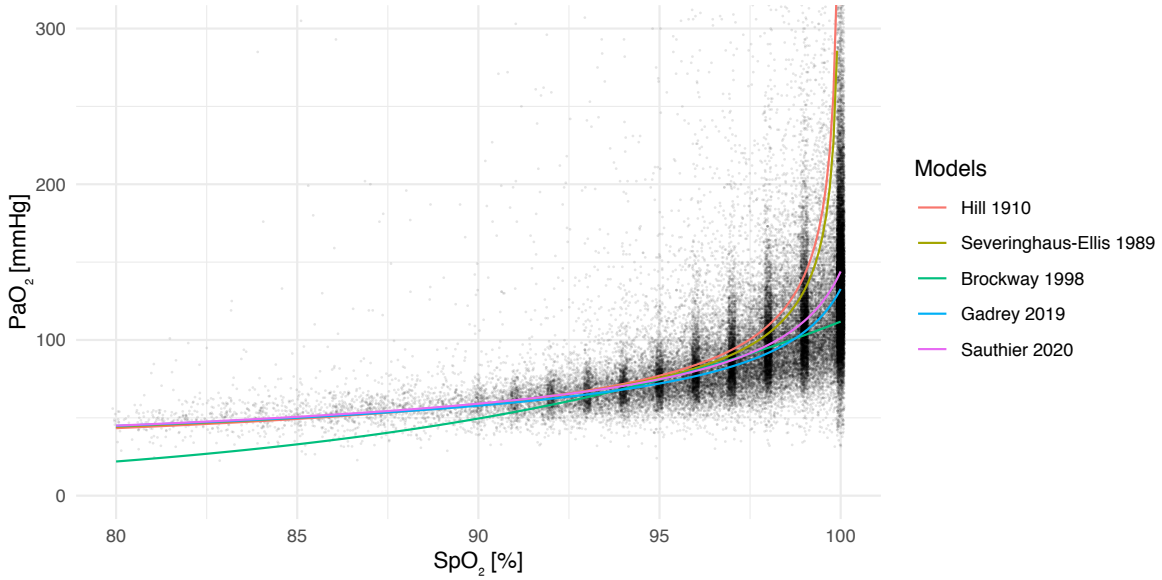


Figure 3: PaO_2 estimation on the derivation dataset using different models.

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

Neural Networks

The three networks had approximately similar performance (**Figure 2** and **SDC 5**), but the LSTM model was slightly less correlated to the observed PaO_2 than the others (ICC 0.32 vs. 0.35, P<0.01). Bias was significantly lower (P<0.01) for the neural networks compared to the mathematical equations (**Figure 2**). The neural network structures are detailed in SDC.

Hypoxemia Severity

Pediatric hypoxemia is determined by the OI or the OSI if no arterial sample is available. We compared the estimated OI (eOI) based on $_{Sauthier}ePaO_2$ to the OSI (**Table 2**). We used the same filtered SpO_2 by pulse rate and heart rate differences for both models. We found that eOI had better hypoxemia severity classification than the OSI for both $SpO_2 \leq 97\%$ (Kappa 0.79 vs 0.60, P<0.001, respectively) and $SpO_2 > 97\%$ (Kappa 0.58 vs 0.52, P<0.001, respectively). We also found that the eOI had a higher global agreement (ICC), lower fixed bias, higher correlation (slope) and better precision (limits of agreements difference and coefficient of determination) than OSI for both SpO_2 ≤97% and >97%. Moreover, the $_{Sauthier}ePaO_2$ had a similar classification performance (Kappa) for $SpO_2 > 97\%$ than OSI has for $SpO_2 \leq 97\%$ (0.58 vs. 0.60, P=0.06). Discrimination for mild hypoxemia or more severe (OI ≥ 4) and for severe hypoxemia (OI ≥ 16) was similar for $ePaO_2$ and OSI. Bland Altman plots and AUROC plots are shown in **Figure 4** and **SDC 6** respectively.

Metric (95% CI)	SpO2 ≤ 97%			SpO2 > 97%		
	OI using OSI	OI using $ePaO_2$	P value	OI using OSI	OI using $ePaO_2$	P value
Weighted Kappa	0.60 (0.58 – 0.62)	0.79 (0.77 – 0.81)	<0.001	0.52 (0.50 – 0.54)	0.58 (0.56 – 0.59)	<0.001
Intraclass Correlation Coefficient	0.91 (0.89 – 0.92)	0.93 (0.92 – 0.94)	0.029	0.62 (0.60 – 0.64)	0.72 (0.69 – 0.75)	<0.001
Mean Absolute Error	3.68 (3.57 – 3.79)	1.90 (1.80 – 2.01)	<0.001	3.58 (3.47 – 3.68)	1.60 (1.54 – 1.67)	<0.001
Coefficient of determination (R^2)	0.84 (0.82 – 0.86)	0.87 (0.85 – 0.89)	0.043	0.51 (0.48 – 0.55)	0.55 (0.51 – 0.59)	0.170
Slope	0.81 (0.78 – 0.84)	1.02 (0.98 – 1.06)	<0.001	0.42 (0.39 – 0.44)	0.93 (0.88 – 0.99)	<0.001
Fixed Bias	2.01 (1.85 – 2.17)	0.15 (0.02 – 0.28)	<0.001	-0.48 (-0.63 – -0.34)	0.13 (0.05 – 0.21)	<0.001
Proportional Bias	-0.12 (-0.16 – -0.09)	0.09 (0.06 – 0.13)	<0.001	-0.61 (-0.67 – -0.55)	0.26 (0.20 – 0.34)	<0.001
Upper and lower limits of agreements difference	0.32 (0.30 – 0.34)	0.25 (0.23 – 0.28)	<0.001	0.29 (0.28 – 0.31)	0.16 (0.15 – 0.18)	<0.001
AUROC for mild hypoxemia (OI≥4)	0.92 (0.91 – 0.93)	0.93 (0.92 – 0.94)	0.122	0.87 (0.86 – 0.88)	0.88 (0.87 – 0.89)	0.052
AUROC for severe hypoxemia (OI≥16)	0.98 (0.97 – 0.98)	0.98 (0.98 – 0.99)	0.055	0.96 (0.95 – 0.97)	0.96 (0.95 – 0.97)	0.986

CI: Confidence interval, OI: Oxygenation Index, OSI: Oxygenation Saturation Index, AUROC: area under the receiving operating characteristic curve

Table 2 : Hypoxemia severity assessment

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

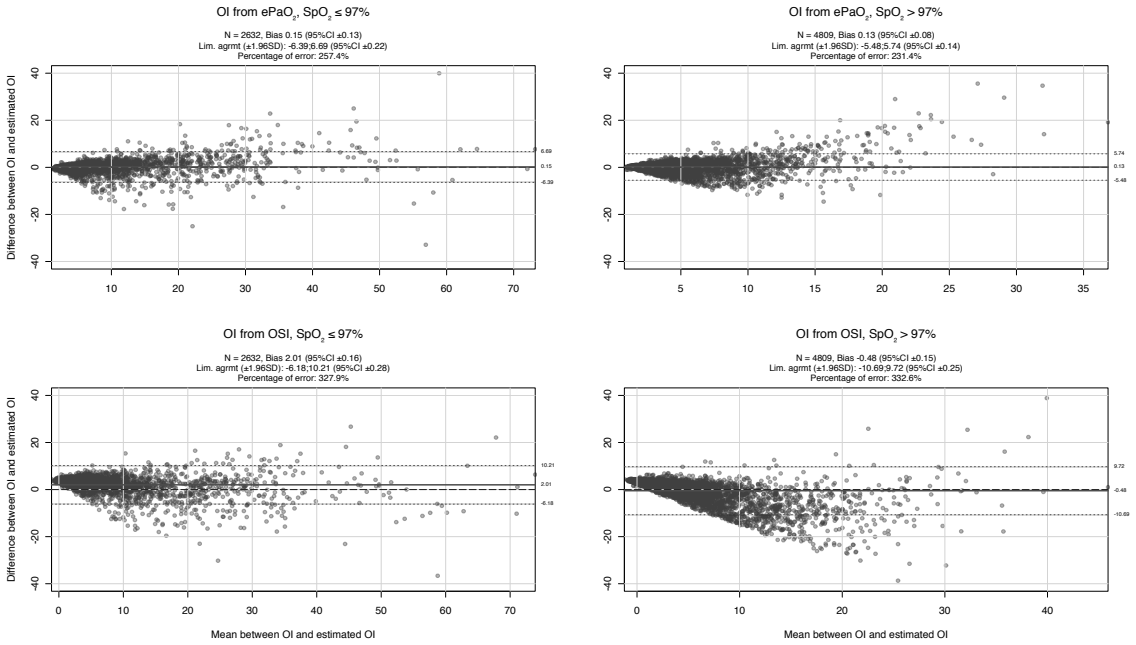
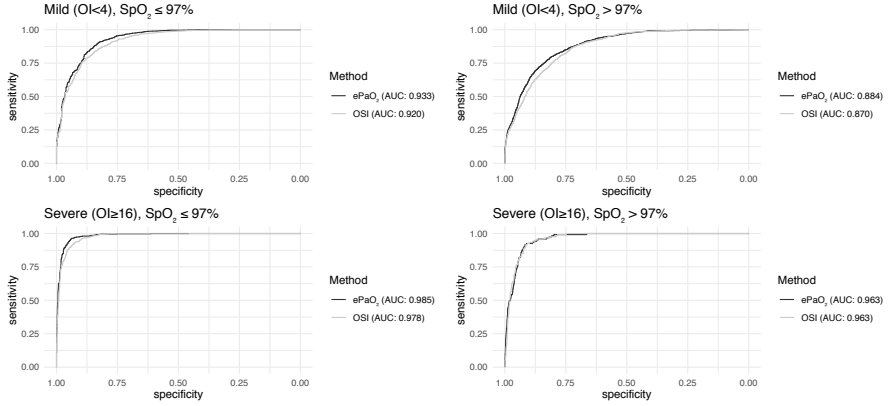


Figure 4: Bland Altman plots for estimated Oxygenation Index (OI) using ɘPaO2 and Oxygenation Saturation Index (OSI).



Supplemental Digital Content 6: ɘPaO2 and Oxygenation Saturation Index (OSI) Discrimination for Mild (OI<4) and Severe (OI≥16) Hypoxemia

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

DISCUSSION

We have developed and validated a predictive model that can automatically and continuously estimate the PaO_2 using only noninvasive data that are commonly measured in ICU (SpO_2, pulse rate and electrical heart rate) across two independent cohorts. A difference between the pulse read by the oximeter and the electrical heart rate was a good indicator of erroneous values, showing lower SaO_2-SpO_2 agreement. SpO_2 is optimally aggregated with a mean function on a one-minute time window. A new mathematical equation ($_{Sauthier}ePaO_2$) outperformed three existing mathematical equations for estimating PaO_2 and application of deep learning algorithms were unable to further improve performance. At SpO_2 values 99% and above, estimation of PaO_2 was still unbiased, but imprecise across all models. When $ePaO_2$ is used to estimate OI, a better hypoxemia estimation is found than with OSI. Although the improvement was modest, those findings are clinically relevant as they will allow building real-time clinical decision support systems that aim to automatically monitor degree of hypoxemia in patients on life support. Moreover, OI and eOI will have the same thresholds, facilitating the interpretation. They will also allow continuous quantitative analysis of hypoxemia, such as duration and severity of each hypoxemia episode. Those results may also decrease blood draws and resource needs in a vulnerable population to anemia.

This study confirmed Gadrey's contribution (17) to Severinghaus' and Hill's equation (**SDC 5**) that changing m=1 to m=0.99 improved most of the bias seen for SpO_2 values above 95%. However, we found no studies in the literature that tried to optimize Severinghaus' P_{50} value and n_H for a pediatric critical care population. In our study, these new parameters improved the agreement (ICC), the bias and the slope, for all the clinically significant range of SpO_2. As we observed no clear association between age and P_{50}, it was unlikely that age would improve the model. However, precision remained low as a single value of SpO_2 can correspond to a broad range of PaO_2 (6). Future models may integrate patients' previous ABGs to improve continuous PaO_2 estimation precision. To improve reading in higher oxygenation values, other solutions such as the Oxygen Reserve Index (39) used SvO_2 changes when SpO_2 is saturated. However, it requires specific proprietary material. To our knowledge, this study is the first to show different ways to retrospectively improve precision using continuous data: selecting a short time window aggregated by a mean function and exclude erroneous values when pulse rate and electrical heart rate are different.

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

In this study, we showed that three types of deep learning models were able to capture information from the continuous non-invasive input data resulting in the estimations with the lowest bias. However, the difference with the $_{Sauthier}ePaO_2$ equation was small (between 2 mmHg and 4 mmHg) and is probably not clinically significant. Nonetheless, considering that a simple and an explicit mathematical equation can perform similarly, or better, neural networks do not seem to be worth the complexity they imposed, especially when ICC is low.

Our study has strengths. To our knowledge, this is the first and largest study to explore the PaO_2-SpO_2 relationship in the critical care field using only automatic continuously collected data. The ability to continuously assess hypoxemia severity is essential to improve early identification. Second, our model (**Figure 1, equation 5**) is simple, explicit and has been validated in a distinct external cohort. Even though most of the cohort are children, the results did include adult patients (8%) and the $_{Sauthier}ePaO_2$ parameters were close to the ones validated on adults, suggesting a good generalizability. Third, this study suggests that SpO_2 > 97% can still provide significant information regarding the hypoxemia severity. The $_{Sauthier}ePaO_2$ had a similar classification performance (Kappa) for SpO_2 > 97% than OSI has for $SpO_2 \leq 97\%$. As precision is still lower for high SpO_2, it still supports the recommendation to minimize FiO_2 for appropriate SpO_2 targets (4).

However, our study has limitations. First, the retrospective nature of the study precludes the validation of the metadata (especially patient identification, timestamp and arterial labeling). However, those ABGs represented real-life situations and both hospitals are committed to the higher standard of care including the validation of research databases (40). Second, measuring arterial oxygenation is a clinical decision and is not necessarily indicated for all patients. Thus, the patients included in this study are expected to be more severe than those who did not get any arterial measurement. However, overcoming this limitation in prospective trial would likely require invasive procedures in patients that are not warranted clinically. The large number of patients and ABG included in this study should mitigate this limitation. Third, we used a broad pediatric ICU population to derive and to validate the models. One would prefer to revalidate the model in a specific ARDS cohort before using it in a clinical decision support system dedicated to this entity.

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

CONCLUSION

In this two-center study with over 4'000 patients and nearly 65'000 arterial blood gases paired with continuous data, we developed and validated a simple mathematical equation and methods to filter SpO_2 streams accurately and automatically in order to estimate the PaO_2 for patients admitted in critical care units using only continuous and noninvasive data. Despite a limited precision for high SpO_2, those results allow a precise OI estimation even for $SpO_2 > 97\%$ and are important for all future bedside clinical decision tools aiming at improving oxygenation titration in real time in critical care.

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

REFERENCE

1. Matthay MA, Zemans RL, Zimmerman GA, et al.: Acute respiratory distress syndrome. *Nat Rev Dis Prim* 2019; 5:18
2. Chomton M, Brossier D, Sauthier M, et al.: Ventilator-associated pneumonia and events in pediatric intensive care: A single center study. *Pediatr Crit Care Med* 2018; 19:1106–1113
3. Sauthier MS, Jouvet PA, Newhams, MM, et al.: Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor* 2020; 2:e0175
4. Khemani RG, Smith LS, Zimmerman JJ, et al.: Pediatric acute respiratory distress syndrome: definition, incidence, and epidemiology: proceedings from the Pediatric Acute Lung Injury Consensus Conference. *Pediatr Crit Care Med* 2015; 16:S23-40
5. Bellani G, Laffey JG, Pham T, et al.: Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. *JAMA* 2016; 315:788–800
6. Eytan D, Mazwi ML, Goodwin AJ, et al.: Revisiting oxygen dissociation curves and bedside measured arterial saturation in critically ill children. *Intensive Care Med* 2019; 45:1832–1834
7. Khemani RG, Thomas NJ, Venkatachalam V, et al.: Comparison of SpO 2 to PaO 2 based markers of lung disease severity for children with acute lung injury. *Crit Care Med* 2012; 40:1309–1316
8. Riviello ED, Kiviri W, Twagirumugabe T, et al.: Hospital incidence and outcomes of the acute respiratory distress syndrome using the Kigali modification of the Berlin definition. *Am J Respir Crit Care Med* 2016; 193:52–59
9. Tannoury J El, Sauthier M, Jouvet P, et al.: Arterial Partial Pressures of Carbon Dioxide Estimation Using Non-Invasive Parameters in Mechanically Ventilated Children. *IEEE Trans Biomed Eng* 2021; 68:161–169
10. Pollard TJ, Johnson AEW, Raffa JD, et al.: The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci data* 2018; 5:180178
11. Mackle D, Beasley R, Bellomo R, et al.: Conservative oxygen therapy during mechanical ventilation in the ICU. *N Engl J Med* 2020; 382:989–998

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

12. Helmerhorst HJF, Arts DL, Schultz MJ, et al.: Metrics of arterial hyperoxia and associated outcomes in critical care. *Crit Care Med* 2017; 45:187–195
13. Slater A, Straney L, Alexander J, et al.: The Effect of Imputation of PaO2/FIO2 From SpO2/FIO2 on the Performance of the Pediatric Index of Mortality 3. *Pediatr Crit Care Med* 2020; 21:520–525
14. Hill AV: The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. *Proc Physiol Soc* 1910; 40:iv–vii
15. Severinghaus JW: Simple, accurate equations for human blood O2 dissociation computations. *J Appl Physiol Respir Environ Exerc Physiol* 1979; 46:599–602
16. Ellis RK: Determination of PO2 from saturation. *J Appl Physiol* 1989; 67:902–902
17. Gadrey SM, Lau CE, Clay R, et al.: Imputation of partial pressures of arterial oxygen using oximetry and its impact on sepsis diagnosis. *Physiol Meas* 2019; 40:115008
18. Myburgh JA, Webb RK, Worthley LIG: The P50 is reduced in critically ill patients. *Intensive Care Med* 1991; 17:355–358
19. Breuer HWM, Groeben H, Breuer J, et al.: Oxygen saturation calculation procedures: A critical analysis of six equations for the determination of oxygen saturation. *Intensive Care Med* 1989; 15:385–389
20. Pandharipande PP, Shintani AK, Hagerman HE, et al.: Derivation and validation of Spo2/Fio2 ratio to impute for Pao2/Fio2 ratio in the respiratory component of the Sequential Organ Failure Assessment score. *Crit Care Med* 2009; 37:1317–1321
21. Rice TW, Wheeler AP, Bernard GR, et al.: Comparison of the SpO2/FIO2 ratio and the PaO 2/FIO2 ratio in patients with acute lung injury or ARDS. *Chest* 2007; 132:410–417
22. Khemani RG, Patel NR, Bart RD, et al.: Comparison of the pulse oximetric saturation/fraction of inspired oxygen ratio and the Pao2/fraction of inspired oxygen ratio in children. *Chest* 2009; 135:662–668
23. Schwede M, Lee RY, Zhuo H, et al.: Clinician Recognition of the Acute Respiratory Distress Syndrome: Risk Factors for Under-Recognition and Trends over Time∗. *Crit Care Med* 2020; 830–837
24. Sanz F, Dean N, Dickerson J, et al.: Accuracy of PaO2/FiO2 calculated from SpO2 for severity assessment in ED

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

patients with pneumonia. *Respirology* 2015; 20:813–818

25. Johnson AE, Ghassemi MM, Nemati S, et al.: Machine Learning and Decision Support in Critical Care. *Proc IEEE Inst Electr Electron Eng* 2016; 104:444–466
26. Khemani RG, Rubin S, Belani S, et al.: Pulse oximetry vs. PaO2 metrics in mechanically ventilated children: Berlin definition of ARDS and mortality risk. *Intensive Care Med* 2015; 41:94–102
27. Leisman DE, Harhay MO, Lederer DJ, et al.: Development and Reporting of Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and Critical Care Journals. *Crit Care Med* 2020; 1–11
28. Collins GS, Reitsma JB, Altman DG, et al.: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015; 162:55
29. Brossier D, Sauthier M, Mathieu A, et al.: Qualitative subjective assessment of a high-resolution database in a paediatric intensive care unit—Elaborating the perpetual patient's ID card. *J Eval Clin Pract* 2019; 1–6
30. Carpenter J, Bithell J: Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med* 2000; 19:1141–64
31. Liljequist D, Elfving B, Roaldsen KS: Intraclass correlation – A discussion and demonstration of basic features. 2019.
32. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86:420–8
33. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1:307–310
34. Ghazal S, Sauthier M, Brossier D, et al.: Using machine learning models to predict oxygen saturation following ventilator support adjustment in critically ill children: a single center pilot study. *bioRxiv* 2019; 14:334896
35. Chicco D: Ten quick tips for machine learning in computational biology. *BioData Min* 2017; 10:35
36. Ismail Fawaz H, Forestier G, Weber J, et al.: Deep learning for time series classification: a review. *Data Min Knowl Discov* 2019; 33:917–963
37. Choi E, Schuetz A, Stewart WF, et al.: Using recurrent neural network models for early detection of heart failure onset. *J Am Med Informatics Assoc* 2017; 24:361–370

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

38. Catling FJR, Wolff AH: Temporal convolutional networks allow early prediction of events in critical care. *J Am Med Inform Assoc* 2020; 27:355–365

39. Szmuk P, Steiner JW, Olomu PN, et al.: Oxygen Reserve Index: A Novel Noninvasive Measure of Oxygen Reserve--A Pilot Study. *Anesthesiology* 2016; 124:779–784

40. Mathieu A, Sauthier M, Jouvet P, et al.: Validation process of a high-resolution database in a paediatric intensive care unit—Describing the perpetual patient's validation. *J Eval Clin Pract* 2021; 27:316–324

Version non définitive. La version définitive de cet article (Sauthier M, Tuli G, Jouvet P A, Brownstein J. S., Randolph, A. G. Estimated Pao 2: A Continuous and Noninvasive Method to Estimate Pao 2 and Oxygenation Index. *Crit Care Explor*, 2021 Oct;*3*(10):e0546) est disponible au https://doi.org/10.1097/cce.0000000000000546

Discussion

Dans cet article, nous montrons une nouvelle façon d'estimer la PaO_2 ($ePaO_2$) à partir de la SpO_2 en utilisant des données continues. Grâce aux données multimodales couramment surveillées aux soins intensifs (pouls, fréquence cardiaque électrique et SpO_2), nous avons pu filtrer automatiquement les valeurs corrompues de la SpO_2 et ainsi améliorer la précision de l'estimation. Nous avons montré que cette méthode est la meilleure estimation disponible dans la littérature pour reconnaitre et catégoriser l'IRHA en pédiatrie. De plus, même si cette estimation s'avère non biaisée, elle reste moins précise dans les hautes valeurs ($SpO_2 \geq 98$ %). Toutefois, son utilisation afin d'estimer l'OI s'avère plus performante que l'OSI, tout en étant valide pour les $SpO_2 > 97$ % à un degré similaire de ce qu'était l'OSI dans la plage de valeurs de $SpO_2 \leq 97$ %. Cette méthode a également l'avantage d'utiliser les mêmes seuils de gravité que la référence (OI), contrairement à l'OSI qui en utilise d'autres. Ce travail permet une mesure continue de la gravité de l'IRHA permettant de quantifier la durée, la fréquence et l'importance de l'IHRA chez les enfants.

En revanche, ces résultats sont limités par plusieurs aspect. Premièrement, la pigmentation de la peau ainsi que le site anatomique de la mesure n'était pas disponible en continu de façon fiable. Chez les adultes, la littérature rapporte des différences pouvant être significatives, soit 1-2 % absolu (variable selon le degré d'hypoxémie) entre les adultes ayant une pigmentation claire par rapport à ceux ayant une pigmentation foncée, conduisant probablement à une sous-estimation de l'hypoxémie chez les personnes ayant une pigmentation plus foncée (49,76,77). Ceci peut impacter la qualité de l'estimation de la PaO_2 et par conséquent les indices de gravité de l'hypoxémie dans cette sous-population. Les études pédiatriques (nourrissons) sont moins claires, certaines rapportant une légère différence (78), d'autres non (79). Toutefois, ces études sont limitées soit par une catégorisation arbitraire souvent binaire (peaux « blanches » ou « noires »), soit par une auto-identification des patients à une catégorie ou l'autre, soit par un petit nombre de patients. Il n'en résulte pas moins que l'ajout de métadonnées objectives aux bases de données à haute résolution temporelle ne pourra qu'améliorer la qualité de la science qui en découle. Cela souligne aussi une fois de plus que des données sources diversifiées et fiables sont capitales au développement de modèles prédictifs performants et cliniquement pertinents (80,81).

De plus, la littérature semble révéler des différences significatives entre les marques et les générations de saturomètres (60–62,82,83). Il est difficile de trancher sur la supériorité d'une marque par rapport à une autre, car les critères de jugements sont variables (biais ou précision de la SpO_2, délais de la mesure, qualité et pertinence des alarmes, exactitude de la mesure du pouls ou tolérance aux mouvements ou autres artéfacts notamment). En pédiatrie, les lignes de conduite utilisant la SpO_2 pour le ratio SF reposent sur une étude multicentrique dont 3 centres utilisaient des appareils de Masimo et 3 autres des Nellcor (65). Puisque les données de cet article proviennent de saturomètres Masimo, la généralisation de ces résultats à d'autres marques nécessite une étape supplémentaire de validation. Par ailleurs, le site anatomique de la mesure de la SpO_2 est également un élément important à l'estimation de la SaO_2. Il existe une légère variation (biais) entre les sites utilisés chez l'enfant (57). Selon ces auteurs, cette différence semble plus importante dans le cas d'hypoxémie plus prononcée ($SpO_2 < 90$ %). En outre, la valeur de la SpO_2 va être différente selon la région pré- ou postductale de la mesure chez le patient porteur d'un canal artériel persistant et d'une hypertension artérielle pulmonaire, ce qui est également vrai pour la mesure artérielle de l'oxygénation.

Cette étude démontre aussi que, même avec des modèles informatiques spécialisés dans le traitement des données continues (réseaux récurrents et convolutifs) et l'accès à une grande quantité de données, il n'a pas été possible d'améliorer significativement la précision de l'estimation de la PaO_2 lorsque la valeur de SpO_2 était élevée. Étant donné les faibles progrès sur ce sujet dans les dernières années, il est possible qu'on soit proche du maximum possible dans l'estimation de la PaO_2 à partir des données de surveillance standard. D'autres solutions ont été développées pour améliorer l'estimation dans les valeurs élevées de SpO_2, notamment l'indice de réserve en oxygène (ORI, *oxygen reserve index*). Ce système est utile pour des PaO_2 entre 100-200 mmHg (84) et a surtout été utilisé au bloc opératoire, afin d'anticiper une désaturation avant que la SpO_2 ne commence à diminuer. Toutefois, cette solution requiert du matériel supplémentaire. Par ailleurs, on pourrait contester la plus-value de l'utilisation de la PaO_2 comme référence pour déterminer la gravité de l'IRHA par rapport à la SpO_2. Toutefois, tant que les lignes de conduite utiliseront la PaO_2 comme référence, il faudra utiliser cette valeur ou son estimation.

Article 2 – Comparison of the Automated Pediatric Logistic Organ Dysfunction-2 versus Manual Pediatric Logistic Organ Dysfunction-2 Score for Critically Ill Children

Préface

L'évaluation de la gravité d'un patient en IRHA implique une vision globale de la défaillance d'organes. Le score *PEdiatric Logistic Organ Dysfunction-2* (PELOD-2) permet d'évaluer de façon quotidienne la gravité de l'atteinte et son résultat permet d'estimer la mortalité attendue (85–87). Toutefois, puisque l'appareil respiratoire est inclus dans le score, il a été suggéré de le retirer du calcul chez les patients en IRHA (PELOD-2 dit « non respiratoire), afin d'éviter tout biais (88). Il est donc potentiellement utile à l'évaluation de l'impact d'un SADC en IRHA sur l'évolution de la défaillance d'organes non respiratoires. Cependant, son calcul manuel est fastidieux et empêche une utilisation généralisée dans un service de soins intensifs. Dans ce travail réalisé en collaboration avec 5 autres chercheurs, j'ai réalisé plus de 80 % du travail, soit la collecte des données (en collaboration), le développement et la validation de l'algorithme, l'analyse des données et la rédaction du manuscrit. J'ai également supervisé une étudiante (FLH) pour ce travail. Il a été publié dans la revue *Pediatric Critical Care Medicine* en 2020.

Article

Comparison of the Automated Pediatric Logistic Organ Dysfunction-2 Versus Manual Pediatric Logistic Organ Dysfunction-2 Score for Critically Ill Children

Michaël Sauthier, MD[1,2]; Florence Landry-Hould[2]; Stéphane Leteurtre, MD, PhD[3]; Atsushi Kawaguchi, MD, PhD[1,2]; Guillaume Emeriaud, MD, PhD[1,2]; Philippe Jouvet, MD, PhD[1,2]

Objectives: The Pediatric Logistic Organ Dysfunction-2 is a validated score that quantifies organ dysfunction severity and requires complex data collection that is time-consuming and subject to errors. We hypothesized that a computer algorithm that automatically collects and calculates the Pediatric Logistic Organ Dysfunction-2 (aPELOD-2) score would be valid, fast and at least as accurate as a manual approach (mPELOD-2).
Design: Retrospective cohort study.
Setting: Single center tertiary medical and surgical pediatric critical care unit (Sainte-Justine Hospital, Montreal, Canada).
Patients: Critically ill children participating in four clinical studies between January 2013 and August 2018, a period during which mPELOD-2 data were manually collected.
Interventions: None.
Measurements and Main Results: The aPELOD-2 was calculated for all consecutive admissions between 2013 and 2018 (n = 5,279) and had a good survival discrimination with an area under the receiver operating characteristic curve of 0.84 (95% CI, 0.81–0.88). We also collected data from four single-center studies in which mPELOD-2 was calculated (n = 796, 57% medical, 43% surgical) and compared these measurements to those of the aPELOD-2. For those patients, median age was 15 months (interquartile range, 3–73 mo), median ICU stay was 5 days (interquartile range, 3–9 d), mortality was 3.9% (n = 28). The intraclass correlation coefficient between mPELOD-2 and aPELOD-2 was 0.75 (95% CI, 0.73–0.77). The Bland-Altman showed a bias of 1.9 (95% CI, 1.7–2) and limits of agreement of −3.1 (95% CI, −3.4 to −2.8) to 6.8 (95% CI, 6.5–7.2). The highest agreement (Cohen's Kappa) of the Pediatric Logistic Organ Dysfunction-2 components was noted for lactate level (0.88), invasive ventilation (0.86), and creatinine level (0.82) and the lowest for the Glasgow Coma Scale (0.52). The proportion of patients with multiple organ dysfunction syndrome was higher for aPELOD-2 (78%) than mPELOD-2 (72%; p = 0.002). The aPELOD-2 had a better survival discrimination (area under the receiver operating characteristic curve, 0.81; 95% CI, 0.72–0.90) over mPELOD-2 (area under the receiver operating characteristic curve, 0.70; 95% CI, 0.59–0.82; p = 0.01).
Conclusions: We successfully created a freely available automatic algorithm to calculate the Pediatric Logistic Organ Dysfunction-2 score that is less labor intensive and has better survival discrimination than the manual calculation. Use of an automated system could greatly facilitate integration of the Pediatric Logistic Organ Dysfunction-2 score at the bedside and within clinical decision support systems. (*Pediatr Crit Care Med* 2020; XX:00–00)
Key Words: automatic data processing; children; clinical decision support systems; critical care; hospital mortality; organ dysfunction scores

[1]Pediatric Intensive Care Unit, Department of Pediatrics, Sainte-Justine Hospital, Montreal, QC, Canada.
[2]Department of Pediatrics, Université de Montréal, Montreal, QC, Canada.
[3]Univ. Lille, CHU Lille, EA 2694 – Santé Publique: épidémiologie et qualité des soins, Service de réanimation pédiatrique, F-59000 Lille, France.

DOI: 10.1097/PCC.0000000000002235

Organ dysfunction assessment is central in critical care medicine (1). Even though mortality has substantially decreased in PICUs (2, 3) and is not the only outcome of interest, it is still the reference to build PICU severity scores (4–7). Although these scores are reliable and accurately reflect severity of illness, they are generally not used at the bedside because data collection is time-consuming and human error is very likely given the numerous elements of information required. These scores are frequently used as surrogate outcome measures in randomized clinical trials and are necessary to compare groups of patients. The Pediatric Logistic Organ Dysfunction-2 (PELOD-2) is a well-validated score based on 10 variables corresponding to five organ systems that is able to measure the severity of organ dysfunction (4, 8–11). Because all the required variables are stored in electronic medical records (EMRs), automation of data collection

Article 2 - *Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children*

OBJECTIVE: The Pediatric Logistic Organ Dysfunction-2 (PELOD-2) is a validated score that quantifies organ dysfunction severity and requires complex data collection that is time-consuming and subject to errors. We hypothesized that a computer algorithm that automatically collects and calculates the PELOD-2 (aPELOD-2) score would be valid, fast and at least as accurate and possibly more accurate than a manual approach (mPELOD-2).

DESIGN: Retrospective cohort study

SETTING: Single center tertiary medical and surgical pediatric critical care unit (Sainte-Justine Hospital, Montreal, Canada).

PATIENTS: Critically ill children participating in four clinical studies between January 2013 to August 2018, a period during which mPELOD-2 data was manually collected.

INTERVENTIONS: None.

MEASUREMENTS AND MAIN RESULTS: The aPELOD-2 was calculated for all consecutive admissions between 2013 and 2018 (n=5279) and had a good survival discrimination with an area under the ROC curve (AUROC) of 0.84 (95% CI: 0.81–0.88). We also collected data from 4 single center studies in which mPELOD-2 was calculated (n=796, 57% medical, 43% surgical and compared these measurements to those of the aPELOD-2). For those patients, median age was 15 months (IQR 3-73 months), median ICU stay was 5 days (IQR 3-9 days), mortality was 3.9% (n=28). The intraclass correlation coefficient between m- and aPELOD-2 was 0.75, 95% CI 0.73–0.77. The Bland-Altman showed a bias of 1.9 (95% CI 1.7–2) and limits of agreement of -3.1 (95% CI -3.4 – -2.8) to 6.8 (95% CI 6.5–7.2). The highest agreement (Cohen's Kappa) of the PELOD-2 components was noted for lactate level (0.88), invasive ventilation (0.86) and creatinine level (0.82) and the lowest for the Glasgow Coma Scale (0.52). The proportion of patients with multiple organ dysfunction syndrome was higher for aPELOD-2 (78%) than mPELOD-2 (72%, p=0.002). The aPELOD-2 had a better survival discrimination (AUROC: 0.81, 95% CI 0.72–0.90) over mPELOD-2 (AUROC: 0.70, 95% CI 0.59–0.82, p=0.01).

CONCLUSION: We successfully created a freely available automatic algorithm to calculate the PELOD-2 score that is less labor intensive and has better survival discrimination than the manual calculation. Use of an automated system could greatly facilitate integration of the PELOD-2 score at the bedside and within clinical decision support systems.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

INTRODUCTION

Organ dysfunction assessment is central in critical care medicine (1). Even though mortality has substantially decreased in pediatric intensive care units (PICUs) (2, 3) and is not the only outcome of interest, it is still the reference to build PICU severity scores (4–7). Although these scores are reliable and accurately reflect severity of illness, they are generally not used at the bedside because data collection is time-consuming and human error is very likely given the numerous elements of information required. These scores are frequently used as surrogate outcome measures in randomized clinical trials and are necessary to compare groups of patients. The Pediatric Logistic Organ Dysfunction-2 (PELOD-2) is a well-validated score based on ten variables corresponding to five organ systems that is able to measure the severity of organ dysfunction (4, 8–11). Because all the required variables are stored in electronic medical records (EMR), automation of data collection would save time, provide perpetually updated information on patient clinical course and perhaps allow for estimation of therapeutic response. A few adult scores, specifically the SOFA (12–14) and APACHE scores (15), have been automated with a performance that is comparable to that of manually collected scores. Even if EMRs have been used to calculate PELOD-2 in clinical studies (7, 10), no studies have compared manual calculation of the PELOD-2 to an automatic algorithm. The primary aim of this study was to validate an algorithm able to automatically calculate the PELOD-2 (aPELOD-2) based on the survival discrimination and the proportion of multiple organ dysfunction syndromes (MODS). We hypothesized that aPELOD-2 has a survival discrimination that is similar to other PELOD-2 external validation studies. The secondary aim was to compare the pragmatic performance of the aPELOD-2 to a manually calculated PELOD-2 (mPELOD2) measured in several research studies. Our hypothesis was that the performance of the aPELOD-2 is good and at least equivalent to that of the mPELOD-2.

MATERIALS AND METHODS

For the primary aim, we included data for all consecutive patients admitted to the PICU of Sainte-Justine University Hospital (Montreal, Canada) between January 8, 2013, and August 3, 2018. For the secondary aim, we included all patients that had a mPELOD-2 calculated for clinical studies undertaken at the PICU of Sainte-Justine University Hospital during the same

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

period. For both the primary and secondary aims, we excluded patients aged 18 years or older on admission. We followed the TRIPODS validation guidelines (16). The institutional review board approved this retrospective cohort study (reference number 2018-1587) and waived the need for individual consent.

Manually collected PELOD-2 scores came from four clinical studies: two were prospective studies (study one and three: ClinicalTrials.gov number NCT02613377 and NCT01977547, respectively) and two were retrospective studies with data collected by medical students (study two and four: manuscripts currently in preparation). The four studies comprised a broad good diversity of PICU patients: transfused patients, patients with respiratory failure, patients having undergone surgery for congenital heart disease and patients with delirium.

Data Collection

Both mPELOD-2 and aPELOD-2 used the same EMR (IntelliSpace Critical Care and Anesthesia, Philips, USA, version F.01) as the data source. All PELOD-2 related fields were either directly recorded in the EMR (e.g. laboratory values and respiratory data) or typed in with an error checking mechanism that prevented physiologically incompatible values from being entered. Calculation of the Glasgow Coma Scale (GCS) was automatic and used drop-down menus to measure each function. The mPELOD-2 was collected by trained personnel that included medical students and research staff. Standard training for research clerks consisted of a cross validation of five to ten subjects before they were allowed to collect data independently; all research staff were involved in multiple research studies requiring score calculations. Medical students had no experience with ICU scoring systems before they began data collection; their training consisted of basic education on the PELOD-2 score (as compared to other scores) and included hands on data acquisition while supervised to ensure accuracy of data collection and full understanding of the data elements required.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

Algorithm Description

The elements required to compute aPELOD-2 were collected using successive Structured Query Language (SQL) queries (**Figure 1**). The algorithm is freely available at https://github.com/sauthiem/aPELOD2 under the open-source GNU AGPL v3.0 license. The first step was to identify where the 10 variables required to compute the aPELOD-2 were stored in the database and import them into a temporary table structured on an entity-attribute-value (EAV) model that is robust to synonyms and can easily handle heterogenous data (17, 18). To calculate the PaO_2/FiO_2 ratios, we extracted PaO_2 values from the EAV table, and looked for the last FiO_2 value available in the 60 minutes preceding PaO_2 measurement. The PaO_2/FiO_2 ratio was then inserted as a new element in the EAV table. The 60-minutes maximum time lapse was based on local practice to calculate the PaO_2/FiO_2 ratio for the mPELOD-2 (step 2). For each data row, the day number was calculated on 24-hour intervals from the time of admission beginning with day one (step 3). Then, the algorithm pivoted the EAV table into a column-based structure, i.e., one column per variable (step 4). During the pivoting process, data were agglomerated with the most abnormal value per patient and per day number. Finally, the most abnormal value was converted into points following the PELOD-2 pointing system and a summation was done to calculate the aPELOD-2 (step 5). As indicated in the original PELOD-2 methodology (4), missing values were considered normal (no point). Because studies 1, 3 and 4 used the calendar day to calculate the day number (from midnight to midnight), we adjusted the calculation of the aPELOD-2 to follow the same collection method as the mPELOD-2.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

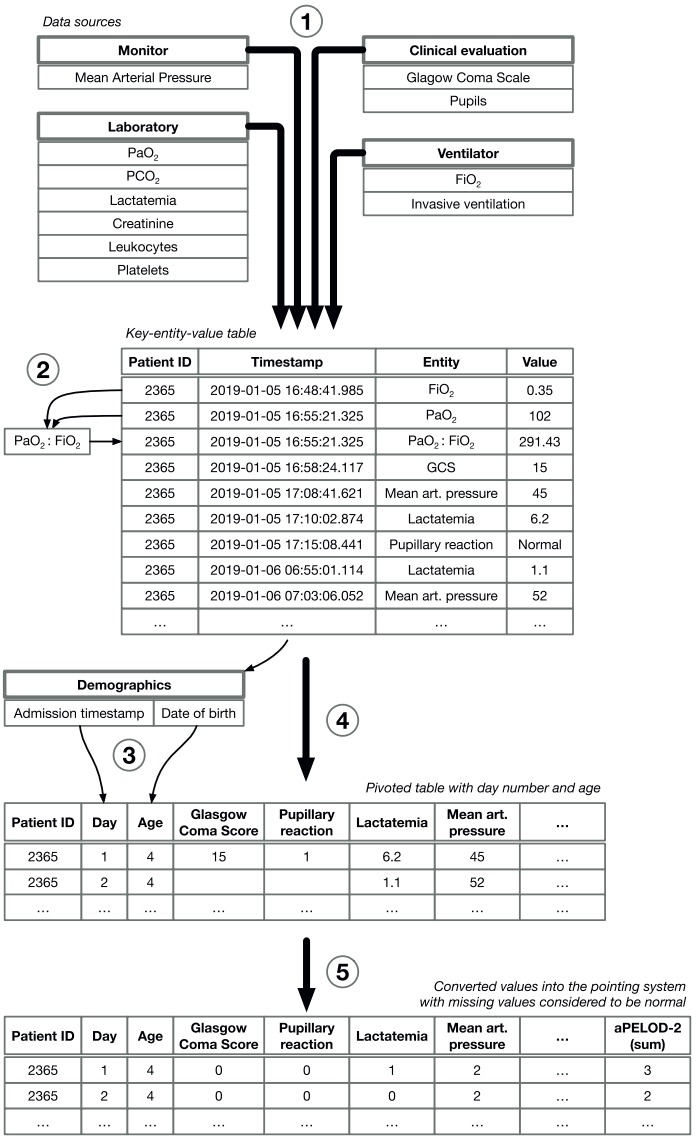


Figure 1: aPELOD-2 algorithm data flow schema. Step 1: data are collected from different sources and centralized into a key-entity-value structure. Step 2: PaO_2/FiO_2 calculation. Step 3: Age and day number calculation for each collected element. Step 4: Pivoting the keys into columns with selection of the most abnormal value per patient and per day. Step 5: The most abnormal value is converted into points and all categories are summed.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

Statistical Analysis

We described the patient population using median and interquartile range (IQR) for continuous variables (age, ICU days, length of ventilation) and count with percentages for categorical variables and mortality. Statistical analysis was conducted in R 3.5.2 with the pROC package (19). We estimated survival discrimination with the area under the receiver operating characteristic curve (AUROC). The 95% confidence interval (95% CI) and *p*-value were calculated using the DeLong method (20). We compared the proportion of MODS present at admission; MODS was defined as the presence of two or more organs with one point or more (1).

We estimated the correlation between aPELOD-2 and mPELOD-2 using intraclass correlation coefficients (ICC) with 95% CIs. We also calculated ICC among mPELOD-2 scores when a patient was evaluated more than once. Because the overall correlation between aPELOD-2 and mPELOD-2 involved multiple "judges" (clinical studies in our case) and different evaluations (the mPELOD-2), we used a one-way random effect model (ICC1,1) (21, 22). In all other cases, only two evaluations were compared (aPELOD-2 and mPELOD-2 or two mPELOD-2 studies); these methods were constant throughout all evaluations and a two-way random-effect model (ICC2,1) was used in those cases. The level of clinical significance of ICC was considered fair if between 0.4 and 0.59, good if between 0.6 and 0.74 and excellent if between 0.75 and 1 (23). We calculated inter-rater agreement for the different components of the PELOD-2 (categorical variables) with a linearly weighted Cohen's Kappa coefficient (24). Because Cohen's Kappa may not be reliable for rare observations or even impossible to calculate if agreement is perfect in a single category, we also reported overall and specific agreements for each component (25, 26). To illustrate specific agreements, we plotted the confusion matrix showing agreement proportion for each PELOD-2 variable between aPELOD-2 and mPELOD-2 (supplemental data, Supplemental Digital Content 2). Except in the case of a very low prevalence, Kappa agreement was interpreted as moderate if between 0.41 and 0.60, substantial if between 0.61 and 0.80 and almost perfect if between 0.81 and 1 (27). We also measured agreement between aPELOD-2 and mPELOD-2 using a Bland-Altman plot (28). Accuracy was estimated with bias (mean of the differences with 95% CIs) and precision was assessed with limits of agreement (± 1.96 × standard deviations of the differences with 95% CIs) and percentage error (29).

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

We compared aPELOD-2 and mPELOD-2 performance based on survival discrimination (AUROC). If more than one mPELOD-2 was collected for the same patient (i.e., the patient was included in two clinical studies), the average mPELOD-2 was used for comparison to the aPELOD-2. We compared aPELOD-2 and mPELOD-2 MODS estimation with a McNemar test. Statistical significance was defined as a p value < 0.05.

RESULTS

We included data from 5279 patients admitted to PICU between January 8, 2013, and August 3, 2018 (**Table 1**). A total of 796 admission day mPELOD-2 calculations were collected in 725 children admitted between May 2013 and June 2018 who had been included in four different clinical studies. Median age was 15 months (IQR 3 – 73 months), female proportion was 46% and overall mortality was 3.9%. The most frequent reasons for admission was elective surgery requiring postoperative care in PICU (40%), admission from the emergency department (28%) and admission from inpatient wards (22%).

Variables	Validation on a PICU database	aPELOD-2 comparison to PELOD 2 scores manually collected in clinical trials				
		All studies	Study 1	Study 2	Study 3	Study 4
Patients – n (%)	5279 (100)	725 (100)	387 (49)[b]	171 (21)[b]	157 (20)[b]	81 (10)[b]
Mortality – n (%)	155 (2.9)	28[a] (3.9)	27 (7.0)	-	2 (1.3)	-
Females – n (%)	2293 (43)	331 (46)	195 (50)	68 (40)	69 (44)	31 (38)
Age (months) – median (IQR)	30 (6-111)	15 (3-73)	19 (5-93)	26 (5-103)	12 (2-41)	0 (0-1)
ICU days – median (IQR)	2 (1-5)	5 (3-9)	6 (4-12)	4 (3-6)	5 (3-7)	7 (4-11)
Invasive ventilation days – median (IQR)	0 (0-1)	1 (0-4)	1 (0-5)	0 (0-1)	1 (1-4)	3 (1-7)
Admission origin:						
Planned surgery – n (%)	1477 (28)	293 (40)	116 (30)	46 (27)	117 (75)	69 (85)
Emergency department – n (%)	2066 (39)	204 (28)	118 (30)	70 (41)	21 (13)	-
Inpatient wards – n (%)	1230 (23)	170 (24)	114 (29)	38 (22)	18 (11)	9 (11)
Other hospitals – n (%)	261 (5)	28 (4)	22 (6)	7 (4)	1 (1)	-
Unplanned surgery – n (%)	206 (4)	23 (3)	13 (3)	7 (4)	-	3 (4)
Outpatient clinic – n (%)	39 (1)	7 (1)	4 (1)	3 (2)	-	-

IQR: Interquartile range, PICU: pediatric intensive care unit.
[a] One deceased patient was enrolled in study one and three
[b] Proportion on the sum of the four studies (796 patients)

Table 1: Demographical data.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

aPELOD-2 Validation

The aPELOD-2 AUROC calculated on all consecutive admissions to PICU during the study period (n = 5279 consecutive encounters, 2.9% mortality) was 0.84 (95% CI: 0.81 – 0.88) (**Figure 2**). The proportion of MODS at admission was 62% (n=3250).

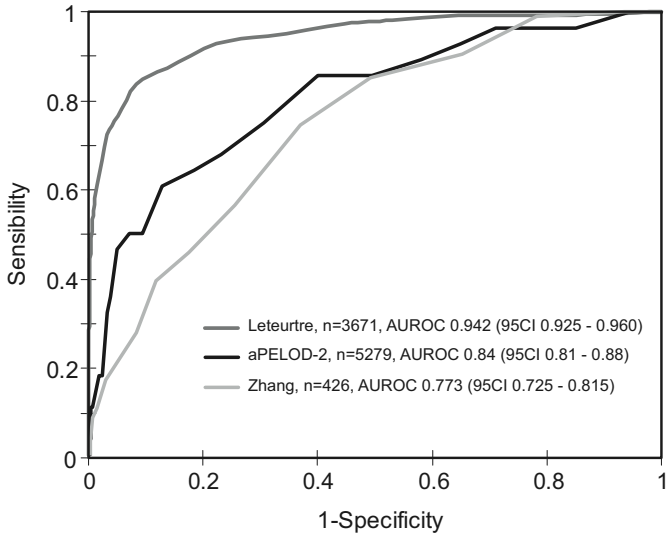


Figure 2: Comparison of the area under the ROC (AUROC) between the original PELOD-2 study (4), the aPELOD-2 and another external validation study by Zhang et al. (11) with 95% confidence interval (95CI).

aPELOD-2 Comparison to mPELOD-2 from Four Clinical Trials

The median value for mPELOD-2 was 5 (IQR 3 – 7) and for aPELOD-2 was 7 (IQR 4 – 9.5) with a Pearson R^2 correlation coefficient of 0.68 (95% CI: 0.64 – 0.72, $p < 0.001$). Bland-Altman analysis (**Figure 3**) showed a bias of +1.9 (95% CI: 1.7 – 2) for the aPELOD-2 over the mPELOD-2 The limits of agreements were calculated from -3.1 to 6.8 (95% CI: -3.4 – -2.8 and 6.5 – 7.2, respectively). The percentage error was 180%. Bland-Altman among pairs of manually calculated PELOD-2 (studies 1-3 and 3-4 with 29 and 25 patients respectively) revealed a bias of 0.1 (95% CI: -0.4 – 0.6) and -0.4 (95% CI: -

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

1.1 – 0.3), limits of agreements ±2.6 and ±3.3 and percentage error 144% and 99%, respectively. The ICC between aPELOD-2 and mPELOD-2 (**Table 2**) was 0.75 (95% CI: 0.73 – 77) with variability among studies: ICC values were 0.75, 0.85, 0.62 and 0.20 for studies 1, 2, 3 and 4, respectively. The ICC among pairs of mPELOD-2 varied from 0.58 to 0.92.

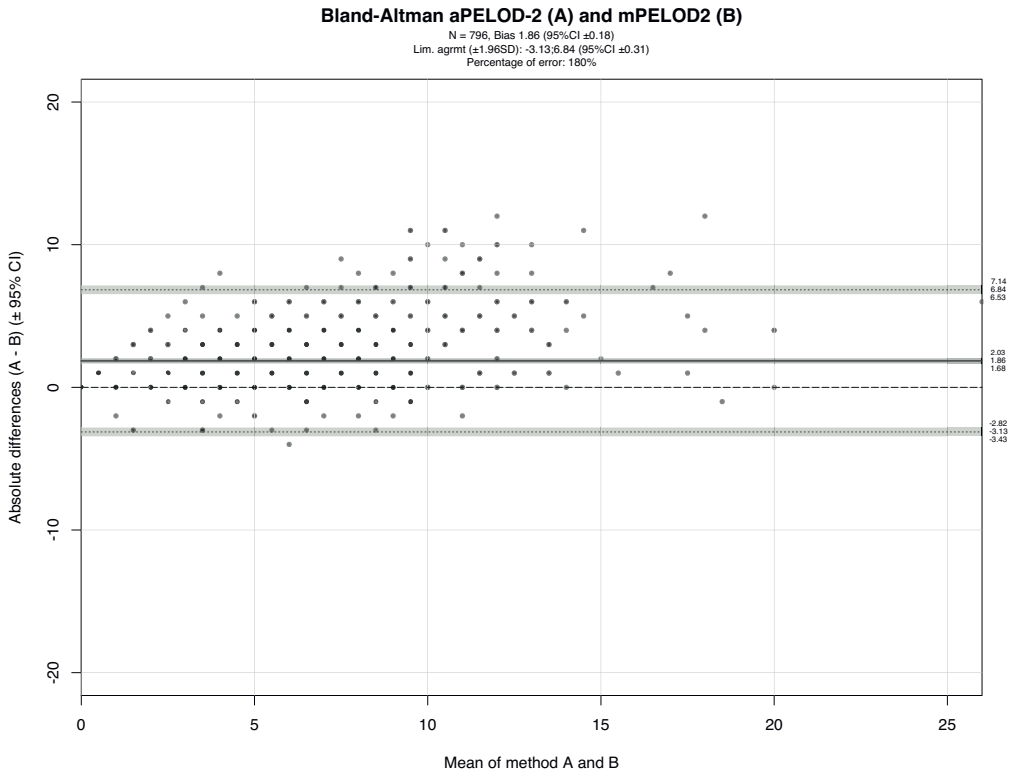


Figure 3: Bland-Altman plot between the aPELOD-2 (method A) and m PELOD-2 (method B) with bias, limits of agreement and 95% confidence interval.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

Compared Groups	Number of compared scores	ICC	95% CI
aPELOD-2 vs mPELOD2	796	0.75	0.73 - 0.77
aPELOD-2 vs study 1	387	0.75	0.41 - 0.87
aPELOD-2 vs study 2	171	0.85	0.78 - 0.90
aPELOD-2 vs study 3	157	0.62	0.13 - 0.86
aPELOD-2 vs study 4	80	0.20	0 - 0.46
mPELOD2 study 1 vs 3	29	0.92	0.85 - 0.96
mPELOD2 study 3 vs 4	25	0.71	0.46 - 0.86

ICC: Intraclass coefficients correlations. CI: confidence intervals.

Table 2: Intraclass coefficients correlations

The weighted Cohen's Kappa coefficient was calculated for each PELOD-2 component for studies 2 and 3 (data were unavailable for studies 1 and 4) (**Table 3**). The highest coefficients were observed for lactate level (0.88), use of invasive ventilation (0.86) and creatinine level (0.82) while the lowest coefficients were noted for platelet count (0.64), leukocyte count (0.52) and the GCS (0.52). Because of the low prevalence of certain classes, there was a discrepancy between the overall agreement and the Kappa for leukocytes (agreement 98%, Kappa 0.53) and the PaO_2:FiO_2 ratio (agreement 95%, Kappa 0). The Kappa was impossible to calculate for pupillary reaction, because there was a complete agreement on a single class (i.e., normal results only). Specific agreements (supplemental data, **Supplemental Digital Content 1**) were higher for near-normal results and decreased for more abnormal values. The confusion matrix plot (supplemental data, **Supplemental Digital Content 2**) showed a trend for the aPELOD-2 to overrate the severity of the organ dysfunction as compared to the mPELOD-2. Less values were reported as missing by the aPELOD-2 than the mPELOD-2 (supplemental data, **Supplemental Digital Content 3**).

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

PELOD-2 Component	Study	Overall agreement [%]	Cohen's kappa	95% CI
Glasgow Coma Score	2	96	0.84	0.72 - 0.95
	3	77	0.06	-0.02 - 0.14
	2+3	87	0.52	0.39 - 0.64
Pupillary reaction	2	100	– [a]	
	3	100	– [a]	
	2+3	100	– [a]	
Lactatemia	2	98	0.88	0.76 - 1
	3	96	0.88	0.79 - 0.98
	2+3	97	0.88	0.81 - 0.95
Mean arterial pressure	2	84	0.77	0.69 - 0.85
	3	73	0.59	0.49 - 0.7
	2+3	78	0.68	0.62 - 0.75
Creatinine	2	88	0.76	0.67 - 0.86
	3	94	0.88	0.81 - 0.96
	2+3	91	0.82	0.76 - 0.88
Invasive ventilation	2	94	0.86	0.78 - 0.94
	3	93	0.78	0.65 - 0.9
	2+3	93	0.86	0.81 - 0.92
PaO_2 : FiO_2	2	99	0 [b]	
	3	90	-0.02	-0.05 - 0.01
	2+3	95	-0.02	-0.03 - 0
PCO_2	2	92	0.75	0.61 - 0.88
	3	87	0.67	0.55 - 0.79
	2+3	90	0.71	0.62 - 0.8
Leukocytes	2	99	0.5	-0.11 - 1
	3	97	0.53	0.17 - 0.89
	2+3	98	0.52	0.22 - 0.83
Platelets	2	92	0.75	0.64 - 0.86
	3	74	0.53	0.42 - 0.65
	2+3	84	0.64	0.56 - 0.72

CI: confidence interval
[a] Complete agreement with a null variance
[b] Null variance in one group

Table 3: Cohen's Kappa and overall agreement on the ten components of the PELOD-2 score.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

Article 2 - *Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children*

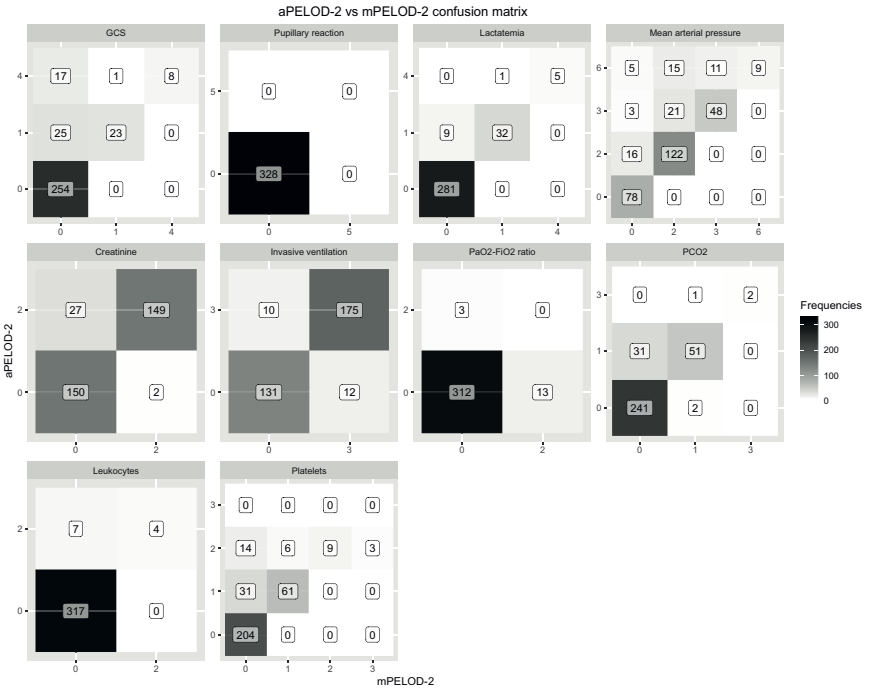
	Study	Overall agreement [%]	Specific agreement per point categories [%]				Patients number				Cohen's kappa (95% CI)
			1st	2nd	3rd	4th	1st	2nd	3rd	4th	
Glasgow Coma Score	2	96	98	95	73		276	44	22		0.84 (0.72-0.95)
	3	77	87	14	0		274	28	12		0.06 (-0.02-0.14)
	2+3	87	92	64	47		550	72	34		0.52 (0.39-0.64)
Pupillary reaction	2	100	100				342				– [a]
	3	100	100				314				– [a]
	2+3	100	100				656				– [a]
Lactate	2	98	99	88	100		308	32	2		0.88 (0.76-1)
	3	96	98	86	89		263	42	9		0.88 (0.79-0.98)
	2+3	97	98	86	91		571	74	11		0.88 (0.81-0.95)
Mean arterial pressure	2	84	91	87	77	42	97	156	70	19	0.77 (0.69-0.85)
	3	73	82	77	69	33	83	140	61	30	0.59 (0.49-0.7)
	2+3	78	87	82	73	37	180	296	131	49	0.68 (0.62-0.75)
Creatinine	2	88	90	86			198	144			0.76 (0.67-0.86)
	3	94	93	95			131	183			0.88 (0.81-0.96)
	2+3	91	91	91			329	327			0.82 (0.76-0.88)
Invasive ventilation	2	94	95	91			223	119			0.86 (0.78-0.94)
	3	93	82	96			61	253			0.78 (0.65-0.9)
	2+3	93	92	94			284	372			0.86 (0.81-0.92)
$PaO_2 : FiO_2$	2	99	100	0			341	1			0 [b]
	3	90	95	0			299	15			-0.02 (-0.05-0.01)
	2+3	95	98	0			640	16			-0.02 (-0.03-0)
PCO_2	2	92	95	75	100		285	53	4		0.75 (0.61-0.88)
	3	87	91	75	0		230	83	1		0.67 (0.55-0.79)
	2+3	90	94	75	80		515	136	5		0.71 (0.62-0.8)
Leukocytes	2	99	99	50			338	4			0.5 (-0.11-1.1)
	3	97	98	55			303	11			0.53 (0.17-0.89)
	2+3	98	99	53			641	15			0.52 (0.22-0.83)
Platelets	2	92	97	80	0	0	289	45	5	3	0.75 (0.64-0.86)
	3	74	78	75	50		164	114	36		0.53 (0.42-0.65)
	2+3	84	90	77	44	0	453	159	41	3	0.64 (0.56-0.72)

[a] Complete agreement with a null variance
[b] Null variance in one group

Supplemental Digital Content 1: Similar as Table 3, but with specific agreements on the different components of the PELOD-2. Categories represents the number of points that are possible for each organ.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

Article 2 - *Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children*



Supplemental Digital Content 2: Plotted confusion matrix of the aPELOD-2 and mPELOD-2

	Missing Values		Normal Values		Abnormal Values	
	aPELOD2	**mPELOD2**	**aPELOD2**	**mPELOD2**	**aPELOD2**	**mPELOD2**
Glasgow Coma Score	29 (18.5%)	129 (82.2%)	62 (39.5%)	0 (0%)	66 (42%)	28 (17.8%)
Pupillary reaction	47 (29.9%)	75 (47.8%)	109 (69.4%)	0 (0%)	1 (0.6%)	82 (52.2%)
Lactate	4 (2.5%)	22 (14%)	117 (74.5%)	0 (0%)	36 (22.9%)	135 (86%)
Mean arterial pressure	0 (0%)	5 (3.2%)	11 (7%)	0 (0%)	146 (93%)	152 (96.8%)
Creatinine	0 (0%)	6 (3.8%)	56 (35.7%)	0 (0%)	101 (64.3%)	151 (96.2%)
Invasive ventilation	16 (10.2%)	0 (0%)	6 (3.8%)	35 (22.3%)	135 (86%)	122 (77.7%)
PaO2 : FiO2	22 (14%)	13 (8.3%)	133 (84.7%)	0 (0%)	2 (1.3%)	144 (91.7%)
PCO2	0 (0%)	0 (0%)	96 (61.1%)	10 (6.4%)	61 (38.9%)	147 (93.6%)
Leukocytes	2 (1.3%)	8 (5.1%)	146 (93%)	0 (0%)	9 (5.7%)	149 (94.9%)
Platelets	2 (1.3%)	8 (5.1%)	56 (35.7%)	0 (0%)	99 (63.1%)	149 (94.9%)

Supplemental Digital Content 3: Missing, normal and abnormal values for aPELOD-2 and mPELOD-2 (study 3)

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

MODS Screening

The proportion of patients (studies 2 and 3, n=328) with two or more organ dysfunctions on admission day was higher (p=0.002) when evaluated by the aPELOD-2 (78%, n=237) than by the mPELOD-2 (72%, n=255).

aPELOD-2 and mPELOD-2 Survival Discrimination

Survival discrimination was similar between the aPELOD-2 (AUROC 0.74, 95% CI: 0.63 – 0.85) and mPELOD-2 (AUROC 0.70, 95% CI: 0.59 – 0.81) (p = 0.15) (**Figure 4**, part A). However, when the aPELOD-2 was calculated using the first 24 hours after the admission as recommended in the original PELOD-2 methodology **(Figure 4**, part B), the aPELOD-2 AUROC increased (0.81, 95% CI: 0.72 – 0.90) and became significantly higher than the mPELOD-2 (p=0.01).

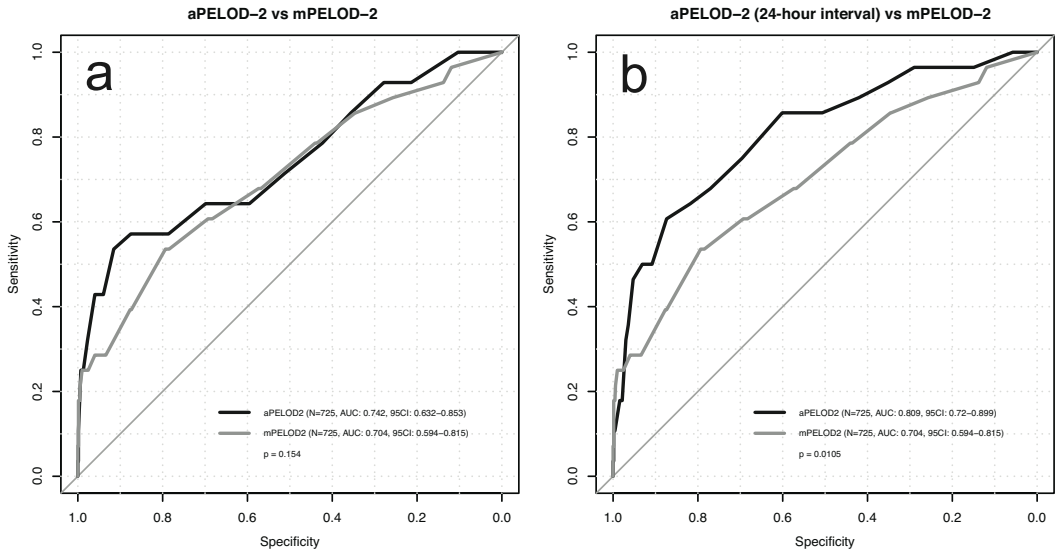


Figure 4: Death discrimination using the area under the ROC curve for day-one PELOD-2. mPELOD2 curve is the same in both panels. Panel A: aPELOD-2 was calculated using data from the day of the admission (from 0:00 am to 11:59 pm). Panel B: aPELOD-2 was calculated using data from the first 24-hour after the admission.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

mPELOD-2 Computation

In order to explore the causes of disagreement between mPELOD-2 and aPELOD-2, we verified the manual calculation done for the only study that provided detailed data on manual collection (study 3). We found 10% disagreement between the indicated mPELOD-2 and the verified one; correction did not significantly improve the ICC with aPELOD-2 (0.65, 95% CI: 0.20 – 0.82 vs. 0.62, 95% CI: 0.13 – 0.81).

Time Estimated to Calculate aPELOD-2

The algorithm was able to calculate a single aPELOD-2 score in approximatively 0.03 second.

DISCUSSION

The performance of the aPELOD-2 algorithm was good (AUROC 0.84) and was similar to AUROC values between 0.76 and 0.94 reported in the literature (**Figure 1**) (7–11, 30, 31). Proportion of MODS (62%) was also similar to that reported in the literature (55%) (1). The intraclass correlation between aPELOD-2 and mPELOD-2 was between good and excellent, but with important variation among studies. Data collection was done by medical students for both the study with the lowest correlation (study 4) and the study with the highest correlation (study 2). Because both studies observed no mortality, it was impossible to use survival discrimination as a surrogate for the quality of data gathering. We interpret this as an example of interrater variability when non-professional raters collect data. In the literature, the interrater correlation has not been formally studied for the PELOD-2 score. However, the data collection process for the first version of the PELOD, which had an ICC of 0.79 and 0.86 on a subset of the original cohort, was very similar to that which occurred for PELOD-2 data collection in our study (32). The APACHE II score ICC has been evaluated in a dedicated prospective blinded study comparing three specifically trained raters (33) and had an excellent overall score (ICC 0.9); however, clinical interpretation for some components of the score was as low as 0.40. The Simplified Acute Physiological Scores (SAPS) II and 3 have also been evaluated with trained

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

medical personnel; authors reported an overall ICC of 0.84 and 0.80 for SAPS II and 3, respectively (34). This is congruent with our findings in which the ICC between professional research clerks (study 1 and 3) was excellent (0.92) and decreased when less experienced raters were involved. In this context, an ICC of 0.75 between aPELOD-2 and mPELOD-2 strengthens the validity of the aPELOD-2. Moreover, the aPELOD-2 does not need any specific training to have perfect reproducibility. On the other hand, if the aPELOD-2 cannot be used, this data suggests there might be benefit in using highly qualified personnel with standardized training for mPELOD-2 data collection.

The aPELOD-2 algorithm significantly outperformed the manual PELOD-2 score. However, the superiority of the aPELOD-2 discrimination may be due to the calculation by 24-hour intervals starting at admission time. The only study that collected mPELOD-2 based on the 24-hour definition (study 2) had a nil mortality rate, preventing survival discrimination comparison.

Overall aPELOD-2 scores were slightly higher than mPELOD-2 (supplemental data, Supplemental Digital Content 1 and 2). This can be explained, at least partially, by either a better sensitivity or a lower specificity. The algorithm may be more prone to include possible erroneous data because it cannot disregard abnormal data based on the clinical context. For example, laboratory results are involved in seven of the ten components of the PELOD-2 score. Clinically suspected erroneous laboratory results are usually repeated or amended in clinical practice, but never erased from the EMR. Research assistants may decide during manual PELOD-2 scoring to keep or ignore laboratory results based on their clinical value. Thus, these laboratory results could be picked by the algorithm as the most abnormal value without consideration of whether the data is clinically valid. Regarding continuous data such as blood pressure, even EMRs with clinical validation may be subject to errors (35). To limit this risk, future scores could base their selection of abnormal continuous data on medians or percentiles, known to be simple and robust to minimize erroneous data, raw signal analysis with filtering (35) or machine learning algorithms (36). Furthermore, missing values are imputed as normal values for most severity scores including PELOD-2, but future scores may want to distinguish normal from missing values (supplemental data, Supplemental Digital Content 3). The original PELOD-2 definition

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

did not account for the increased amount of data collected in modern ICUs such as that provided by continuous data stream from monitors (37) and the associated risk of having a single outlier that needs to be validated as clinically relevant data. On the other hand, the algorithm will not miss any abnormal result recorded in the EMR, regardless of the amount of data to analyze. As part of this study, we extracted approximatively 3.3 million values required by the 10 components of the PELOD-2 in 5279 subjects in about 15 minutes. This amount of data is impossible to process for humans in a reasonable time period. Based on previous internal data, our institution estimates that about 20 minutes per day will be required for a research assistant to collect data for a daily PELOD score. Thus, if the PELOD-2 was systematically calculated on a daily basis in a center similar to ours (1100 admissions per year, 6700 patient-days per year), this could save 2200 work hours (1.2 full time equivalent). Moreover, all the other steps required by the PELOD-2 score (age and day calculation, PaO_2/FiO_2 calculation, normal values that depend on the patient age and summation of the different components) all comprise a risk for error that a computer could easily avoid. For example, we found 10% of disagreement among mPELOD-2 calculation and also noticed that some mPELOD-2 scores reported three points in a category (platelets) for which the maximum is two (supplemental data, Supplemental Digital Content 2).

Components of severity scores that need clinical input are known to have a lower agreement between human raters or between algorithm and human (13, 14, 33, 38). Indeed, the lowest agreement in this study was for the GCS, possibly because only a human can ascertain whether the clinical context in which the GCS is measured truly represents the actual severity of the neurological status and is not affected by sedation or the need for invasive ventilation.

The strengths of our study are the number of patients included in the survival discrimination analysis and the quality of the validation process. To our knowledge, this study is the largest that has compared an automatic calculation of a severity score to the manual equivalent in pediatric critical care. Moreover, the open-source license makes the algorithm available for integration into constantly updated clinical decision support systems.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

Our study has limitations. First, data come from specific clinical studies that may bring a selection bias. Nevertheless, the performance of the aPELOD-2 on all our consecutive encounters is very good and comparable to that in the literature. To minimize the possible impact of the latter limitation, we plan to conduct a similar study in several PICUs. Second, the percentage error in the Bland-Altman analysis is high (180%). This indicates that the limits of agreement are proportionally high compared to the mean value of the referenced method. An upper limit of 30% was suggested for adult cardiac output studies (39). In other contexts, this requires careful interpretation, especially when the variable is discrete and broad as seen in the PELOD-2 (28, 40). We compared this result to the percentage error among the different mPELOD-2 studies and found that they were high as well (99% and 144%). Therefore, the interpretation of the percentage error is limited in the aPELOD-2 validation process. Third, there is underrepresentation of mortality and patients with severe organ failure that certainly limits parts of the validation (such as the pupillary reaction). This is seen as a consequence of global PICU mortality improvement; a large multicenter study would be required to address this limitation. Finally, the algorithm does not account for cyanotic status in children with congenital heart disease. Therefore, the PaO_2/FiO_2 ratio is calculated instead of being set to normal. Fortunately, the impact of this limitation is minimized by the PELOD-2 strict threshold of 60 which implies that all normally oxygenated cyanotic patient (PaO_2 40 mmHg) with a FiO_2 below 67% would still be counted as normal. Future upgrades of the EMR will correct this limitation.

CONCLUSIONS

The aPELOD-2 provides a valid estimation of the PELOD-2 score that is fast, less labor intensive and well correlated with the mPELOD-2. Moreover, the algorithm is freely available. Use of the aPELOD-2 could occupy an important place within clinical decision support systems in pediatric critical care as well as serve for research purposes at the bedside. We have found that the aPELOD-2 has a better survival discrimination than the mPELOD-2 but recognize that some components, such as the GCS, may be better evaluated by the mPELOD-2. Our next steps will be to use the algorithm within a larger multicenter dataset in order to improve the algorithm on the component that requires clinical judgment and to increase robustness against erroneous data.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

REFERENCES

1. Leteurtre S, Duhamel A, Deken V, et al.: Daily estimation of the severity of organ dysfunctions in critically ill children by using the PELOD-2 score. *Crit Care* 2015; 19:2–7
2. Namachivayam P, Shann F, Shekerdemian L, et al.: Three decades of pediatric intensive care: Who was admitted, what happened in intensive care, and what happened afterward*. *Pediatr Crit Care Med* 2010; 11:549–555
3. Pinto NP, Rhinesmith EW, Kim TY, et al.: Long-Term Function After Pediatric Critical Illness. *Pediatr Crit Care Med* 2017; 18:e122–e130
4. Leteurtre S, Duhamel A, Salleron J, et al.: PELOD-2: An update of the PEdiatric logistic organ dysfunction score. *Crit Care Med* 2013; 41:1761–1773
5. Pollack MM, Holubkov R, Funai T, et al.: The Pediatric Risk of Mortality Score: Update 2015. *Pediatr Crit Care Med* 2016; 17:2–9
6. Straney L, Clements A, Parslow RC, et al.: Paediatric index of mortality 3: An updated model for predicting mortality in pediatric intensive care. *Pediatr Crit Care Med* 2013; 14:673–681
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
8. Gonçalves J-P, Severo M, Rocha C, et al.: Performance of PRISM III and PELOD-2 scores in a pediatric intensive care unit. *Eur J Pediatr* 2015; 174:1305–1310
9. El-Nawawy A, Mohsen AA, Abdel-Malik M, et al.: Performance of the pediatric logistic organ dysfunction (PELOD) and (PELOD-2) scores in a pediatric intensive care unit of a developing country. *Eur J Pediatr* 2017; 176:849–855
10. Schlapbach LJ, Straney L, Bellomo R, et al.: 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; 44:179–188

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

11. Zhang L, Huang H, Cheng Y, et al.: [Predictive value of four pediatric scores of critical illness and mortality on evaluating mortality risk in pediatric critical patients]. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue* 2018; 30:51–56
12. Harrison AM, Yadav H, Pickering BW, et al.: Validation of computerized automatic calculation of the sequential organ failure assessment score. *Crit Care Res Pract* 2013; 2013:975672
13. Aakre C, Franco PM, Ferreyra M, et al.: Prospective validation of a near real-time EHR-integrated automated SOFA score calculator. *Int J Med Inform* 2017; 103:1–6
14. Huerta LE, Wanderer JP, Ehrenfeld JM, et al.: Validation of a Sequential Organ Failure Assessment Score using Electronic Health Record Data. *J Med Syst* 2018; 42:199
15. Beck BF, Kaboli PJ, Perencevich EN, et al.: An automated computerized critical illness severity scoring system derived from APACHE III: modified APACHE. *J Crit Care* 2018; 48:237–242
16. Collins GS, Reitsma JB, Altman DG, et al.: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015; 162:55
17. Nadkarni PM, Brandt C: Data extraction and ad hoc query of an entity-attribute-value database. *J Am Med Informatics Assoc* 1998; 5:511–527
18. Murphy SN, Weber G, Mendis M, et al.: Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Informatics Assoc* 2010; 17:124–130
19. Robin X, Turck N, Hainard A, et al.: pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12:77
20. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44:837–45
21. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86:420–8
22. McGraw KO, Wong SP: Forming Inferences about Some Intraclass Correlation Coefficients. *Psychol Methods* 1996; 1:30–46
23. Cicchetti D V.: Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

Instruments in Psychology. *Psychol Assess* 1994; 6:284–290

24. Cohen J: Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70:213–220
25. Cicchetti D V, Feinstein AR: High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990; 43:551–8
26. Hripcsak G, Heitjan DF: Measuring agreement in medical informatics reliability studies. *J Biomed Inform* 2002; 35:99–110
27. Cohen J: A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas* 1960; 20:37–46
28. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1:307–310
29. McLean AS, Needham A, Stewart D, et al.: Estimation of cardiac output by noninvasive echocardiographic techniques in the critically ill subject. *Anaesth Intensive Care* 1997; 25:250–254
30. Karam O, Demaret P, Duhamel A, et al.: Performance of the PEdiatric Logistic Organ Dysfunction-2 score in critically ill children requiring plasma transfusions. *Ann Intensive Care* 2016; 6
31. Wong JJ, Hornik CP, Mok YH, et al.: Performance of the Paediatric Index of Mortality 3 and Paediatric Logistic Organ Dysfunction 2 Scores in Critically Ill Children. *Ann Acad Med Singapore* 2018; 47:285–290
32. Leteurtre S, Martinot A, Duhamel A, et al.: Validation of the paediatric logistic organ dysfunction (PELOD) score: Prospective, observational, multicentre study. *Lancet* 2003; 362:192–197
33. Kho ME, McDonald E, Stratford PW, et al.: Interrater reliability of APACHE II scores for medical-surgical intensive care patients: A prospective blinded study. *Am J Crit Care* 2007; 16:378–383
34. Strand K, Strand LI, Flaatten H: The interrater reliability of SAPS II and SAPS 3. *Intensive Care Med* 2010; 36:850–853
35. Hug CW, Clifford GD, Reisner AT: Clinician blood pressure documentation of stable intensive care patients: An intelligent archiving agent has a higher association with future hypotension. In: Critical Care Medicine. 2011.

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

36. Johnson AE, Ghassemi MM, Nemati S, et al.: Machine Learning and Decision Support in Critical Care. *Proc IEEE Inst Electr Electron Eng* 2016; 104:444–466

37. Brossier D, El Taani R, Sauthier M, et al.: Creating a High-Frequency Electronic Database in the PICU: The Perpetual Patient. *Pediatr Crit Care Med* 2018; 19:e189–e198

38. Arts DGT, De Keizer NF, Vroom MB, et al.: Reliability and accuracy of Sequential Organ Failure Assessment (SOFA) scoring. *Crit Care Med* 2005; 33:1988–1993

39. Critchley LAH, Critchley JAJH: A meta-analysis of studies using bias and precision statistics to compare cardiac output measurement techniques. *J Clin Monit Comput* 1999; 15:85–91

40. Giavarina D: Understanding Bland Altman analysis. *Biochem Medica* 2015; 25:141–151

Version non définitive. La version définitive de cet article (Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric Logistic Organ Dysfunction-2 versus manual Pediatric Logistic Organ Dysfunction-2 score for critically ill children. *Pediatr Crit Care Med*. 2020 Feb;21(4):E160–9) est disponible au https://doi.org/10.1097/PCC.0000000000002235.

Discussion

L'évaluation continue de la gravité de l'IRHA chez les survivants est complexe. Le score PELOD-2 non respiratoire (88) permet d'évaluer au quotidien l'amélioration ou la détérioration globale, indépendamment de l'évolution l'IRHA. Il peut donc servir de critère de jugement continu pour évaluer les impacts de l'IRHA. Toutefois, ce score a été validé pour une utilisation manuelle, ce qui limite fortement son utilisation automatique ou à grande échelle. Ce travail vient lever cette limitation et permettre son utilisation au sein d'un SADC.

Très peu d'études ont validé l'automatisation d'un processus initialement conçu pour être réalisé manuellement, tenant probablement pour acquis que les données informatiques sont nécessairement fiables. Dans cet article, nous montrons que non seulement l'automatisation est possible, et probablement plus performante, mais nous identifions également quelques limites au calcul automatique, afin de cibler les prochaines améliorations. Nous avons montré qu'en dépit de la formation des personnes collectant les données, de nombreuses erreurs « humaines » survenaient, autant dans la recherche d'information dans le dossier médical que dans le calcul du score à proprement parler. Ces erreurs humaines sont facilement corrigibles par un programme informatique. Toutefois, ce travail a aussi montré le bénéfice qu'offre une interprétation humaine aux données extraites. En particulier, l'échelle de coma de Glasgow saisie dans le score doit représenter l'état neurologique avant que le patient ne soit intubé ou anesthésié. Or, cette échelle est couramment documentée dans le dossier électronique dans le suivi du patient une fois qu'il a été intubé et anesthésié, influençant négativement le score. Ce degré d'analyse est plus complexe à intégrer dans un système automatique, car la simple présence de médicaments anesthésiants au dossier du patient ne suffit pas à conclure à leur effet.

Par ailleurs, le score de PELOD-2 a été conçu pour être calculé par blocs de 24 h en commençant à l'admission du patient aux soins intensifs. Lorsque calculé manuellement, il est parfois calculé en fonction des jours civils par souci de simplicité. Cette étude démontre la supériorité de la méthode originale. Nous avons rendu l'algorithme accessible à tous (licence ouverte, AGPL-3).

L’automatisation de ce score donnera une perspective indispensable à un SADC dédié à l’IRHA pour l’évaluation globale de la gravité du patient.

Article 3 – Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza

Préface

Ce travail avait pour but de développer un modèle prédictif pour identifier précocement les patients les plus à risque d'évoluer vers une IRHA prolongée. La persistance d'une IRHA implique, par définition, la nécessité d'une assistance respiratoire et donc d'une hospitalisation en soins intensifs pédiatriques. Ce travail utilise différentes méthodes d'apprentissage automatique supervisées pour d'abord prédire le sous-groupe de patients le plus à risque. Ensuite, ces modèles permettent d'améliorer notre compréhension de l'IHRA par l'identification des facteurs les plus contributifs au modèle et d'évaluer leur potentiel comme cible thérapeutique. De plus, l'utilisation d'un modèle prédictif peut aussi s'avérer utile pour anticiper les ressources nécessaires en cas de vagues épidémiques (ce travail a été réalisé avant la pandémie COVID-19). Dans cet article réalisé en collaboration avec 3 autres chercheurs, j'ai réalisé plus de 85 % du travail, soit le formatage et l'analyse des données, le développement et la validation des modèles prédictifs et la rédaction du manuscrit. Il a été publié dans la revue *Critical Care Explorations* en 2020.

Article

Predictive Modeling Report

Critical Care Explorations

OPEN

Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza

Michaël S. Sauthier, MD, MBI[1–3]; Philippe A. Jouvet, MD, PhDMBA[3]; Margaret M. Newhams, MPH[1]; Adrienne G. Randolph , MD, MSc[1,4]; for the Pediatric Acute Lung Injury and Sepsis Investigators (PALISI) Pediatric Intensive Care Influenza (PICFLU) Network Investigators

Background: Influenza virus is a major cause of acute hypoxemic respiratory failure. Early identification of patients who will suffer severe complications can help stratify patients for clinical trials and plan for resource use in case of pandemic.

Objective: We aimed to identify which clinical variables best predict prolonged acute hypoxemic respiratory failure in influenza-infected critically ill children. Acute hypoxemic respiratory failure was defined using hypoxemia cutoffs from international consensus definitions of acute respiratory distress syndrome in patients with ventilatory support. Prolonged acute hypoxemic respiratory failure was defined by acute hypoxemic respiratory failure criteria still present at PICU day 7.

Derivation Cohort: In this prospective multicenter study across 34 PICUs from November 2009 to April 2018, we included children (< 18 yr) without comorbid risk factors for severe disease.

Validation Cohort: We used a Monte Carlo cross validation method with N^2 random train-test splits at a 70–30% proportion per model.

Prediction Model: Using clinical data at admission (day 1) and closest to 8 AM on PICU day 2, we calculated the area under the receiver operating characteristic curve using random forests machine learning algorithms and logistic regression.

Results: We included 258 children (median age = 6.5 yr) and 11 (4.2%) died. By day 2, 65% ($n = 165$) had acute hypoxemic respiratory failure dropping to 26% ($n = 67$) with prolonged acute hypoxemic respiratory failure by day 7. Those with prolonged acute hypoxemic respiratory failure had a longer ICU stay (16.5 vs 4.0 d; $p < 0.001$) and higher mortality (13.4% vs 1.0%). A multivariable model using random forests with 10 admission and eight day 2 variables performed best (0.93 area under the receiver operating characteristic curve; 95 CI%: 0.90–0.95) where respiratory rate, F_{IO_2}, and pH on day 2 were the most important factors.

Conclusions: In this prospective multicentric study, most children with influenza virus–related respiratory failure with prolonged acute hypoxemic respiratory failure can be identified early in their hospital course applying machine learning onto routine clinical data. Further validation is needed prior to bedside implementation.

Key Words: acute respiratory distress syndrome; automatic data processing; children; clinical decision support systems; critical care; machine learning

About one in 10 children for influenza virus infection require admission to a PICU for acute hypoxemic respiratory failure (AHRF), and up to 9% of the critically ill children will not survive (1–4). In the event of an outbreak of a novel influenza A virus, PICUs are at risk to be overwhelmed by the number of patients who require mechanical ventilation and advanced rescue therapies (5, 6). Acute respiratory distress syndrome (ARDS) is a major subgroup of AHRF. Using consensus definitions, ARDS is stratified into a mild, moderate, or severe disease essentially based on the level of hypoxemia and its relationship to mortality (6, 7).

Predictive models have been designed to predict ARDS (8) or its mortality (9–13). Nearly all published ARDS predictive models used logistic regression (LR), which is fairly easy to understand, but must follows several assumptions and has limited abilities to

[1]Department of Anesthesiology, Critical Care and Pain Medicine, Boston Children's Hospital, Boston, MA.

[2]Department of Biomedical Informatics, Harvard Medical School, Boston, MA.

[3]Departments of Pediatrics, Sainte-Justine Hospital, Montreal, QC, Canada.

[4]Departments of Anaesthesia and Pediatrics, Harvard Medical School, Boston, MA.

Crit Care Expl 2020; 2:e0175

DOI: 10.1097/CCE.0000000000000175

Article 3 - *Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza*

BACKGROUND: Influenza virus is a major cause of acute hypoxemic respiratory failure (AHRF). Early identification of patients who will suffer severe complications can help stratify patients for clinical trials and plan for resource use in case of pandemic.

OBJECTIVE: We aimed to identify which clinical variables best predict prolonged AHRF in influenza infected critically ill children. AHRF was defined using hypoxemia cutoffs from international consensus definitions of acute respiratory distress syndrome in patients with ventilatory support. Prolonged AHRF was defined by AHRF criteria still present at PICU day 7.

DERIVATION COHORT: In this prospective multicenter study across 34 PICUs from November 2009 to April 2018, we included children (<18 years) without comorbid risk factors for severe disease.

VALIDATION COHORT: We used a Monte Carlo cross validation method with N^2 random train-test splits at a 70%-30% proportion per model.

PREDICTION MODEL: Using clinical data on admission (day 1) and closest to 8am on PICU day 2, we calculated the area under the ROC curve (AUROC) using random forests machine learning algorithms and logistic regression.

RESULTS: We included 258 children (median age 6.5 years) and 11 (4.2%) died. By day 2, 65% (n=165) had AHRF dropping to 26% (n=67) with prolonged AHRF by day 7. Those with prolonged AHRF had a longer ICU stay (16.5 days vs. 4.0 days, p<0.001) and higher mortality (13.4% vs. 1.0%). A multivariable model using random forests with 10 admission and 8 day 2 variables performed best (0.93 AUROC, 95 CI% 0.90 – 0.95) where respiratory rate, FiO_2 and pH on day 2 were the most important factors.

CONCLUSION: In this prospective multicentric study, most children with influenza virus-related respiratory failure with prolonged AHRF can be identified early in their hospital course applying machine learning onto routine clinical data. Further validation is needed prior to bedside implementation.

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

INTRODUCTION

About one in ten children for influenza virus infection require admission to a pediatric intensive care unit for acute hypoxemia respiratory failure (AHRF), and up to 9% of the critically ill children won't survive (1–4). In the event of an outbreak of a novel influenza A virus, PICUs are at risk to be overwhelmed by the number of patients who require mechanical ventilation and advanced rescue therapies (5, 6). Acute respiratory distress syndrome (ARDS) is a major subgroup of AHRF. Using consensus definitions, ARDS is stratified into a mild, moderate or severe disease essentially based on the level of hypoxemia and its relationship to mortality (6, 7).

Predictive models have been designed to predict ARDS (8) or its mortality (9–13). Nearly all published ARDS predictive models used logistic regression, which is fairly easy to understand, but must follows several assumptions and has limited abilities to exploit non-linear data. However, more recent machine learning models do not have those restrictions in finding the best pathway to a pre-specified outcome and has been shown to be superior to simple logistic regression for some clinical cohorts (13, 14) and may help with treatment response interpretation (15, 16).

Initial hypoxemia severity has been associated with a longer ventilation duration (17) and multivariable scores have been developed to predict prolonged mechanical ventilation (18, 19). However, none were specifically built to predict prolonged AHRF in influenza infected patients. We hypothesized that in a group of children with minimal risk factors for developing AHRF from influenza infection, using commonly available clinical and laboratory data available early on in their hospital course we could develop a model that would accurately predict children with a prolonged AHRF. Such model would be helpful for future clinical trials that would aim to target the sickest patients early in their clinical course.

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

MATERIALS AND METHODS

Data were prospectively collected by the PALISI Pediatric Intensive Care Influenza (PICFLU) Investigators across 34 international PICUs from November 2009 to April 2018. Detailed methods have been previously reported (20, 21). Children (<18 years) were admitted to a PICU with severe acute respiratory infection symptoms and microbiologically confirmed influenza virus infection. Children with underlying heart, lung, immune and other disorders that would predispose them to influenza-related complications were excluded. For example, children with mild asthma not on daily controller medications, mild eczema and those with other conditions that did not impair respiratory function were eligible. We also excluded children suffering a prehospital cardiac arrest with early death from neurologic complications and those where hypoxemia was thought to be due to left atrial hypertension (6). This study was approved by the respective Institutional Review Boards of the participating centers.

Data were collected closest to the PICU admission time and daily closest to 8 a.m. to reflect the time periods of admission and daily clinical rounds when intensive clinical assessments were common. Hypoxemia cutoffs used for classification were stratified according to **Figure 1** using primarily the pediatric acute lung injury consensus conference (PALICC) cutoffs and secondarily the Berlin ARDS definition cutoffs (**Figure 1**) when PALICC cutoffs could not be calculated due to missing data (6, 7). Only patients with ventilatory support (invasive or noninvasive) could be considered to have AHRF. Following PALICC recommendations, if noninvasive ventilation was used, hypoxemia was defined by a PaO_2/FiO_2 (PF) ratio ≤ 300 mmHg or by SpO_2/FiO_2 (SF) ≤ 264 mmHg if no arterial sample were available. In case of invasive ventilation, hypoxemia was defined by an oxygenation index (OI) ≥ 4 or by an oxygenation saturation index (OSI) ≥ 5 if no arterial sample was available. OSI and SF were analyzed using only measurements where SpO_2 ≤ 97% as PALICC recommends and using all recorded SpO_2 values.

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

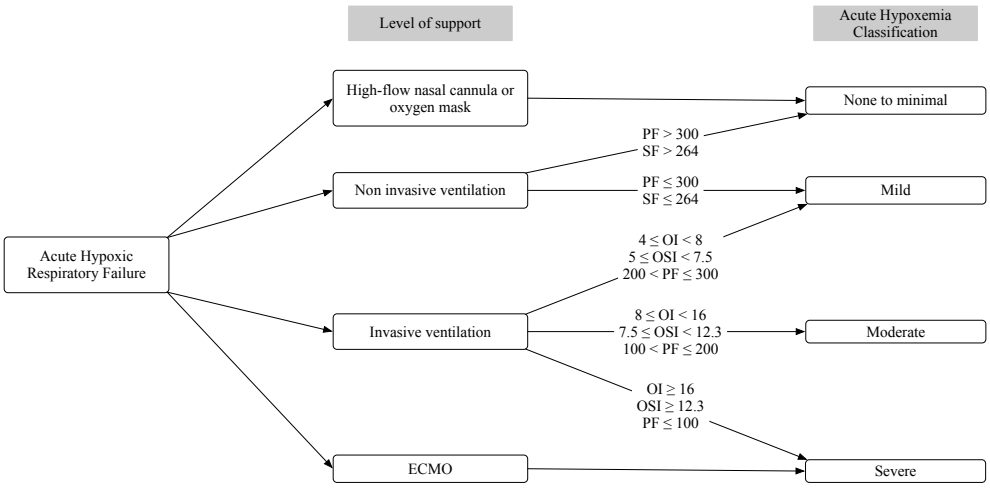


Figure 1: Acute hypoxemia respiratory failure severity classification. If the arterial-based metrics are not available, they may be replaced by their non-invasive equivalent.

Patients supported by extracorporeal membrane oxygenation (ECMO) were considered to have severe AHRF, regardless of their oxygenation measurement and patients supported by non-invasive ventilation were classified as mild or no AHRF, even if a more severe hypoxemia was assessed. The main outcome was the presence of AHRF seven days after the PICU admission (prolonged AHRF) while still in the PICU.

Missing Data Imputation and Statistical Analysis

Descriptive statistics included medians and interquartile ranges (IQR) for continuous variables and frequencies with percentages for categorical variables. Wilcoxon and Fisher exact tests were used for continuous and discrete variable comparisons, respectively. Missing values were inferred when possible (e.g., ventilation mode based on other available information). If inference was not possible, we assumed normality for pH (7.40), PCO_2 (40 mmHg) or respiratory rate using age reference (22),

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

as frequently done in ICU studies (23–26). Because no normal value exists for mean airways pressure (MAwP), we used the median value of day one and day two, respectively. If no inference could be made (e.g., if no PaO_2 was measured), we censored the observation. As non-invasive metrics (OSI and SF ratio) were directly tested, we did not use them to estimate OI or PF.

Data analyses were conducted using Python language (Python Software Foundation, USA, version 3.7.6) and R (R Foundation for Statistical Computing, Austria, version 3.6.2) with the packages "pROC" and "ggalluvial" (27, 28). We used a Monte Carlo cross validation method with N^2 random train-test splits (i.e. 66,564 repetitions in our case) at a 70%-30% proportion per model (29). The Monte Carlo methods is a bootstrap-based method that provide a robust empirical distribution in order to compare different models (29). We validated the model by calculating the averaged area under the ROC curve (AUROC) on all test groups to estimate the model discrimination (23). This methods is appropriate for small datasets and has the advantage to use all the observations, to be relatively robust against overfitting and is able to estimate 95% confidence intervals (95% CI) and P values for model comparison (23). Calibration was assessed using the Hosmer-Lemeshow Goodness of Fit test;(30) $P > 0.05$ suggested that the model was well calibrated. We followed the 2020 standards for prediction models in critical care (31) and the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines for development and validation of predictive models (32).

Predictive Models

We chose the random forests (RF) algorithm, a method based on decision tree, for machine learning based on its prior performance and low risk of overfitting (33, 34). Random forests algorithm is an ensemble method, *i.e.* it is based on multiple small decision subtrees (33). Each subtree is randomly built and is able to come up to a conclusion, but the final result is determined by the sum of all the subtrees in a similar way to a democratic process. Models were built using the R package "randomForest" set with 1000 trees maximum depth (35). Results were compared to a multivariable logistic regression (LR). We compared common hypoxemia metrics (OI, PF, OSI and SF) to multivariable models that included respiratory and clinical

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

variables. We used data recorded on day one (PICU admission day), day two (8 a.m. the day after) and both days. In multivariable models, we estimated the importance of each predictor using the error-rate on classification after permutation (35).

RESULTS

Of the 260 eligible patients, we excluded two that died early after cardiac arrest and resulting neurologic sequelae. No patients were excluded due to suspected left atrial hypertension. Median age was 6.2 years (IQR 2.1 – 10.6), female ratio 42% (n=109) and hospital mortality 4.2% (n=11). By day two, 65 % (n=165) met the criteria for AHRF and 26% (n=67) on day seven. Patients' characteristics are summarized in **Table 1**. About 51% (n=132) were invasively ventilated on day one and 38% (n=97) had an arterial sample. Hypoxemia metrics (OI, PF, OSI and SF) were applicable to 29%, 38%, 33% and 62% on day one. The availability of the main components of each commonly used oxygenation assessment metrics is illustrated in **Figure 2**.

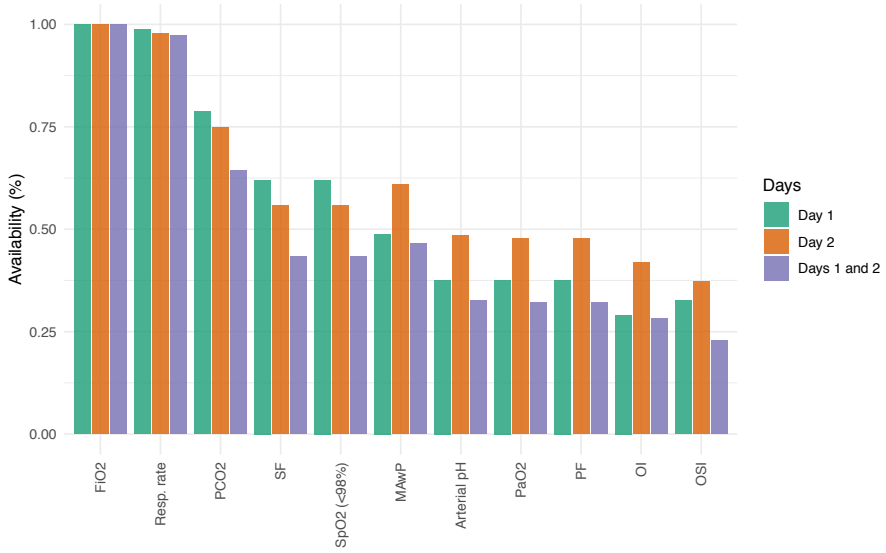


Figure 2: Availability of common oxygenation metrics.

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

Variables	No AHRF at day 7 (n=191)	Prolonged AHRF (n=67)	P value
Age, years, median [IQR]	5.5 [1.7 – 9.6]	8.6 [4.2 – 13.5]	<0.001
Gender, female (%)	84 (44 %)	25 (37.3 %)	0.39
ICU days, median [IQR]	4 [2.4 – 6.8]	16.5 [12.2 – 27.6]	<0.001
PRISM III, median [IQR]	4 [0 – 8]	14 [6.5 – 24.5]	<0.001
Hospital mortality, n (%)	2 (1 %)	9 (13.4 %)	<0.001
Highest hypoxemia severity (day 1 and 2), n (%)			
None or minimal	90 (47.1 %)	3 (4.5 %)	<0.001
Mild	58 (30.4 %)	11 (16.4 %)	0.036
Moderate	29 (15.2 %)	11 (16.4 %)	0.85
Severe	14 (7.3 %)	42 (62.7 %)	<0.001
Arterial blood sample availability, n (%)			
Day 1	51 (26.7 %)	46 (68.7 %)	<0.001
Day 2	65 (34 %)	58 (86.6 %)	<0.001
Lung infiltrates (day 1 or 2), n (%)	137 (71.7 %)	66 (98.5 %)	<0.001
Admission invasive ventilation, n (%)	82 (42.9 %)	50 (74.6 %)	<0.001
Admission mean airway pressure (cmH_2O), median [IQR]	11 [9 – 14]	16.5 [13 – 20]	<0.001
Admission oxygenation markers, median [IQR]			
Oxygenation index	7.6 [4 – 13.2]	22.7 [11.6 – 32.9]	<0.001
PaO_2/FiO_2	170 [87 – 285]	87 [65 – 149]	<0.001
Oxygenation saturation index	7.9 [4.2 – 12.6]	15.1 [10.1 – 21.7]	<0.001
SpO_2/FiO_2	190 [121 – 318]	97 [96 – 151]	<0.001
Mechanical ventilation hours, median [IQR]	88 [55 – 133]	274 [200 – 485]	<0.001
Ventilator free days, median [IQR]	25.6 [23.4 – 27.2]	14.6 [0 – 19]	<0.001
ECMO support, n (%)	4 (2.1 %)	30 (44.8 %)	<0.001
Hours between admission and day 2, median [IQR]	17.7 [13.6 – 26.7]	16.5 [13.2 – 23.9]	0.36

Table 1: Characteristics of 258 children with influenza virus-related critical illness and no preexisting risk factors for severe disease.

The evolution of respiratory modalities over PICU days 1-7 (**Figure 3**) showed that proportion of patients requiring invasive ventilation increased until PICU day 3, but the number of patients meeting AHRF criteria (**Figure 1**) constantly decrease from PICU day 1 to 7. The hypoxemia severity evolution per patient was represented in an alluvial plot (**Figure 4**), showing that although 48% of severe patients on PICU day 1 remained severe on PICU day 7, there were many patients undergoing changes of severity classification over the first PICU week.

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

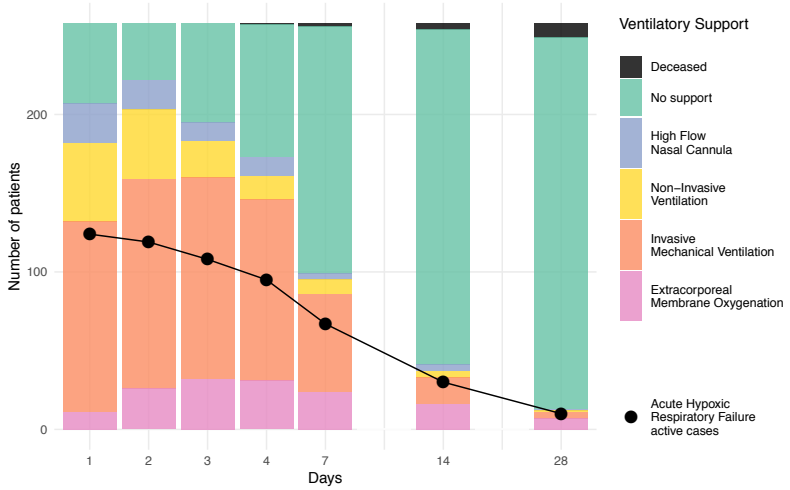


Figure 3: Daily ventilation modalities and acute hypoxic respiratory failure proportion over time.

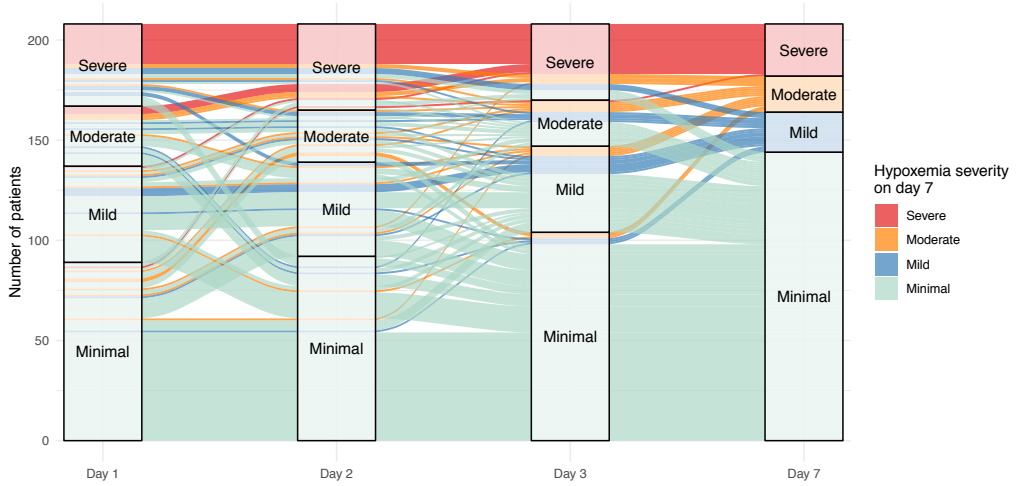


Figure 4: Alluvial plot showing the hypoxemia severity evolution between admission and day 7.

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

Predictive Models: Hypoxemia Markers

Many patients did not have an available MAwP to calculate OI and OSI, an arterial blood gas for OI and PF, or $SpO_2 \leq 97\%$ for OSI and SF (**Figure 2**). However, common oxygenation markers (OI, PF, OSI and SF) were found to be discriminant for prolonged AHRF in those patients with available data for calculation (**Figure 5**). When using only admission data, SF predicted better than OSI (AUROC 0.79 vs. 0.69, P = 0.04), but when both admission and day 2 were provided to the model (**Figure 5**), the difference between OSI and SF was no longer significant (P = 0.65). Using both day 1 and day 2 values in the model improved the discrimination for OI, PF and SF (P = 0.04, 0.009 and 0.002, respectively).

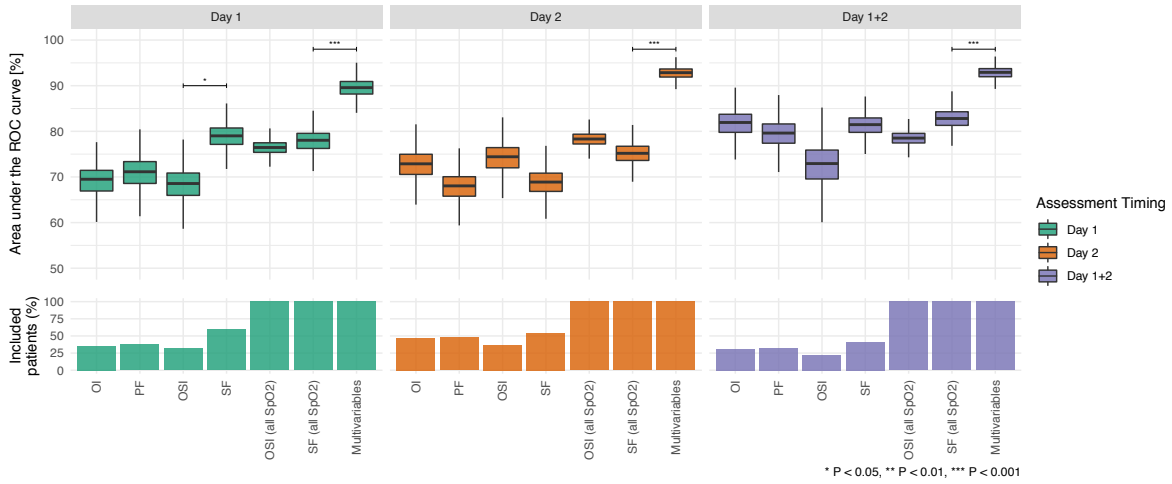


Figure 5: Random forests empirical distributions of the area under the ROC curve obtained after Monte Carlo simulation using common hypoxemia markers and multivariable models. Boxplot indicate the median value, interquartile range and 95% confidence interval. OI: Oxygenation index, PF: PaO_2/FiO_2, OSI: Oxygenation Saturation Index and SF: SpO_2/FiO_2.

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

Predictive Models: Multivariable Models

Predictors included continuous age and pediatric risk of mortality score (PRISM) III (36) at PICU admission and SpO_2, FiO_2, mean airway pressure, invasive ventilation, respiratory rate, pH and PCO_2 on admission and the day after. Random forests multivariable models outperformed all models using only oxygenation markers (all $P < 0.001$). Models using both day 1 and 2 achieved a 0.93 AUROC (95% CI 0.90 – 0.95) and was similar to model using only day 1 (AUROC 0.90, 95% CI 0.85 – 0.93, P = 0.17) or day 2 (AUROC 0.92, 95% CI 0.90 – 0.95, P=0.96) data (**Figure 5**). Models using day 2 and both days had good calibration (P = 0.13 and 0.07, respectively) but the model using only day one data was borderline (P = 0.05). Using a probability threshold of 0.5, model based on both day 1 and day 2 had 71% sensitivity (95% CI 60% – 82%), 93% specificity (95% CI 88% – 0.97%), 78% positive predictive value (95% CI 67% – 88%), 91% negative predictive value (95% CI 87% – 94%) and 88% accuracy (95% CI 85% – 91%) (**Supplemental Digital Content 1**). The analysis of the predictors' importance after adjusting for PRISM III score (**Supplemental Digital Content 2**) revealed that respiratory rate, FiO_2 and pH on day 2 were the most important factors for the model to predict prolonged AHRF.

Models	Timing	Algorithm	AUC	Sensitivity	Specificity	Pos. pred. value	Neg pred value	Accuracy
OI	Day 1	LR	0.78 (0.74–0.81)	0.56 (0.44–0.68)	0.87 (0.76–0.97)	0.76 (0.63–0.92)	0.74 (0.68–0.79)	0.74 (0.7–0.77)
OI	Day 1	RF	0.69 (0.62–0.74)	0.51 (0.41–0.62)	0.79 (0.72–0.84)	0.63 (0.55–0.71)	0.69 (0.64–0.74)	0.67 (0.63–0.71)
PF	Day 1	LR	0.72 (0.68–0.75)	0.75 (0.65–0.85)	0.56 (0.39–0.69)	0.57 (0.48–0.66)	0.74 (0.69–0.8)	0.64 (0.57–0.69)
PF	Day 1	RF	0.71 (0.62–0.77)	0.77 (0.65–0.86)	0.61 (0.36–0.71)	0.61 (0.48–0.68)	0.78 (0.67–0.85)	0.68 (0.54–0.73)
OSI	Day 1	LR	0.8 (0.76–0.84)	0.77 (0.6–0.89)	0.62 (0.48–0.75)	0.62 (0.55–0.7)	0.77 (0.67–0.88)	0.68 (0.64–0.72)
OSI	Day 1	RF	0.69 (0.61–0.79)	0.67 (0.54–0.9)	0.56 (0.46–0.66)	0.55 (0.48–0.61)	0.68 (0.6–0.87)	0.61 (0.56–0.69)
SF	Day 1	LR	0.75 (0.72–0.78)	0.67 (0.56–0.76)	0.72 (0.62–0.79)	0.49 (0.41–0.56)	0.84 (0.79–0.88)	0.7 (0.65–0.74)
SF	Day 1	RF	0.79 (0.73–0.84)	0.64 (0.29–0.83)	0.76 (0.66–0.89)	0.52 (0.42–0.61)	0.84 (0.74–0.92)	0.73 (0.67–0.78)
OSI (all SpO2)	Day 1	LR	0.77 (0.74–0.79)	0.51 (0.42–0.59)	0.92 (0.88–0.97)	0.7 (0.6–0.82)	0.85 (0.82–0.88)	0.82 (0.8–0.84)
OSI (all SpO2)	Day 1	RF	0.76 (0.73–0.79)	0.52 (0.39–0.62)	0.89 (0.84–0.93)	0.61 (0.51–0.7)	0.85 (0.81–0.88)	0.8 (0.76–0.83)
SF (all SpO2)	Day 1	LR	0.77 (0.74–0.8)	0.27 (0.01–0.62)	0.92 (0.79–1)	0.7 (0.45–1)	0.8 (0.74–0.87)	0.75 (0.73–0.78)

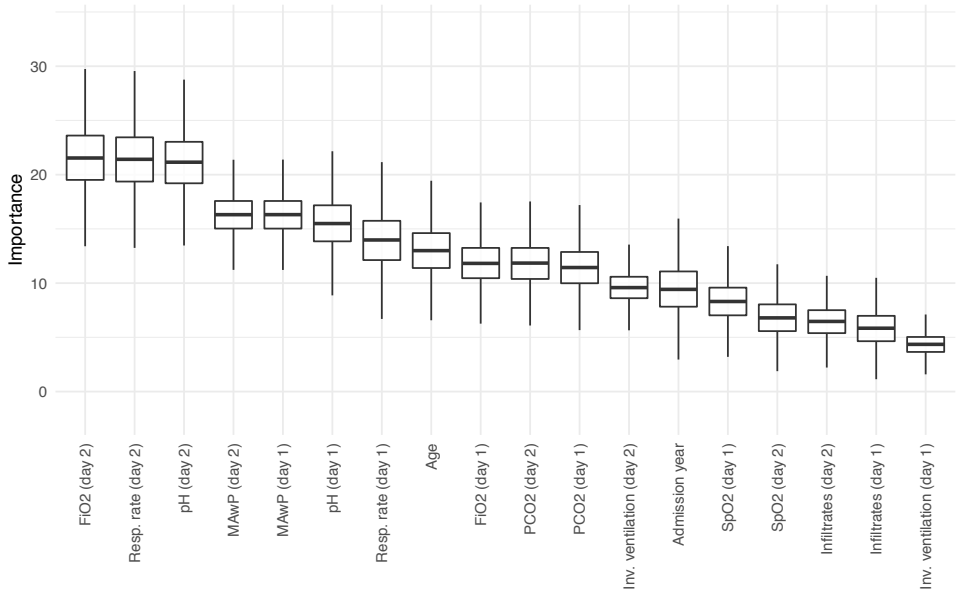
Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

SF (all SpO2)	Day 1	RF	0.78 (0.72–0.82)	0.59 (0.34–0.72)	0.83 (0.75–0.9)	0.53 (0.45–0.62)	0.86 (0.79–0.9)	0.77 (0.72–0.8)
Multivariables	Day 1	LR	0.82 (0.79–0.85)	0.55 (0.48–0.64)	0.89 (0.85–0.93)	0.63 (0.55–0.72)	0.86 (0.83–0.88)	0.81 (0.78–0.83)
Multivariables	Day 1	RF	0.9 (0.85–0.93)	0.67 (0.57–0.78)	0.93 (0.88–0.97)	0.76 (0.66–0.86)	0.9 (0.86–0.93)	0.86 (0.83–0.89)
OI	Day 2	LR	0.76 (0.73–0.79)	0.55 (0.4–0.68)	0.79 (0.71–0.93)	0.69 (0.62–0.85)	0.68 (0.62–0.75)	0.68 (0.65–0.71)
OI	Day 2	RF	0.73 (0.66–0.78)	0.63 (0.52–0.72)	0.74 (0.64–0.81)	0.67 (0.59–0.74)	0.7 (0.64–0.76)	0.69 (0.63–0.73)
PF	Day 2	LR	0.72 (0.69–0.75)	0.61 (0–0.9)	0.7 (0.29–1)	0.66 (0.48–0.84)	0.69 (0.51–0.8)	0.65 (0.5–0.71)
PF	Day 2	RF	0.68 (0.61–0.73)	0.67 (0.54–0.75)	0.63 (0.51–0.73)	0.61 (0.54–0.68)	0.68 (0.6–0.74)	0.65 (0.58–0.69)
OSI	Day 2	LR	0.84 (0.79–0.88)	0.68 (0.44–0.9)	0.84 (0.71–0.99)	0.78 (0.66–0.97)	0.78 (0.67–0.91)	0.77 (0.72–0.81)
OSI	Day 2	RF	0.74 (0.66–0.8)	0.58 (0.42–0.73)	0.78 (0.68–0.85)	0.67 (0.56–0.74)	0.71 (0.62–0.79)	0.69 (0.61–0.76)
SF	Day 2	LR	0.77 (0.74–0.8)	0.66 (0.44–0.76)	0.76 (0.64–0.87)	0.57 (0.46–0.65)	0.83 (0.75–0.87)	0.73 (0.67–0.77)
SF	Day 2	RF	0.69 (0.63–0.74)	0.47 (0.35–0.61)	0.77 (0.66–0.84)	0.49 (0.39–0.58)	0.76 (0.71–0.8)	0.67 (0.61–0.72)
OSI (all SpO_2)	Day 2	LR	0.78 (0.75–0.8)	0.49 (0.41–0.57)	0.95 (0.89–0.98)	0.77 (0.62–0.9)	0.85 (0.82–0.87)	0.83 (0.81–0.86)
OSI (all SpO_2)	Day 2	RF	0.78 (0.75–0.81)	0.56 (0.44–0.67)	0.92 (0.87–0.95)	0.69 (0.59–0.8)	0.86 (0.83–0.9)	0.83 (0.79–0.86)
SF (all SpO_2)	Day 2	LR	0.79 (0.76–0.81)	0.53 (0.36–0.64)	0.87 (0.77–0.94)	0.58 (0.46–0.67)	0.85 (0.8–0.88)	0.78 (0.74–0.81)
SF (all SpO_2)	Day 2	RF	0.75 (0.71–0.8)	0.43 (0.31–0.57)	0.87 (0.79–0.93)	0.54 (0.42–0.66)	0.82 (0.79–0.86)	0.76 (0.72–0.8)
Multivariables	Day 2	LR	0.86 (0.83–0.88)	0.61 (0.52–0.7)	0.87 (0.81–0.92)	0.62 (0.52–0.71)	0.87 (0.84–0.9)	0.81 (0.77–0.83)
Multivariables	Day 2	RF	0.93 (0.89–0.95)	0.69 (0.59–0.79)	0.94 (0.89–0.97)	0.79 (0.68–0.88)	0.9 (0.87–0.93)	0.88 (0.85–0.9)
OI	Day 1+2	LR	0.81 (0.77–0.84)	0.66 (0.54–0.81)	0.77 (0.69–0.88)	0.74 (0.67–0.84)	0.71 (0.63–0.82)	0.72 (0.66–0.77)
OI	Day 1+2	RF	0.82 (0.75–0.86)	0.71 (0.54–0.84)	0.79 (0.6–0.89)	0.76 (0.62–0.87)	0.74 (0.64–0.84)	0.75 (0.65–0.83)
PF	Day 1+2	LR	0.73 (0.67–0.77)	0.79 (0.68–0.9)	0.52 (0.35–0.68)	0.63 (0.54–0.72)	0.71 (0.64–0.81)	0.66 (0.6–0.72)
PF	Day 1+2	RF	0.79 (0.73–0.85)	0.82 (0.64–0.96)	0.63 (0.49–0.77)	0.7 (0.61–0.77)	0.79 (0.64–0.95)	0.73 (0.65–0.8)
OSI	Day 1+2	LR	0.82 (0.74–0.87)	0.79 (0.55–1)	0.63 (0.27–0.91)	0.7 (0.54–0.88)	0.77 (0.61–1)	0.71 (0.6–0.77)
OSI	Day 1+2	RF	0.72 (0.61–0.8)	0.64 (0.44–0.83)	0.59 (0.41–0.83)	0.62 (0.51–0.77)	0.61 (0.49–0.76)	0.61 (0.52–0.71)
SF	Day 1+2	LR	0.82 (0.78–0.85)	0.73 (0.58–0.86)	0.74 (0.66–0.8)	0.59 (0.52–0.64)	0.85 (0.77–0.91)	0.74 (0.7–0.76)
SF	Day 1+2	RF	0.81 (0.76–0.85)	0.68 (0.47–0.86)	0.8 (0.7–0.9)	0.64 (0.54–0.75)	0.84 (0.75–0.92)	0.76 (0.71–0.81)
OSI (all SpO_2)	Day 1+2	LR	0.78 (0.75–0.8)	0.51 (0.43–0.61)	0.94 (0.88–0.98)	0.75 (0.62–0.87)	0.85 (0.83–0.88)	0.83 (0.81–0.86)
OSI (all SpO_2)	Day 1+2	RF	0.78 (0.75–0.81)	0.58 (0.46–0.7)	0.91 (0.87–0.95)	0.69 (0.59–0.8)	0.87 (0.83–0.9)	0.83 (0.79–0.86)

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

SF (all SpO_2)	Day 1+2	LR	0.8 (0.78–0.83)	0.62 (0.49–0.71)	0.82 (0.75–0.88)	0.53 (0.46–0.61)	0.87 (0.82–0.9)	0.77 (0.73–0.79)
SF (all SpO_2)	Day 1+2	RF	0.83 (0.78–0.87)	0.62 (0.48–0.76)	0.86 (0.79–0.92)	0.6 (0.51–0.71)	0.87 (0.83–0.91)	0.8 (0.76–0.83)
Multivariables	Day 1+2	LR	0.86 (0.83–0.89)	0.64 (0.52–0.75)	0.86 (0.81–0.91)	0.61 (0.53–0.69)	0.88 (0.84–0.91)	0.81 (0.77–0.83)
Multivariables	Day 1+2	RF	0.93 (0.9–0.95)	0.71 (0.6–0.82)	0.93 (0.88–0.97)	0.78 (0.67–0.88)	0.91 (0.87–0.94)	0.88 (0.85–0.91)
Multivariables (reduced)	Day 1+2	LR	0.84 (0.81–0.87)	0.59 (0.48–0.69)	0.9 (0.85–0.94)	0.68 (0.57–0.77)	0.87 (0.83–0.9)	0.83 (0.79–0.85)
Multivariables (reduced)	Day 1+2	RF	0.91 (0.88–0.94)	0.72 (0.61–0.83)	0.91 (0.86–0.96)	0.74 (0.64–0.85)	0.91 (0.87–0.94)	0.86 (0.83–0.89)
Multivariables (reduced+year)	Day 1+2	LR	0.84 (0.8–0.86)	0.6 (0.48–0.71)	0.9 (0.84–0.94)	0.67 (0.55–0.76)	0.87 (0.83–0.9)	0.82 (0.78–0.85)
Multivariables (reduced+year)	Day 1+2	RF	0.92 (0.88–0.94)	0.72 (0.61–0.84)	0.92 (0.87–0.97)	0.76 (0.66–0.87)	0.91 (0.87–0.94)	0.87 (0.84–0.9)

Supplemental Digital Content 1: Performance of all models with 95% confidence interval. Sensitivity, specificity, positive predictive value, negative predictive value and accuracy were calculated using probability cutoff of 0.5. OI: Oxygenation Index, OSI: Oxygenation Saturation Index, PF: PaO_2/FiO_2 ratio, SF: SpO_2/FiO_2 ratio



Supplemental Digital Content 2: Importance of predictors in random forest models using the error rate classification after permutation of the predictors. MAwP: mean airway pressure.

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

Using a backward elimination method, a simpler model was found to include only age, pH, SpO_2 and PCO_2 on day one, FiO_2, invasive ventilation and respiratory rate on day 2 and MAwP on both days. This model achieved a similar performance (AUROC 0.91, 95% CI 0.87 – 0.94, P = 0.4) than the more complex models.

Logistic Regression and Random Forests Comparison

Multivariable random forests models were superior (P < 0.001) to all logistic regression models. The best logistic regression model achieved an 0.86 AUROC (95% CI 0.83 – 0.88), but its calibration was low (P = 0.007).

Model Robustness to Time

In logistic regression, admission year was not a significant coefficient (P>0.6), either in univariate or multivariate analysis. In the random forests' models, the importance of the year was among the lowest values (**Supplemental Digital Content 2**), suggesting that this variable does not contain discriminant information. When the year was removed from the multivariate model, the discrimination (AUC) was similar (P>0.7) for both random forests and logistic regression.

DISCUSSION

Children who develop prolonged AHRF that is still present on or after PICU day 7 can be identified fairly accurately by the morning of PICU day 2 by applying machine learning to common respiratory variables collected clinically. These children have high morbidity and mortality and high use of ICU resources such as mechanical ventilation. Our final parsimonious model included age, pH, SpO_2 and PCO_2 on day one, FiO_2, invasive ventilation and respiratory rate on day two and MAwP on both days and it had an AUROC of 0.91. Although missing data due to rules related to use of oxygenation values (e.g. $SpO_2 > 97\%$) were frequent, it could be overcome with either simple rules or decrease in the SpO_2 thresholds. Machine learning using random forests outperformed logistic regression for all multivariable models.

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

AHRF is a severe condition that includes ARDS, among other diagnoses. ARDS has a stricter definition (6), but this may not be applicable to all patients, especially in pediatrics where the arterial samples are infrequent and high SpO_2 (>97%) tolerated limiting its use (**Figure 2**). The restriction on the SpO_2 to be ≤ 97% rely on its close relationship with oxyhemoglobin between 80% and 97% (37, 38). Due to inability to use SpO2 >97% and because availability of PaO2 is infrequent and limited to the most severe cases, the PALICC and the Berlin definition of hypoxemia have a limited applicability in clinical research practice in children. As shown in our study, none of the common hypoxemia markers were usable to the whole cohort to adequately classify patients. As other have suggested, it may be better to use the available SpO_2 even if >97% than censoring the observation (39) but how to do this requires further investigation. Our findings support that AHRF severity includes oxygenation and mean airway pressure as is in the PALICC but not the Berlin definition

In prior studies, data collected later after admission were slightly more discriminant for mortality than those collected at ICU presentation, although discrimination remained low (<0.70) (11, 17). Our results don't support a significant benefit of using only day 2 data instead of admission day data as other studies also found (12). However, our data support incorporating the temporal evolution of respiratory variables for optimal prognostication (**Figure 5**). Coupling modern machine learning algorithms that don't rely on specific assumption like traditional statistical models (40) to high temporal resolution databases that stores raw waveforms, future studies may be able to estimate to assess subtle changes in patients trajectories (41).

The presence of infiltrates was not predictive in either the machine learning or logistic regression models. This is likely because a high proportion of children without prolonged AHRF had infiltrates noted by PICU day 2 (72%) even though all children with prolonged AHRF had infiltrates by day 2 (**Table 1**, 99%, $P < 0.001$). Similarly, although the PRISM III score was found to initially be informative, model discrimination was relatively unchanged once it was removed from the model likely because pH, hypoxemia and age which were included in the final model are also part of the PRISM III score. We also showed that year was not an important predictor, suggesting that our model to predict severe influenza infection prediction is robust to time.

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

Death was uncommon and in the 11 patients that died who were analyzed (2 were censored due to early death from neurologic injury) it occurred three to four weeks after PICU admission (**Figure 3**) with prolonged ventilator support. Although this may offer a window of opportunity for new therapeutic strategies, early identification of the most severe patients may decrease death and hasten recovery. The PALICC definition does not specify how to categorize severity for non-invasive ventilation or ECMO. This is problematic in severe pediatric influenza infection, where ECMO use is frequent as seen in the literature (42) and in our cohort where nearly 45% of the prolonged AHRF cohort had ECMO support. As ARDS severity definitions have in the past been based solely on their association with hospital mortality, we believe that ECMO should be classified as severe AHRF or ARDS.

Our study has strengths. First, it is one of the largest pediatric cohorts with AHRF related to influenza virus infection. Our model had low bias with good generalizability and excellent discrimination (AUROC > 0.90). Although we found no direct comparison in the literature for sustained PARDS prediction, the AUROC for ARDS mortality in the literature ranges from 0.60 to 0.84 (8–13, 17, 43). Second, because many clinical variables are missing that limit the ability to diagnose PARDS, our study provides a clinically useful definition of AHRF that can be done using the available variables in clinical practice.

Our study also has limitations. First, we do not have an external validation cohort. This limitation is mitigated in part by the multicentric design and use of Monte Carlo cross validation methods. Second, bacterial co-infection is a known risk factor for more severe and sustained PARDS (21). However, we did not include co-infection in the predictive models, because it is a difficult diagnosis to make prospectively at day 2 (21, 44). Third, the outcome was treated as a binary variable, without incorporating mortality. Because we documented only two deaths (0.8%) before the seventh day, a composite outcome would probably have been influenced minimally by survival. Furthermore, patients with significant comorbidities that were associated with lung disease were excluded from this study, precluding the use of the model in this population. However, those patients were already known to be more at risk of prolonged AHRF (1).

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

CONCLUSION

In our observational prospective multicentric study, prolonged AHRF at one week for children with severe influenza is strongly associated with the initial respiratory severity and its evolution by day 2 and is associated with a significantly higher mortality, morbidity. Our model may help future trials to target the most severe group in the first 24h after admission and may guide PICU managers to anticipate resources that would be required in case of the emergence of a novel influenza virus. External validation is needed before it can be used at the bedside.

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

REFERENCES

1. Jouvet P, Hutchison J, Pinto R, et al.: Critical illness in children with influenza A/pH1N1 2009 infection in Canada*. *Pediatr Crit Care Med* 2010; 11:603–609
2. Randolph AG, Vaughn F, Sullivan R, et al.: Critically ill children during the 2009-2010 influenza pandemic in the United States. *Pediatrics* 2011; 128:e1450-8
3. Ampofo K, Gesteland PH, Bender J, et al.: Epidemiology, Complications, and Cost of Hospitalization in Children With Laboratory-Confirmed Influenza Infection. *Pediatrics* 2006; 118:2409–2417
4. Schrag SJ, Shay DK, Gershman K, et al.: Multistate surveillance for laboratory-confirmed, influenza-associated hospitalizations in children 2003-2004. *Pediatr Infect Dis J* 2006; 25:395–400
5. Kumar A, Zarychanski R, Pinto R, et al.: Critically ill patients with 2009 influenza A(H1N1) infection in Canada. *JAMA - J Am Med Assoc* 2009; 302:1872–1879
6. Khemani RG, Smith LS, Zimmerman JJ, et al.: Pediatric acute respiratory distress syndrome: definition, incidence, and epidemiology: proceedings from the Pediatric Acute Lung Injury Consensus Conference. *Pediatr Crit Care Med* 2015; 16:S23-40
7. Ranieri VM, Rubenfeld GD, Thompson BT, et al.: Acute respiratory distress syndrome: The Berlin definition. *JAMA - J Am Med Assoc* 2012; 307:2526–2533
8. Gajic O, Dabbagh O, Park PK, et al.: Early Identification of Patients at Risk of Acute Lung Injury. *Am J Respir Crit Care Med* 2011; 183:462–470
9. Villar J, Ambrós A, Soler JA, et al.: Age, Pao2/Fio2, and plateau pressure score: A proposal for a simple outcome score in patients with the acute respiratory distress syndrome. *Crit Care Med* 2016; 44:1361–1369
10. Chen W-L, Lin W-T, Kung S-C, et al.: The Value of Oxygenation Saturation Index in Predicting the Outcomes of Patients with Acute Respiratory Distress Syndrome. *J Clin Med* 2018; 7
11. Lai C-C, Sung M-I, Liu H-H, et al.: The Ratio of Partial Pressure Arterial Oxygen and Fraction of Inspired

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

Oxygen 1 Day After Acute Respiratory Distress Syndrome Onset Can Predict the Outcomes of Involving Patients. *Medicine (Baltimore)* 2016; 95:e3333

12. Spicer AC, Calfee CS, Zinter MS, et al.: A Simple and Robust Bedside Model for Mortality Risk in Pediatric Patients With Acute Respiratory Distress Syndrome. *Pediatr Crit Care Med* 2016; 17:907–916
13. Hu C-AA, Chen C-M, Fang Y-C, et al.: Using a machine learning approach to predict mortality in critically ill sectional influenza patients : a cross- - retrospective multicentre study in Taiwan. *BMJ Open* 2020; 10:1–10
14. Johnson AE, Ghassemi MM, Nemati S, et al.: Machine Learning and Decision Support in Critical Care. *Proc IEEE Inst Electr Electron Eng* 2016; 104:444–466
15. Zampieri FG, Costa EL, Iwashyna TJ, et al.: Heterogeneous effects of alveolar recruitment in acute respiratory distress syndrome: a machine learning reanalysis of the Alveolar Recruitment for Acute Respiratory Distress Syndrome Trial. *Br J Anaesth* 2019; 123:88–95
16. Goligher EC, Tomlinson G, Hajage D, et al.: Extracorporeal Membrane Oxygenation for Severe Acute Respiratory Distress Syndrome and Posterior Probability of Mortality Benefit in a Post Hoc Bayesian Analysis of a Randomized Clinical Trial. *JAMA - J Am Med Assoc* 2018; 320:2251–2259
17. Khemani RG, Smith L, Lopez-Fernandez YM, et al.: Paediatric acute respiratory distress syndrome incidence and epidemiology (PARDIE): an international, observational study. *Lancet Respir Med* 2018; 2600:1–14
18. Payen V, Jouvet P, Lacroix J, et al.: Risk factors associated with increased length of mechanical ventilation in children. *Pediatr Crit Care Med* 2012; 13:152–157
19. Seneff MG, Zimmerman JE, Knaus WA, et al.: Predicting the duration of mechanical ventilation: The importance of disease and patient characteristics. *Chest* 1996; 110:469–479
20. Hall MW, Geyer SM, Guo CY, et al.: Innate immune function and mortality in critically III children with influenza: A multicenter study. *Crit Care Med* 2013; 41:224–236
21. Randolph AG, Xu R, Novak T, et al.: Vancomycin monotherapy may be insufficient to treat methicillin-resistant staphylococcus aureus coinfection in children with influenza-related critical illness. *Clin Infect Dis* 2019; 68:365–

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

372

22. Fleming S, Thompson M, Stevens R, et al.: Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies. *Lancet* 2011; 377:1011–1018
23. Labarère J, Bertrand R, Fine MJ, et al.: How to derive and validate clinical prediction models for use in intensive care medicine. *Intensive Care Med* 2014; 40:513–527
24. Leteurtre S, Duhamel A, Salleron J, et al.: PELOD-2: An update of the PEdiatric logistic organ dysfunction score. *Crit Care Med* 2013; 41:1761–1773
25. 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
26. Pollack MM, Holubkov R, Funai T, et al.: The Pediatric Risk of Mortality Score: Update 2015. *Pediatr Crit Care Med* 2016; 17:2–9
27. Robin X, Turck N, Hainard A, et al.: pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12:77
28. Brunson JC: ggalluvial: Alluvial Plots in "ggplot2." 2019;
29. Zhang P: Model Selection Via Multifold Cross Validation Author (s): Ping Zhang Source : The Annals of Statistics , Vol . 21 , No . 1 (Mar ., 1993), pp . 299-313 Published by : Institute of Mathematical Statistics Stable URL : http://www.jstor.org/stable/303559. *Ann Stat* 1993; 21:299–313
30. Hosmer DW, Hosmer T, Le Cessie S, et al.: A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997; 16:965–980
31. Leisman DE, Harhay MO, Lederer DJ, et al.: Development and Reporting of Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and Critical Care Journals. *Crit Care Med* 2020; 1–11
32. Collins GS, Reitsma JB, Altman DG, et al.: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015; 162:55
33. Breiman L: Random Forrests. *Mach Learn* 2001; 45:5–32

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

34. Fernández-Delgado M, Cernadas E, Barro S, et al.: Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res* 2014; 15:3133–3181
35. Liaw A, Wiener M: Classification and Regression by randomForest. *R news* 2002; 2:18–22
36. Pollack MM, Patel KM, Ruttimann UE: PRISM III: An updated pediatric risk of mortality score. *Crit Care Med* 1996; 24:743–752
37. Khemani RG, Rubin S, Belani S, et al.: Pulse oximetry vs. PaO2 metrics in mechanically ventilated children: Berlin definition of ARDS and mortality risk. *Intensive Care Med* 2015; 41:94–102
38. Khemani RG, Thomas NJ, Venkatachalam V, et al.: Comparison of SpO 2 to PaO 2 based markers of lung disease severity for children with acute lung injury. *Crit Care Med* 2012; 40:1309–1316
39. Slater A, Straney L, Alexander J, et al.: The Effect of Imputation of PaO2/FIO2 From SpO2/FIO2 on the Performance of the Pediatric Index of Mortality 3. *Pediatr Crit Care Med* 2020; 1
40. Hyland SL, Faltys M, Hüser M, et al.: Machine learning for early prediction of circulatory failure in the intensive care unit. *Nat Med* 2019; 26
41. Brossier D, Sauthier M, Mathieu A, et al.: Qualitative subjective assessment of a high-resolution database in a paediatric intensive care unit—Elaborating the perpetual patient's ID card. *J Eval Clin Pract* 2019; 1–6
42. Zangrillo A, Biondi-Zoccai G, Landoni G, et al.: Extracorporeal membrane oxygenation (ECMO) in patients with H1N1 influenza infection: a systematic review and meta-analysis including 8 studies and 266 patients receiving ECMO. *Crit Care* 2013; 17:R30
43. Santos RS, Silva PL, Rocco JR, et al.: A mortality score for acute respiratory distress syndrome: Predicting the future without a crystal ball. *J Thorac Dis* 2016; 8:1872–1876
44. Chomton M, Brossier D, Sauthier M, et al.: Ventilator-associated pneumonia and events in pediatric intensive care: A single center study. *Pediatr Crit Care Med* 2018; 19:1106–1113

Version non définitive. La version définitive de cet article (Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. *Crit Care Explor*. 2020;2(8):e0175) est disponible au https://doi.org/10.1097/cce.0000000000000175.

Discussion

Dans ce travail, nous avons développé et validé un modèle prédictif de la persistance de l'IRHA dans une population d'enfants atteints d'une infection grave à l'influenza, permettant de valider un critère de jugement de gravité différent de la mortalité. Le modèle s'est avéré très performant dans le domaine (aire sous la courbe ROC [AUROC] 0,93) où généralement la discrimination est < 0,85 (89). L'utilisation de modèles basés sur l'apprentissage automatique (forêts aléatoires) a permis d'améliorer les performances du modèle par rapport à une approche conventionnelle (régression logistique).

Cette étude a aussi permis d'améliorer notre compréhension des facteurs de risques associés à une IRHA prolongée. Contrairement à la régression où les coefficients répondent directement à cette question, les autres modèles d'apprentissage automatique doivent utiliser des stratégies comme la permutation ou le masquage pour déterminer la contribution de chaque variable (90). Nous avons ainsi montré que les données les plus importantes étaient la FiO_2, la fréquence respiratoire et le pH mesurés au matin du deuxième jour (<24 h après l'admission). Ces trouvailles sont congruentes avec d'autres indices décrits dans la littérature principalement adulte, comme l'indice ROX (91) ou EWS.O2 (92), qui tiennent compte de la fréquence respiratoire du patient en plus de l'hypoxémie. De plus, cette étude a aussi montré que l'application stricte des critères diagnostiques de l'IRHA, notamment l'utilisation de $SpO_2 \leq 97$ %, génère énormément de données manquantes et sélectionne de façon biaisée une sous-population. Dans cet article nous avons montré plusieurs stratégies d'imputation pour pallier ces limitations. Ces résultats sont congruents avec notre travail sur l'estimation de la PaO_2 (article 1) et le travail d'autres chercheurs qui suggère qu'il vaut mieux utiliser une $SpO_2 > 97$ % malgré la diminution de la précision que de la considérer manquante (93).

Nous avons aussi montré le bénéfice de la temporalité sur les marqueurs hypoxémiques (OI, PF et SF). En effet, l'utilisation combinée des jours 1 et 2 était supérieure à l'utilisation isolée des jours 1 ou 2. Alors que la totalité des scores de gravité en soins intensifs se contente de ne retenir que la valeur la plus grave dans une certaine période (86,94–96), cette étude suggère donc un bénéfice à l'intégration de données temporelles continues dans un SADC dédié à l'IRHA.

Article 4 – Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria

Préface

La ventilation mécanique est un traitement très fréquent de l'IRHA et les formes graves d'IRHA sont associées à un recours à la ventilation mécanique prolongée (97). Son utilisation dite « prolongée » est un concept très subjectif en fonction de son contexte (postopératoire ou centre de soins de longue durée) et aucune définition pédiatrique standardisée n'existait avant ce travail. Par conséquent, il est difficile d'étudier et d'améliorer un phénomène dans ce contexte. Outre les décès, les IRHA les plus graves seront celles qui nécessiteront une ventilation prolongée. En prévision du développement d'un modèle prédictif dans la population pédiatrique la plus fragile, les nouveau-nés, nous avons revu la littérature dans le domaine pédiatrique et proposé une définition robuste aux différents types de ventilation et applicables, quel que soit l'âge ou le terme de l'enfant. Il a été réalisé en collaboration avec 2 autres chercheurs experts dans le domaine. J'ai réalisé plus de 80 % du travail ayant mené à cette publication, soit la planification de l'étude, la revue de la littérature (en collaboration), l'analyse des données et la rédaction du manuscrit. L'étude a été publiée dans le journal *Respiratory Care* en 2017 et avait été citée par plus de 25 publications en 2021.

Article

Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria

Michaël Sauthier MD, Louise Rose RN PhD, and Philippe Jouvet MD PhD

BACKGROUND: A 2005 consensus conference led by the National Association for Medical Direction of Respiratory Care (NAMDRC) defined prolonged mechanical ventilation (PMV) for adults as invasive and/or noninvasive mechanical ventilation (NIV) for ≥ 21 consecutive days for ≥ 6 h/d. In children, no such consensus definition exists. This results in substantial variability in definitional criteria, making study of the impact and outcomes of PMV across and within settings problematic. The objective of this work was to identify how PMV for children and neonates is described in the literature and to outline pediatric/neonatal considerations related to PMV, with the goal of proposing a pediatric/neonatal adaptation to the NAMDRC definition. METHODS: We searched electronic databases for studies describing PMV in children. We extracted definitional criteria and developed recommendations based on the literature review and our clinical experience. RESULTS: Of the 416 citations obtained, 87 met inclusion criteria, totaling 34,255 subjects. Identified criteria for the pediatric PMV definition included: number of consecutive days of mechanical ventilation (ranging from 6 h to 3 months), inclusion of NIV, time spent off the ventilator during weaning (considered as same ventilation episode), and importance of chronological age (term neonates) and postmenstrual age for preterm neonates. We considered high-flow nasal cannula; however, we determined that its current role as a weaning adjunct is unclear. CONCLUSIONS: Therefore, we developed the following recommendations for the pediatric PMV definition: ≥ 21 consecutive days (after 37 weeks postmenstrual age) of ventilation for ≥ 6 h/d considering invasive ventilation and NIV and including short interruptions (< 48 h) of ventilation during the weaning process as the same episode of ventilation. We propose a definition of pediatric PMV that incorporates the number of consecutive days of mechanical ventilation while taking into account use of NIV and lung maturity and including short interruptions during the weaning process. *Key words: mechanical ventilation; prolonged mechanical ventilation; neonates; children; intensive care; critical care.* [Respir Care 2017;62(1):49–53. © 2017 Daedalus Enterprises]

Introduction

Mechanical ventilation is a common treatment in intensive care, whether for neonates, children, or adults. Independent of the underlying disease, this supportive treatment is associated with many complications that may prolong its duration, such as ventilator-associated lung in-

Drs Sauthier and Jouvet are affiliated with the Research Center of Sainte-Justine Hospital and the Pediatric ICU, Sainte-Justine Hospital, Montreal, Québec, Canada. Dr Rose is affiliated with the Department of Critical Care, Sunnybrook Health Sciences Centre, the Sunnybrook Research Institute, the Lawrence S Bloomberg Faculty of Nursing and Interdepartmental Division of Critical Care Medicine, University of Toronto, the Provincial Centre of Weaning Excellence/Prolonged Ventilation Weaning Centre, Toronto East General Hospital, and the Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Ontario, Canada.

Dr Rose has disclosed a relationship with the Canadian Institutes of Health. Dr Jouvet has disclosed a relationship with the Health Research Agency of Quebec (FRQS). Dr Sauthier has disclosed no conflicts of interest.

Supplementary material related to this paper is available at http://www.rcjournal.com.

Correspondence: Philippe Jouvet MD PhD, Soins Intensifs Pédiatriques, Hôpital Sainte-Justine, 3175 Chemin Côte Sainte-Catherine, Montréal, Québec H3T 1C5, Canada. E-mail: philippe.jouvet@umontreal.ca.

DOI: 10.4187/respcare.04881

Article 4 - *Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria*

Background: A 2005 consensus conference led by the National Association for Medical Direction of Respiratory Care (NAMDRC) defined prolonged mechanical ventilation (PMV) for adults as invasive and/or non-invasive mechanical ventilation (NIV) for 21 consecutive days or more for at least 6 hours per day. In children, no such consensus definition exists. This results in substantial variability in definitional criteria making study of the impact and outcomes of PMV across and within settings problematic.

Objective: To identify how PMV for children and neonates is described in the literature and to outline paediatric/neonatal considerations related to PMV, with the goal of proposing a paediatric/neonatal adaptation to the NAMDRC definition.

Methods: We searched electronic databases for studies describing PMV in children. We extracted definitional criteria and developed recommendations based on the literature review and our clinical experience.

Results: Of the 416 citations obtained, 87 met inclusion criteria totalling 34255 patients. Identified criteria for the paediatric PMV definition included: number of consecutive days of mechanical ventilation (ranging from 6 hours to 3 months), inclusion of NIV, time spent off the ventilator during weaning (considered as same ventilation episode), and importance of chronological age (term neonates) and postmenstrual age (PMA) for preterm neonates. We considered high flow nasal cannula, however determined its current role as a weaning adjunct as unclear. Therefore, we developed the following recommendations for the paediatric PMV definition: 21 consecutive days or more (after 37 weeks PMA) of ventilation for 6 hours or more each day considering invasive and NIV and including short interruptions (<48h) of ventilation during the weaning process as the same episode of ventilation.

Conclusions: We propose a definition of paediatric PMV that incorporates the number of consecutive days of mechanical ventilation while taking into account use of NIV, lung maturity, and including short interruptions during the weaning process.

Version non définitive. La version définitive de cet article (Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. *Respir Care*. 2017 Jan 1;62(1):49–53) est disponible au http://rc.rcjournal.com/cgi/doi/10.4187/respcare.04881

Introduction

Mechanical ventilation is a common treatment in intensive care, whether for neonates, children or adults. Independent of the underlying disease, this supportive treatment is associated with many complications that may prolong its duration such as ventilator-associated lung injury and pneumonia. Healthcare costs associated with intensive care are projected to increase as a result of more patients requiring prolonged mechanical ventilation (PMV).[1] In 2005, a consensus conference led by the National Association for Medical Direction of Respiratory Care (NAMDRC) defined PMV for adults as mechanical ventilation for 21 consecutive days or more for at least 6 hours per day of invasive (via endotracheal tube or tracheostomy) and/or non-invasive (facial/nasal interface) methods of delivery.[2] In children, ten years after the published NAMDRC definition, no such consensus definition exists making it difficult to interpret study results describing a PMV population and to determine the impact of PMV across and within settings.

The absence of a paediatric PMV definition has resulted in: *1)* substantial variability in duration of ventilation described as PMV in the published literature with duration generally ranging from 2-7 days[3-5] to 21-28 days.[6-9] These two groups may be derived from the perspective of pediatric intensivists who want to differentiate between patients that are able to be extubated quickly versus more longer term ICU patients, and pediatric pulmonologists who may be making decision about home mechanical ventilation support; *2)* lack of standardization on the inclusion of non-invasive ventilation (NIV) in the definition; *3)* lack of standardization on the inclusion of times when the child is ventilator free during weaning (i.e., should these contribute to the number of days of consecutive ventilation?); and *4)* no specific considerations for neonates in relation to when to commence counting the number of consecutive days that define PMV (i.e., at which age, chronological or postmenstrual).[7-10] Therefore, there is an urgent need to standardize the definition of PMV in children to: *1)* describe the epidemiology of PMV; *2)* assist clinicians with reprioritizing goals of care and the type information to share with patients and their families;[10] and *3)* identify effective interventions that will reduce the number of children requiring PMV. In this perspective, we propose an operational definition of paediatric PMV based on relevant literature and the authors' clinical experience.

Version non définitive. La version définitive de cet article (Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. *Respir Care*. 2017 Jan 1;62(1):49–53) est disponible au http://rc.rcjournal.com/cgi/doi/10.4187/respcare.04881

Methods:

In February 2016, we searched the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE (2000 to February 2016) and EMBASE (2000 to February 2016) using the following search terms: prolonged, protracted, chronic, sustained, increased length or long term mechanical ventilation. Criteria were determined *a priori* and studies were eligible for inclusion if they included mechanically ventilated (invasive and NIV) children or neonates, including mixed adult and paediatric cohorts. Exclusion criteria included: case reports, case series of fewer than 10 patients, commentary or editorials, and non-peer reviewed articles. Articles written in a language other than English were not included. We extracted the definitional criteria used to identify the study cohort as PMV and any rationale reported for the definitional criteria used.

Results and discussion

Of the 416 citations obtained, 87 met inclusion criteria (appendix 1) totalling 34'255 patients. Criteria used to define PMV were heterogeneous. Most of the articles (77%) defined PMV by a time criterion i.e. the duration of mechanical ventilation. However, there was no consensus as to the duration, with ranges varying from 6 hours to 3 months (**Figure 1**), with shorter duration (≥ 4 days) for surgical as opposed to medical patients (≥ 7 or 21 days, with 21 days used by the most recent studies). Other studies defined PMV as the presence of a tracheostomy, being ventilated at home, or in a care centre dedicated to chronic ventilation.

Version non définitive. La version définitive de cet article (Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. *Respir Care*. 2017 Jan 1;62(1):49–53) est disponible au http://rc.rcjournal.com/cgi/doi/10.4187/respcare.04881

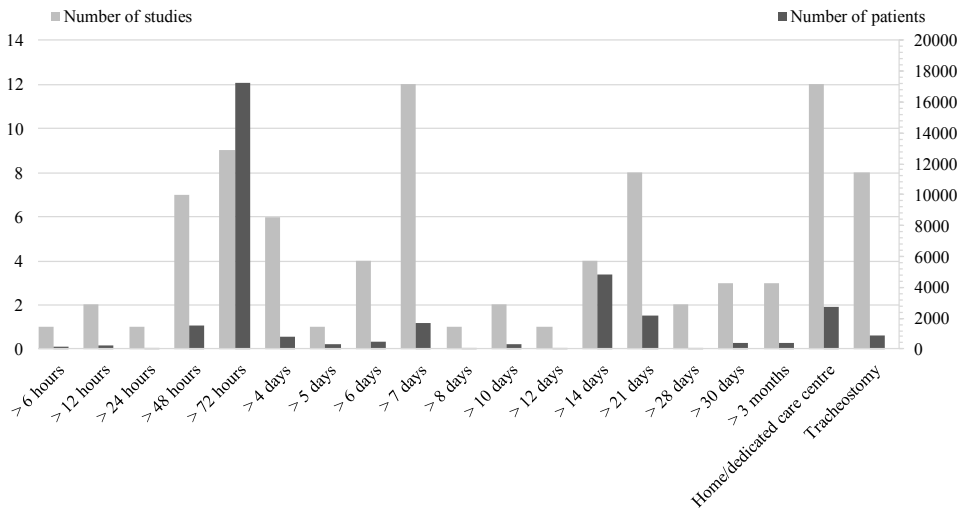


Figure 1: Duration of mechanical ventilation used to describe PMV

What duration of ventilation should be used for PMV in children?

In the NAMDRC PMV definition,[2] 21 consecutive days of mechanical ventilation was chosen for the adult population. Moreover, 21 days of ventilatory support has been used to define PMV in recent paediatric literature[6, 7] and may correspond to a turning point in the goals of care and type of weaning strategies to be used. Another advantage of this definition is that it offers objectivity, uniformity and simplicity in terms of identifying the patient cohort.[11] It corresponds to 2.5% – 3% proportion of the population admitted to PICUs[4, 7] and represents a cohort of ventilated children that need rehabilitation strategies[12] to promote weaning success and generally have a tracheotomy in place if they require invasive ventilation.[13, 14]

Version non définitive. La version définitive de cet article (Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. *Respir Care*. 2017 Jan 1;62(1):49–53) est disponible au http://rc.rcjournal.com/cgi/doi/10.4187/respcare.04881

How should short episodes off the ventilator during weaning affect establishment of PMV as defined as 21 days of consecutive ventilation?

In our systematic review, few studies[4, 7] discussed the time off the ventilator during weaning. The NAMDRC PMV definition recommends this term be applied to patients receiving mechanical ventilation for 21 *consecutive* days or more for at least 6 hours per day. However, if clinically indicated during these 21 days, weaning trials may occur with patients removed from any form of ventilatory support for up to 48 hours prior to the establishment of weaning success.[2] Patients that fail weaning will be returned to ventilatory support, however this means there has been an interruption in the number of consecutive days of mechanical ventilation. Therefore, we propose that in the case of weaning failure (defined as the need to recommence ventilation within 48 hours before 21 days of ventilation), the days off ventilatory support should be included in the number of consecutive days of mechanical ventilation used to define paediatric PMV (see **figure 2** panel C). Infants with lung fragility (e.g. bronchopulmonary dysplasia) may be at risk of more than one PMV episode in their life. For simplicity and uniformity reasons, we argue that the duration on and off the ventilator in the paediatric PMV definition should not be modulated by patient history and comorbidities.

Version non définitive. La version définitive de cet article (Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. *Respir Care*. 2017 Jan 1;62(1):49–53) est disponible au http://rc.rcjournal.com/cgi/doi/10.4187/respcare.04881

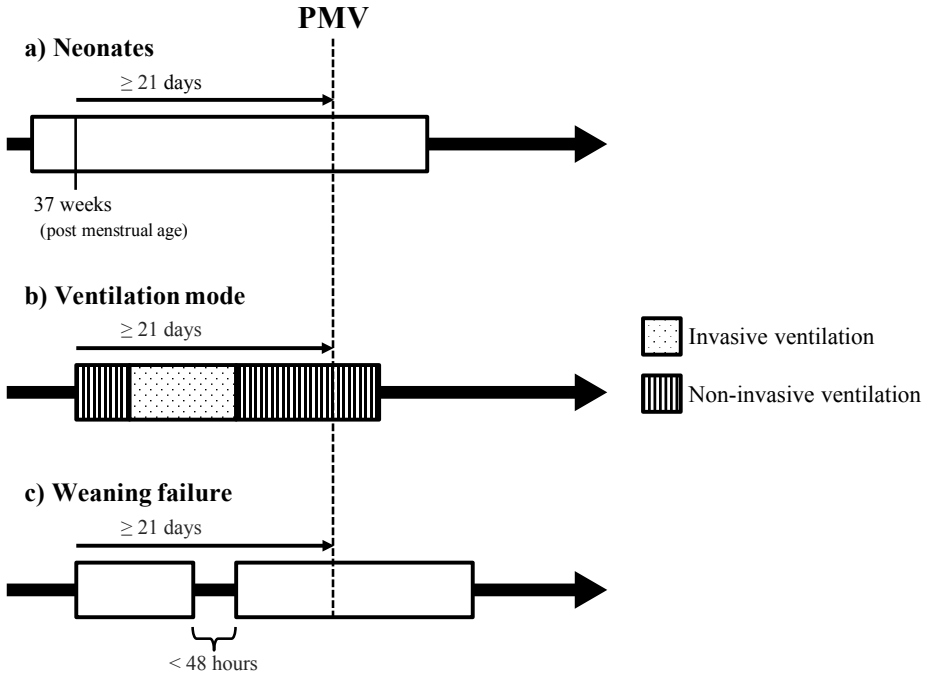


Figure 2: Schematic representation of the proposed definition of prolonged mechanical ventilation (PMV) in neonates and children. Panel a: Definition of PMV in neonates older than 37 weeks post menstrual age. Panel b: Different ventilation modes included in the PMV definition in children. Panel c: Weaning failure less than 48 hours before 21 days counted as the same episode of respiratory support.

Should high flow nasal cannulae be included in the number of consecutive days of ventilatory support to establish paediatric PMV?

The NAMDRC definition of adult PMV recommended NIV be included in the number of consecutive days of mechanical ventilation which we support for inclusion of paediatric PMV. However, the role of high-flow nasal cannulae (HFNC) was not discussed and at that time, little data were available on the physiological benefits of HFNC. At present, although there are limited and equivocal data suggesting HFNC generates positive end-expiratory pressure (PEEP) in adults, the evidence seems stronger in the paediatric population. In children, HFNC has been demonstrated to generate PEEP between 4 and 6 cmH_2O with

Version non définitive. La version définitive de cet article (Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. *Respir Care*. 2017 Jan 1;62(1):49–53) est disponible au http://rc.rcjournal.com/cgi/doi/10.4187/respcare.04881

gas flows higher than 1.7 L/min/Kg.[15-18] In adults, HFNC generates a more unpredictable and less marked increase of PEEP that ranges from 1.5 to 5 cmH_2O.[19, 20] HFNC decreases work of breathing in all age groups, whether it is measured by the electrical activity of the diaphragm,[16] pressure rate product,[21] clinical scores or vital signs,[15] even when no respiratory failure is present.[16] The risk of intubation with HFNC is similar to NIV and significantly lower than using standard nasal cannula for neonates.[22] Despite all of these data, inclusion of the HFNC in the paediatric PMV definition is still a source of debate and we advocate that this issue be discussed in a consensus conference.

When defining paediatric PMV do we use the chronological age or the corrected age for neonates?

Preterm birth is defined by the World Health Organization as neonates born alive before 37 weeks of pregnancy. In neonates, age can be counted according to chronological age (age from birth) or using the corrected age defined as the chronological age reduced by the number of weeks born before 40 weeks of gestation; the term should be used only for children up to 3 years of age who were born preterm.[23] Corrected age takes into account the maturation stage of the newborn that includes neurological, cardiovascular, and respiratory development and is considered important as a defining feature of other definitions. For example, bronchopulmonary dysplasia (BPD) is diagnosed by the ongoing need of supplemental oxygen at different time points depending on the postmenstrual age (PMA; GA plus chronological age) and GA.[24] Neonates born before 32 weeks GA may be diagnosed with BPD at 36 weeks PMA; those born at, or after, 32 weeks GA may be diagnosed using the chronological age from 4 to 8 weeks of life. We propose the definition of paediatric PMV uses chronological age when the cohort is homogeneous in terms of GA (e.g. preterm neonates of 28 weeks GA). However, if a cohort includes premature and term neonates, the usual case in PICUs, the corrected age should be used to standardize the maturation stage of neonates. In such a heterogeneous cohort, the duration of ventilation before 37 weeks GA would not count towards the 21 consecutive days of ventilation. This means the definition of paediatric PMV could not be applied before 40 weeks GA (see figure 2 panel A). To illustrate this operational definition of paediatric PMV, here are three cases:

Version non définitive. La version définitive de cet article (Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. *Respir Care*. 2017 Jan 1;62(1):49–53) est disponible au http://rc.rcjournal.com/cgi/doi/10.4187/respcare.04881

a) (**Fig 2**, panel a). A 5-month-old girl who was born at 24 weeks (1-month-old corrected age or 44 weeks PMA) is transferred from the neonatal unit to the PICU. She was intubated from birth to 38 weeks PMA and was successfully weaned from invasive ventilation to NIV after a second hydrocortisone treatment at 40 weeks PMA. She is still on NIV 22 hours on 24 (Total of 5 months of mechanical ventilation; 7 weeks (44 minus 37) if corrected age used).

b) (**Fig 2**, panel b). A 7-year-old boy was admitted to the ICU for pneumonia and required NIV support. On the second day, he was intubated for severe oxygenation failure. After 7 days of invasive mechanical ventilation, he was extubated. Two hours later, NIV was commenced due to increased work of breathing. He was subsequently reintubated because of severe CO_2 retention. Ten days later, he was extubated to a NIV and was successfully weaned from mechanical ventilation after 4 days (Total 23 days of mechanical ventilation).

c) (**Fig 2**, panel c). A 10-year-old boy who was intubated for a Guillain-Barré syndrome was extubated on the sixth day of mechanical ventilation. Thirty-six hours later, he was unable to protect his airway and required reintubated. Two weeks later, he was successfully extubated (Total 22 days of mechanical ventilation).

Summary

We have highlighted important issues for consideration in the establishment of a paediatric PMV definition, namely determination of 21 consecutive days or more of mechanical ventilation (after 37 weeks PMA) of ventilation for 6 hours or more each day considering invasive and NIV and including short interruptions (<48h) of ventilation during the weaning process as the same episode of ventilation (Table 1). To gain consensus and subsequent adoption of this definition, we plan to host a consensus conference with international experts representing paediatric intensive care, neonatal intensive care and the paediatric pulmonology community. We will refine this definition by discussing the relevance to different patient populations (e.g. chronic lung disease, neuromuscular disease, congenital heart defect), and debating the role of weaning adjuncts such as HFNC to come to a final consensus.

Version non définitive. La version définitive de cet article (Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. *Respir Care*. 2017 Jan 1;62(1):49–53) est disponible au http://rc.rcjournal.com/cgi/doi/10.4187/respcare.04881

References

1. Zilberberg MD, Shorr AF. Prolonged acute mechanical ventilation and hospital bed utilization in 2020 in the United States: implications for budgets, plant and personnel planning. BMC Health Serv Res 2008;8:242.

2. MacIntyre NR, Epstein SK, Carson S, Scheinhorn D, Christopher K, Muldoon S, et al. Management of patients requiring prolonged mechanical ventilation: report of a NAMDRC consensus conference. Chest 2005;128(6):3937-3954.

3. Wakeham MK, Kuhn EM, Lee KJ, McCrory MC, Scanlon MC. Use of tracheostomy in the PICU among patients requiring prolonged mechanical ventilation. Intensive Care Med 2014;40(6):863-870.

4. Payen V, Jouvet P, Lacroix J, Ducruet T, Gauvin F. Risk factors associated with increased length of mechanical ventilation in children. Pediatr Crit Care Med 2012;13(2):152-157.

5. Polito A, Patorno E, Costello JM, Salvin JW, Emani SM, Rajagopal S, et al. Perioperative factors associated with prolonged mechanical ventilation after complex congenital heart surgery. Pediatr Crit Care Med 2011;12(3):e122-126.

6. Monteverde E, Fernandez A, Poterala R, Vidal N, Siaba Serrate A, Castelani P, et al. Characterization of pediatric patients receiving prolonged mechanical ventilation. Pediatr Crit Care Med 2011;12(6):e287-291.

7. Traiber C, Piva JP, Fritsher CC, Garcia PC, Lago PM, Trotta EA, et al. Profile and consequences of children requiring prolonged mechanical ventilation in three Brazilian pediatric intensive care units. Pediatr Crit Care Med 2009;10(3):375-380.

8. Thomas M, Greenough A, Morton M. Prolonged ventilation and intact survival in very low birth weight infants. Eur J Pediatr 2003;162(2):65-67.

9. Fraser J, Henrichsen T, Mok Q, Tasker RC. Prolonged mechanical ventilation as a consequence of acute illness. Arch Dis Child 1998;78(3):253-256.

Version non définitive. La version définitive de cet article (Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. *Respir Care*. 2017 Jan 1;62(1):49–53) est disponible au http://rc.rcjournal.com/cgi/doi/10.4187/respcare.04881

10. Rose L, Fowler RA, Fan E, Fraser I, Leasa D, Mawdsley C, et al. Prolonged mechanical ventilation in Canadian intensive care units: A national survey. J Crit Care 2015;30(1):25-31.

11. Rose L, Fowler RA, Goldstein R, Katz S, Leasa D, Pedersen C, et al. Patient transitions relevant to individuals requiring ongoing ventilatory assistance: A Delphi study. Can Respir J 2014;21(5):287-292.

12. Jubran A, Grant BJ, Duffner LA, Collins EG, Lanuza DM, Hoffman LA, et al. Effect of pressure support vs unassisted breathing through a tracheostomy collar on weaning duration in patients requiring prolonged mechanical ventilation: a randomized trial. JAMA 2013;309(7):671-677.

13. Overman AE, Liu M, Kurachek SC, Shreve MR, Maynard RC, Mammel MC, et al. Tracheostomy for infants requiring prolonged mechanical ventilation: 10 years' experience. Pediatrics 2013;131(5):e1491-1496.

14. Holloway AJ, Spaeder MC, Basu S. Association of timing of tracheostomy on clinical outcomes in PICU patients. Pediatr Crit Care Med 2015;16(3):e52-58.

15. Spentzas T, Minarik M, Patters AB, Vinson B, Stidham G. Children with respiratory distress treated with high-flow nasal cannula. J Intensive Care Med 2009;24(5):323-328.

16. Pham TM, O'Malley L, Mayfield S, Martin S, Schibler A. The effect of high flow nasal cannula therapy on the work of breathing in infants with bronchiolitis. Pediatr Pulmonol 2015;50(7):713-720.

17. Milesi C, Baleine J, Matecki S, Durand S, Combes C, Novais AR, et al. Is treatment with a high flow nasal cannula effective in acute viral bronchiolitis? A physiologic study. Intensive Care Med 2013;39(6):1088-1094.

18. Hough JL, Pham TM, Schibler A. Physiologic effect of high-flow nasal cannula in infants with bronchiolitis. Pediatr Crit Care Med 2014;15(5):e214-219.

19. Sotello D, Rivas M, Mulkey Z, Nugent K. High-flow nasal cannula oxygen in adult patients: a narrative review. Am J Med Sci 2015;349(2):179-185.

Version non définitive. La version définitive de cet article (Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. *Respir Care*. 2017 Jan 1;62(1):49–53) est disponible au http://rc.rcjournal.com/cgi/doi/10.4187/respcare.04881

20. Sztrymf B, Messika J, Bertrand F, Hurel D, Leon R, Dreyfuss D, et al. Beneficial effects of humidified high flow nasal oxygen in critical care patients: a prospective pilot study. Intensive Care Med 2011;37(11):1780-1786.

21. Rubin S, Ghuman A, Deakers T, Khemani R, Ross P, Newth CJ. Effort of breathing in children receiving high-flow nasal cannula. Pediatr Crit Care Med 2014;15(1):1-6.

22. Lee JH, Rehder KJ, Williford L, Cheifetz IM, Turner DA. Use of high flow nasal cannula in critically ill infants, children, and adults: a critical review of the literature. Intensive Care Med 2013;39(2):247-257.

23. Engle WA, American Academy of Pediatrics Committee on F, Newborn. Age terminology during the perinatal period. Pediatrics 2004;114(5):1362-1364.

24. Jobe AH, Bancalari E. Bronchopulmonary dysplasia. Am J Respir Crit Care Med 2001;163(7):1723-1729.

Version non définitive. La version définitive de cet article (Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. *Respir Care*. 2017 Jan 1;62(1):49–53) est disponible au http://rc.rcjournal.com/cgi/doi/10.4187/respcare.04881

Discussion

Dans cet article, nous montrons que la ventilation prolongée en pédiatrie est un terme qui n'avait pas de définition claire et uniforme, menant à une très grande variabilité dans les études allant de 6 h à plusieurs mois de durée de ventilation. Ce travail est important, car il propose une définition standardisée, facilement applicable, moderne et basée sur les données probantes disponibles pour une problématique qui est probablement en croissance (98,99). Cette définition permettra d'établir des outils de pronostication pour l'IRHA et pourra être utilisée comme critère de jugement dans des études interventionnelles visant à montrer l'efficacité d'un SADC pour la reconnaissance et la prise en charge de l'IRHA.

Toutefois, cette définition devra faire l'objet d'une validation à plus large échelle puis d'une conférence de consensus avant de prétendre être la définition officielle en pédiatrie. L'étude observationnelle multicentrique LongVentKids (263 centres internationaux), en cours de réalisation, a été initiée par notre groupe de recherche à la suite de mon travail (https://longventkids.ca/) et validera entre autres cette définition et les conséquences de la ventilation prolongée chez l'enfant.

Article 5 – Long-Term Mechanical Ventilation in Neonates: A 10-Year Overview and Predictive Model

Préface

Ce travail est la suite logique de l'article présenté précédemment. Il vise à valider la définition de la ventilation prolongée en pédiatrie chez une population particulièrement vulnérable : les nouveau-nés, dont le manque de données dans la littérature est également très important. Cette population a la particularité d'être hétérogène en matière de prématurité et de malformations congénitales, ce qui rend l'utilisation d'une définition uniciste complexe. Dans ce travail, un volet descriptif présente le devenir à 18 mois d'âge corrigé sur les plans respiratoire, neurologique et digestif. Un deuxième volet présente un modèle prédictif pour identifier précocement le sous-groupe de patients ventilés le plus longtemps. Dans cet article réalisé en collaboration avec 6 autres chercheurs, j'ai réalisé plus de 85 % du travail, soit la conception de l'étude, la collecte de données (en collaboration), la supervision d'une étudiante (KBG), l'analyse des données, le développement et la validation des modèles et la rédaction du manuscrit. L'étude a été publiée dans le journal *Frontiers in Pediatrics* en 2021.

Article

frontiers in Pediatrics

ORIGINAL RESEARCH
published: 13 July 2021
doi: 10.3389/fped.2021.689190

Long-Term Mechanical Ventilation in Neonates: A 10-Year Overview and Predictive Model

Michaël Sauthier[1,2]*, *Nicolas Sauthier*[3], *Krystale Bergeron Gallant*[1], *Gregory A. Lodygensky*[1,2], *Atsushi Kawaguchi*[1,4], *Guillaume Emeriaud*[1,2] *and Philippe Jouvet*[1,2]

[1] Research Center of Sainte-Justine Hospital, Centre Hospitalier Universitaire Sainte-Justine, Université de Montréal, Montréal, QC, Canada, [2] Department of Pediatrics, Centre Hospitalier Universitaire Sainte-Justine, Université de Montréal, Montréal, QC, Canada, [3] Department of Anesthesia, Centre Hospitalier de l'Université de Montréal, Université de Montréal, Montréal, QC, Canada, [4] Department of Intensive Care Medicine, Pediatric Critical Care Medicine, Tokyo Women's Medical University, Tokyo, Japan

OPEN ACCESS

Edited by:
Courtney Rowan,
Indiana University, United States

Reviewed by:
Serge Grazioli,
Geneva University Hospitals (HUG), Switzerland
Renee Potera,
University of Texas Southwestern Medical Center, United States

***Correspondence:**
Michaël Sauthier
michael.sauthier@umontreal.ca

Specialty section:
This article was submitted to Pediatric Critical Care, a section of the journal Frontiers in Pediatrics

Received: 31 March 2021
Accepted: 18 May 2021
Published: 13 July 2021

Citation:
Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Kawaguchi A, Emeriaud G and Jouvet P (2021) Long-Term Mechanical Ventilation in Neonates: A 10-Year Overview and Predictive Model. Front. Pediatr. 9:689190. doi: 10.3389/fped.2021.689190

Objectives: Significant resources are devoted to neonatal prolonged mechanical ventilation (NPMV), but little is known about the outcomes in those children. Our primary objective was to describe the NPMV respiratory, digestive, and neurological outcomes at 18 months corrected age. Our second objective was on the early identification of which patients, among the NPMV cohort, will need to be ventilated for ≥125 days, which corresponded to the 75th percentile in the preliminary data, and to describe that subgroup.

Methods: In this retrospective cohort study, we included all children born between 2004 and 2013 who had a NPMV (≥21 days of invasive or noninvasive respiratory support reached between 40 and 44 weeks of postconceptional age). We used random forests, logistic regression with penalization, naive Bayes, and XGBoost to predict which patients will need ≥125 days of ventilation. We used a Monte Carlo cross validation.

Results: We included 164 patients. Of which, 40% ($n = 66$) were female, and the median gestational age was 29 weeks [interquartile range (IQR): 26–36 weeks] with a bimodal distribution. Median ventilation days were 104 (IQR: 66–139 days). The most frequently associated diagnoses were pulmonary hypertension (43%), early pulmonary dysplasia (41%), and lobar emphysema (37%). At 18 months corrected age, 29% ($n = 47$) had died, 59% ($n = 97$) were free of any respiratory support, and 45% ($n = 74$) were exclusively orally fed. A moderate area under the ROC curve of 0.65 (95% CI: 0.54–0.72) for identifying patients in need of ≥125 days of ventilation at inclusion was achieved by random forests classifiers. Among the 26 measured at inclusion, the most contributive ones were PCO_2, inspired O_2 concentration, and gestational age. At 18 months corrected age, patients ventilated for ≥125 days had a lower respiratory weaning success (76 vs. 87%, $P = 0.05$), lower exclusive oral feeding proportion (51 vs. 84%, $P < 0.001$), and a higher neurological impairment (median Pediatric Cerebral Performance Category score 3 vs. 2, $P = 0.008$) than patients ventilated for <125 days.

Article 5 - *Long-Term Mechanical Ventilation in Neonates: A 10-Year Overview and Predictive Model*

Objectives: Significant resources are devoted to neonatal prolonged mechanical ventilation (NPMV), but little is known about the outcomes of those children. Our primary objective was to describe NPMV respiratory, digestive and neurological outcomes at 18-month corrected age. Our second objective was to early identify among the NPMV cohort, which patients will need to be ventilated ≥125 days, which corresponded to the 75th percentile in preliminary data and describe that subgroup.

Methods: In this retrospective cohort study, we included all children born between 2004 and 2013 who had a NPMV (≥21 days of invasive or noninvasive respiratory support reached between 40 and 44 weeks of postconceptional age). We used random forests, logistic regression with penalization, naive Bayes and XGBoost to predict which patients will need ≥125 days of ventilation. We used a Monte Carlo cross validation.

Results: We included 164 patients with 40% (n=66) female, median gestational age was 29 weeks (interquartile range [IQR] 26-36 weeks) with a bimodal distribution. Median ventilation days were 104 (IQR 66-139 days). Most frequent associated diagnoses were pulmonary hypertension (43%), early pulmonary dysplasia (41%) and lobar emphysema (37%). At 18 months corrected age, 29% (n=47) had died, 59% (n=97) were free of any respiratory support and 45% (n=74) were exclusively orally fed. A moderate area under the ROC curve of 0.65 (95% CI: 0.54-0.72) for identifying patients ≥125 days of ventilation at inclusion was achieved by random forests classifiers. Among the 26 measured at inclusion, the most contributive ones were PCO_2, inspired O_2 concentration and gestational age. At 18 months corrected age, patients ventilated ≥125 days had a lower respiratory weaning success (76% vs. 87%, P=0.05), lower exclusive oral feeding proportion (51% vs. 84%, P<0.001) and a higher neurological impairment (median Pediatric Cerebral Performance Category score 3 vs. 2, P=0.008) than patients ventilated <125 days.

Conclusion: NPMV is a severe condition with a high risk of mortality, neurological impairment and oral feed delay at 18 months. Most survivors were weaned of any respiratory support. We identified risk factors that allow early identify the most at-risk children of long-term ventilation with a moderate discrimination.

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

Introduction

Neonates requiring prolonged mechanical ventilation requires significant resources and are at high risk of multiple and serious long-term complications (1). The exact incidence of neonatal prolonged mechanical ventilation (NPMV) is unknown, but around 3% of ventilated children are supported for more than 21 days (2–4). Despite this group represents a minority of patients, it is responsible for the majority of the economic burden and is increasing in the last decade (4–6). Except for isolated prematurity where the risk factors and outcomes of bronchopulmonary dysplasia have been well described (7–11), infants with congenital anomalies that cannot be discharged through a home ventilation program have been less described in the recent literature (12). In 2017, our team found the most widely used pediatric definition for prolonged mechanical ventilation that included all children still ventilated who had at least 21 consecutive days of ventilation after 37 weeks postmenstrual age (4). The main advantage of this definition is to be applicable to a broad population of newborns (preterm, congenital anomalies, infections, etc.). Those patients are at high-risk of impaired development (13,14). Earlier multidisciplinary interventions aim to improve long-term outcomes, but no data or tools are available to help the clinician identify the most at-risk patients.

Our primary objective of this study was to describe the respiratory, neurological and digestive functional status of the NPMV population at 18-month corrected age, an important milestone in children follow-up in our institution. Our second objective was to build a predictive model to early identify which patients will be ventilated more than the 75th percentile and the outcomes of this subgroup.

Material and Methods

We included all neonates born between April 2004 and December 2013 and admitted in the NICU (65-bed level III unit) of Sainte-Justine hospital (Montreal, Canada), with a NPMV diagnosis. Using our proposition of pediatric prolonged mechanical ventilation (4), NPMV was defined as at least 21 consecutive days of any ventilatory support (invasive, noninvasive and high-flow nasal cannula) reached between 40- and 44- weeks postmenstrual age. The ventilatory support had to be more than just a

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

nocturnal application (i.e. ≥ 6 hours a day) to be counted. A ventilation episode was considered continuous if the interruption in mechanical ventilation support was less than 48 hours. Patients with a neurologic death diagnosis or transferred to another institution before inclusion were excluded. The institutional review board approved this retrospective cohort study and waived the need for individual consent (reference number 3872).

Data Collection

The charts were reviewed by MS and KBG. Data collection quality process was validated on the first ten patients who were reviewed by both MS and KBG to standardize data collection between both researchers. The charts were reviewed for demographic data, primary and secondary diagnoses, perinatal history in the child's and mother's chart and all ventilatory episodes. Data were manually collected and validated from paper charts into an electronic case report form (eCRF). The eCRF had alerts for common typographical errors or impossible physiological values. It also automatically calculated observation times (inclusion, 18 months corrected age) based on the date of birth and gestational age. Data were collected at inclusion (when NPMV diagnosis was first met) and at 18-month corrected age (± 1 month). Individual birth weight Z-scores were automatically calculated using Olsen's growth curves (15) and inclusion and 18-month weights Z-scores used the World Health Organization growth curve using the corrected age (16,17). Neurologic impairment was estimated at 18 months corrected age using the Pediatric Overall Performance Category (POPC) and Pediatric Cerebral Performance Category (PCPC) scores (18). These scores are well validated for functional neurological status in the pediatric critical care population. Although they can provide only a global assessment, they can be done retrospectively using the patient's chart.

Statistical Analysis and Missing Values

We described the patient population using median and interquartile range (IQR) for continuous variables and count with percentages for categorical variables and mortality. Considering the small sample size, we conducted the analyses using Fisher's exact test for nominal variables and Mann-Whitney U test for continuous variables. We used a linear regression to estimate the slope of the number of patients per year. We considered $P < 0.05$ to be significant. We performed multivariate

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

logistic regressions adjusted for gestational age and gender and reported the adjusted odds ratio (aOR). The statistical analyses were conducted in R 4.0.3 (R Project for Statistical Computing, RRID:SCR_001905) using the tidyverse package (tidyverse, RRID:SCR_019186).

Missing values were limited by alerts provided by the eCRF. If the information remained missing, missing values were inferred if possible (e.g., ventilation mode based on the other available parameters). If missing values remains, observation were censored in the descriptive analysis and imputed as physiologically normal values for age for predictive models, as done for most ICU severity scores (19–22).

Predictive Models

We aimed to early predict in the process which patient will need the longest ventilation time. We followed the TRIPOD (23) and the 2020 ICU predictive model guidelines (24) to build the models. Previous work at our institution showed 75th percentile of total ventilation time (invasive or not) was at 125 days (25). We built four machine learning classifiers to identify the most severe quartile based on the available data at inclusion. We selected four algorithms that can efficiently learn from a small dataset and with few hyperparameters to set: logistic regression with penalization (Elastic Net), Naive Bayes methods and random forests and eXtreme Gradient Boosting (26). All models were built using Python v3.9 (Python Programming Language, RRID:SCR_008394), scikit-learn v0.24 library (scikit-learn, RRID:SCR_002577) and XGBoost library (27). XGBoost and random forests are both decision tree ensemble algorithms. However, random forests rely on bagging, which is a democratic process to “elect” the best decision among the subgroups of trees) (28). XGBoost is based on a boosting process, which is an ensemble of weak learners that is reinforced depending on the quality of the assessment. Both are effective in the medical fields (29). The hyperparameters, if applicable, were optimized to minimize the error and reported.

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

We consulted the PICU and NICU specialists in our institution to establish the list of variables to include in the models. After an item generation process completed with data in the literature (1), elements with full agreements were selected to be finally included in the models.

We used a Monte Carlo cross validation method to estimate an empiric distribution from our dataset. This method allows for the validation of our models without an external cohort and estimates 95% confidence intervals (30). We randomly divided with stratification into a train (70%) and a test (30%) set. We repeated the split processes (bootstraps) for N^2 times (N=number of patients) (31) and calculated the discrimination ability of the models on the test cohort with the area under the receiver operating characteristic curve (AUROC), its 95% confidence interval (95% CI) and the *P*-values when comparing different algorithms (32,33). We determined the importance of each variable using the permutation method (28). Convergence of the best models was assessed using the cumulative mean of the AUROC for each consecutive iteration.

Results

During the 10-year period, 9,726 infants were admitted to the NICU and 52% (n=5,087) had at least one ventilatory episode. Of these, 164 met the inclusion criteria after a full review of the charts. No patients were excluded. Forty percent (n=66) of the patients were female. Median gestational age was 29 weeks (IQR 26-36), with a bimodal distribution (**Figure 1** and **Supplementary Table 1**). Birth weights were approximately normal (median Z-score: -0.3, IRQ -1.3 to 0.3). Other demographic and ventilation data are presented in **Table 1** and **Table 2**. Clinical definitions used to define diagnoses are presented in **Supplementary Table 2**.

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

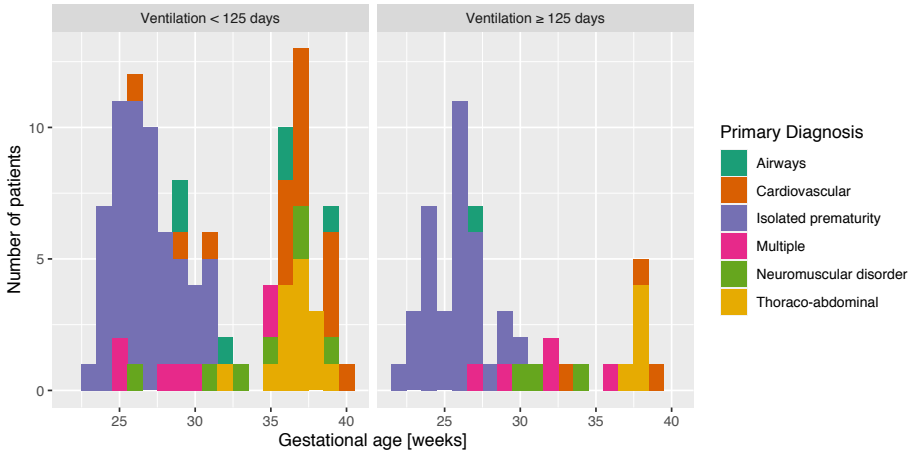


Figure 1: Primary diagnosis categories distributed per gestational age.

Ventilation	Primary Diagnosis	Gestational age (weeks)																		
		22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
< 125 days	Airways	0	0	0	0	0	0	0	2	0	0	1	0	0	0	2	0	0	1	0
< 125 days	Cardiovascular	0	0	0	0	1	0	0	1	0	1	0	0	0	0	4	6	0	4	1
< 125 days	Isolated prematurity	0	1	7	9	10	10	5	4	3	4	0	0	0	0	0	0	0	0	0
< 125 days	Multiple	0	0	0	2	0	0	1	1	1	0	0	0	0	2	0	0	0	0	0
< 125 days	Neuromuscular disorder	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	2	0	1	0
< 125 days	Thoraco-abdominal	0	0	0	0	0	0	0	0	0	0	1	0	0	1	4	5	3	1	0
≥ 125 days	Airways	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
≥ 125 days	Cardiovascular	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0
≥ 125 days	Isolated prematurity	1	3	7	3	11	5	1	2	1	0	0	0	0	0	0	0	0	0	0
≥ 125 days	Multiple	0	0	0	0	0	1	0	1	0	0	2	0	0	0	1	0	0	0	0
≥ 125 days	Neuromuscular disorder	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0
≥ 125 days	Thoraco-abdominal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	0	0

Supplementary Table 1: Number of Patients per Gestational Age and per Primary Diagnosis

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

Variables	**All Children (n=164)**	**Ventilation Days ≥125 days (n=52)**	**Ventilation Days <125 days (n=112)**	***P***
	n (%) or median [IQR]			
Female, n	66 (40)	15 (29)	51 (46)	0.059
Gestational age, wk	29 [26–36]	27 [25–31.2]	30 [26–36]	0.005
Gestational age, strata				
<28 wk	73 (45)	32 (62)	41 (37)	0.004
28 to 32 wk	36 (22)	9 (17)	27 (24)	0.42
33 to 36 wk	24 (15)	4 (8)	20 (18)	0.10
≥37 wk	31 (19)	7 (14)	24 (21)	0.29
Birth Weight, g	1065 [770–2155]	810 [690–1343]	1205 [820–2496]	<0.001
Birth Weight, Z-score	-0.3 [-1.3–0.3]	-0.6 [-1.3–0]	-0.2 [-1.2–0.4]	0.09
Delivery, caesarean	112 (68)	35 (67)	77 (69)	0.86
Delivery room intubation, n	89 (54)	31 (60)	58 (52)	0.40
Apgar scores, strata				
1-minute Apgar score	4 [2–7]	4 [2–6]	4 [2–7]	0.42
5-minute Apgar score	6 [5–8]	6 [5–7.5]	6 [5–8]	0.40
10-minute Apgar score	7 [6–9]	7 [6–8.5]	7.5 [6–9]	0.08
Arterial coord pH	7.3 [7.2–7.3]	7.3 [7.2–7.3]	7.3 [7.2–7.3]	0.48
Total ventilation time, days	104 [66–139]	172 [141–236]	84 [51–106]	<0.001
High-frequency oscillatory ventilation, days and n	22.8 [10.5–40.5] 131 (80)	35.1 [17.7–45] 45 (87)	18.1 [8.1–35] 86 (77)	0.005 0.21
Conventional Mechanical Ventilation, days and n	23 [12.1–50.8] 161 (98)	37.2 [16.1–116.7] 51 (98)	21.4 [10.5–40.3] 110 (98)	0.007 >0.99
Noninvasive Ventilation, days and n	29.6 [11–52] 154 (94)	52.7 [29.4–88.1] 51 (98)	22.9 [6–40.9] 103 (92)	<0.001 0.17
High Flow Nasal Cannula, days and n	31.6 [14.7–44.7] 75 (46)	48.8 [34.8–98] 31 (60)	20.4 [13.8–32] 44 (39)	<0.001 0.019

Table 1: Patient Characteristics. IQR: Interquartile Range.

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

Variables [a]	**All children (n=164)**	**Ventilation Days ≥ 125 days (n=52)**	**Ventilation Days < 125 days (n=112)**	**Adjusted Odds Ratios** [c] **(95% CI)**	***P***
	n (%) or median (IQR)				
Weight at inclusion, Z-score	-0.5 [-1 – 0.3]	-0.8 [-2 – 0.1]	-0.5 [-1 – 0.3]	0.67 (0.48 – 0.89)	0.009
PCO_2, mmHg	62 [53 – 68]	67 [61 – 71]	60 [52 – 66]	1.02 (0.99 – 1.06)	0.23
Bicarbonate, mEq/L	31 [28 – 34]	33 [30 – 35]	30 [28 – 33]	1.03 (0.96 – 1.11)	0.39
Invasive ventilation	61 (37)	18 (35)	43 (38)	1.44 (0.65 – 3.24)	0.37
PEEP, cmH_2O	6 [5 – 6]	6 [5 – 6]	5 [5 – 6]	1.31 (0.91 – 1.94)	0.16
FiO_2, %	30 [25 – 40]	40 [30 – 46]	28 [23 – 40]	1.01 (0.99 – 1.03)	0.41
SpO_2, % [b]	94 [93 – 96]	94 [92 – 96]	94 [93 – 96]	1.03 (0.96 – 1.11)	0.47
SpO_2/FiO_2 ratio	313 [224 – 384]	248 [191 – 322]	333 [238 – 401]	1 (0.99 – 1)	0.04
Diagnosis [b]					
Pulmonary hypertension	70 (43)	24 (46)	46 (41)	1.55 (0.77 – 3.18)	0.22
Early pulmonary dysplasia [c]	67 (41)	20 (38)	47 (42)	0.85 (0.42 – 1.7)	0.64
Lobar emphysema	60 (37)	16 (31)	44 (39)	0.62 (0.3 – 1.28)	0.20
Patent ductus arteriosus	28 (17)	8 (15)	20 (18)	0.91 (0.34 – 2.23)	0.84
Cardiovascular disease	26 (16)	6 (12)	20 (18)	1.27 (0.4 – 3.81)	0.67
CDH or pulmonary hypoplasia	16 (10)	5 (10)	11 (10)	1.98 (0.52 – 7)	0.29
Neuromuscular disease	13 (8)	3 (6)	10 (9)	0.81 (0.17 – 2.9)	0.76
Esophageal atresia	10 (6)	3 (6)	7 (6)	1.62 (0.32 – 6.8)	0.52
Polymalformative syndrome with heart disease	10 (6)	2 (4)	8 (7)	0.92 (0.13 – 4.2)	0.92
Sepsis	10 (6)	3 (6)	7 (6)	0.75 (0.17 – 2.44)	0.65
Tracheobronchomalacia	10 (6)	1 (2)	9 (8)	0.31 (0.02 – 1.82)	0.28
Surgical necrotizing enterocolitis	9 (5)	3 (6)	6 (5)	1.03 (0.2 – 4.33)	0.97
Polymalformative syndrome with airways lesion	8 (5)	2 (4)	6 (5)	1.07 (0.15 – 5.27)	0.94

Table 2: Characteristics at Inclusion. IQR: Interquartile range. CI: confidence interval. Odds ratio were adjusted for gestational age and sex. CDH: Congenital diaphragmatic hernia. PEEP: positive end-expiratory pressure. [a] Most recent value before inclusion. [b] Total is not equal to 100% because each patient can have more than one diagnosis. [c] Adjusted for gender and gestational age.

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

Variables	Definition
Bicarbonates	Measures in an arterial. capillary or venous blood sample
Birth weight z-score	Normalized with Olsen's growth curves
Cardiovascular disease	Congenital heart or vessels structural malformation
Congenital diaphragmatic hernia or pulmonary hypoplasia	Pulmonary hypoplasia relied on a volumetric radiological modality.
Early pulmonary dysplasia [c]	Based on the radiological report during the first week of life.
Esophageal atresia	
Female	
FiO_2	Inspired oxygen concentration
Gestational age	Expressed in weeks
Inclusion weight (Z)	Weight z-score at inclusion using World Health Organization growth curves
Invasive ventilation	Ventilation through an infraglottic device
Lobar emphysema	Lobar emphysema on at least two radiological reports before inclusion
Mean airways pressure	Mean airways pressure read by the ventilator
Neuromuscular disease	Neurological (central or peripheral) disease and/or muscular pathology diagnosed by a pediatric neurologist
PCO_2	CO_2 partial pressure in blood gas (arterial. capillary or venous)
PEEP	Positive end-expiratory pressure set on a ventilator (invasive or not)
Persistent ductus arteriosus	Ductus arteriosus that required an interventional or surgical ligation
Polymalformative syndrome with airways lesion	Congenital malformation touching two organs or more. including the airways
Polymalformative syndrome with heart disease	Congenital malformation touching two organs or more. including the heart
Pulmonary hypertension [b]	Defined by at least two echocardiograms with type II or III septum or a valve regurgitation that allowed for an estimation of the pulmonary pressures to be more than half of systemic pressures (39)
Sepsis	clinically diagnosed with a positive culture
Severe intraventricular hemorrhage	Intraventricular hemorrhage grade 3 and 4 (40)
SpO_2	Pulsed oximetry measuring oxygen hemoglobin saturation
Surgical necrotizing enterocolitis	Necrotizing enterocolitis that required a surgical treatment
Tracheal hypoplasia	Diagnosis proven by direct endoscopy
Tracheobronchomalacia	Diagnosis proven by direct endoscopy

Supplementary Table 1: Number of Patients per Gestational Age and per Primary Diagnosis

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

Outcomes at 18 months are presented in **Table 3** and showed that 29% (n=47) of the patients had died. Among survivors, 83% (n=97) were completely weaned of any respiratory or oxygen support. Three patients (2%) were still tracheotomized at this point. Exclusive oral feeding was achieved for 72% of the patients. However, they were slightly underweight with a -0.9 Z-score (IQR -2 – 0). Their neurological functional status was also impacted with a median PCPC and POPC score of 3 (IQR 2 – 3), corresponding to "moderate disability" (i.e., special classrooms).

Variables	**All children (n=164)**	**Ventilation Days ≥ 125 days (n=52)**	**Ventilation Days < 125 days (n=112)**	**Adjusted Odds Ratios[b] (95% CI)**	***P***
	n (%) or median [IQR]				
Mortality	47 (29)	11 (21)	36 (32)	0.68 (0.3 – 1.5)	0.355
Lost to follow-up or inadequate information to assess	8 (5)	1 (2)	7 (6)	0.39 (0.02 – 2.42)	0.396
Hospitalized at 18 months of corrected age	4 (3)	3 (7)	1 (1)	4.56 (0.53 – 96)	0.204
Respiratory status [a]					
No respiratory support	97 (83)	31 (76)	66 (87)	0.35 (0.12 – 1)	0.052
Invasive ventilation	2 (2)	1 (2)	1 (1)	4.55 (0.16 – 127)	0.307
Noninvasive ventilation	3 (3)	3 (7)	0	-	-
High flow nasal cannula	1 (1)	1 (2)	0	-	-
Standard nasal cannula	6 (5)	4 (10)	2 (3)	5.35 (0.89 – 45)	0.079
Tracheostomy	3 (2)	1 (2)	2 (2)	0.66 (0.03 – 7.7)	0.745
Nutritional status [a]					
Exclusive oral feeding	74 (72)	20 (51)	54 (84)	0.09 (0.02 – 0.27)	<0.001
Weight, kg	10 [9 – 11]	10 [9 – 11]	10 [9 – 12]	0.75 (0.53 – 1.02)	0.087
Weight, Z-score	-0.9 [-2 – 0]	-0.9 [-2 – -0.4]	-0.9 [-1 – 0.7]	0.69 (0.46 – 1)	0.063
Neurological status					
PCPC	3 [2 – 3]	3 [2 – 3]	2 [2 – 3]	2.28 (1.29 – 4.36)	0.008
POPC	3 [2 – 3]	3 [2 – 3]	2 [2 – 3]	1.96 (1.21 – 3.33)	0.008

Table 3: Outcomes at 18 Months of Corrected Age

Subgroup of Patients Ventilated ≥125 days

Patients ventilated ≥ 125 days had lower female proportion (29%, n=15) than the subgroup ventilated < 125 days (46%, n=51), but it did not reach statistical significance (P = 0.06). The weight Z-score at inclusion was significantly lower in the

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

subgroup ≥125 days (aOR 0.67, 95% CI 0.48 – 0.89, *P=0*.009). However, the ventilatory characteristics and diagnoses were not significantly different when adjusted for gestational age and gender. A comparison of patients' characteristics is shown in **Table 1** and **Table 2**. Unadjusted odds ratios are presented in **Supplementary Tables 3** and **4**.

Variables [a]	**All children (n=164)**	**Ventilation Days ≥ 125 days (n=52)**	**Ventilation Days < 125 days (n=112)**	**Unadjusted Odds Ratios** [c] **(95% CI)**	***P***
	n (%) or median (IQR)				
Weight at inclusion, Z-score	-0.5 [-1 – 0.3]	-0.8 [-2 – 0.1]	-0.5 [-1 – 0.3]	0.86 (0.66 – 1.12)	0.27
PCO_2, mmHg	62 [53 – 68]	67 [61 – 71]	60 [52 – 66]	1.04 (1.01 – 1.07)	0.01
Bicarbonate, mEq/L	31 [28 – 34]	33 [30 – 35]	30 [28 – 33]	1.07 (1 – 1.14)	0.06
Invasive ventilation	61 (37)	18 (35)	43 (38)	0.85 (0.42 – 1.68)	0.64
PEEP, cmH_2O	6 [5 – 6]	6 [5 – 6]	5 [5 – 6]	1.44 (1.04 – 2.09)	0.04
FiO_2, %	30 [25 – 40]	40 [30 – 46]	28 [23 – 40]	1.01 (0.99 – 1.03)	0.16
SpO_2, % [b]	94 [93 – 96]	94 [92 – 96]	94 [93 – 96]	1.01 (0.96 – 1.09)	0.68
SpO_2/FiO_2 ratio	313 [224 – 384]	248 [191 – 322]	333 [238 – 401]	1 (0.99 – 1)	<0.01
Diagnosis [b]					
Pulmonary hypertension	70 (43)	24 (46)	46 (41)	1.23 (0.63 – 2.39)	0.54
Early pulmonary dysplasia [c]	67 (41)	20 (38)	47 (42)	0.86 (0.44 – 1.69)	0.67
Lobar emphysema	60 (37)	16 (31)	44 (39)	0.69 (0.33 – 1.37)	0.29
Patent ductus arteriosus	28 (17)	8 (15)	20 (18)	0.84 (0.32 – 1.99)	0.70
Cardiovascular disease	26 (16)	6 (12)	20 (18)	0.6 (0.21 – 1.52)	0.31
CDH or pulmonary hypoplasia	16 (10)	5 (10)	11 (10)	0.98 (0.29 – 2.85)	0.97
Neuromuscular disease	13 (8)	3 (6)	10 (9)	0.62 (0.14 – 2.15)	0.49
Esophageal atresia	10 (6)	3 (6)	7 (6)	0.92 (0.19 – 3.46)	0.90
Polymalformative syndrome with heart disease	10 (6)	2 (4)	8 (7)	0.52 (0.08 – 2.17)	0.42
Sepsis	10 (6)	3 (6)	7 (6)	0.74 (0.17 – 2.35)	0.63
Tracheobronchomalacia	10 (6)	1 (2)	9 (8)	0.22 (0.01 – 1.24)	0.16
Surgical necrotizing enterocolitis	9 (5)	3 (6)	6 (5)	1.08 (0.22 – 4.28)	0.91
Polymalformative syndrome with airways lesion	8 (5)	2 (4)	6 (5)	0.71 (0.1 – 3.19)	0.68

Supplementary Table 3: Characteristics at Inclusion - Unadjusted Odds Ratios. IQR: Interquartile range. CI: confidence interval. Odds ratio were adjusted for gestational age and sex. CDH: Congenital diaphragmatic hernia. PEEP: positive end-expiratory pressure. [a] Most recent value before inclusion. [b] Total is not equal to 100% because each patient can have more than one diagnosis. [c] Adjusted for gender and gestational age.

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

Variables	All children (n=164)	Ventilation Days ≥ 125 days (n=52)	Ventilation Days < 125 days (n=112)	Unadjusted Odds Ratios[b] (95% CI)	*P*
	n (%) or median [IQR]				
Mortality	47 (29)	11 (21)	36 (32)	0.57 (0.25 – 1.2)	0.15
Lost to follow-up or inadequate information to assess	8 (5)	1 (2)	7 (6)	0.29 (0.02 – 1.72)	0.26
Hospitalized at 18 months of corrected age	4 (3)	3 (7)	1 (1)	5.92 (0.73 – 121.91)	0.13
Respiratory status [a]					
No respiratory support	97 (83)	31 (76)	66 (87)	0.47 (0.17 – 1.26)	0.13
Invasive ventilation	2 (2)	1 (2)	1 (1)	1.87 (0.07 – 48.26)	0.66
Noninvasive ventilation	3 (3)	3 (7)	0	-	-
High flow nasal cannula	1 (1)	1 (2)	0	-	-
Standard nasal cannula	6 (5)	4 (10)	2 (3)	4 (0.75 – 29.81)	0.12
Tracheostomy	3 (2)	1 (2)	2 (2)	1.08 (0.05 – 11.51)	0.95
Nutritional status [a]					
Exclusive oral feeding	74 (72)	20 (51)	54 (84)	0.19 (0.07 – 0.48)	<0.001
Weight. kg	10 [9 – 11]	10 [9 – 11]	10 [9 – 12]	0.84 (0.63 – 1.11)	0.23
Weight. Z-score	-0.9 [-2 – 0]	-0.9 [-2 – -0.4]	-0.9 [-1 – 0.7]	0.77 (0.53 – 1.09)	0.15
Neurological status					
PCPC	3 [2 – 3]	3 [2 – 3]	2 [2 – 3]	2.02 (1.19 – 3.63)	0.01
POPC	3 [2 – 3]	3 [2 – 3]	2 [2 – 3]	1.62 (1.04 – 2.59)	0.04

Supplementary Table 4: Outcomes at 18 Months of Corrected Age – Unadjusted Odds Ratios. IQR: Interquartile range. PCPC: Pediatric Cerebral Performance Category. POPC: Pediatric Overall Performance Category. [a] Only available data were used to calculate percentages. [b] Adjusted for gender and gestational age.

There is no significant difference in mortality at 18 months corrected age (**Table 3**) in the subgroup ventilated <125 days (32%) compared to those ventilated ≥125 days (21%, aOR 0.68, 95% CI 0.3-1.5). Survivors ventilated ≥125 days subgroup had a lower proportion of being weaned off any respiratory support (aOR 0.35, 95% CI 0.12 – 1, P = 0.06) when adjusted for gestational age and gender. Exclusive oral feeding was higher for patients ventilated < 125 days (aOR 0.09, 95 CI 0.02 – 0.27, $P < 0.001$). Neurological impairment was also less severe in the < 125 days subgroup (median PCP 2, IQR 2 – 3 and median PCPC 3, IQR 2 – 3 with $P=0.008$, respectively).

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

Evolution of Ventilation Modalities

Ventilation modalities evolved over time. Our study covered from 2004 to 2013 and showed the introduction of high-flow nasal cannula (HFNC) in 2010 (**Figure 2**). We compared the median duration for each modality before (2004 to 2008) and after HFNC was introduced (2009 to 2013). The median noninvasive ventilation (NIV) duration was 21 days (IQR 5-57) before and 34 days (IQR 16-52) after HFNC implementation (P = 0.20). The conventional mechanical ventilation had a median duration of 23 days before 2009 (IQR 17-79) and 23 days from 2009 to 2013 (IQR 9-42, P=0.09). The high frequency oscillatory ventilation (HFOV) had median duration of 15 days before 2009 (IQR 6-34) and 27 days after (IQR 12-41 days, P=0.009). The median total ventilation time went from 94 days (IQR 42-119) to 110 days (IQR 83-142, P=0.02) after 2009. If HFNC were excluded after 2009, the median time was 85 days (IQR 63-110) which was not different from the period before 2009 (P=0.79). Furthermore, the number of patients included per year (**Figure 2** and **3**) linearly increased by about 2 patients per year (P=0.008).

We also showed that the number of preterms born before <28 weeks increased in the observation period (**Figure 3**).

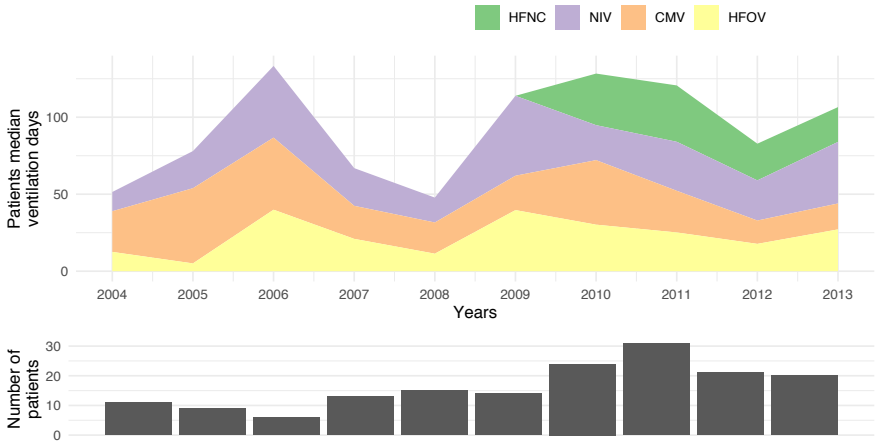


Figure 2: Median ventilation time per patient for each modality. HFNC, high-flow nasal canula; NIV, noninvasive ventilation; CMV, conventional mechanical ventilation; HFOV, high frequency oscillatory ventilation.

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

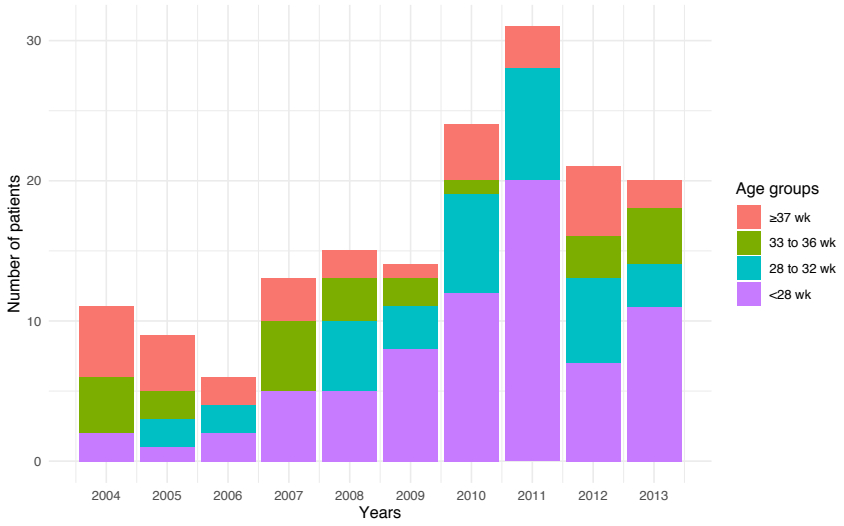


Figure 3: Gestational age distribution per year of inclusion. wk, weeks.

Predictive Models

All models included 26 variables (Supplementary Table 2). Among the four tested models, random forests and XGBoost showed the highest discrimination with an AUROC 0.65 (95% CI 0.54 – 0.72, P=0.008) and 0.62 (95% CI 0.50 – 0.70, P=0.025), respectively. The difference was not statistically significant (P = 0.41). Logistic regression with penalization and naive Bayes showed a weak discrimination with an AUROC of 0.58 (95% CI: 0.46 – 0.66, *P*=0.09) and AUROC of 0.53 (95% CI: 0.45 – 0.62, *P*=0.19).

The importance analysis (**Figure 4**) of the random forests and XGBoost models showed that blood gas results close to the inclusion (PCO_2 and bicarbonate) provided important features for the models, so as the inspired oxygen concentration and gestational age, birth weight and weight at inclusion Z-score. Diagnoses were generally less contributive to the model. The most contributive variables were similar among random forest and XGBoost algorithms.

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

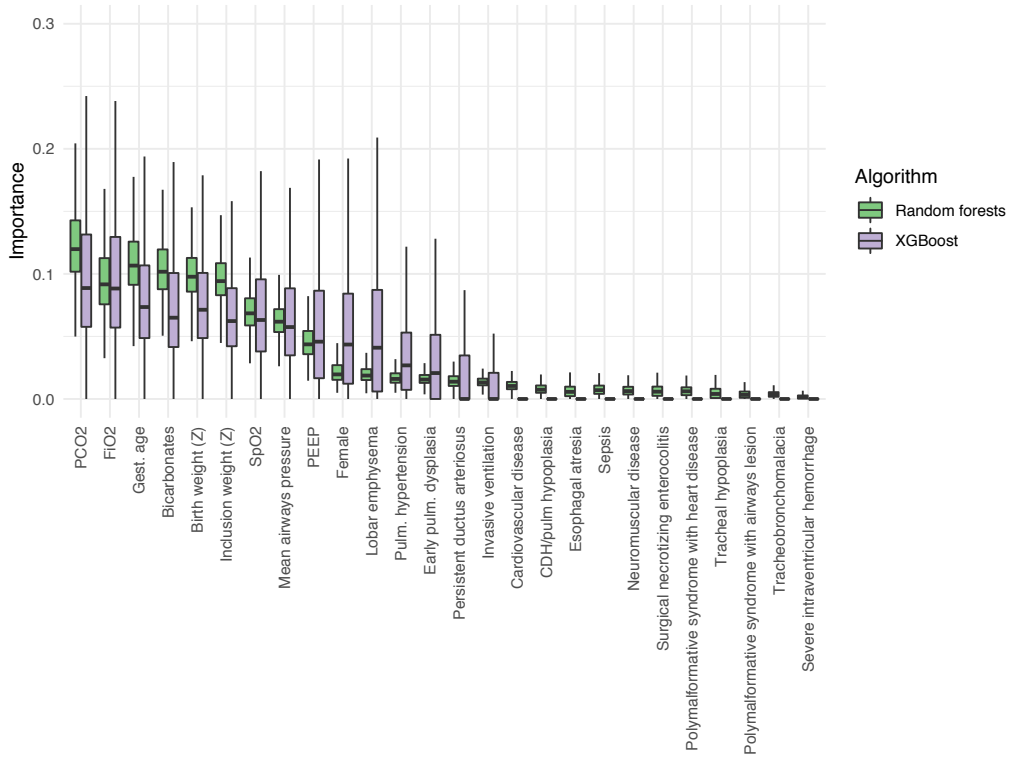
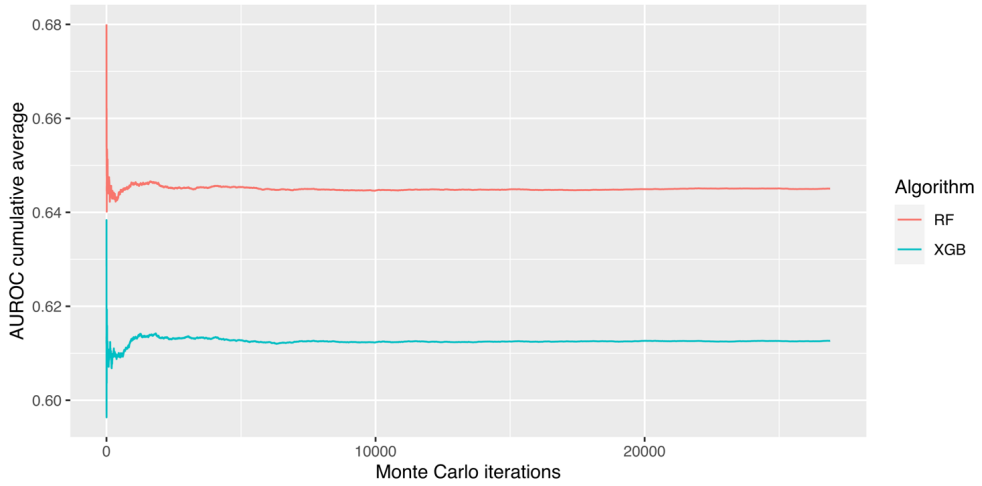


Figure 4: Relative importance of the different variables included in the random forests and in the extreme gradient boosting models (XGBoost). PEEP, positive end expiratory pressure; CDH, congenital diaphragmatic hernia; Z, Z-score.

Convergence of the best models (random forests and XGBoost) was achieved after about 10'000 Monte Carlo repetitions (**Supplementary Figure 1**). Random forest was optimal with a maximum depth of 300 trees, logistic regression used an L1 ratio of 0.5 and XGBoost used a maximum of 100 boosting rounds. By definition, no hyperparameter optimization was required for the naive Bayes classifier.

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.



Supplementary Figure 1: Model convergence over the Monte Carlo iterations. RF: random forests. XGB: eXtreme Gradient Boosting

Discussion

We described a diversified cohort of 164 hospitalized newborns who had NPMV with their outcomes at 18 months corrected age. Although this condition is rare, this study indicates that it is a severe one with multiple complex associated comorbidities. This study also suggests that the number of patients and ventilation duration are increasing over the studied period. As previously shown, these children required complex multidisciplinary care and have a significant economic and logistical burden with high mortality rate. Moreover, most survivors developed a significant neurological impairment, even if most of them are successfully weaned of any respiratory or nutritional support. Several causes can contribute to the neurological impairment in this heterogenous population (for example, intra-ventricular hemorrhage, hypoxic ischemic encephalopathy, genetic susceptibility, low cardiac output, etc.). The problem was already described more than three decades

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

ago, but very few recent studies presented an updated description of the situation, including the increase use of modern ventilation modes, such as high flow nasal cannula (12).

About 80% of the cohort was born prematurely, but with similar proportion in the subgroup ventilated <125 days and ≥ 125 days. Once adjusted for gestational age and gender, no diagnoses were shown to be significantly associated with both subgroups. At inclusion, the patients who received ≥125 days of ventilation were significantly smaller (weight Z-score) than those who received less than 125 days. Nutrition has been closely linked to respiratory failure in newborns, but the causal relationship is highly complex to establish (34). In our study, birth weight Z-scores were similar between groups, suggesting that the difference be acquired between birth and inclusion. This can be the consequence of the underlying disease severity or the direct cause of the sustained ventilation. This study also showed that our center rarely used the tracheostomies in the <125 days subgroup, although many benefits of this intervention, including an improved nutrition and growth, have been suggested (35). Moreover, we found no difference between the groups on the use of invasive ventilation or other ventilator parameters. However, one should note that more than half of the patients in the ≥125 days of total ventilation had a hypoxemia level at inclusion (SF median 248, IQR 191 – 322) comparable to the pediatric acute respiratory distress syndrome which has an SF threshold of 264 (36). Blood gases were not different between the groups at inclusion, and both had high PCO_2 levels (median PCO_2 62, IQR 53 – 68).

The outcomes at 18-month corrected age showed a high mortality rate (29%), which is one of the highest mortality rates seen in modern PICUs that usually have an overall mortality below 5% (37). In 1987, a similar study described a mortality of 25% at two years of age (12). However, the authors described that 25% of the survivors still required a ventilation support even after 18 years of follow-up. In our study, we observed that 17% of the patients still needed a support at 18 months and most of them were at home with standard nasal cannula. Only 2% of the patients still required invasive ventilation at this point. At 18 months, patients were generally underweight (median Z-score -0.9) for both groups (P = 0.06). Exclusive oral feeding was achieved in 51% (n=20) and 84% (n=54) of the ≥125 days and <125 days of ventilation groups, respectively (P < 0.001).

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

Neurological impairment evaluation is usually difficult and limited in retrospective studies. However, PCPC and POPC scores have been validated for that purpose (18). Overall, the cohort showed a moderate disability (median 3, IQR 2-3) which corresponds to "sufficient cerebral function for age-appropriate independent activities of daily life" but children usually require "special education classroom and/or learning deficit present" (18). The disability was less marked in the <125 days of ventilation subgroup (median 2) compared to the ≥125 days of ventilation subgroup (median 3, P=0.008). An association between a longer ventilation time and a worse neurological outcome is known for newborns (13,14). However, the causality is once again complex to determine, since multiple confounding variables are present.

Ventilation modalities have evolved in the last decade with an increasing NIV and HFNC as well as the total ventilation time per patient as shown in our data (**Figure 2**) and in the literature (38). As HFNC were introduced in 2010, one may ask if the increase was only due to this new modality or if this mode replaces another one. Our data showed that the total duration is no longer different if HFNC is removed from the equation and that invasive and NIV median time were similar before and after the HFNC introduction in this population. Although HFNC are known to have a real respiratory support effect (4), without any objective measurement of the lung function to compare these patients, we can only speculate between a lower clinical threshold or an increased respiratory failure severity. Our data also showed that the median time on HFOV nearly doubled before and after 2010. It may be associated with the increase of preterms <28 weeks over the years (**Figure 3**) as, in our institution, this population was more frequently ventilated with HFOV.

The study has strengths. First, the number of patients is large for the field. Very few studies focused on NPMV for patients that cannot be discharged to a home ventilation program. Because both the number of patients and the length of ventilation are increasing, having more data on the topic is essential. Second, we chose to take a pragmatic point of view when describing the diversified cohort of preterm and term neonates (**Figure 1**). Because these patients are sharing the same resources in the hospital, we argue that they will benefit from a global and comprehensive analysis. Third, this study is the first to our knowledge to publish a predictive model to early identify among the NPMV neonates the one that will be

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

ventilated ≥125 days. Fourth, we used modern algorithms to isolate the most important variables that guided the model to identify the most severe patients. These data will help with the hypothesis generation.

This study has limitations. First, NPMV is a rare entity and the underlying diagnoses even more so. Thus, it is difficult to detect a statistically significant difference between the subgroups of patients ventilated ≥ 125 days or less. To overcome this limitation, our center is piloting a prospective cross-sectional multicentric study on the topic (Long VentKid: https://longventkids.ca) with the collaboration of many international pediatric intensive care societies. The second limitation is the retrospective nature of the study, especially for the accuracy of some clinical diagnoses such as pulmonary hypertension, patent ductus arteriosus or tracheobronchomalacia. As these diagnoses were only assessed if clinically relevant, prevalence data may be biased. Furthermore, some diagnoses are subjective, especially those that relies on radiological data. We limited the inter-observer variability by requiring that a minimum of two reports confirmed the diagnosis or to be proven by the gold standard (e.g. tracheobronchomalacia had to be proven by endoscopy). Nevertheless, the only differences between the groups were about objective data.

Conclusion

We described the main characteristics and the outcomes at 18 months corrected age of neonates that required a prolonged mechanical ventilation. We showed that mortality is high, most patients have a significant neurological impairment and half of those with the longest ventilation duration were exclusively orally fed at 18 months. However, only 2% of the survivors were still requiring invasive ventilation at 18 months corrected age. Future interventional studies should answer if an early and multidisciplinary intensive therapy on the most at-risk patients can improve their outcomes.

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

References

1. Koltsida G, Konstantinopoulou S. Long term outcomes in chronic lung disease requiring tracheostomy and chronic mechanical ventilation. Semin Fetal Neonatal Med (2019) 24:101044. doi:10.1016/j.siny.2019.101044
2. Traiber C, Piva JP, Fritsher CC, Garcia PC, Lago PM, Trotta EA, Ricachinevsky CP, Bueno FU, Baecker V, Lisboa BD. Profile and consequences of children requiring prolonged mechanical ventilation in three Brazilian pediatric intensive care units. Pediatr Crit Care Med (2009) 10:375–380. doi:10.1097/PCC.0b013e3181a3225d
3. Payen V, Jouvet P, Lacroix J, Ducruet T, Gauvin F. Risk factors associated with increased length of mechanical ventilation in children. Pediatr Crit Care Med (2012) 13:152–157. doi:10.1097/PCC.0b013e3182257a24
4. Sauthier M, Rose L, Jouvet P. Pediatric Prolonged Mechanical Ventilation: Considerations for Definitional Criteria. Respir Care (2017) 62:49–53. doi:10.4187/respcare.04881
5. Zilberberg MD, Shorr AF. Prolonged acute mechanical ventilation and hospital bed utilization in 2020 in the United States: implications for budgets, plant and personnel planning. BMC Heal Serv Res (2008) 8:242. doi:10.1186/1472-6963-8-242
6. Wagner DP. Economics of prolonged mechanical ventilation. Am Rev Respir Dis (1989) 140:S14-8. doi:10.1164/ajrccm/140.2_Pt_2.S14
7. Cristea AI, Carroll AE, Davis SD, Swigonski NL, Ackerman VL. Outcomes of Children With Severe Bronchopulmonary Dysplasia Who Were Ventilator Dependent at Home. Pediatrics (2013) 132:e727–e734. doi:10.1542/peds.2012-2990
8. Van Marter LJ, Dammann O, Allred EN, Leviton A, Pagano M, Moore M, Martin C. Chorioamnionitis, mechanical ventilation, and postnatal sepsis as modulators of chronic lung disease in preterm infants. J Pediatr (2002) 140:171–176. doi:10.1067/MPD.2002.121381
9. Farstad T, Bratlid D, Medbo S, Markestad T, Norwegian Extreme Prematurity Study G. Bronchopulmonary dysplasia - prevalence, severity and predictive factors in a national cohort of extremely premature infants. Acta Paediatr (2011) 100:53–58. doi:10.1111/j.1651-2227.2010.01959.x

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

10. Choi Y-B, Lee J, Park J, Jun YH. Impact of Prolonged Mechanical Ventilation in Very Low Birth Weight Infants: Results From a National Cohort Study. J Pediatr (2018) 194:34-39.e3. doi:10.1016/j.jpeds.2017.10.042
11. Walsh MC, Morris BH, Wrage LA, Vohr BR, Poole WK, Tyson JE, Wright LL, Ehrenkranz RA, Stoll BJ, Fanaroff AA, et al. Extremely low birthweight neonates with protracted ventilation: mortality and 18-month neurodevelopmental outcomes. J Pediatr (2005) 146:798–804. doi:10.1016/j.jpeds.2005.01.047
12. Schreiner MS, Downes JJ, Kettrick RG, Ise C, Voit R. Chronic Respiratory Failure in Infants With Prolonged Ventilator Dependency. JAMA J Am Med Assoc (1987) 258:3398. doi:10.1001/jama.1987.03400230058032
13. Gaillard EA, Cooke RW, Shaw NJ. Improved survival and neurodevelopmental outcome after prolonged ventilation in preterm neonates who have received antenatal steroids and surfactant. Arch Dis Child Fetal Neonatal Ed (2001) 84:F194-6. doi:10.1136/fn.84.3.f194
14. Wheater M, Rennie JM. Poor prognosis after prolonged ventilation for bronchopulmonary dysplasia. Arch Dis Child Fetal Neonatal Ed (1994) 71:F210-1. doi:10.1136/fn.71.3.f210
15. Olsen IE, Groveman SA, Lawson ML, Clark RH, Zemel BS. New Intrauterine Growth Curves Based on United States Data. Pediatrics (2010) 125:e214–e224. doi:10.1542/peds.2009-0913
16. Organization WH, others. WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development. (2006)
17. Villar J, Giuliani F, Bhutta ZA, Bertino E, Ohuma EO, Ismail LC, Barros FC, Altman DG, Victora C, Noble JA, et al. Postnatal growth standards for preterm infants: the Preterm Postnatal Follow-up Study of the INTERGROWTH-21 st Project. Lancet Glob Heal (2015) 3:e681–e691. doi:10.1016/S2214-109X(15)00163-1
18. Fiser DH, Long N, Roberson PK, Hefley G, Zolten K, Brodie-Fowler M. Relationship of pediatric overall performance category and pediatric cerebral performance category scores at pediatric intensive care unit discharge with outcome measures collected at hospital discharge and 1- and 6-month follow-up assessments. Crit Care Med (2000) 28:2616–2620. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10921604

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

19. Leteurtre S, Duhamel A, Salleron J, Grandbastien B, Lacroix J, Leclerc F. PELOD-2: An update of the PEdiatric logistic organ dysfunction score. Crit Care Med (2013) 41:1761–1773. doi:10.1097/CCM.0b013e31828a2bbd
20. 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: doi:10.1001/jamapediatrics.2017.2352
21. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med (2006) 34:1297–310. doi:10.1097/01.CCM.0000215112.84523.F0
22. Pollack MM, Holubkov R, Funai T, Dean JM, Berger JT, Wessel DL, Meert K, Berg RA, Newth CJL, Harrison RE, et al. The Pediatric Risk of Mortality Score: Update 2015. Pediatr Crit Care Med (2016) 17:2–9. doi:10.1097/PCC.0000000000000558
23. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Ann Intern Med (2015) 162:55. doi:10.7326/M14-0697
24. Leisman DE, Harhay MO, Lederer DJ, Abramson M, Adjei AA, Bakker J, Ballas ZK, Barreiro E, Bell SC, Bellomo R, et al. Development and Reporting of Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and Critical Care Journals. Crit Care Med (2020)1–11. doi:10.1097/CCM.0000000000004246
25. Sauthier M, Bergeron-Gallant K, Lodygensky G, Emeriaud G, Jouvet P. Risk factors and prognosis for newborns with chronic mechanical ventilation: preliminary results of a retrospective cohort [Abstract]. Chest J. (2015). Available online at: https://journal.chestnet.org/article/S00123692(16)36225-0/fulltext
26. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. (2011) 12:2825–30. Available online at: https://www.jmlr.org/papers/v12/pedregosa11a.html

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

27. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '16. (New York, NY, USA: Association for Computing Machinery), 785–794. doi:10.1145/2939672.2939785
28. Breiman L. Random Forrests. Mach Learn (2001) 45:5–32. Available at: https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf
29. Hu C-AA, Chen C-M, Fang Y-C, Liang S-J, Wang H-C, Fang W-F, Sheu C-C, Perng W-C, Yang K-Y, Kao K-C, et al. Using a machine learning approach to predict mortality in critically ill sectional influenza patients : a cross- - retrospective multicentre study in Taiwan. BMJ Open (2020) 10:1–10. doi:10.1136/bmjopen-2019-033898
30. Sauthier MS, Jouvet PA, Newhams, MM, Randolph AG. Machine Learning Predicts Prolonged Acute Hypoxemic Respiratory Failure in Pediatric Severe Influenza. Crit Care Explor (2020) 2:e0175. doi:10.1097/cce.0000000000000175
31. Zhang P. Model Selection Via Multifold Cross Validation Author (s): Ping Zhang Source : The Annals of Statistics , Vol . 21 , No . 1 (Mar ., 1993), pp . 299-313 Published by : Institute of Mathematical Statistics Stable URL : http://www.jstor.org/stable/303559. Ann Stat (1993) 21:299–313.
32. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat Med (2000) 19:1141–64. doi:10.1007/BF00743033
33. Khosravi A, Nahavandi S, Creighton D, Atiya AF. Comprehensive review of neural network-based prediction intervals and new advances. IEEE Trans Neural Networks (2011) 22:1341–1356. doi:10.1109/TNN.2011.2162110
34. Dutta S, Singh B, Chessell L, Wilson J, Janes M, McDonald K, Shahid S, Gardner VA, Hjartarson A, Purcha M, et al. Guidelines for feeding very low birth weight infants. Nutrients (2015) 7:423–42. doi:10.3390/nu7010423

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

35. Overman AE, Liu M, Kurachek SC, Shreve MR, Maynard RC, Mammel MC, Moore BM. Tracheostomy for Infants Requiring Prolonged Mechanical Ventilation: 10 Years' Experience. Pediatrics (2013) doi:10.1542/peds.2012-1943

36. Khemani RG, Smith LS, Zimmerman JJ, Erickson S, Pediatric Acute Lung Injury Consensus Conference G. Pediatric acute respiratory distress syndrome: definition, incidence, and epidemiology: proceedings from the Pediatric Acute Lung Injury Consensus Conference. Pediatr Crit Care Med (2015) 16:S23-40. doi:10.1097/PCC.0000000000000432

37. Namachivayam P, Shann F, Shekerdemian L, Taylor A, van Sloten I, Delzoppo C, Daffey C, Butt W. Three decades of pediatric intensive care: Who was admitted, what happened in intensive care, and what happened afterward*. Pediatr Crit Care Med (2010) 11:549–555. doi:10.1097/PCC.0b013e3181ce7427

38. Kawaguchi A, Yasui Y, deCaen A, Garros D. The Clinical Impact of Heated Humidified High-Flow Nasal Cannula on Pediatric Respiratory Distress. Pediatr Crit care Med a J Soc Crit Care Med World Fed Pediatr Intensive Crit Care Soc (2017) 18:112–119. doi:10.1097/PCC.0000000000000985

Version non définitive. La version définitive de cet article (Sauthier M, Sauthier N, Bergeron Gallant K, Lodygensky GA, Atsushi K, Emeriaud G, et al. Long Term Mechanical Ventilation in Neonates: a 10-year Overview and a Predictive Model. *Front Pediatr*. 2021;9(July):1–10.) est disponible au https://doi.org/10.3389/fped.2021.689190.

Discussion

Dans cet article, nous montrons l'application de la définition proposée de la ventilation prolongée dans une population diversifiée de nouveau-nés. Cette étude apporte une importante contribution descriptive dans un domaine très peu documenté avec une seule autre étude retrouvée sur une population similaire dans les 35 dernières années (100). D'abord, on remarque que la définition de ventilation prolongée que nous avions proposée pour une population néonatale, soit l'utilisation d'un support respiratoire entre 40 et 44 semaines d'âge postmenstruel depuis au moins 21 jours de ventilation, sélectionne des nouveau-nés avec une atteinte particulièrement grave et généralement ventilés de façon beaucoup plus prolongée que 21 jours (100). Ensuite, cette étude met en évidence que cette définition correspond à une importante mortalité et morbidité à moyen terme, tant sur les plans respiratoire, neurologique et digestif. En particulier, on rapporte une mortalité de 29 %, largement au-dessus de la mortalité observée en soins intensifs pédiatriques qui est inférieure à 5 % (101). On rapporte également chez les survivants à 18 mois d'âge corrigé que 41 % dépendent encore d'un support respiratoire, 55 % nécessitent une alimentation entérale et que la moitié des enfants avaient besoin de classes spécialisées.

Ces résultats sont importants, car certains traitements invasifs, comme la trachéostomie, ont des indications encore fragiles dans le domaine de la ventilation prolongée pédiatrique. Cela est possiblement en raison du manque de définition et de pronostic de la ventilation prolongée pédiatriques. Chez les adultes, la trachéostomie est reconnue pour diminuer les médicaments anesthésiants et permettre une meilleure mobilité et nutrition (102,103). En pédiatrie, des bénéfices similaires ont été rapportés par des équipes expérimentées (104), mais sont également plusieurs effets indésirables dans une population néonatale, notamment l'hypertension pulmonaire, le reflux gastro-œsophagien, l'alimentation entérale et les trachéites (105,106). De plus, leur séjour hospitalier était significativement prolongé. Dans le contexte d'étude rétrospective, il est difficile d'identifier la contribution de la trachéostomie par rapport à la maladie de base, mais il est tout à fait envisageable que ces éléments soient plutôt liés à la gravité de l'atteinte respiratoire (et globale) sous-jacente. De plus, les bénéfices démontrés l'étaient dans des centres qui, en plus de la trachéostomie, soutenaient le développement et avaient un programme de réadaptation pour ces patients. Le développement de modèles prédictifs performants permettrait de cibler les enfants les

plus à risque de ventilation prolongée et de considérer des interventions précoces sur les plans respiratoires, neurologiques et digestifs.

Dans notre étude, nous avons exploré les facteurs de risque associé avec les temps les plus longs de ventilation de notre cohorte (le 4^{e} quartile, soit $\geq$ 125 jours) et comme retrouvé dans la littérature, ce sous-groupe était également associé aux atteintes les plus importantes. Afin d'identifier précocement ce sous-groupe le plus grave, soit à maximum 4 semaines d'âge corrigé dès le diagnostic de la ventilation prolongée, nous avons utilisé plusieurs méthodes d'apprentissage automatique (régression pénalisée, forêts aléatoires et XGBoost), tous reconnus performant dans le domaine de la classification médicale (107,108). Toutefois, le petit échantillon (164 patients) dont nous disposions a possiblement limité l'apprentissage avec une discrimination du modèle qui s'est avérée faible (AUROC 0,65). Une validation croisée par une méthode de Monte-Carlo a pu établir des intervalles de confiance à partir d'une population relativement limitée. Au niveau explicatif, cette étude a pu identifier quelques facteurs contributifs, notamment la pression partielle en gaz carbonique dans le sang (PCO_2), la FiO_2 et le terme. Cependant, ces facteurs ne sont pas modifiables et ne sont donc pas de potentielles cibles thérapeutiques.

De plus, bien que cette étude soit limitée à une population de nouveau-né, d'autres études ont montré qu'une définition semblable était applicable chez les enfants plus âgés (97) et comportait également une importante comorbidité. La ventilation prolongée est donc un critère de jugement grave qui représente un indicateur important à suivre et à prédire pour l'IRHA dans un SADC.

Discussion générale

Cette thèse fournit différents outils informatiques nécessaires à la réalisation de SADC automatiques permettant de reconnaitre l'IRHA en pédiatrie et d'en évaluer la gravité automatiquement. Elle contribue à améliorer le champ de connaissance dans le domaine, notamment par la réalisation d'un système continu de collecte et de gestion des données à haute résolution temporelle en soins intensifs. Elle fournit également une méthode continue d'évaluation de l'IRHA utilisable sans limite supérieure de SpO_2, un outil d'évaluation automatique de la gravité globale des patients en IRHA, démontre un modèle prédictif de l'IRHA prolongée et présente une définition pédiatrique de ventilation prolongée en pédiatrie, utilisable comme critère de jugement pour l'IRHA.

Ce travail apporte une contribution dans l'automatisation des processus autour du diagnostic de l'IRHA en levant plusieurs barrières, notamment le problème des données corrompues, soit les valeurs erronées, manquantes ou imprécises, qui sont un frein majeur aux développements des SADC (18). La reconnaissance précoce de l'IRHA est un élément clé pour plusieurs entités, mais particulièrement pour sa forme la plus grave : le SDRA. La reconnaissance précoce pourrait permettre de commencer des traitements plus adaptés à la sous-population la plus à risque, et ainsi éviter d'entrer dans un cercle vicieux ou les traitements, notamment la ventilation mécanique, est également responsable des lésions pulmonaires secondaires (1). Cette hypothèse doit encore être démontrée et certains outils présentés dans cette thèse vont permettre le développement de modèles prédictifs spécifiquement pour cette question. De plus, le développement de critères de jugement différents de la mortalité vient également contribuer positivement au domaine de l'IRHA. Le fait que les seuils de gravité de l'IHRA (légère, modérée ou grave) aient été exclusivement basés sur la mortalité est aujourd'hui inadapté à une population de plus en plus complexe, avec une mortalité en baisse, mais dont les séquelles chez les survivants commencent à peine à être reconnus (6,7). Les travaux présentés montrent également que la ventilation prolongée dans une population vulnérable de nouveau-né est un problème majeur avec une consommation de ressources disproportionnées par rapport aux autres patients. L'impact à long terme est encore inconnu, ce qui

est d'autant plus inquiétant dans une très jeune population de nouveau-nés. Le besoin de pronostication pour cibler des interventions précoces dans ces populations est donc aussi évident.

Ce travail ouvre également la voie à l'importance d'utiliser des données continues à haute résolution temporelle. Au moment d'écrire cette thèse (2021), l'étude de l'évolution continue des patients avec SDRA en est à ses tout débuts avec 1 point par 6 h (4,17), mais la présence de $SpO_2 > 97$ % ont fortement limité l'interprétation de ces données et l'étude des trajectoires des patients. Grâce à la réalisation de la base de données à haute résolution et au marqueur continu reposant sur la $ePaO_2$, une vision plus complète et plus juste de l'évolution de l'IRHA pourrait être appréciée par rapport à des synthèses agressives (par exemple, la valeur la plus anormale en 24 h qui ne distingue pas la progression de la dégradation). Les travaux inclus dans cette thèse rendent possible l'étude de l'oxygénation à chaque seconde et sans limites supérieures de SpO_2 ainsi que la mise en relation de plusieurs signaux bruts avec le degré d'hypoxémie. Toutefois, l'utilisation et l'interprétation de données continues sont plus complexes et une collaboration sera nécessaire entre des experts, entre autres, en sciences des données, en physiologie et en médecine.

Toutefois, malgré tous les efforts déployés pour que les algorithmes développés dans le cadre de cette thèse soient généralisables au plus grand nombre de patients possible et applicables en toute circonstance, certaines limitations subsistent. D'abord, certaines métadonnées autour de la SpO_2, notamment le site anatomique de la mesure, la pigmentation de la peau, la présence de certaines hémoglobinopathies et la qualité de la perfusion sanguine, n'étaient pas disponibles en continu de façon suffisamment fiable pour être intégré dans les modèles développé dans ce travail (en particulier la pigmentation de la peau). Ceci est important, car comme discuté plus haut, une pigmentation plus foncée de la peau semble associée à une sous-estimation de l'hypoxémie mesurée par SpO_2 chez l'adulte et probablement chez l'enfant aussi (76–79). De plus, dans l'étude de la $ePaO_2$, les 2 unités de soins intensifs pédiatriques ayant participé à l'étude utilisaient des saturomètre Masimo limitant la généralisation à d'autres marques.

L'évaluation de la gravité de l'hypoxémie est également très complexe chez les enfants porteurs d'une cardiopathie cyanogènes, même si la lecture de la SpO_2 est généralement fiable dans les valeurs basses (60 % – 80 %) (3,109). Il est difficile, voire impossible, de faire la différence entre le shunt droit-gauche et une diminution de la capacité d'oxygénation des poumons seulement en mesurant l'oxygénation périphérique. Bien que nous ayons développé les algorithmes nécessaires à la reconnaissance automatique des cardiopathies cyanogènes (11), il est probable qu'il faille développer des marqueurs indépendants de l'oxygénation pour évaluer la gravité de l'insuffisance respiratoire dans cette population. Par exemple, la puissance mécanique (110) ou la compliance pulmonaire (5,9) seraient des alternatives qui seront étudiées par les experts révisant les lignes de conduite. Également, la population avec insuffisance respiratoire chronique développant une exacerbation aigüe est aussi un défi à stratifier (5). Les lignes de conduite actuelles mentionnent simplement que la dégradation doit être « significative » par rapport à leur niveau de base. Grâce à l'indice d'oxygénation continu développé dans l'article 1, un seuil de détérioration relatif au niveau de base pourrait être déterminé et facilement quantifiable. Par ailleurs, ce travail ne prend pas en compte les assistances circulatoires extracorporelles, comme l'ECMO (*extracorporeal membrane oxygenation*). Ces thérapies de dernier recours viennent assister ou remplacer le poumon, faussant l'évaluation de la gravité par les indices d'hypoxémie mesurés en périphérie sur le patient. Dans l'article 3, et même si non spécifiés dans les lignes de conduite, nous avons classé d'emblée les patients assistés par ECMO au niveau maximal de gravité, quels que soient leurs indices d'oxygénation. Par ailleurs, les assistances ventriculaires, pulsatiles ou non, vont créer une discordance entre le pouls et le rythme cardiaque, ce qui implique que les mécanismes de validation de la SpO_2 développée dans l'article 1 ne sont pas utilisables dans cette population. Ces patients restent une minorité sur l'ensemble des enfants traités aux soins intensifs et ne touchent pas la phase précoce de l'IRHA à laquelle le projet de SADC se destine. Dernièrement, les développements proposés dans cette thèse présupposent l'accès à un dossier médical informatique, dont l'implantation est encore très variable dans le monde.

Toutes ces limitations renforcent le fait qu'un système complètement autonome pour la reconnaissance et la prise en charge de l'IRHA pédiatrique n'est pas encore possible et que seul un système d'*aide* au diagnostic de l'IRHA pour le soignant est réalisable aujourd'hui. Toutefois, cette

étape est cruciale pour la suite qui est de développer un valider un système d'aide à la décision pour respecter les recommandations en fonction du diagnostic fait automatiquement et potentiellement d'améliorer significativement les soins prodigués à cette population.

Perspectives

Cette thèse fournit plusieurs outils nécessaires à l'évaluation continue de l'IRHA en pédiatrie. La suite logique est d'abord d'étudier cet aspect continu de l'hypoxémie, encore très peu documentée dans la littérature. L'utilisation de la $ePaO_2$ continue ouvre 3 champs principaux qui visent à mieux préciser la gravité de l'IRHA et donc à pouvoir personnaliser les interventions. Le premier est la nécessité de décrire précisément les variations de l'hypoxémie et de pouvoir faire la différence sur le devenir, par exemple, entre une atteinte hypoxémique profonde, mais relativement isolée et des atteintes légères répétées ou prolongées. Autrement formulée, est-ce que les atteintes hypoxémiques sont cumulatives ? Est-ce qu'il existe un seuil de tolérance et est-ce que celui-ci varie selon les particularités du patient ? Ces questions permettront de personnaliser les recommandations qui sont aujourd'hui encore très générales. Cela permettra également de déterminer des cibles de conformité aux lignes de conduite d'abord en proportion du temps où la recommandation est respectée (**Figure 5**), qui sera ensuite modulée par l'importance de la déviation à la recommandation. Le second champ est le développement de profils informatiques (*computable phenotypes* (111)) pour l'IRHA. La réversibilité du dérecrutement pulmonaire, l'aspect très aigu de l'obstruction endotrachéale ou l'hypoxémie plus grave du SDRA ont des profils certainement différents, en particulier lorsque des données multimodales continues sont utilisées (par exemple, variations spontanées ou après intervention des pressions inspiratoires ou des volumes). Il est également probable que même parmi le SDRA, différents profils existent. Les technologies d'apprentissage automatique non supervisées pourront aider dans ce domaine. Le troisième champ est l'anticipation des évènements. Une fois les profils établis, il deviendra possible de les prédire (apprentissage supervisé) et éventuellement d'agir en amont. Dans le cadre du traitement de signal, anticiper un évènement fait appel aux mêmes technologies que l'identification d'artéfacts ou l'imputation de données manquantes (18). La $ePaO_2$ et les indices continus qui en découlent pourraient permettre de solutionner plusieurs problèmes d'un coup.

La seconde étape sera le développement d'un SADC spécifique aux pathologies les plus graves : le SDRA et la pneumonie acquise en ventilation mécanique (5,112). Ces derniers comportent d'autres critères diagnostics que l'hypoxémie (**Tableau 1**), mais l'IHRA est le critère majeur et

sine qua none de ces entités. Parmi les autres critères, certains seront plus faciles à automatiser que d'autres. Par exemple, l'interprétation de la radiographie est déjà possible de façon automatique dans un contexte autre que le SDRA (113) et sera certainement utilisable de façon pleinement automatique vu les progrès réalisés avec l'apprentissage profond dans le diagnostic automatique de la radiographie de thorax (114). Toutefois, le défi reste d'obtenir des jeux de données assez grands et correctement étiquetés, sachant que l'étalon d'or dans le domaine, la lecture par la radiologie, ne bénéficie pas d'une congruence parfaite parmi les experts (113). En revanche, d'autres critères du SDRA pédiatrique nécessiteront un travail pour définir certains concepts d'une façon plus précise avant de pouvoir les automatiser. Par exemple, dans le SDRA pédiatrique, il est dit que l'œdème pulmonaire « ne doit pas être pleinement expliqué par l'insuffisance cardiaque ou la surcharge liquidienne » (5). Une telle définition laisse place à l'interprétation qui nécessitera d'être clarifié, possiblement par la révision en cours des lignes directrices. Dernièrement, le SADC doit évoluer et laisser une place pour interagir avec le soignant qui sera en mesure de préciser les critères les plus subjectifs. La réalisation de l'interface du SADC est déjà en cours (**Figure 5**) et repose en premier lieu sur l'évaluation et la classification en temps réel de la gravité de l'IRHA, afin d'établir la conformité des autres paramètres. On retrouve dans la littérature plusieurs exemples d'automatisation dans le domaine de l'hypoxémie et du SDRA. Ces systèmes augmentent souvent les performances cliniques (diagnostiques ou thérapeutique) en contexte de recherche. Toutefois, ils sont soit inapplicables en temps réel (analyse des notes médicales ou résultats d'hypoxémie en basse résolution), ou ne fournissent qu'une analyse parcellaire (analyse isolée de l'imagerie), ou requiert une intervention humaine avant toute suggestion (12–15,23,24,26–29,38,115). De plus, la grande majorité été développé pour une population adulte et aucun ne fournit de solution applicable en continu ou permet de traiter les hautes valeurs de SpO_2.

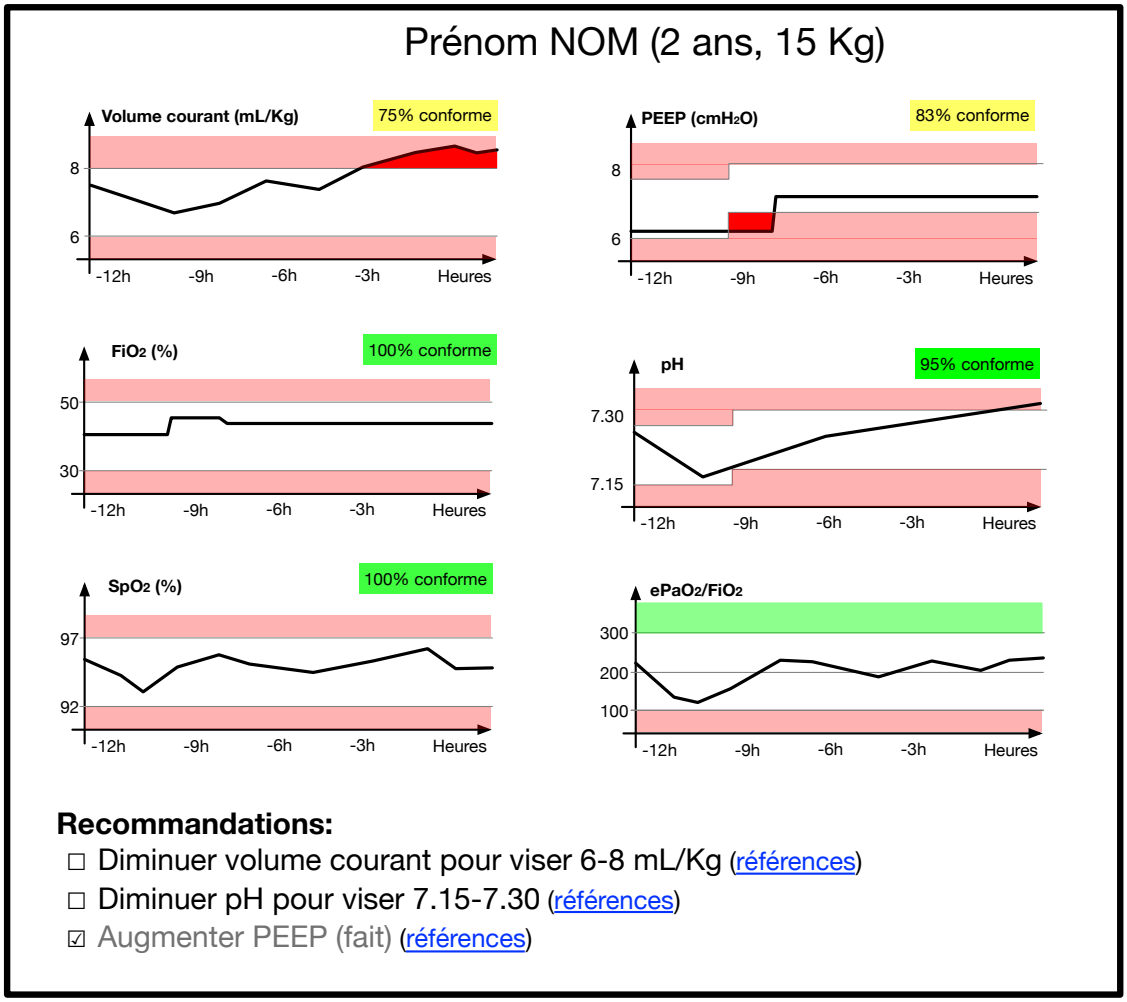


Figure 5. – Maquette d'interface montrant la conformité en temps réel selon la gravité de l'hypoxémie

La validation d'un SADC est complexe et se fait en plusieurs étapes. Le processus nécessite une équipe multidisciplinaire qui est déjà regroupée au CHU Sainte-Justine : médecins experts de la pathologie, pédagogues, informaticiens, ergonomes, biostatisticiens et ingénieurs. La phase pilote est essentielle pour construire un SADC dont l'ergonomie soit acceptée par l'équipe traitante. Comme les SADC sont destinés à interagir avec le comportement des soignants, il est essentiel qu'ils soient acceptés par l'équipe qui devra fournir des efforts pour intégrer l'outil à leur pratique dans le cadre de l'étude de validation. En revanche, une fois l'étude terminée, si les résultats sont

favorables, l'implantation de l'outil est immédiate, contrairement aux études cliniques qui nécessitent plusieurs années supplémentaires (20). La diffusion du SADC à d'autres centres sera assurée par l'étude d'efficacité multicentrique. En effet, la validation de l'efficacité d'un SADC nécessite de séparer les bénéfices apportés par l'outil de ceux résultant de l'apprentissage par le soignant. Pour diminuer la contamination entre les 2 groupes (avec et sans SADC), on réalisera une étude multicentrique avec une randomisation en grappes, soit des périodes aléatoires avec et sans SADC pour l'ensemble de l'équipe (116). Les critères de jugements seront ceux développés dans cette thèse : aPELOD-2 et ventilation prolongée à l'échelle de l'unité de soins clinique.

Une fois que le SADC, permettant de reconnaitre précocement le SDRA et d'augmenter la conformité aux lignes directrice, sera réalisé, le deuxième champ de développement sera l'anticipation de la gravité et des évènements indésirables. Les lignes de conduite actuelles expliquent comment se positionner en fonction de la gravité de l'IRHA à l'instant même. Une prochaine étape, dont le bénéfice sera à démontrer, sera d'agir avant la survenue d'un IRHA grave. Plusieurs approches ont déjà été développées pour prédire le résultat à la suite d'un changement des paramètres de l'oxygénation (10,11). En particulier, l'équation d'Al-Otaibi (**Figure 6**) (10) peut être solutionnée pour estimer la FiO_2 pour une PaO_2 cible (**Figure 7**). Ces résultats permettraient possiblement d'ajuster la pression expiratoire positive selon les grilles des lignes de conduite (16) avant la dégradation de l'IHRA et de diminuer les séquelles pulmonaires. Toutefois, ces équations devront d'abord être revalidées en pédiatrie et en utilisant la $ePaO_2$ continue plutôt que la PaO_2.

$$P_{new} = \frac{F_{new} \cdot P_{old}}{F_{old}} \cdot \left[1 - \frac{F_{old} - F_{new}}{2}\right]$$

P : PaO_2, F : FiO_2

Figure 6. – Équation originale d'Al-Otaibi (2011)

$$F_{new} = \frac{(-2 \cdot P_{old}) + (P_{old} \cdot F_{old}) + \sqrt{(4 \cdot P_{old}^2 - (4 \cdot P_{old}^2 \cdot F_{old} + P_{old}^2 \cdot F_{old}^2) + (8 \cdot P_{old} \cdot P_{new} \cdot F_{old}))}}{2 \cdot P_{old}}$$

P : PaO_2, F : FiO_2

Figure 7. – Solution proposée de l'équation d'Al-Otaibi pour estimer la FiO_2 nécessaire pour une certaine PaO_2 cible

L'anticipation est la clé pour envisager des traitements précoces et personnalisés, mais plusieurs étapes de validation sont encore nécessaires. D'abord, l'obtention de modèles hautement performants est capitale à une personnalisation adéquate des soins, en particulier quand les traitements (ventilation mécanique) comportent des risques. Ensuite, l'hypothèse d'un bénéfice à agir plus tôt dans une population à risque de dégradation de l'IRHA doit encore être démontrée, par rapport à un système simplement réactif.

Conclusion

Cette thèse présente des outils informatiques nécessaires à la réalisation d'un SADC en temps réel permettant de reconnaitre l'IRHA, de classifier sa gravité et de fournir un pronostic. D'abord, on a expliqué la réalisation et le déploiement d'une base de données à haute résolution temporelle, nécessaire à la réalisation de SADC en temps réel. Ensuite, on a détaillé le développement et la validation d'un outil d'estimation continue de la gravité de l'IHRA à partir de valeurs couramment mesurées en soins intensifs et valides sans limites supérieures de SpO_2. Nous avons validé l'automatisation d'un score permettant de quantifier la gravité globale des patients en IRHA et avons proposé et validé la première définition de la ventilation prolongée en pédiatrie. Chaque élément présenté dans cette thèse a été développé avec le souci de pouvoir être automatisé et inclus dans un SADC visant à améliorer le devenir des patients en IRHA.

Références bibliographiques

1. Matthay MA, Zemans RL, Zimmerman GA, Arabi YM, Beitler JR, Mercat A, et al. Acute respiratory distress syndrome. Nature reviews Disease primers. 2019 Mar 14;5(1):18.

2. Collins J-A, Rudenski A, Gibson J, Howard L, O'Driscoll R. Relating oxygen partial pressure, saturation and content: the haemoglobin–oxygen dissociation curve. Breathe. 2015 Sep;11(3):194.

3. Eytan D, Mazwi ML, Goodwin AJ, Goodfellow SD, Assadi A, Greer RW, et al. Revisiting oxygen dissociation curves and bedside measured arterial saturation in critically ill children. Intensive Care Medicine. 2019;45(12):1832–4.

4. Khemani RG, Smith L, Lopez-Fernandez YM, Kwok J, Morzov R, Klein MJ, et al. Paediatric acute respiratory distress syndrome incidence and epidemiology (PARDIE): an international, observational study. The Lancet Respiratory Medicine. 2018;2600(18):1–14.

5. Khemani RG, Smith LS, Zimmerman JJ, Erickson S, Pediatric Acute Lung Injury Consensus Conference G. Pediatric acute respiratory distress syndrome: definition, incidence, and epidemiology: proceedings from the Pediatric Acute Lung Injury Consensus Conference. Pediatr Crit Care Med. 2015;16(5 Suppl 1):S23-40.

6. Ducharme-Crevier L, La K-A, Francois T, Gerardis G, Beauchamp M, Harrington K, et al. PICU Follow-Up Clinic: Patient and Family Outcomes 2 Months After Discharge. Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies. 2021 Jun;

7. Boucher V, Mathy C, Lacroix J, Émériaud G, Jouvet P, Tse SM. Post-discharge respiratory outcomes of children with acute respiratory distress syndrome. Pediatric Pulmonology. 2020;55(2):468–73.

8. Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, et al. Report of the American-European consensus conference on ARDS: Definitions, mechanisms, relevant outcomes and clinical trial coordination. Intensive Care Medicine. 1994;20(3):225–32.

9. Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, et al. Acute respiratory distress syndrome: The Berlin definition. JAMA - Journal of the American Medical Association. 2012/07/17. 2012;307(23):2526–33.

10. Al-Otaibi HM, Hardman JG. Prediction of arterial oxygen partial pressure after changes in FI O2: Validation and clinical application of a novel formula. British Journal of Anaesthesia. 2011;107(5):806–12.

11. Ghazal S, Sauthier M, Brossier D, Bouachir W, Jouvet PA, Noumeir R. Using machine learning models to predict oxygen saturation following ventilator support adjustment in critically ill children: a single center pilot study. bioRxiv. 2019;14(2):334896.

12. Koenig HC, Finkel BB, Khalsa SS, Lanken PN, Prasad M, Urbani R, et al. Performance of an automated electronic acute lung injury screening system in intensive care unit patients. Critical Care Medicine. 2011;39(1):98–104.

13. Herasevich V, Yilmaz M, Khan H, Hubmayr RD, Gajic O. Validation of an electronic surveillance system for acute lung injury. Intensive care medicine. 2009 Jun;35(6):1018–23.

14. Ahmed A, Thongprayoon C, Pickering BW, Akhoundi A, Wilson G, Pieczkiewicz D, et al. Towards prevention of acute syndromes: Electronic identification of at-risk patients during hospital admission. Applied Clinical Informatics. 2014;5(1):58–72.

15. Azzam HC, Khalsa SS, Urbani R, Shah C V., Christie JD, Lanken PN, et al. Validation Study of an Automated Electronic Acute Lung Injury Screening Tool. Journal of the American Medical Informatics Association. 2009;16(4):503–8.

16. Khemani RG, Parvathaneni K, Yehya N, Bhalla AK, Thomas NJ, Newth CJL. Positive end-expiratory pressure lower than the ards network protocol is associated with higher pediatric acute respiratory distress syndrome mortality. American Journal of Respiratory and Critical Care Medicine. 2018;

17. Bhalla AK, Klein MJ, Emeriaud G, Lopez-Fernandez YM, Napolitano N, Fernandez A, et al. Adherence to Lung-Protective Ventilation Principles in Pediatric Acute Respiratory Distress Syndrome: A Pediatric Acute Respiratory Distress Syndrome Incidence and Epidemiology Study. Critical Care Medicine. 9000;

18. Johnson AE, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine Learning and Decision Support in Critical Care. Proc IEEE Inst Electr Electron Eng. 2016/10/22. 2016;104(2):444–66.

19. Bates DW, Gawande AA. Improving Safety with Information Technology. New England Journal of Medicine. 2003;

20. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. Lancet. 2003/10/22. 2003;362(9391):1225–30.

21. Kwan JL, Lo L, Ferguson J, Goldberg H, Diaz-Martinez JP, Tomlinson G, et al. Computerised clinical decision support systems and absolute improvements in care: Meta-analysis of controlled clinical trials. The BMJ. 2020;370:m3216.

22. Newth CJL, Khemani RG, Jouvet PA, Sward KA. Mechanical Ventilation and Decision Support in Pediatric Intensive Care. Vol. 64, Pediatric Clinics of North America. 2017. p. 1057–70.

23. Finlay-Morreale HE, Louie C, Toy P. Computer-generated Automatic Alerts of Respiratory Distress after Blood Transfusion. Journal of the American Medical Informatics Association. 2008;15(3):383–5.

24. Afshar M, Joyce C, Oakey A, Formanek P, Yang P, Churpek MM, et al. A Computable Phenotype for Acute Respiratory Distress Syndrome Using Natural Language Processing and Machine Learning. AMIA . Annual Symposium proceedings AMIA Symposium. 2018;2018:157–65.

25. Yang P, Wu T, Yu M, Chen F, Wang C, Yuan J, et al. A new method for identifying the acute respiratory distress syndrome disease based on noninvasive physiological parameters. PLoS ONE. 2020;15(2):1–21.

26. Taoum A, Mourad-Chehade F, Amoud H. Evidence-based model for real-time surveillance of ARDS. Biomedical Signal Processing and Control [Internet]. 2019;50:83–91. Available from: https://doi.org/10.1016/j.bspc.2019.01.016

27. Solti I, Cooke CR, Xia F, Wurfel MM. Automated Classification of Radiology Reports for Acute Lung Injury: Comparison of Keyword and Machine Learning Based Natural

Language Processing Approaches. Proceedings IEEE International Conference on Bioinformatics and Biomedicine. 2009 Nov;2009:314–9.

28. Sabeti E, Drews J, Reamaroon N, Warner E, Sjoding MW, Gryak J, et al. Learning Using Partially Available Privileged Information and Label Uncertainty: Application in Detection of Acute Respiratory Distress Syndrome. IEEE Journal of Biomedical and Health Informatics. 2021;25(3):784–96.

29. Mayampurath A, Churpek MM, Su X, Shah S, Munroe E, Patel B, et al. External Validation of an Acute Respiratory Distress Syndrome Prediction Model Using Radiology Reports. Critical Care Medicine. 2020;(14):791–8.

30. Bourdeaux CP, Thomas MJC, Gould TH, Malhotra G, Jarvstad A, Jones T, et al. Increasing compliance with low tidal volume ventilation in the ICU with two nudge-based interventions: Evaluation through intervention time-series analyses. BMJ Open. 2016;6(5):6–11.

31. Castellanos I, Martin M, Kraus S, Bürkle T, Prokosch HU, Schüttler J, et al. Effects of staff training and electronic event monitoring on long-term adherence to lung-protective ventilation recommendations. Journal of Critical Care. 2018;43:13–20.

32. Eslami S, de Keizer NF, Abu-Hanna A, de Jonge E, Schultz MJ. Effect of a clinical decision support system on adherence to a lower tidal volume mechanical ventilation strategy. Journal of Critical Care. 2009;24(4):523–9.

33. Eslami S, Abu-Hanna A, Schultz MJ, de Jonge E, de Keizer NF. Evaluation of consulting and critiquing decision support systems: Effect on adherence to a lower tidal volume mechanical ventilation strategy. Journal of Critical Care. 2012;27(4):425.e1-425.e8.

34. Bagga S, Paluzzi DE, Chen CY, Riggio JM, Nagaraja M, Marik PE, et al. Better ventilator settings using a computerized clinical tool. Respiratory Care. 2014;59(8):1172–7.

35. Blum JM, Stentz MJ, Maile MD, Jewell E, Raghavendran K, Engoren M, et al. Automated alerting and recommendations for the management of patients with preexisting hypoxia and potential acute lung injury: A pilot study. Anesthesiology. 2013;119(2):295–302.

36. McKinley BA, Moore FA, Sailors RM, Cocanour CS, Marquez A, Wright RK, et al. Computerized decision support for mechanical ventilation of trauma induced ARDS: Results of a randomized clinical trial. Journal of Trauma - Injury, Infection and Critical Care. 2001;50(3):415–25.

37. Walsh BK, Smallwood C, Rettig J, Kacmarek RM, Thompson J, Arnold JH. Daily goals formulation and enhanced visualization of mechanical ventilation variance improves mechanical ventilation score. Respiratory Care. 2017;62(3):268–78.

38. Hotz JC, Bornstein D, Kohler K, Smith E, Suresh A, Klein M, et al. Real-Time Effort Driven Ventilator Management: A Pilot Study∗. Pediatric Critical Care Medicine. 2020;933–40.

39. Herasevich V, Tsapenko M, Kojicic M, Ahmed A, Kashyap R, Venkata C, et al. Limiting ventilator-induced lung injury through individual electronic medical record surveillance. Critical Care Medicine. 2011;39(1):34–9.

40. Jouvet PA, Payen V, Gauvin F, Emeriaud G, Lacroix J. Weaning children from mechanical ventilation with a computer-driven protocol: a pilot trial. Intensive care medicine. 2013/01/31. 2013 May;39(5):919–25.

41. Arnal J-M, Wysocki M, Novotni D, Demory D, Lopez R, Donati S, et al. Safety and efficacy of a fully closed-loop control ventilation (IntelliVent-ASV®) in sedated ICU patients with acute respiratory failure: a prospective randomized crossover study. Intensive care medicine. 2012 May;38(5):781–7.

42. Jouvet P, Hernert P, Wysocki M. Development and implementation of explicit computerized protocols for mechanical ventilation in children. Ann Intensive Care. 2011/12/23. 2011;1(1):51.

43. Lellouche F, Bouchard P-A, Simard S, L'Her E, Wysocki M. Evaluation of fully automated ventilation: a randomized controlled study in post-cardiac surgery patients. Intensive care medicine. 2013 Mar;39(3):463–71.

44. Rose L, Schultz MJ, Cardwell CR, Jouvet P, McAuley DF, Blackwood B. Automated versus non-automated weaning for reducing the duration of mechanical ventilation for critically ill

adults and children. The Cochrane database of systematic reviews. 2014 Jun;2014(6):CD009235.

45. Lellouche F, Bouchard PA, Roberge M, Simard S, L'Her E, Maltais F, et al. Automated oxygen titration and weaning with FreeO2 in patients with acute exacerbation of COPD: A pilot randomized trial. International Journal of COPD. 2016;11(1):1983–90.

46. Khemani RG, Patel NR, Bart RD, Newth CJLL, Bart 3rd RD, Newth CJLL. Comparison of the pulse oximetric saturation/fraction of inspired oxygen ratio and the Pao2/fraction of inspired oxygen ratio in children. Chest. 2008/11/26. 2009;135(3):662–8.

47. Gadrey SM, Lau CE, Clay R, Rhodes GT, Lake DE, Moore CC, et al. Imputation of partial pressures of arterial oxygen using oximetry and its impact on sepsis diagnosis. Physiological Measurement. 2019 Oct 25;40(11):115008.

48. Khemani RG, Rubin S, Belani S, Leung D, Erickson S, Smith LS, et al. Pulse oximetry vs. PaO2 metrics in mechanically ventilated children: Berlin definition of ARDS and mortality risk. Intensive Care Med. 2014/09/19. 2015;41(1):94–102.

49. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial Bias in Pulse Oximetry Measurement. New England Journal of Medicine [Internet]. 2020;383(25):2477–8. Available from: https://doi.org/10.1056/NEJMc2029240

50. Bizzarro MJ, Li FY, Katz K, Shabanova V, Ehrenkranz RA, Bhandari V. Temporal quantification of oxygen saturation ranges: an effort to reduce hyperoxia in the neonatal intensive care unit. Journal of perinatology : official journal of the California Perinatal Association. 2014 Jan;34(1):33–8.

51. Jouvet P, Farges C, Hatzakis G, Monir A, Lesage F, Dupic L, et al. Weaning children from mechanical ventilation with a computer-driven system (closed-loop protocol): a pilot study. Pediatr Crit Care Med. 2007/08/19. 2007;8(5):425–32.

52. Rizzi M, Goldenberg N, Bonduel M, Revel-Vilk S, Amankwah E, Albisetti M. Catheter-Related Arterial Thrombosis in Neonates and Children: A Systematic Review. Thrombosis and haemostasis. 2018 Jun;118(6):1058–66.

53. Weiss IK, Fink S, Edmunds S, Harrison R, Donnelly K. Continuous arterial gas monitoring: initial experience with the Paratrend 7 in children. Intensive care medicine. 1996 Dec;22(12):1414–7.

54. Coule LW, Truemper EJ, Steinhart CM, Lutin WA. Accuracy and utility of a continuous intra-arterial blood gas monitoring system in pediatric patients. Critical care medicine. 2001 Feb;29(2):420–6.

55. Abraham E, Gallagher TJ, Fink S. Clinical evaluation of a multiparameter intra-arterial blood-gas sensor. Intensive care medicine. 1996 May;22(5):507–13.

56. Gamo M, Hirose Y, Matsuo K. [Problems in clinical use of intraarterial blood gas monitoring system, Paratrend 7]. Masui The Japanese journal of anesthesiology. 2000 Dec;49(12):1387–90.

57. Das J, Aggarwal A, Aggarwal NK. Pulse oximeter accuracy and precision at fifi ve different sensor locations in infants and children with cyanotic heart disease. Indian Journal of Anaesthesia [Internet]. 2010 Nov [cited 2021 Nov 11];54(6):531–4. Available from: www.ijaweb.org

58. Fitzgerald RK, Johnson A. Pulse oximetry in sickle cell anemia. Critical Care Medicine [Internet]. 2001;29(9). Available from: https://journals.lww.com/ccmjournal/Fulltext/2001/09000/Pulse_oximetry_in_sickle_cell_anemia.25.aspx

59. Zur B, Bagci S, Ludwig M, Stoffel-Wagner B. Oxygen Saturation in Pulse Oximetry in Hemoglobin Anomalies TT - Sauerstoffsättigungen in der Pulsoximetrie bei Hämoglobinanomalien. Klin Padiatr. 2012;224(04):259–65.

60. Giuliano KK, Higgins TL. New-Generation Pulse Oximetry in the Care of Critically Ill Patients. American Journal of Critical Care. 2005 Jan 1;14(1):26–37.

61. Hay WW, Rodden DJ, Collins SM, Melara DL, Hale KA, Fashaw LM. Reliability of Conventional and New Pulse Oximetry in Neonatal Patients. Journal of Perinatology 2002 22:5 [Internet]. 2002 Jun 27 [cited 2021 Nov 11];22(5):360–6. Available from: https://www.nature.com/articles/7210740

62. Dawson JA, Saraswat A, Simionato L, Thio M, Kamlin COF, Owen LS, et al. Comparison of heart rate and oxygen saturation measurements from Masimo and Nellcor pulse oximeters in newly born term infants. Acta Paediatrica [Internet]. 2013 Oct 1 [cited 2021 Nov 11];102(10):955–60. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/apa.12329

63. Hill AV. The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. PROCEEDINGS OF THE PHYSIOLOGICAL SOCIETY. 1910;40(supp):iv–vii.

64. Riviello ED, Kiviri W, Twagirumugabe T, Mueller A, Banner-Goodspeed VM, Officer L, et al. Hospital incidence and outcomes of the acute respiratory distress syndrome using the Kigali modification of the Berlin definition. American Journal of Respiratory and Critical Care Medicine. 2016 Jan;193(1):52–9.

65. Khemani RG, Thomas NJ, Venkatachalam V, Scimeme JP, Berutti T, Schneider JB, et al. Comparison of SpO 2 to PaO 2 based markers of lung disease severity for children with acute lung injury. Vol. 40, Critical Care Medicine. 2012. p. 1309–16.

66. Brown SM, Grissom CK, Moss M, Rice TW, Schoenfeld D, Hou PC, et al. Nonlinear Imputation of PaO2/FIO2 From SpO2/FIO2 Among Patients With Acute Respiratory Distress Syndrome. Chest. 2016;150(2):307–13.

67. Rice TW, Wheeler AP, Bernard GR, Hayden DL, Schoenfeld DA, Ware LB. Comparison of the SpO2/FIO2 ratio and the PaO 2/FIO2 ratio in patients with acute lung injury or ARDS. Chest. 2007;132(2):410–7.

68. Pandharipande PP, Shintani AK, Hagerman HE, St Jacques PJ, Rice TW, Sanders NW, et al. Derivation and validation of Spo2/Fio2 ratio to impute for Pao2/Fio2 ratio in the respiratory component of the Sequential Organ Failure Assessment score. Critical Care Medicine. 2009 Apr;37(4):1317–21.

69. Mackle D, Beasley R, Bellomo R, Bailey M, McGuinness S, Deane A, et al. Conservative oxygen therapy during mechanical ventilation in the ICU. New England Journal of Medicine. 2020;382(11):989–98.

70. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Scientific data. 2018;5:180178.

71. Mathieu A, Sauthier M, Jouvet P, Emeriaud G, Brossier D. Validation process of a high-resolution database in a paediatric intensive care unit—Describing the perpetual patient's validation. Journal of Evaluation in Clinical Practice. 2021 May 5;27(2):316–24.

72. Wetzel RC. First Get the Data, Then Do the Science!*. Pediatric Critical Care Medicine. 2018;19(4).

73. Brossier D, El Taani R, Sauthier M, Roumeliotis N, Emeriaud G, Jouvet P. Creating a high-frequency electronic database in the PICU: The perpetual patient. Pediatric Critical Care Medicine. 2018/02/07. 2018;19(4):e189–98.

74. Brossier D, Sauthier M, Mathieu A, Goyer I, Emeriaud G, Jouvet P. Qualitative subjective assessment of a high-resolution database in a paediatric intensive care unit—Elaborating the perpetual patient's ID card. Journal of Evaluation in Clinical Practice. 2020;26(1):86–91.

75. Nadkarni PM, Brandt C. Data extraction and ad hoc query of an entity-attribute-value database. Journal of the American Medical Informatics Association. 1998;5(6):511–27.

76. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. Anesthesiology. 2005;102(4):715–9.

77. Feiner JR, Severinghaus JW, Bickler PE. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: The effects of oximeter probe type and gender. Anesthesia and Analgesia. 2007;105(SUPPL. 6).

78. Vesoulis Z, Tims A, Lodhi H, Lalos N, Whitehead H. Racial discrepancy in pulse oximeter accuracy in preterm infants. Journal of Perinatology [Internet]. 2021;(September):1–7. Available from: http://dx.doi.org/10.1038/s41372-021-01230-3

79. Foglia EE, Whyte RK, Chaudhary A, Mott A, Chen J, Propert KJ, et al. The Effect of Skin Pigmentation on the Accuracy of Pulse Oximetry in Infants with Hypoxemia. The Journal of pediatrics [Internet]. 2016/12/09. 2017 Mar;182:375-377.e2. Available from: https://pubmed.ncbi.nlm.nih.gov/27939107

80. Leisman DE, Harhay MO, Lederer DJ, Abramson M, Adjei AA, Bakker J, et al. Development and Reporting of Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and Critical Care Journals. Critical Care Medicine [Internet]. 2020 Mar [cited 2020 Mar 7];1–11. Available from: http://journals.lww.com/10.1097/CCM.0000000000004246

81. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Quality and Safety. 2019;28(3):231–7.

82. Pena-Juarez RA, Chavez-Saenz JA, Garcfa-Canales A, Medina-Andrade MA, Martmez-Gonzalez MT, Gutierrez-Cobian L, et al. Comparison of oximeters for the detection of critical congenital heart diseases. Archivos de cardiologia de Mexico [Internet]. 2019 Apr 1 [cited 2021 Nov 12];89(1):159–66. Available from: https://pubmed.ncbi.nlm.nih.gov/31702739/

83. Thilo EH, Andersen D, Wasserstein ML, Schmidt J, Luckey D. Saturation by pulse oximetry: comparison of the results obtained by instruments of different brands. The Journal of pediatrics [Internet]. 1993 [cited 2021 Nov 12];122(4):620–6. Available from: https://pubmed.ncbi.nlm.nih.gov/7681875/

84. Szmuk P, Steiner JW, Olomu PN, Ploski RP, Sessler DI, Ezri T. Oxygen Reserve Index: A Novel Noninvasive Measure of Oxygen Reserve--A Pilot Study. Anesthesiology. 2016 Apr;124(4):779–84.

85. Leteurtre S, Duhamel A, Grandbastien B, Proulx F, Cotting J, Gottesman R, et al. Daily estimation of the severity of multiple organ dysfunction syndrome in critically ill children. Cmaj. 2015/09/16. 2010;182(11):1181–7.

86. Leteurtre S, Martinot A, Duhamel A, Proulx F, Grandbastien B, Cotting J, et al. Validation of the paediatric logistic organ dysfunction (PELOD) score: Prospective, observational, multicentre study. Lancet. 2003;362(9379):192–7.

87. Leteurtre S, Duhamel A, Salleron J, Grandbastien B, Lacroix J, Leclerc F. PELOD-2: An update of the PEdiatric logistic organ dysfunction score. Critical Care Medicine. 2013;41(7):1761–73.

88. Leclerc F, Duhamel A, Deken V, Le Reun C, Lacroix J, Leteurtre S, et al. 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(7):590–3.

89. Yehya N, Harhay MO, Klein MJ, Shein SL, Piñeres-Olave BE, Izquierdo L, et al. Predicting Mortality in Children with Pediatric Acute Respiratory Distress Syndrome: A Pediatric Acute Respiratory Distress Syndrome Incidence and Epidemiology Study. Critical Care Medicine. 2020;E514–22.

90. Liaw A, Wiener M. Classification and Regression by randomForest. R news. 2002;2(3):18–22.

91. Roca O, Caralt B, Messika J, Samper M, Sztrymf B, Hernández G, et al. An Index Combining Respiratory Rate and Oxygenation to Predict Outcome of Nasal High-Flow Therapy. American journal of respiratory and critical care medicine. 2019 Jun;199(11):1368–76.

92. Viglino D, L'her E, Maltais F, Maignan M, Lellouche F. Evaluation of a new respiratory monitoring tool "Early Warning ScoreO(2)" for patients admitted at the emergency department with dyspnea. Resuscitation. 2020 Mar;148:59–65.

93. Slater A, Straney L, Alexander J, Schell D, Millar J. The Effect of Imputation of PaO2/FIO2 From SpO2/FIO2 on the Performance of the Pediatric Index of Mortality 3. Pediatric Critical Care Medicine. 2020 Feb;21(6):520–5.

94. Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. Critical care medicine. 1988 Nov;16(11):1110–6.

95. 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 Pediatrics. 2017;171(10).

96. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. Critical care medicine. 1981;9(8):591–7.

97. Payen V, Jouvet P, Lacroix J, Ducruet T, Gauvin F. Risk factors associated with increased length of mechanical ventilation in children. Pediatr Crit Care Med. 2012;13(2):152–7.

98. MacIntyre NR, Epstein SK, Carson S, Scheinhorn D, Christopher K, Muldoon S, et al. Management of patients requiring prolonged mechanical ventilation: report of a NAMDRC consensus conference. Chest. 2005;128(6):3937–54.

99. Rose L, McKim DA, Katz SL, Leasa D, Nonoyama M, Pedersen C, et al. Home mechanical ventilation in Canada: A national survey. Respiratory Care. 2015;60(5):695–704.

100. Schreiner MS, Downes JJ, Kettrick RG, Ise C, Voit R. Chronic Respiratory Failure in Infants With Prolonged Ventilator Dependency. JAMA: The Journal of the American Medical Association. 1987 Dec 18;258(23):3398.

101. Pollack MM, Holubkov R, Funai T, Dean JM, Berger JT, Wessel DL, et al. The Pediatric Risk of Mortality Score: Update 2015. Pediatr Crit Care Med. 2016;17(1):2–9.

102. Andriolo BN, Andriolo RB, Saconato H, Atallah ÁN, Valente O. Early versus late tracheostomy for critically ill patients. Cochrane Database of Systematic Reviews. 2015;2017(6):CD007271.

103. Freeman BD. Tracheostomy Update: When and How. Crit Care Clin. 2017;33(2):311–22.

104. Overman AE, Liu M, Kurachek SC, Shreve MR, Maynard RC, Mammel MC, et al. Tracheostomy for infants requiring prolonged mechanical ventilation: 10 Years' experience. Pediatrics. 2013;131(5).

105. Spentzas T, Auth M, Hess P, Minarik M, Storgion S, Stidham G. Natural course following pediatric tracheostomy. Journal of Intensive Care Medicine. 2010;25(1):39–45.

106. Sauthier M, Bergeron-Gallant K, Emeriaud G, Jouvet P. Prognosis for newborns with prolonged mechanical ventilation: preliminary results of a retrospective single center cohort study [Abstract S42]. Annals of Intensive Care. 2017 Jan;7(1):7.

107. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

New York, NY, USA: Association for Computing Machinery; 2016. p. 785–794. (KDD '16).

108. Breiman L. Random Forrests. Machine learning. 2001;45(1):5–32.

109. Harris BU, Char DS, Feinstein JA, Verma A, Shiboski SC, Ramamoorthy C. Accuracy of Pulse Oximeters Intended for Hypoxemic Pediatric Patients. Pediatric Critical Care Medicine. 2016;17(4):315–20.

110. Gattinoni L, Tonetti T, Cressoni M, Cadringher P, Herrmann P, Moerer O, et al. Ventilator-related causes of lung injury: the mechanical power. Intensive Care Medicine. 2016;42(10):1567–75.

111. Geva A, Gronsbell JL, Cai T, Cai T, Murphy SN, Lyons JC, et al. A Computable Phenotype Improves Cohort Ascertainment in a Pediatric Pulmonary Hypertension Registry. J Pediatr. 2017/06/20. 2017;188:224-231 e5.

112. Chomton M, Brossier D, Sauthier M, Vallières E, Dubois J, Emeriaud G, et al. Ventilator-associated pneumonia and events in pediatric intensive care: A single center study. Pediatric Critical Care Medicine. 2018 Sep;19(12):1106–13.

113. Zaglam N, Jouvet P, Flechelles O, Emeriaud G, Cheriet F. Computer-aided diagnosis system for the Acute Respiratory Distress Syndrome from chest radiographs. Computers in Biology and Medicine. 2014/07/08. 2014;52:41–8.

114. Chen Y, Roberts CS, Ou W, Petigara T, Goldmacher G V, Fancourt N, et al. Deep learning for classification of pediatric chest radiographs by WHO's standardized methodology. PloS one. 2021;16(6):e0253239.

115. Yang Y, Huang Y, Gao F, Yuan L, Wang Z. Lung ultrasonography versus chest CT in COVID - 19 pneumonia : a two - centered retrospective comparison study from China. Intensive Care Medicine [Internet]. 2020;10–2. Available from: https://doi.org/10.1007/s00134-020-06096-1

116. Giraudeau B. L'essai clinique randomisé par grappes. Médecine/Sciences. 2004;20(3):363–6.