



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ

ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ



ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

**«ΜΕΘΟΔΟΛΟΓΙΑ ΒΙΟΪΑΤΡΙΚΗΣ ΕΡΕΥΝΑΣ, ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΚΑΙ
ΚΛΙΝΙΚΗ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»**

Μεταπτυχιακή Διπλωματική Εργασία

«Προγνωστική ικανότητα των δεικτών PIM-3 και PELOD-2 για τη θνητότητα σε Ελληνική ΜΕΘ Παίδων:
Ανάλυση διακριτικής ικανότητας και βαθμονόμησης»

“Predictive performance of PIM-3 and PELOD-2 scales for mortality in a Greek PICU: Discrimination and
calibration analysis”

ΓΡΗΓΟΡΙΟΣ ΚΟΥΡΤΗΣ

Επιβλέπων: Μπατσίδης Απόστολος

Αναπληρωτής Καθηγητής, Τμήμα Μαθηματικών, Πανεπιστήμιο Ιωαννίνων

Δεύτερος Επιβλέπων: Μπρότσης Θεόδωρος MSc, PhD

Εντεταλμένος Διδάσκων, Πανεπιστήμιο Θεσσαλίας

Τρίτος Επιβλέπων: Ζιντζαράς Ηλίας

Καθηγητής Βιομετρίας-Βιομαθηματικών, Τμήμα Ιατρικής, Πανεπιστήμιο Θεσσαλίας

ΛΑΡΙΣΑ, ΣΕΠΤΕΜΒΡΙΟΣ 2025

Ευχαριστίες

Ευχαριστώ θερμά τον επιβλέποντα κ. Απόστολο Μπατσίδα, Αναπληρωτή Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, για την αδιάλειπτη καθοδήγηση, τις καίριες παρατηρήσεις, τον χρόνο που αφιέρωσε, και τη σταθερή διαθεσιμότητά του σε κάθε στάδιο της μεταπτυχιακής διατριβής. Ευχαριστώ επίσης την κα Αριστούλα Πατσούρα, Διευθύντρια της Μονάδας Εντατικής Θεραπείας Παιδών του Γενικού Νοσοκομείου Παιδών Αθηνών «Παναγιώτης & Αγλαΐα Κυριακού», για την έγκριση και τη διευκόλυνση της συλλογής των δεδομένων και για την έμπρακτη στήριξη της ερευνητικής διαδικασίας. Τέλος, ευχαριστώ θερμά το ιατρικό και νοσηλευτικό προσωπικό της ίδιας Μονάδας για τη συνεργασία και τη συμβολή τους.

Περιεχόμενα

Περίληψη.....	σελ. 3
Abstract.....	σελ. 5
Εισαγωγή.....	σελ. 7
Ερευνητική Μεθοδολογία.....	σελ. 9
Αποτελέσματα.....	σελ. 14
Συζήτηση.....	σελ. 28
Περιορισμοί.....	σελ. 31
Συμπεράσματα.....	σελ. 31
Βιβλιογραφικές Αναφορές.....	σελ. 33

Προγνωστική ικανότητα των δεικτών PIM-3 και PELOD-2 για τη θνητότητα σε Ελληνική ΜΕΘ Παιδών: Ανάλυση διακριτικής ικανότητας και βαθμονόμησης

Περίληψη

Εισαγωγή: Η αξιόπιστη εκτίμηση του κινδύνου σε Παιδιατρική Μονάδα Εντατικής Θεραπείας (ΠΜΕΘ) αποτελεί θεμέλιο για τη λήψη κλινικών αποφάσεων, την επικοινωνία με τις οικογένειες και την αξιολόγηση της ποιότητας της παρεχόμενης φροντίδας.

Στόχοι: Η παρούσα μεταπτυχιακή διατριβή στοχεύει στην αξιολόγηση της απόδοσης των PIM-3 και PELOD-2 σε ελληνική ΠΜΕΘ ως προς τη διακριτική ικανότητα, τη βαθμονόμηση και στη συγκριτική ανάλυση διαγνωστικών δεικτών σε κλινικά χρήσιμα κατώφλια.

Μέθοδοι: Αναδρομική μονοκεντρική μελέτη σε 545 συνεχόμενες εισαγωγές παιδιών 0–16 ετών από τον Ιανουάριο 2020 έως Δεκέμβριο 2021. Η αξιολόγηση της διακριτικής ικανότητας πραγματοποιήθηκε με τη χρήση των καμπυλών ROC και της περιοχής κάτω από την καμπύλη (AUC). Η βαθμονόμηση ελέγχθηκε με πίνακες και διαγράμματα calibration, τον έλεγχο Hosmer–Lemeshow, καθώς και με τους δείκτες Brier score και Log-loss. Εφαρμόστηκε λογιστική επαναβαθμονόμηση ως προς τον σταθερό όρο και την κλίση. Τα βέλτιστα κατώφλια καθορίστηκαν με βάση τον δείκτη Youden J. **Αποτελέσματα:** Τα σκορ PIM-3 και PELOD-2 παρουσίασαν καλή διακριτική ικανότητα: PIM-3 AUC= 0,83 (95% ΔΕ: 0,76–0,91), PELOD-2 AUC= 0,85 (95% ΔΕ: 0,77–0,92), χωρίς στατιστικά σημαντική διαφορά μεταξύ τους (DeLong $p=0,67$). Πριν την επαναβαθμονόμηση παρατηρήθηκε υποεκτίμηση του κινδύνου, με λόγο παρατηρούμενων προς αναμενόμενων θανάτων (O/E) περίπου 1,8. Το τεστ Hosmer–Lemeshow ήταν στατιστικά σημαντικό τόσο για το PIM-3 ($\chi^2=24,40$ $df=3$ $p<0,001$) όσο και για το PELOD-2 ($\chi^2=42,07$ $df=3$ $p<0,001$), υποδεικνύοντας κακή βαθμονόμηση. Μετά την επαναβαθμονόμηση (ως προς τον σταθερό όρο και την κλίση), ο λόγος O/E βελτιώθηκε σε 1,00, ενώ η βαθμονόμηση κρίθηκε ικανοποιητική: PIM-3 ($\chi^2=2,41$, $df=3$, $p=0,493$), PELOD-2 ($\chi^2=2,70$, $df=3$, $p=0,440$). Επιπλέον, βελτιώθηκαν οι συνολικοί δείκτες ακρίβειας: Brier score 0,0568 και 0,0515, Log-loss 0,2065 και 0,1919 για τα PIM-3 και PELOD-2, αντίστοιχα.. Στο βέλτιστο κατώφλι βάσει του δείκτη Youden J, το PIM-3 εμφάνισε ευαισθησία 78% και ειδικότητα 77%,

ενώ το PELOD-2 είχε ευαισθησία 74% και υψηλότερη ειδικότητα 84%. Το PELOD-2 παρουσίασε μεγαλύτερο θετικό λόγο πιθανοφάνειας (LR^+) και οριακά υψηλότερη θετική προγνωστική αξία (PPV), ενώ και τα δύο σκορ διατήρησαν πολύ υψηλή αρνητική προγνωστική αξία ($NPV \approx 0,97$).

Συμπέρασμα: Τα PIM-3 και PELOD-2 διακρίνουν επαρκώς τους ασθενείς και μετά την επαναβαθμονόμηση παρέχουν αξιόπιστες, κλινικά χρήσιμες προβλέψεις θνησιμότητας. Συνιστάται η αναφορά ζωνών κινδύνου αντί για χρήση μοναδικού διαγνωστικού κατωφλίου. Επιπλέον, προτείνεται λειτουργική αξιοποίηση των εργαλείων: το PIM-3 ως εργαλείο πρώιμης διαλογής κατά την εισαγωγή, και το PELOD-2 για επιβεβαίωση και δυναμική επανεκτίμηση της πρόγνωσης κατά τη διάρκεια της νοσηλείας.

Λέξεις Κλειδιά: PIM-3, PELOD-2, Διακριτική ικανότητα, Βαθμονόμηση, Λογιστική επαναβαθμονόμηση, Θνητότητα

Predictive performance of PIM-3 and PELOD-2 scores for mortality in a Greek Pediatric ICU: Discrimination and calibration analysis

Abstract

Introduction: Reliable risk estimation in the Pediatric Intensive Care Unit (PICU) underpins clinical decision-making, communication with families, and evaluation of the quality of care delivered.

Objectives: This postgraduate thesis aims to evaluate the performance of PIM-3 and PELOD-2 in a Greek PICU in terms of discrimination and calibration, and to conduct a comparative analysis of diagnostic indices at clinically useful thresholds.

Methods: Retrospective single-center study of 545 consecutive admissions of children aged 0–16 years from January 2020 to December 2021. Discrimination was assessed using ROC curves and the area under the curve (AUC). Calibration was examined with calibration tables and plots, the Hosmer–Lemeshow test, and the indices Brier score and Log-loss. Logistic recalibration was applied with respect to intercept and slope. Optimal cut-offs were determined by Youden’s J index. **Results:** The PIM-3 and PELOD-2 scores demonstrated good discrimination ability: PIM-3 AUC = 0.83 (95% CI, 0.76–0.91) and PELOD-2 AUC = 0.85 (95% CI, 0.77–0.92), with no statistically significant difference between them (DeLong $p = 0.67$). Before recalibration, risk was underestimated, with an observed-to-expected mortality ratio (O/E) of approximately 1.8. The Hosmer–Lemeshow test was statistically significant for both PIM-3 ($\chi^2 = 24.40$, $df = 3$, $p < 0.001$) and PELOD-2 ($\chi^2 = 42.07$, $df = 3$, $p < 0.001$), indicating poor calibration. After recalibration of intercept and slope, the O/E ratio improved to 1.00 and calibration was acceptable: PIM-3 ($\chi^2 = 2.41$, $df = 3$, $p = 0.493$) and PELOD-2 ($\chi^2 = 2.70$, $df = 3$, $p = 0.440$). Overall accuracy also improved: Brier score 0.0568 and 0.0515, and Log-loss 0.2065 and 0.1919 for PIM-3 and PELOD-2, respectively. At the Youden J-optimal threshold, PIM-3 yielded sensitivity 78% and specificity 77%, while PELOD-2 showed sensitivity 74% and higher specificity 84%. PELOD-2 had a larger positive likelihood ratio (LR⁺) and a marginally higher positive predictive value (PPV), whereas both scores maintained a very high negative predictive value (NPV ≈ 0.97).

Conclusion: PIM-3 and PELOD-2 adequately discriminate patients and, after recalibration, provide reliable and clinically useful mortality predictions. Reporting risk zones rather than relying on a single diagnostic threshold is recommended. In practice,

PIM-3 can support early triage at admission, whereas PELOD-2 can be used for confirmation and dynamic reassessment of prognosis during length of stay.

Keywords: PIM-3, PELOD-2, Discrimination, Calibration, Logistic recalibration, Mortality

Εισαγωγή

Η έγκαιρη και αξιόπιστη εκτίμηση του κινδύνου θανάτου και της κλινικής βαρύτητας σε Παιδιατρική Μονάδα Εντατικής Θεραπείας (ΠΜΕΘ) αποτελεί θεμέλιο για ποιοτική, ασφαλή και λογοδοτούσα περίθαλψη. Ειδικότερα, υποστηρίζει την κλινική λήψη αποφάσεων, την επικοινωνία με τις οικογένειες, την ορθολογική κατανομή πόρων και τη συγκριτική αξιολόγηση των υπηρεσιών υγείας (Recher et al., 2022, Malhotra et al., 2019).

Στην κλινική πράξη, η αντικειμενική αποτίμηση επιτυγχάνεται με δομημένες και τυποποιημένες κλίμακες. Μεταξύ των διαθέσιμων συστημάτων-μοντέλων για τον βαρέως πάσχοντα παιδιατρικό πληθυσμό, τα PIM-3 και PELOD-2 συγκαταλέγονται στα ευρύτερα χρησιμοποιούμενα (Recher et al., 2022). Το PIM-3 εκτιμά την πιθανότητα νοσοκομειακού θανάτου με βάση δεδομένα από την εισαγωγή έως και την πρώτη ώρα νοσηλείας στην ΠΜΕΘ (Stranney et al., 2013). Το PELOD-2 ποσοτικοποιεί τη βαρύτητα της πολυοργανικής δυσλειτουργίας και υπολογίζεται με βάση τις δυσμενέστερες τιμές του πρώτου 24ώρου και μπορεί να εφαρμόζεται και σειριακά ώστε να αποτυπώνεται η κλινική πορεία (Leteurtre et al., 2013).

Η έγκυρη χρήση των βαθμολογικών κλιμάκων προϋποθέτει εξωτερική επικύρωση σε τοπικά δεδομένα, καθώς έχουν αναπτυχθεί σε διαφορετικούς πληθυσμούς και συστήματα υγείας (Alba et al., 2017). Η αξιολόγηση εστιάζει πρωτίστως σε δύο διαστάσεις: (α) στη διακριτική ικανότητα, δηλαδή στην ικανότητα του εργαλείου να διαχωρίζει αποτελεσματικά τα παιδιά που θα καταλήξουν από εκείνα που θα επιβιώσουν, κατατάσσοντας τα με συνέπεια από χαμηλό σε υψηλό κίνδυνο, και (β) στη βαθμονόμηση, δηλαδή στη συμφωνία ανάμεσα στις εκτιμώμενες πιθανότητες θανάτου και στα πραγματικά παρατηρούμενα ποσοστά σε όλο το φάσμα κινδύνου (Demler et al., 2015, Alba et al., 2017). Πέραν αυτών, όταν οι αποφάσεις λαμβάνονται σε συγκεκριμένα κλινικά κατώφλια, αξιοποιούνται συμπληρωματικά διαγνωστικά μέτρα: η ευαισθησία και η ειδικότητα, ώστε να περιγράφεται η ικανότητα έγκαιρου εντοπισμού υψηλού κινδύνου με ελεγχόμενο ποσοστό ψευδώς θετικών περιπτώσεων, καθώς και οι θετικές και αρνητικές προγνωστικές αξίες, οι οποίες εξαρτώνται από τον επιπολασμό της υπό μελέτη έκβασης στον εκάστοτε πληθυσμό (Monaghan et al., 2021).

Σε μετα-ανάλυση 29 μελετών, τα PIM-3 και PELOD-2 εμφάνισαν καλή διακριτική ικανότητα, με συνοπτικό εμβασμό κάτω από την καμπύλη λειτουργικού χαρακτηριστικού δέκτη (AUC /SROC) 0,82 (95% ΔΕ: 0,78–0,85) και 0,83 (95% ΔΕ: 0,80–0,86), αντίστοιχα (Shen and Jiang, 2021). Στο κατώφλι που μεγιστοποιεί τον δείκτη Youden J οι συνολικές τιμές ευαισθησίας/ειδικότητας αναφέρθηκαν ως 0,75/0,76 για το PIM-3 και 0,78/0,75 για το PELOD-2 (Shen and Jiang, 2021).

Ως προς τη βαθμονόμηση, τα ευρήματα διαφέρουν μεταξύ πληθυσμών: σε ιταλική πολυκεντρική μελέτη (TIPNet) (Wolfler et al., 2016), το PIM-3 εμφάνισε καλή προσαρμογή με λόγο τυποποιημένης θνητότητας (SMR=0,98), ενώ σε άλλη προοπτική μελέτη καταγράφηκε φτωχή βαθμονόμηση (SMR=0,53), υποδηλώνοντας συστηματική υπερεκτίμηση του κινδύνου (Malhotra et al., 2019). Για το PELOD-2, σε προοπτική μελέτη καταγράφηκε καλή βαθμονόμηση, SMR=1,00 (95% ΔΕ: 0,75–1,30) (El-Nawawy et al., 2017). Επιπλέον, στη βιβλιογραφία αναφέρονται περιπτώσεις τοπικής λογιστικής επαναβαθμονόμησης με στόχο την ευθυγράμμιση των εκτιμώμενων πιθανοτήτων θανάτου με τα παρατηρούμενα ποσοστά (Soto-Campos and Palumbo, 2018, Soto-Campos et al., 2018).

Στο ελληνικό πλαίσιο, έχουν πραγματοποιηθεί ελάχιστες στοχευμένες προσπάθειες αξιολόγησης της απόδοσης προγνωστικών σκορ σε ΠΜΕΘ. Ενδεικτικά, το PRISM III-24 έχει μελετηθεί σε ελληνικό πληθυσμό με καλή διακριτική ικανότητα και ικανοποιητική βαθμονόμηση, τεκμηριώνοντας τη χρησιμότητά του σε επίπεδο μονάδας (Volakli et al., 2013). Αντίθετα, για τα PIM-3 και PELOD-2, παρότι έχουν χρησιμοποιηθεί σε ελληνικές δημοσιεύσεις ως δείκτες βαρύτητας/οργανικής δυσλειτουργίας (Kiekkas et al., 2023) και σε συναφή ερευνητικά πλαίσια (Nieri et al., 2019) δεν υπάρχουν, από όσο είμαστε σε θέση να γνωρίζουμε, δημοσιευμένες, εστιασμένες αξιολογήσεις της προγνωστικής τους απόδοσης σε ελληνικούς πληθυσμούς ΠΜΕΘ.

Σκοπός της μελέτης είναι η συστηματική αποτίμηση της επίδοσης των PIM-3 και PELOD-2 σε ελληνικό πληθυσμό ΠΜΕΘ, με επιμέρους στόχους: (i) τη συγκριτική εκτίμηση της διακριτικής ικανότητας, (ii) τη διερεύνηση της βαθμονόμησης στο πλήρες φάσμα κινδύνου και (iii) τον υπολογισμό διαγνωστικών μέτρων βάσει του δείκτη Youden J. Κατόπιν διαπίστωσης ασυμφωνίας βαθμονόμησης, εφαρμόστηκε λογιστική επαναβαθμονόμηση (σταθερός όρος/κλίση) και επανεκτιμήθηκαν η

διακριτική ικανότητα, η βαθμονόμηση και τα διαγνωστικά μέτρα στα αναθεωρημένα κατώφλια. Διερευνητικά, αξιολογήθηκε η επανεκτίμηση συντελεστών στον ίδιο πληθυσμό για το PELOD-2 (με το αρχικό σκορ ως μοναδικό παράγοντα πρόβλεψης).

Ερευνητική Μεθοδολογία

Συλλογή δεδομένων

Διενεργήθηκε αναδρομική μελέτη παρατήρησης με συλλογή δεδομένων από έντυπους ιατρικούς φακέλους της πολυδύναμης ΠΜΕΘ 13 κλινών του Γενικού Νοσοκομείου Παίδων Αθηνών Π.κ.Α. Κυριακού (Ιανουάριος 2020–Δεκέμβριος 2021). Συμπεριλήφθηκαν διαδοχικά όλες οι εισαγωγές παιδιών 0–16 ετών που πληρούσαν τα προκαθορισμένα κριτήρια ένταξης και διέθεταν τα αναγκαία στοιχεία για τον υπολογισμό των σκορ PIM-3 και PELOD-2. Αποκλείστηκαν εισαγωγές με παραμονή <12 ώρες στην ΠΜΕΘ ή με ελλιπή καταγραφή των απαιτούμενων δεδομένων. Σε περιπτώσεις επανεισαγωγής, κάθε νέο επεισόδιο θεωρήθηκε ξεχωριστή εισαγωγή και αναλύθηκε ανεξάρτητα. Στην τελική ανάλυση περιλήφθηκαν 545 εισαγωγές.

Μεταβλητές ενδιαφέροντος

Ως έκβαση ορίστηκε η ενδονοσοκομειακή θνητότητα (θνητότητα στη ΠΜΕΘ), κωδικοποιημένη ως δυαδική μεταβλητή (0=επιζών, 1=μη επιζών). Ως μεταβλητές ενδιαφέροντος χρησιμοποιήθηκαν οι τιμές των πιθανοτήτων των PIM-3 και PELOD-2, ενώ καταγράφηκαν βασικά δημογραφικά και κλινικά στοιχεία εισαγωγής, συμπεριλαμβανομένων του φύλου, της ηλικίας, της διάγνωσης εισαγωγής, της χρήσης επεμβατικού μηχανικού αερισμού και της διάρκειάς του, καθώς και της συνολικής διάρκειας νοσηλείας. Το PIM-3 αποτελεί επικαιροποιημένη εκδοχή του PIM-2, ανεπτυγμένη σε διεθνές, πολυκεντρικό δείγμα, με στόχο την εκτίμηση της πιθανότητας ενδονοσοκομειακού θανάτου κατά την εισαγωγή στην ΠΜΕΘ. Για τον υπολογισμό του χρησιμοποιείται η πρώτη διαθέσιμη τιμή κάθε μεταβλητής από την πρώτη επαφή με το προσωπικό της ΠΜΕΘ έως και μία ώρα μετά την άφιξη. Οι προβλεπτικοί παράγοντες περιλαμβάνουν τη συστολική αρτηριακή πίεση, την αντιδραστικότητα των κορών στο φως, την αναλογία $100 \times \text{FiO}_2 / \text{PaO}_2$, την απόλυτη τιμή του ελλείμματος βάσης, τη χρήση μηχανικού αερισμού εντός της πρώτης ώρας, την εκλεκτική εισαγωγή, τη μετεγχειρητική ανάνηψη και την κατηγορία διαγνωστικού κινδύνου (χαμηλού/υψηλού/πολύ υψηλού). Η εκτιμώμενη πιθανότητα (p) θανάτου, η οποία

ανήκει στο διάστημα (0,1) και δύναται να εκφραστεί σε ποσοστό, υπολογίζεται από τη σχέση $p=1/(1+\exp(-\text{logit}))$, όπου το logit (φυσικός λογάριθμος του λόγου πιθανοτήτων) αποτελεί το γραμμικό άθροισμα των συντελεστών του μοντέλου επί τις αντίστοιχες τιμές των μεταβλητών (Straney et al., 2013).

Το PELOD-2 είναι επικαιροποιημένη εκδοχή του PELOD και αποτυπώνει τη βαρύτητα πολυοργανικής δυσλειτουργίας. Υπολογίζεται από τις δυσμενέστερες τιμές του πρώτου 24ώρου μετά την εισαγωγή στην ΠΜΕΘ, με δυνατότητα σειριακής εφαρμογής σε προκαθορισμένες χρονικές στιγμές κατά τη διάρκεια της νοσηλείας. Συγκροτείται από δέκα μεταβλητές που αντιστοιχούν σε πέντε οργανικά συστήματα [νευρικό: Κλίμακα Γλασκώβης, αντιδραστικότητα κορών, καρδιαγγειακό: γαλακτικό οξύ, μέση αρτηριακή πίεση προσαρμοσμένη για την ηλικία, αναπνευστικό: μερική πίεση οξυγόνου στο αρτηριακό αίμα/κλάσμα εισπνεόμενου οξυγόνου ($\text{PaO}_2/\text{FiO}_2$), μερική πίεση διοξειδίου στο αρτηριακό αίμα (PaCO_2), επεμβατικός μηχανικός αερισμός, ουροποιητικό: κρεατινίνη προσαρμοσμένη για την ηλικία, αιματολογικό: λευκά αιμοσφαίρια, αιμοπετάλια]. Κάθε μεταβλητή βαθμολογείται σε επίπεδα βαρύτητας και οι βαθμοί αθροίζονται σε συνολικό σκορ με εύρος 0–33. Υψηλότερες τιμές υποδηλώνουν μεγαλύτερη βαρύτητα και υψηλότερο κίνδυνο. Η πιθανότητα θανάτου μπορεί να εκτιμηθεί, σύμφωνα με τη μελέτη των Leteurtre et al., 2013, από το συνολικό σκορ μέσω της σχέσης $p = 1/(1+\exp(-\text{logit}))$, όπου $\text{logit} = -6,61 + 0,47 \times (\text{PELOD-2})$.

Στατιστική Ανάλυση

Οι κατηγορικές μεταβλητές παρουσιάστηκαν ως απόλυτες και σχετικές συχνότητες, n (%). Οι συνεχείς μεταβλητές συνοψίζονται με τη διάμεσο και το ενδοτεταρτημοριακό εύρος ($\text{IQR}=\text{Q3}-\text{Q1}$), καθώς ο έλεγχος κανονικότητας με διαγράμματα Q–Q και τη δοκιμασία Kolmogorov–Smirnov εμφάνισε απόκλιση. Για τα σκορ PIM-3 και PELOD-2 παρατίθεται συμπληρωματικά η μέση τιμή $\pm$ τυπική απόκλιση ($\text{MT} \pm \text{TA}$) προς διευκόλυνση της ερμηνείας. Οι μονοπαραγοντικές συγκρίσεις μεταξύ επιζώντων και μη επιζώντων πραγματοποιήθηκαν με χ^2 τεστ ανεξαρτησίας για κατηγορικές και με Mann–Whitney U για συνεχείς μεταβλητές. Όλοι οι έλεγχοι ήταν αμφίπλευροι με επίπεδο σημαντικότητας $\alpha=0,05$.

Η διακριτική ικανότητα των μοντέλων ορίστηκε ως η πιθανότητα ένα τυχαίο ζεύγος ασθενών (μη επιζών/επιζών) να καταταγεί σωστά, δηλαδή το μοντέλο να αποδίδει

υψηλότερη εκτιμώμενη πιθανότητα θανάτου στον μη επιζώντα έναντι του επιζώντα. Η αξιολόγηση πραγματοποιήθηκε με ROC καμπύλες και το εμβαδόν κάτω από την καμπύλη (AUC), συνοδευόμενο από 95% διαστήματα εμπιστοσύνης. Τιμή AUC=0,50 δηλώνει απουσία διακριτικής ικανότητας (τυχαία ταξινόμηση), 0,70–0,79 ικανοποιητική, 0,80–0,89 πολύ καλή και $\geq 0,90$ εξαιρετική. Η σύγκριση των AUC μεταξύ PIM-3 και PELOD-2 πραγματοποιήθηκε με τον έλεγχο DeLong για συζευγμένες ROC (DeLong et al., 1988).

Η βαθμονόμηση (calibration) ορίστηκε ως ο βαθμός συμφωνίας μεταξύ των προβλεπόμενων πιθανοτήτων θανάτου και της παρατηρούμενης θνητότητας. Για ισόρροπη εκπροσώπηση σε όλο το φάσμα κινδύνου, οι εισαγωγές κατανεμήθηκαν σε 5 ισοπληθείς ομάδες, προκειμένου να εξασφαλιστεί στατιστική σταθερότητα και να αποφευχθούν πολύ μικρές αναμενόμενες συχνότητες. Ανά ομάδα παρουσιάστηκαν: αριθμός ασθενών (n), παρατηρούμενοι θάνατοι (O), αναμενόμενοι θάνατοι (E=άθροισμα των προβλεπόμενων πιθανοτήτων), παρατηρούμενη θνητότητα (O/n), μέση προβλεπόμενη πιθανότητα (E/n) και λόγος O/E, ο οποίος στη μελέτη της θνητότητας ταυτίζεται με τον λόγο τυποποιημένης θνητότητας (SMR=O/E). Τιμή ίση με τη μονάδα υποδηλώνει καλή βαθμονόμηση, >1 υποεκτίμηση του κινδύνου και <1 υπερεκτίμηση από το μοντέλο. Η καταλληλότητα της βαθμονόμησης ελέγχθηκε με τη δοκιμασία καλής προσαρμογής (goodness-of-fit) Hosmer–Lemeshow (HL) σε 5 ισοπληθείς ομάδες. Για την οπτική εκτίμηση χρησιμοποιήθηκαν διαγράμματα βαθμονόμησης με 10 ισοπληθείς ομάδες, όπου ανά ομάδα απεικονίζονται τα σημεία (μέση προβλεπόμενη πιθανότητα, παρατηρούμενη θνητότητα) και η διακεκομμένη διαγώνιος 45° ως γραμμή τέλει βαθμονόμησης. Παρουσιάστηκαν δύο κλίμακες: 0–0,06 για την περιοχή χαμηλών πιθανοτήτων (όπου συγκεντρώνεται η πλειονότητα των ασθενών) και 0–0,50 για συνολική απεικόνιση (Demler et al., 2015, Alba et al., 2017).

Η συνολική ακρίβεια των προβλέψεων κινδύνου αξιολογήθηκε με το Brier score και το Log-loss. Το Brier score αποτυπώνει πόσο κοντά βρίσκονται οι προβλεπόμενες πιθανότητες στα πραγματικά αποτελέσματα. Το Log-loss περιγράφει την ίδια πληροφορία σε λογαριθμική κλίμακα, ‘τιμωρώντας’ περισσότερο τις λανθασμένες προβλέψεις με υψηλή βεβαιότητα. Ωστόσο, επισημαίνεται ότι και οι δύο δείκτες αντανακλούν συνδυαστικά διακριτική ικανότητα και βαθμονόμηση. Κατά κανόνα, όσο μικρότερη η τιμή και για τους δύο, τόσο καλύτερη η συνολική απόδοση.

Τα διαγνωστικά μέτρα υπολογίστηκαν στο κατώφλι του δείκτη Youden J κάθε μοντέλου, δηλαδή στο σημείο που μεγιστοποιεί τον δείκτη $J = \text{Ευαισθησία} + \text{Ειδικότητα} - 1$. Στο κατώφλι αυτό παρουσιάζονται, για κάθε μοντέλο, η ευαισθησία (ποσοστό παιδιών που πέθαναν που ταξινομήθηκαν ως μη επιζώντες), η ειδικότητα (ποσοστό όσων επέζησαν που ταξινομήθηκαν ως ασθενείς που έλαβαν εξιτήριο), η ακρίβεια (ποσοστό συνολικά σωστών ταξινομήσεων) και η ισορροπημένη ακρίβεια (μέσος όρος ευαισθησίας και ειδικότητας). Αναφέρονται επίσης η θετική και η αρνητική προγνωστική αξία (PPV/NPV: πιθανότητα θανάτου ή επιβίωσης αντίστοιχα, δεδομένης της ταξινόμησης «υψηλού» ή «χαμηλού» κινδύνου). Για την κλινική ερμηνεία της μεταβολής του κινδύνου παρατίθενται οι λόγοι πιθανοφάνειας LR+ και LR- (πόσο αυξομειώνονται τα μεταγενέστερα odds/η πιθανότητα μετά από «θετικό» ή «αρνητικό» αποτέλεσμα) (Monaghan et al., 2021). Συμπληρωματικά παρουσιάζονται ο δείκτης F1 (αρμονικός μέσος ευαισθησίας και PPV), ο συντελεστής συμφωνίας Cohen's κ (συμφωνία πέραν της τύχης), ο διαγνωστικός λόγος πιθανοτήτων DOR (συνοπτική ισχύς που ισοδυναμεί με LR+/LR-) (Glas et al., 2003) και η p-τιμή του ελέγχου McNemar για τυχόν ασυμμετρία σφαλμάτων (ψευδώς θετικά έναντι ψευδώς αρνητικών).

Πραγματοποιήθηκε λογιστική επαναβαθμονόμηση των αρχικών πιθανοτήτων των PIM-3 και PELOD-2. Για κάθε μοντέλο υπολογίστηκε το logit της εκτιμώμενης πιθανότητας ($\text{logit}(p) = \ln[p/(1-p)]$) και εφαρμόστηκε λογιστική παλινδρόμηση με εξαρτημένη την έκβαση (0/1) και ανεξάρτητη μεταβλητή τον $\text{logit}(p)$: $\text{logit}(p_n) = \alpha + \beta * \text{logit}(p)$. Οι νέες πιθανότητες προέκυψαν ως $p_n = \exp(\alpha + \beta * \text{logit}(p)) / \{1 + \exp(\alpha + \beta * \text{logit}(p))\}$ (Binuya et al., 2022). Με βάση τις νέες πιθανότητες αξιολογήθηκαν εκ νέου η βαθμονόμηση, η συνολική ακρίβεια με Brier score και Log-loss και τα διαγνωστικά μέτρα σε κατώφλι που προέκυψε εκ νέου από τον δείκτη Youden J.

Συμπληρωματικά, αναπτύχθηκε εναλλακτικό μοντέλο πρόγνωσης ως απλοποιημένη επαναπροσαρμογή, με μοναδικό προβλεπτικό παράγοντα το συνολικό σκορ PELOD-2: $\text{logit}(p) = \alpha + \beta * (\text{PELOD-2})$ (Binuya et al., 2022). Οι επιδόσεις του νέου μοντέλου αξιολογήθηκαν με, διαγράμματα βαθμονόμησης και δοκιμασία Hosmer–Lemeshow σε 10 ισοπληθείς ομάδες, καθώς και με Brier score και Log-loss. Τα διαγνωστικά μέτρα παρουσιάζονται στο κατώφλι που μεγιστοποιεί τον δείκτη Youden J.

Όλες οι αναλύσεις πραγματοποιήθηκαν στο στατιστικό περιβάλλον R (έκδ. 4.4.1 . R Foundation for Statistical Computing, Vienna, Austria), μέσω του RStudio (Posit Software, PBC), καθώς και στο λογισμικό IBM SPSS Statistics (έκδ. 29.0· IBM Corp., Armonk, NY, USA).

Ηθική και δεοντολογία

Η μελέτη διεξήχθη σύμφωνα με τη Διακήρυξη του Ελσίνκι και τη νομοθεσία προστασίας δεδομένων. Τα δεδομένα προήλθαν από ιατρικούς φακέλους, κωδικοποιήθηκαν με στόχο τη διασφάλισης ανωνυμίας πριν από την ανάλυση και χρησιμοποιήθηκαν αποκλειστικά για ερευνητικούς σκοπούς. Δεν ζητήθηκε ατομική συναίνεση λόγω αναδρομικού και αποπροσωποποιημένου χαρακτήρα της μελέτης. Η συλλογή και χρήση των δεδομένων εγκρίθηκε από το Επιστημονικό Συμβούλιο του Γενικού Νοσοκομείου Παιδών Αθηνών Π.κ.Α. Κυριακού (αρ. πρωτ. Ε.Σ. 5998/02-04-2021, πρακτικό 11η/19-05-2021) και έλαβε θετική γνωμοδότηση από την Επιτροπή Ηθικής και Δεοντολογίας.

Αποτελέσματα

Από τις 746 εισαγωγές που καταγράφηκαν στην περίοδο της μελέτης, 545 πληρούσαν τα προκαθορισμένα κριτήρια ένταξης και συμπεριλήφθηκαν στην ανάλυση. Στο συγκεκριμένο δείγμα η θνητότητα ήταν 7,7% (42/545).

Τα δημογραφικά και κλινικά χαρακτηριστικά του δείγματος παρουσιάζονται στον Πίνακα 1. Η ηλικία δεν ανέδειξε στατιστικά σημαντική διαφορά μεταξύ επιζώντων και μη επιζώντων ($U=10379$, $p=0,851$). Ομοίως, δεν παρατηρήθηκε στατιστικά σημαντική διαφοροποίηση ως προς το φύλο [$\chi^2(1)=0,434$, $p=0,510$].

Όσοι απεβίωσαν είχαν στατιστικά σημαντικά μεγαλύτερη διάρκεια παραμονής στη ΜΕΘ συγκριτικά με τους επιζώντες ($U=7456$, $p=0,002$). Επεμβατικός μηχανικός αερισμός εφαρμόστηκε συχνότερα στους μη επιζώντες (δειγματικό ποσοστό 92,9% έναντι 45,7%, $\chi^2(1)=34,449$, $p<0,001$) και μεταξύ όσων τέθηκαν υπό επεμβατικό μηχανικό αερισμό, η διάρκειά του ήταν επίσης αυξημένη στους μη επιζώντες ($U=4521,5$, $p<0,001$).

Η κατανομή ανά διάγνωση εισαγωγής διέφερε στατιστικά σημαντικά ($\chi^2(6)=87,435$, $p<0,001$). Η σήψη/πολυοργανική ανεπάρκεια ήταν συχνότερη στους μη επιζώντες (28,6% έναντι 2,0%), ενώ στην κατηγορία μετεγχειρητικής παρακολούθησης δεν καταγράφηκαν θάνατοι.

Οι δείκτες κλινικής βαρύτητας και πολυοργανικής ανεπάρκειας ήταν υψηλότεροι στους μη επιζώντες (PIM-3: εκτιμώμενη πιθανότητα θανάτου, $U=3424,5$, $p<0,001$, PELOD-2 σκορ: $U=3093$, $p<0,001$ και PELOD-2: εκτιμώμενη πιθανότητα θανάτου, $U=3109,5$, $p<0,001$).

Πίνακας 1. Δημογραφικά και κλινικά χαρακτηριστικά κατά έκβαση (επιζώντες/μη επιζώντες)

Χαρακτηριστικό	Σύνολο (n=545)	Επιζώντες (n=503)	Μη επιζώντες (n=42)	p-value
Φύλο, n (%)				
Άρρεν	299 (54,9)	278 (55,3)	21 (50,0)	0,510*
Θήλυ	246 (45,1)	225 (44,7)	21 (50,0)	
Ηλικία (έτη), διάμεσος (IQR)	3,3 (0,66–11,0)	3,3 (0,65–11,0)	3,2 (0,9–8,0)	0,851°
Ηλικιακή ομάδα, n (%)				
Βρεφική (0–12 μην.)	178 (32,7)	165 (32,8)	13 (31,0)	0,520*
Νηπιακή (12–36 μην.)	89 (16,3)	81 (16,1)	8 (19,0)	
Προσχολική (36–72 μην.)	61 (11,2)	54 (10,7)	7 (16,7)	
Σχολική (72–144 μην.)	105 (19,3)	96 (19,1)	9 (21,4)	
Εφηβική (>144 μην.)	112 (20,6)	107 (21,3)	5 (11,9)	
Διάρκεια νοσηλείας (ημέρες), διάμεσος (IQR)	2,2 (1,1–5,0)	2,0 (1,1–4,8)	6,5 (1,7–9,2)	0,002°
Διάγνωση εισαγωγής, n (%)				
Αναπνευστικά νοσήματα	102 (18,7)	95 (18,9)	7 (16,7)	<0,001*
Καρδιολογικά νοσήματα	43 (7,9)	37 (7,4)	6 (14,3)	
Νευρολογικά νοσήματα	111 (20,4)	103 (20,5)	8 (19,0)	
Σήψη / Πολυοργανική ανεπάρκεια	22 (4,0)	10 (2,0)	12 (28,6)	
Μετεγχειρητική παρακολούθηση	176 (32,3)	176 (35,0)	0 (0,0)	
Μεταβολικά / Ενδοκρινολογικά	17 (3,1)	14 (2,8)	3 (7,1)	
Λοιπά	74 (13,6)	68 (13,5)	6 (14,3)	
Επεμβατικός μηχανικός αερισμός, n (%)				
Όχι	276 (50,6)	273 (54,3)	3 (7,1)	<0,001*
Ναι	269 (49,4)	230 (45,7)	39 (92,9)	
Διάρκεια επεμβατικού ΜΑ (ημέρες), διάμεσος (IQR)	0,0 (0,0–1,8)	0,0 (0,0–1,3)	3,7 (0,5–8,4)	<0,001°
PIM-3 (%), διάμεσος (IQR)	1,4 (0,6–2,2)	1,3 (0,6–1,8)	9,1 (4,8–22,2)	<0,001°
PIM-3 (%), MT ± TA	3,0 ± 6,5	3,0 ± 6,5	19,3 ± 25,4	
PELOD-2, διάμεσος (IQR)	4 (2–7)	4 (2–6)	10 (6–13)	<0,001°
PELOD-2, MT ± TA	4,5 ± 3,0	4,1 ± 2,9	9,8 ± 4,6	
PELOD-2 (%), διάμεσος (IQR)	1,0 (0,3–2,5)	0,9 (0,3–2,2)	12,9 (2,2–40,7)	<0,001°
PELOD-2 (%), MT ± TA	3,1 ± 9,5	2,5 ± 7,0	24,4 ± 27,7	

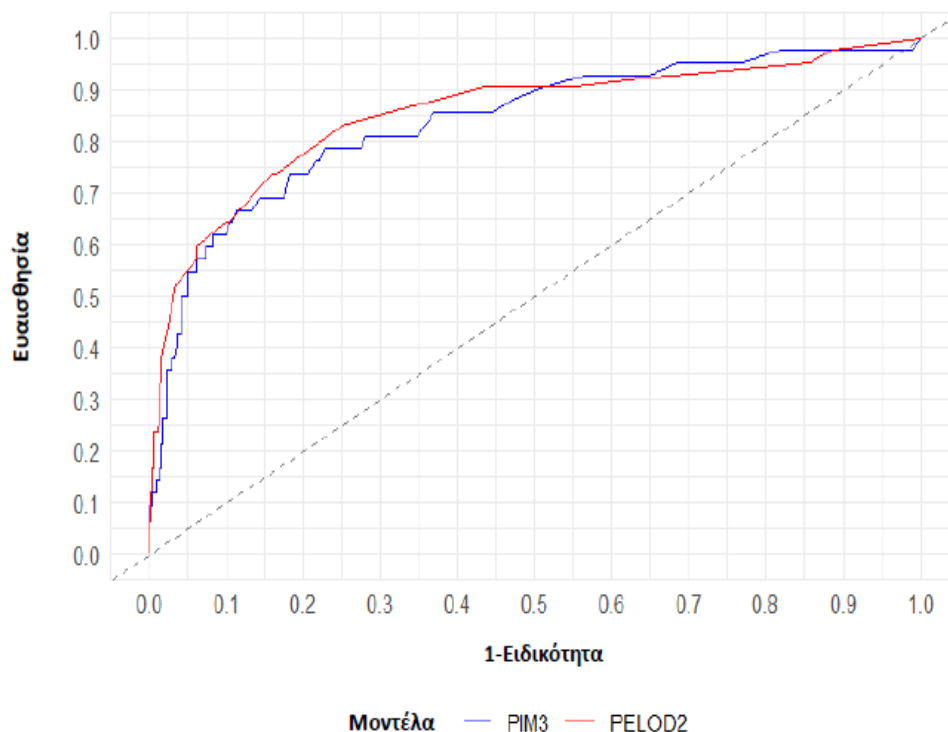
IQR: Ενδοτεταρτημοριακό εύρος, MT ± TA: Μέση τιμή ±Τυπική απόκλιση

*Chi-square test, °Mann–Whitney U test

Διακριτική ικανότητα PIM-3 και PELOD 2

Η διακριτική ικανότητα αξιολογήθηκε με την καμπύλη ROC (AUC). Και τα δύο σκορ εμφάνισαν καλή απόδοση: PIM-3 AUC 0,83 (95% ΔΕ: 0,76–0,91) και PELOD-2 AUC 0,85 (95% ΔΕ: 0,77–0,92). Η διαφορά των AUC ανάμεσα στα σκορ δεν ήταν στατιστικά σημαντική (DeLong: $Z = -0,41$, $p = 0,67$). Τα αποτελέσματα συνοψίζονται στον Πίνακα 2 και απεικονίζονται στο Σχήμα 1.

Σχήμα 1. Καμπύλες ROC για τα PIM3 και PELOD-2



Πίνακας 2. AUC των μοντέλων και στατιστικός έλεγχος διαφοράς (DeLong)

Μοντέλο	AUC	95% ΔΕ (AUC)
PIM-3	0,83	0,76–0,91
PELOD-2	0,85	0,77–0,92

Έλεγχος DeLong για σύγκριση AUC (PIM-3 έναντι PELOD-2): $Z = -0,41$, $p = 0,67$

Βαθμονόμηση PIM-3 και PELOD 2

Με βάση τον έλεγχο καλής προσαρμογής Hosmer–Lemeshow goodness-of-fit test καταδείχθηκε στατιστικά σημαντική απόκλιση μεταξύ των προβλεπόμενων και των παρατηρούμενων τιμών, υποδηλώνοντας ανεπαρκή βαθμονόμηση και στα δύο μοντέλα (PIM-3: $\chi^2=24,40$, $df=3$, $p<0,001$, PELOD-2: $\chi^2=42,07$, $df=3$, $p<0,001$) (Πίνακες 3–4).

Ως προς το PIM-3 (Πίνακας 3), συνολικά παρατηρήθηκαν 42 θάνατοι έναντι 23,15 αναμενόμενων ($O/E=1,81$ με μέση προβλεπόμενη πιθανότητα 4,2% έναντι παρατηρούμενης 7,7%). Διαπιστώθηκε υποεκτίμηση σε όλες τις ομάδες κινδύνου και εντονότερη στην 1η (0,001–0,005, $O/E=5,71$) και στις μεσαίες–υψηλές (3η: $O/E=1,95$ και 4η: $O/E=2,22$). Η μικρότερη απόκλιση παρατηρήθηκε στη 2η ομάδα ($O/E = 1,17$), χωρίς όμως πλήρη συμφωνία. Αναφορικά με το PELOD-2 (Πίνακας 4) συνολικά καταγράφηκαν 42 παρατηρούμενοι έναντι 22,56 αναμενόμενων θανάτων ($O/E=1,86$ με μέση προβλεπόμενη 4,1% έναντι παρατηρούμενης 7,7%). Προέκυψε γενικευμένη υποεκτίμηση σε όλες τις ομάδες, με πολύ έντονη απόκλιση στις δύο χαμηλότερες κατηγορίες πιθανοτήτων (1η: 0,001–0,003, $O/E=10,20$ και 2η: 0,003–0,005, $O/E=4,64$). Καμία ομάδα δεν πλησίασε το θεωρητικό όριο της καλής βαθμονόμησης ($O/E \approx 1$). Η μικρότερη απόκλιση καταγράφηκε στην 5η ομάδα ($O/E = 1,65$). Τα ευρήματα υποδηλώνουν ότι και τα δύο μοντέλα τείνουν να **υποεκτιμούν συστηματικά τον κίνδυνο θανάτου**, με **εντονότερη απόκλιση σε χαμηλές εκτιμώμενες πιθανότητες**, γεγονός που καθιστά τη **βαθμονόμηση τους προβληματική** στον συγκεκριμένο πληθυσμό.

Πίνακας 3. Πίνακας βαθμονόμησης PIM-3 (5 ισοπληθείς ομάδες)

Ομάδα	Εύρος προβλεπόμενης πιθανότητας	n	Παρατηρούμενοι θάνατοι (O)	Αναμενόμενοι θάνατοι (E)	Παρατηρούμενη θνητότητα (%)	Μέση προβλεπόμενη πιθανότητα (%)	O/E
1	0,001–0,005	109	2	0,35	1,8%	0,3%	5,71
2	0,005–0,011	109	1	0,85	0,9%	0,8%	1,17
3	0,011–0,018	109	3	1,53	2,8%	1,4%	1,95
4	0,018–0,045	109	7	3,14	6,4%	2,9%	2,22
5	0,045–0,978	109	29	17,27	26,6%	15,8%	1,67
Σύνολο	—	545	42	23,15	7,7%	4,2%	1,81

Hosmer–Lemeshow: $\chi^2=24,40$, $df=3$, $p<0,001$

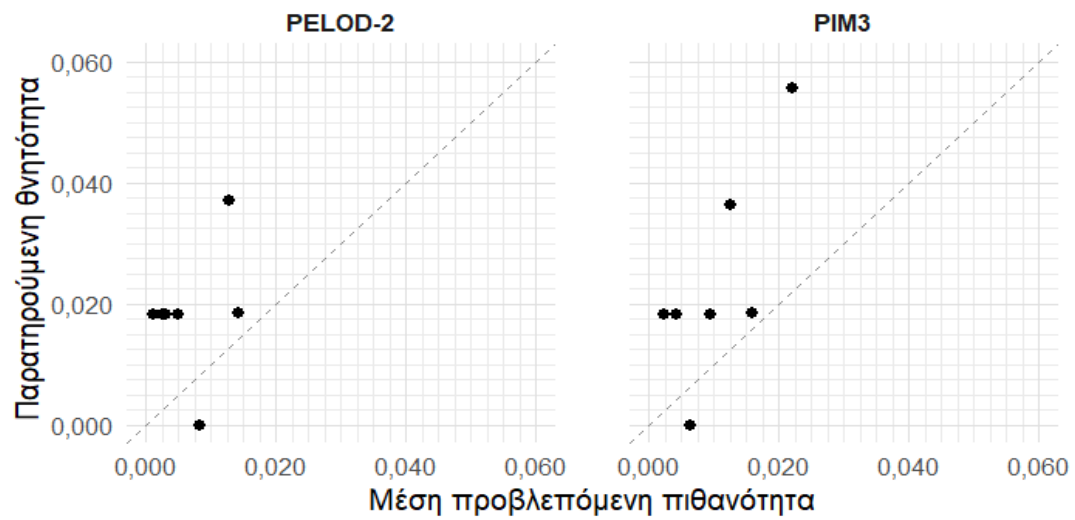
Πίνακας 4. Πίνακας βαθμονόμησης PELOD-2 (5 ισοπληθείς ομάδες)

Ομάδα	Εύρος προβλεπόμενης πιθανότητας	n	Παρατηρούμενοι θάνατοι (O)	Αναμενόμενοι θάνατοι (E)	Παρατηρούμενη θνητότητα (%)	Μέση προβλεπόμενη πιθανότητα (%)	O/E
1	0,001–0,003	109	2	0,19	1,8%	0,2%	10,20
2	0,003–0,005	109	2	0,43	1,8%	0,4%	4,64
3	0,005–0,014	109	2	1,12	1,8%	1,0%	1,77
4	0,014–0,035	109	5	2,00	4,6%	1,8%	2,49
5	0,035–0,864	109	31	18,80	28,4%	17,3%	1,64
Σύνολο	—	545	42	22,56	7,7%	4,1%	1,86

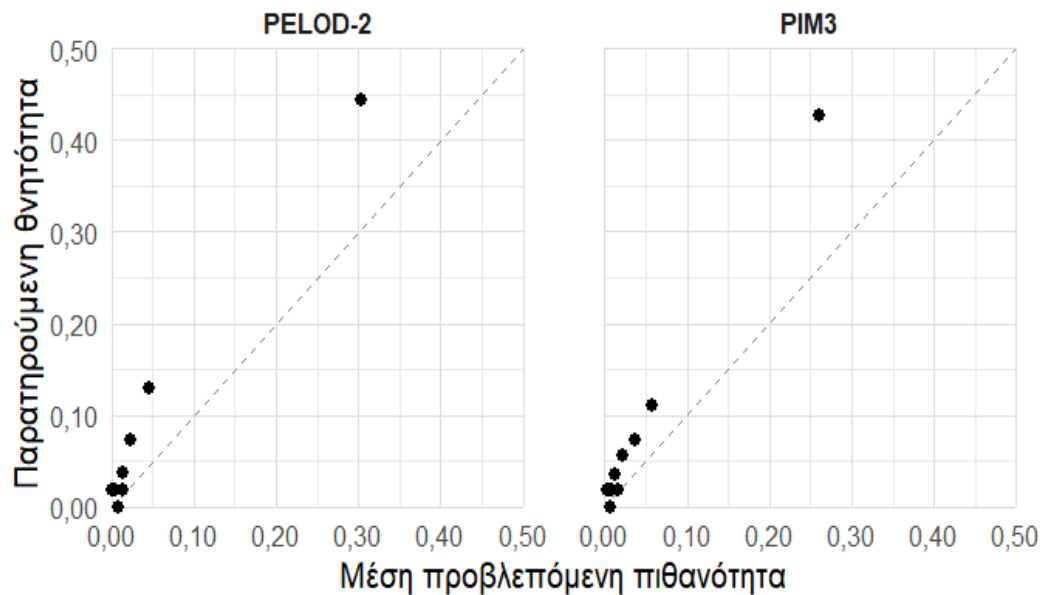
Hosmer–Lemeshow: $\chi^2=42,07$, $df=3$, $p<0,001$

Τα διαγράμματα βαθμονόμησης με 10 ισοπληθείς ομάδες αποτυπώνουν την ίδια εικόνα. Τα σημεία (που αντιστοιχούν στη μέση προβλεπόμενη πιθανότητα θανάτου της ομάδας και την παρατηρούμενη θνητότητα της ίδιας ομάδας) βρίσκονται πάνω από τη διαγώνιο 45°, η οποία αντιπροσωπεύει την ιδανική γραμμή ταύτισης μεταξύ προβλεπόμενου και παρατηρούμενου κινδύνου. Το φαινόμενο αυτό παρατηρείται τόσο στο χαμηλό φάσμα πιθανοτήτων (0–0,06), όπου συγκεντρώνεται η πλειονότητα των ασθενών, όσο και στο ευρύτερο εύρος (0–0,50), υποδηλώνοντας συστηματική υποεκτίμηση του κινδύνου θανάτου και από τα δύο μοντέλα (Σχήματα 2–3).

Σχήμα 2. Διάγραμμα βαθμονόμησης PIM-3, PELOD-2 — 10 ισοπληθείς ομάδες (εστίαση 0–0,06)



Σχήμα 3. Διάγραμμα βαθμονόμησης PIM-3 , PELOD-2 — 10 ισοπληθείς ομάδες (εστίαση 0–0,50)



Συνολικοί δείκτες επίδοσης (Brier score, Log-loss)

Ως συνολικοί δείκτες της ακρίβειας των προβλεπόμενων πιθανοτήτων χρησιμοποιήθηκαν το Brier score και το Log-loss, τα οποία αποτυπώνουν τη συνολική ποιότητα των προβλέψεων. Στο δείγμα μας, το PIM-3 είχε Brier 0,0597 και Log-loss 0,2233, ενώ το PELOD-2 Brier score 0,0549 και Log-loss 0,2135 (Πίνακας 5).

Πίνακας 5. Σύγκριση δεικτών (Brier Score, Log-loss) των μοντέλων PIM3 και PELOD-2

Μοντέλο	Brier	Log-loss
PIM3	0,0597	0,2233
PELOD-2	0,0549	0,2135

Διαγνωστικά μέτρα στο κατώφλι Youden J. δείκτη

Αναλυτικά τα διαγνωστικά μέτρα στο κατώφλι του δείκτη Youden J. παρουσιάζονται στον Πίνακα 6. Στο κατώφλι του δείκτη Youden J. (PIM-3: 0,0325, PELOD-2: 0,0335), μεταξύ όσων κατέληξαν, το ποσοστό που ταξινομήθηκαν ως «θετικοί» (ευαισθησία) ήταν 78% (95% ΔΕ: 63–90) για το PIM-3 και 74% (58–86) για το PELOD-2. Μεταξύ όσων επέζησαν, το ποσοστό που ταξινομήθηκαν ως «αρνητικοί» (ειδικότητα) ήταν 77% (73–81) και 84% (81–87), αντίστοιχα. Το συνολικό ποσοστό σωστών ταξινομήσεων (ακρίβεια) ήταν 77% (73–80) για το PIM-3 και 83,3% (79–86) για το PELOD-2, ενώ ο μέσος όρος ευαισθησίας και ειδικότητας (balanced accuracy) ήταν 0,77 και 0,79. Η πιθανότητα θανάτου δεδομένου θετικού αποτελέσματος (PPV) ήταν 22% (16–30) για το PIM-3 και 28% (20–37) για το PELOD-2, ενώ η πιθανότητα επιβίωσης δεδομένου αρνητικού αποτελέσματος (NPV) ήταν 97% (96–99) και για τα δύο. Οι δείκτες που εκφράζουν την αλλαγή του λόγου των πιθανοτήτων (odds) με θετικό ή αρνητικό αποτέλεσμα (LR+, LR–) ήταν 3,44 (2,74–4,30) και 0,28 (0,16–0,50) για το PIM-3 και 4,64 (3,54–6,08) και 0,31 (0,19–0,52) για το PELOD-2. Επιπλέον, ο εναρμονισμένος μέσος PPV και ευαισθησίας (F1-score) ήταν 0,347 για το PIM-3 και 0,405 για το PELOD-2, ενώ ο δείκτης συμφωνίας πέραν της τύχης (Cohen’s κ) ήταν 0,258 και 0,330, αντίστοιχα.

Πίνακας 6. Διαγνωστικά μέτρα στο κατώφλι του δείκτη Youden και 95% ΔΕ για PIM-3 και PELOD-2

Διαγνωστικό Μέτρο	PIM3	PELOD-2
Κατώφλι	0,0325	0,0335
Ευαισθησία (Se)	0,78 (0,63–0,90)	0,74 (0,58–0,86)
Ειδικότητα (Sp)	0,77 (0,73–0,81)	0,84 (0,81–0,87)
Ακρίβεια (Ac)	0,77 (0,73–0,80)	0,833 (0,79–0,86)
Ισορροπημένη Ακρίβεια (BA)	0,77	0,79
Θετική Προβλεπτική Αξία (PPV)	0,22 (0,16–0,30)	0,28 (0,20–0,37)
Αρνητική Προβλεπτική Αξία (NPV)	0,97 (0,96–0,99)	0,97 (0,96–0,99)
Θετικός Λόγος Πιθανοφάνειας (LR+)	3,44 (2,74–4,30)	4,64 (3,54–6,08)
Αρνητικός Λόγος Πιθανοφάνειας (LR–)	0,28 (0,16–0,50)	0,31 (0,19–0,52)
F1	0,347	0,405
Cohen’s Kappa	0,258	0,330
Διαγνωστικός Λόγος πιθανοτήτων (DOR)	12,37 (5,75–26,61)	14,9 (7,19–30,87)
McNemar	< 0,001	< 0,0001

Επαναβαθμονόμηση

Τα αποτελέσματα της επαναβαθμονόμησης συνοψίζονται στους Πίνακες 7–8 και απεικονίζονται στα Σχήματα 4–5. Για το PIM-3, ο συνολικός λόγος O/E εξισώθηκε σε 1,00 (O=42, E=42,02) και ο έλεγχος Hosmer–Lemeshow δεν ανέδειξε στατιστικά σημαντική απόκλιση ($\chi^2=2,405$, $df=3$, $p=0,493$). Σε επίπεδο ομάδων παρατηρήθηκαν ήπιες αποκλίσεις, με υποεκτίμηση στη χαμηλότερη κατηγορία πιθανοτήτων (0,0036–0,0150, O/E=1,82) και στην ανώτερη (O/E=1,08), ελαφρά υπερεκτίμηση στις μεσαίες (ενδεικτικά 0,0150–0,0302: O/E=0,41 και 0,0302–0,0463: O/E=0,74) και πλησιέστερη ισορροπία στο εύρος 0,0463–0,1015: O/E=0,93). Οι συντελεστές επαναβαθμονόμησης ήταν $\alpha=0,551$ (95% ΔΕ –0,241 - 1,343) και $\beta=0,894$ (0,651–1,137), αντιστοιχώντας στην εξίσωση: $\text{logit}(p_n)=0,551 + 0,894*\text{logit}(p)$.

Πίνακας 7. Πίνακας επαναβαθμονόμησης PIM-3 (5 ισοπληθείς ομάδες)

Ομάδα	Εύρος προβλεπόμενης πιθανότητας	n	Παρατηρούμενοι θάνατοι (O)	Αναμενόμενοι θάνατοι (E)	Παρατηρούμενη θνητότητα (%)	Μέση προβλεπόμενη πιθανότητα (%)	O/E
1	0,0036–0,0150	109	2	1,10	1,8%	1,0%	1,82
2	0,0150–0,0302	109	1	2,43	0,9%	2,2%	0,41
3	0,0302–0,0463	109	3	4,07	2,8%	3,7%	0,74
4	0,0463–0,1015	109	7	7,54	6,4%	6,9%	0,93
5	0,1015–0,9810	109	29	26,89	26,6%	24,7%	1,08
Σύνολο	—	545	42	42,02	7,7%	7,7%	1,00
Hosmer–Lemeshow Test $X^2 = 2.405$, $df = 3$, $p = 0.4928$							

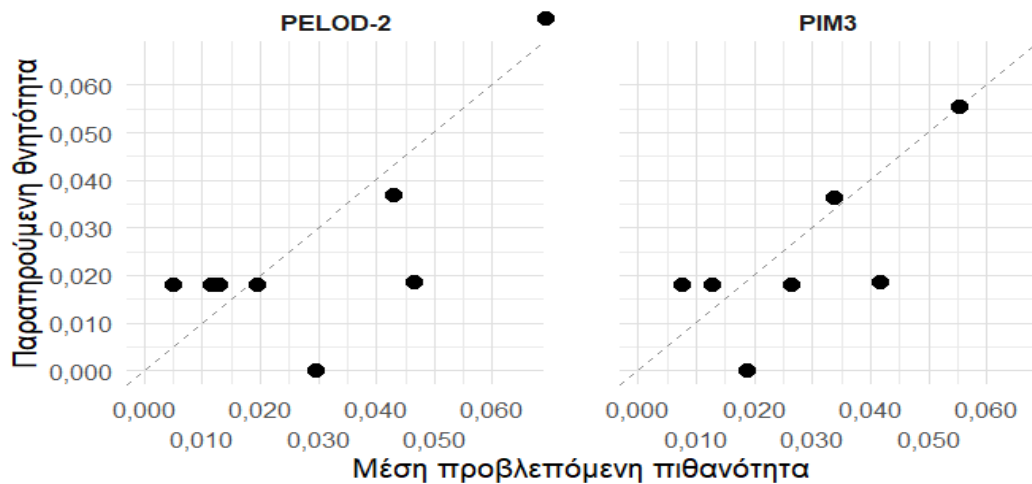
Για το PELOD-2, ο συνολικός O/E ήταν επίσης 1,00 (O=42, E=42,00) με μη στατιστικά σημαντικό Hosmer–Lemeshow ($\chi^2=2,702$, $df=3$, $p=0,440$). Το μοτίβο ανά κατηγορία ήταν παρόμοιο: υποεκτίμηση στη χαμηλότερη (0,0073–0,0154: O/E=1,81) και στην υψηλότερη κατηγορία πιθανοτήτων (O/E=1,08), ισορροπία στη δεύτερη (O/E=1,00) και ελαφρά υπereκτίμηση στις μεσαίες (π.χ. 0,0224–0,0469: O/E=0,51 και 0,0469–0,0955: O/E=0,80). Οι αντίστοιχοι συντελεστές επαναβαθμονόμησης ήταν $\alpha=0,566$ (–0,157 έως 1,289) και $\beta=0,845$ (0,635–1,055), με εξίσωση: $\text{logit}(p_n)=0,566 + 0,845*\text{logit}(p)$.

Πίνακας 8. Πίνακας επαναβαθμονόμησης PELOD-2 (5 ισοπληθείς ομάδες)

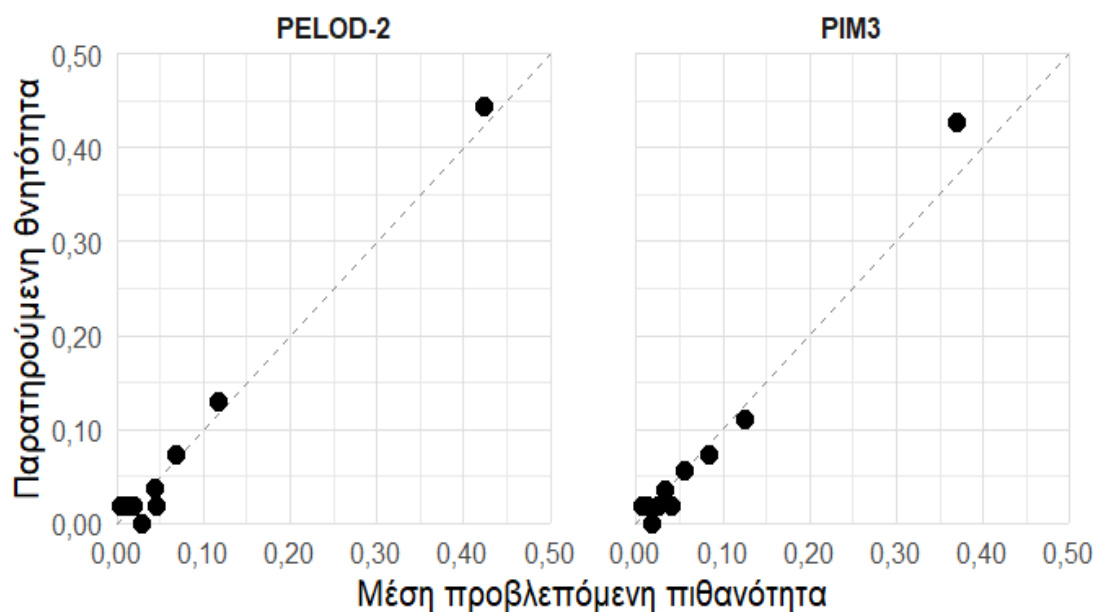
Ομάδα	Εύρος προβλεπόμενης πιθανότητας	n	Παρατηρούμενοι θάνατοι (O)	Αναμενόμενοι θάνατοι (E)	Παρατηρούμενη θνητότητα (%)	Μέση προβλεπόμενη πιθανότητα (%)	O/E
1	0,0073–0,0154	109	2	1,11	1,8%	1,0%	1,81
2	0,0154–0,0224	109	2	2,01	1,8%	1,8%	1,00
3	0,0224–0,0469	109	2	3,96	1,8%	3,6%	0,51
4	0,0469–0,0955	109	5	6,25	4,6%	5,7%	0,80
5	0,0955–0,9375	109	31	28,68	28,4%	26,3%	1,08
Σύνολο	—	545	42	42,00	7,7%	7,7%	1,00
Hosmer–Lemeshow Test $X^2 = 2.702$, $df = 3$, $p = 0.4398$							

Τα διαγράμματα βαθμονόμησης με 10 ισοπληθείς ομάδες (μέση προβλεπόμενη πιθανότητα προς παρατηρούμενη θνητότητα) δείχνουν τα σημεία πλησίον της γραμμής ταύτισης (45°) στα εύρη 0–0,06 και 0–0,50, σε συμφωνία με τα ευρήματα των πινάκων (Σχήματα 4–5).

Σχήμα 4. Διάγραμμα επαναβαθμονόμησης PIM-3 , PELOD-2 — 10 ισοπληθείς ομάδες (εστίαση 0–0,06)



Σχήμα 5. Διάγραμμα επαναβαθμονόμησης PIM-3 , PELOD-2 — 10 ισοπληθείς ομάδες (εστίαση 0–0,5)



Μετά την επαναβαθμονόμηση, οι συνολικοί δείκτες ακριβείας των προβλέψεων βελτιώθηκαν: το Brier score ήταν 0,0568 για το PIM-3 και 0,0515 για το PELOD-2, ενώ το Log-loss ήταν 0,2065 και 0,1919, αντίστοιχα (Πίνακας 9).

Πίνακας 9. Σύγκριση δεικτών Brier Score και Log-loss) των μοντέλων PIM-3 και PELOD-2 μετά την επαναβαθμονόμηση

Μοντέλο	Brier	Log-loss
PIM3	0,0568	0,2065
PELOD2	0,0515	0,1919

Διαγνωστικά μέτρα στο κατώφλι του Youden δείκτη μετά την επαναβαθμονόμηση

Τα διαγνωστικά μέτρα μετά την επαναβαθμονόμηση παρουσιάζονται συγκεντρωτικά στον Πίνακα 10. Στο κατώφλι του δείκτη Youden J. (PIM3: 0,077 και PELOD-2: 0,090), μεταξύ των θανάτων, το ποσοστό που ταξινομήθηκαν ως «θετικοί» (ευαισθησία) ήταν 78,6% (95% ΔΕ: 64–88) για το PIM-3 και 73,8% (59–85) για το PELOD-2, ενώ μεταξύ των επιζώντων το ποσοστό ορθής αρνητικής ταξινόμησης (ειδικότητα) ήταν 77,1% (73–81) και 84,1% (81–87), αντίστοιχα. Η συνολική ορθότητα ταξινόμησης (ακρίβεια) ανήλθε σε 77,2% (74–81) για το PIM-3 και 83,3% (80–86) για το PELOD-2, με ισοσταθμισμένη ακρίβεια (balanced accuracy) 0,779 και 0,790. Η πιθανότητα θανάτου υπό θετικό αποτέλεσμα (PPV) ήταν 22,3% (16–30) για το PIM-3 και 27,9% (20–37), ενώ η πιθανότητα επιβίωσης υπό αρνητικό αποτέλεσμα (NPV) ήταν 97,7% (96–99) και 97,5% (96–99), αντίστοιχα. Οι λόγοι πιθανοφάνειας θετικού και αρνητικού αποτελέσματος (LR+, LR–) ήταν 3,44 (2,74–4,30) και 0,28 (0,16–0,50) για το PIM-3 και 4,64 (3,54–6,08) και 0,31 (0,19–0,52) για το PELOD-2. Ως συμπληρωματικοί δείκτες, ο εναρμονισμένος μέσος PPV και ευαισθησίας (F1) ήταν 0,347 για το PIM-3 και 0,405 για το PELOD-2, ενώ ο δείκτης συμφωνίας πέραν της τύχης (Cohen’s κ) ήταν 0,258 και 0,330, αντίστοιχα.

Πίνακας 10. Διαγνωστικά μέτρα στο κατώφλι του δείκτη Youden και 95% ΔΕ για PIM3 και PELOD-2 (μετά την επαναβαθμονόμηση)

Διαγνωστικό Μέτρο	PIM3	PELOD2
Κατώφλι	0,077	0,090
Ευαισθησία (Se)	0,786 (0,64–0,88)	0,738 (0,59–0,85)
Ειδικότητα (Sp)	0,771 (0,73–0,81)	0,841 (0,81–0,87)
Ακρίβεια (Ac)	0,772 (0,74–0,81)	0,833 (0,80–0,86)
Ισορροπημένη Ακρίβεια (BA)	0,779	0,790
Θετική Προβλεπτική Αξία (PPV)	0,223 (0,16–0,30)	0,279 (0,20–0,37)
Αρνητική Προβλεπτική Αξία (NPV)	0,977 (0,96–0,99)	0,975 (0,96–0,99)
Θετικός Λόγος Πιθανοφάνειας (LR+)	3,44 (2,74–4,30)	4,64 (3,54–6,08)
Αρνητικός Λόγος Πιθανοφάνειας (LR–)	0,28 (0,16–0,50)	0,31 (0,19–0,52)
F1	0,347	0,405
Cohen's Kappa	0,258	0,330
Διαγνωστικός Λόγος πιθανοτήτων (DOR)	12,37 (5,75–26,61)	14,90 (7,19–30,87)
McNemar	< 0,001	< 0,001

Εναλλακτικό μοντέλο με εκ νέου προσαρμογή του PELOD-2: λογιστική παλινδρόμηση με εξαρτημένη μεταβλητή την έκβαση (0=επιβίωση, 1=θάνατος) και ανεξάρτητη μεταβλητή το συνολικό σκορ PELOD-2.

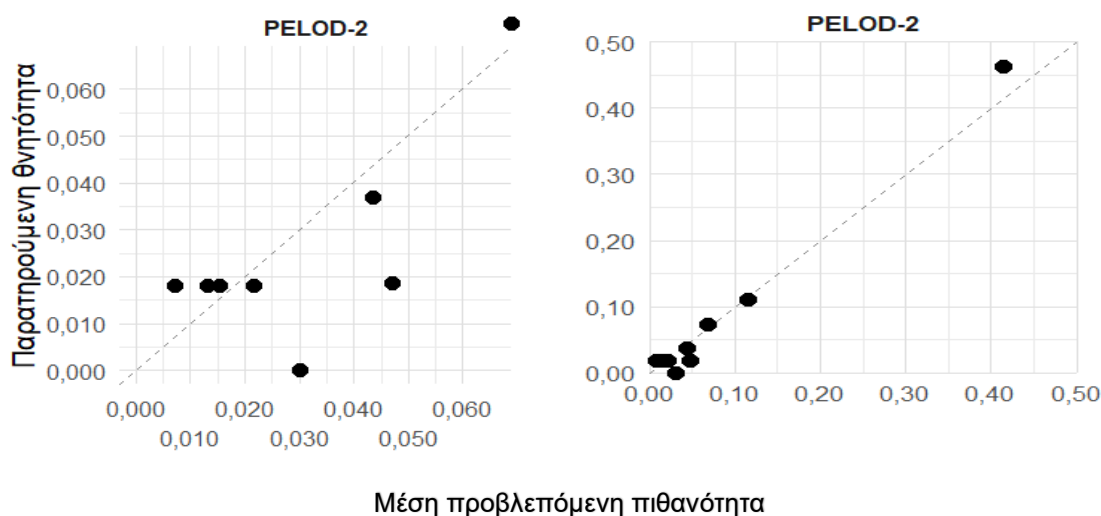
Αναπτύχθηκε διερευνητικό μοντέλο με μοναδικό προβλεπτικό παράγοντα τη συνολική βαθμολογία του PELOD-2. Οι εκτιμήσεις ήταν $\alpha = -4,918$ (95% ΔΕ $-5,737$ έως $-4,099$) και $\beta = 0,381$ (0,285–0,477), άρα προέκυψε το εκτιμώμενο μοντέλο $\text{logit}(p) = -4,918 + 0,381 \cdot \text{PELOD-2score}$. Στο εναλλακτικό μοντέλο με το συνολικό σκορ PELOD-2 ως συνεχή μεταβλητή, ο συνολικός λόγος O/E ήταν 1,00 (O=42, E=42,00) και ο έλεγχος Hosmer–Lemeshow Goodness-of-fit δεν ανέδειξε στατιστικά σημαντική απόκλιση ($\chi^2 = 4.352$, $df = 8$, $p = 0.824$) (Πίνακας 11). Σε επίπεδο ομάδων παρατηρήθηκαν ήπιες αποκλίσεις σε ορισμένες ομάδες: υποεκτίμηση στις χαμηλότερες και υψηλότερες κατηγορίες πιθανοτήτων (ενδεικτικά 1η: 0,0073–0,0073, O/E=2,50, 10η: 0,1845–0,9375, O/E=1,12), και υπερεκτίμηση σε τμήμα του μεσαίου φάσματος (π.χ. 7η: 0,0469–0,0469, O/E=0,39 και 5η: 0,0224–0,0325, O/E=0,00), με πλησιέστερη ισορροπία στις ανώτερες μεσαίες ομάδες (π.χ. 9η: 0,0955–0,1845, O/E=0,97 και 8η: 0,0672–0,0955, O/E=1,08).

Πίνακας 11. Πίνακας βαθμονόμησης του νέου μοντέλου PELOD-2 — 10 ισοπληθείς ομάδες)

Ομάδα	Εύρος προβλεπόμενης πιθανότητας	n	Παρατηρούμενοι θάνατοι (O)	Αναμενόμενοι θάνατοι (E)	Παρατηρούμενη θνητότητα (%)	Μέση προβλεπόμενη πιθανότητα (%)	O/E
1	0,0073–0,0073	55	1	0,40	1,8%	0,7%	2,50
2	0,0073–0,0154	55	1	0,72	1,8%	1,3%	1,38
3	0,0154–0,0154	55	1	0,85	1,8%	1,5%	1,18
4	0,0154–0,0224	55	1	1,19	1,8%	2,2%	0,84
5	0,0224–0,0325	55	0	1,66	0,0%	3,0%	0,00
6	0,0325–0,0469	54	2	2,35	3,7%	4,3%	0,85
7	0,0469–0,0469	54	1	2,53	1,9%	4,7%	0,39
8	0,0672–0,0955	54	4	3,72	7,4%	6,9%	1,08
9	0,0955–0,1845	54	6	6,18	11,1%	11,4%	0,97
10	0,1845–0,9375	54	25	22,41	46,3%	41,5%	1,12
Σύνολο	—	545	42	42,00	7,7%	7,7%	1,00
Hosmer–Lemeshow: $\chi^2=4.352$, df= 8, p=0.824							

Τα διαγράμματα βαθμονόμησης με 10 ισοπληθείς ομάδες (μέση προβλεπόμενη πιθανότητα προς παρατηρούμενη θνητότητα) καταδεικνύουν σημεία πλησίον της διαγώνιου ταύτισης (45°) τόσο στο εύρος 0–0,06 όσο και στο 0–0,50 (Σχήμα 6).

Σχήμα 6. Διάγραμμα βαθμονόμησης του νέου μοντέλου PELOD-2— 10 ισοπληθείς ομάδες (εστίαση 0–0,06 και 0–0,50)



Πίνακας 12. Συνολικοί δείκτες επίδοσης μετά την επαναβαθμονόμηση: Brier score και Log-loss

Μοντέλο	Brier	Log-loss
PELOD2	0,0527	0,1951

Διαγνωστικά μέτρα στο κατώφλι του δείκτη Youden για το νέο μοντέλο PELOD-2 (Πίνακας 13)

Στο κατώφλι μεγιστοποίησης του δείκτη Youden J. (0,047), η ικανότητα ανίχνευσης θανάτων (ευαισθησία) ανήλθε σε 0,833 (95% ΔΕ: 0,69–0,92), ενώ η ικανότητα ορθής αρνητικής ταξινόμησης μεταξύ επιζώντων (ειδικότητα) σε 0,748 (0,71–0,78). Η συνολική ορθότητα ταξινόμησης (ακρίβεια) ήταν 0,754 (0,72–0,79) με ισοσταθμισμένη ορθότητα 0,790. Η θετική και αρνητική προγνωστική πιθανότητα (PPV/NPV) ήταν 0,216 (0,16–0,29) και 0,982 (0,96–0,99), αντίστοιχα. Οι λόγοι πιθανοφάνειας προέκυψαν LR+ 3,30 (2,70–4,04) και LR– 0,22 (0,11–0,44). Συμπληρωματικά, ο δείκτης συμφωνίας πέραν της τύχης (Cohen’s κ) ήταν 0,343 και ο εναρμονισμένος μέσος ευαισθησίας και PPV (F1) 0,34.

Πίνακας 13. Διαγνωστικά μέτρα στο κατώφλι του δείκτη Youden και 95% ΔΕ για το νέο μοντέλο PELOD-2

Διαγνωστικό Μέτρο	PELOD2
Κατώφλι	0,047
Ευαισθησία (Se)	0,833 (0,69–0,92)
Ειδικότητα (Sp)	0,748 (0,71–0,78)
Ακρίβεια (Ac)	0,754 (0,72–0,79)
Ισορροπημένη Ακρίβεια (BA)	0,790
Θετική Προβλεπτική Αξία (PPV)	0,216 (0,16–0,29)
Αρνητική Προβλεπτική Αξία (NPV)	0,982 (0,96–0,99)
Θετικός Λόγος Πιθανοφάνειας (LR+)	3,30 (2,70–4,04)
Αρνητικός Λόγος Πιθανοφάνειας (LR–)	0,22 (0,11–0,44)
F1	0,252
Cohen’s Kappa	0,343
Διαγνωστικός Λόγος πιθανοτήτων (DOR)	14,80 (6,42–34,16)
McNemar	< 0,001

Συζήτηση

Η συγκεκριμένη μελέτη αποτελεί μονοκεντρική έρευνα σε γενικό πληθυσμό παιδιών που νοσηλεύθηκαν σε ΠΜΕΘ (n=545) στην Ελλάδα, με ενδονοσοκομειακή θνητότητα 7,7%. Ως μονοκεντρική μελέτη, τα ευρήματα αντανakλούν πρωτίστως το τοπικό μείγμα ασθενών, τους διαθέσιμους πόρους και τα κλινικά πρωτόκολλα της συγκεκριμένης ΠΜΕΘ. Συνεπώς, η γενίκευση σε άλλες ελληνικές ΠΜΕΘ πρέπει να γίνεται με προσοχή και να επιβεβαιωθεί σε πολυκεντρικά δείγματα. Το πλαίσιο αυτό είναι ουσιώδες για την ερμηνεία των αποτελεσμάτων, καθώς τόσο η επικράτηση της έκβασης όσο και οι διαφορές στο μείγμα των περιστατικών δύνανται να επηρεάσουν τη διακριτική ικανότητα των PIM-3 και PELOD-2, τη βαθμονόμησή τους (Zhang et al., 2022, Genu et al., 2022) και, μέσω της αρχικής πιθανότητας, τις τιμές των προγνωστικών δεικτών (π.χ. PPV/NPV) (Schlapbach et al., 2018, Nguyen et al., 2022). Ως εκ τούτου, οι συγκρίσεις με τη διεθνή βιβλιογραφία πρέπει να γίνονται με αναφορά σε διαφορές στα χαρακτηριστικά των ασθενών και την επικράτηση της έκβασης, ώστε τα ευρήματα να ερμηνεύονται στο κατάλληλο κλινικό πλαίσιο.

Στη μελέτη μας, τα PIM-3 και PELOD-2 εμφάνισαν καλή διακριτική ικανότητα (AUC 0,83 και 0,85, αντίστοιχα), χωρίς στατιστικά σημαντική διαφορά (DeLong $p=0,67$). Πριν την επαναβαθμονόμηση, και τα δύο μοντέλα υποεκτιμούσαν συστηματικά τον κίνδυνο θανάτου ($O/E=1,8$ κατά προσέγγιση). Η λογιστική επαναβαθμονόμηση που εφαρμόστηκε (διόρθωση σταθερού όρου/κλίσης) πέτυχε απόλυτη συνολική ακρίβεια ($O/E = 1,00$) και μη στατιστικά σημαντικό έλεγχο καλής προσαρμογής (Hosmer–Lemeshow), με ταυτόχρονη βελτίωση των συνολικών δεικτών απόδοσης (Brier score και Log-loss). Στο κατώφλι του δείκτη Youden (J), το PELOD-2 παρουσίασε υψηλότερη ειδικότητα (Sp) και LR+, με ελαφρώς μεγαλύτερο PPV, ενώ αμφότερα τα μοντέλα διατήρησαν πολύ υψηλό NPV, εύρημα συμβατό με τη χαμηλή θνητότητα (7,7%). Συνολικά, τα μοντέλα διακρίνουν επαρκώς τους επιζώντες από τους μη επιζώντες και, μετά την επαναβαθμονόμηση, οι απόλυτες πιθανότητες καθίστανται κλινικά αξιόπιστες.

Η διακριτικότητα ικανότητα που παρατηρήσαμε εντάσσεται στο διεθνές αναφερόμενο εύρος και συμφωνεί με τη μετανάλυση των Shen and Jiang (2021) (SROC = 0,82–0,83 για PIM-3/PELOD-2). Σε περιβάλλοντα με περιορισμένους πόρους ή διαφορετικές πρακτικές φροντίδας έχουν αναφερθεί χαμηλότερες AUC (Sankar et al., 2018,

Malhotra et al., 2019), ενώ σε πληθυσμούς υψηλότερης βαρύτητας ή εξειδικευμένα κέντρα, όπου η επικράτηση δυσμενών εκβάσεων είναι μεγαλύτερη, αναφέρονται υψηλότερες AUC (Toteja et al., 2024). Υπό αυτό το πρίσμα, τα αποτελέσματα αυτής της διατριβής τοποθετούνται στο κέντρο του διεθνούς φάσματος και είναι συμβατά με παρατηρήσεις σε γενικούς πληθυσμούς ΠΙΜΕΘ (Genu et al., 2022, Wolfler et al., 2016). Η μικρή, μη στατιστικά σημαντική υπεροχή του PELOD-2 πιθανότατα αντανάκλα τον μεταγενέστερο χρόνο καταγραφής του (χειρότερη τιμή πρώτου 24ώρου), ο οποίος συλλαμβάνει μεγαλύτερη βαρύτητα νόσου και μπορεί να ενισχύει τη διακριτική ικανότητα έναντι της εκτίμησης κατά την εισαγωγή.

Παρά το γεγονός ότι η διακριτική ικανότητα είναι καλή, η κλινική χρησιμότητα κρίνεται από το πόσο καλά «ταιριάζουν» οι προβλεπόμενες με τις παρατηρούμενες εκβάσεις, δηλαδή από τη βαθμονόμηση. Στο δικό μας δείγμα, τα PIM-3 και PELOD-2 παρουσίασαν αρχικά συστηματική υποεκτίμηση του κινδύνου ($O/E=1,8$). Η λογιστική επαναβαθμονόμηση που εφαρμόστηκε (διόρθωση σταθερού όρου/κλίσης), αποκατέστησε την απόλυτη ακρίβεια ($O/E=1,0$, μη στατιστικά σημαντικός Hosmer–Lemeshow) και βελτίωσε τους συνολικούς δείκτες απόδοσης (μικρότερο Brier, Log-loss) χωρίς να μεταβάλει την AUC. Η προσέγγιση αυτή είναι αποδεκτή από τους δημιουργούς του PIM-3 (Straney et al., 2013) ενώ, σύγχρονες εφαρμογές τεκμηριώνουν την πρακτική της τοπικής επαναβαθμονόμησης με εκτίμηση του σταθερού όρου και της κλίσης σε εξωτερικούς πληθυσμούς. Ενδεικτικά, στη Νότια Αφρική εφαρμόστηκε «επαναβαθμονόμηση PIM-3» με έλεγχο του σταθερού όρου και της κλίσης και ουσιαστική βελτίωση της βαθμονόμησης (Pienaar et al., 2022).

Η αρχική κακή βαθμονόμηση είναι συχνή όταν εφαρμόζεται ένα μοντέλο σε συνθήκες διαφορετικές από εκείνες που δημιουργήθηκε. Στη μονοκεντρική μελέτη των Malhotra et al. (2019), το PIM-3 υπερεκτιμούσε τον κίνδυνο ($SMR=0,53$, HL μη αποδεκτό), και οι συγγραφείς συστήνουν τοπική επαναβαθμονόμηση. Σε πολυκεντρική βραζιλιάνικη επικύρωση, η ζώνη βαθμονόμησης (calibration-belt) ανέδειξε αποκλίσεις μεταξύ κέντρων και ενδεχόμενη χρονική μετατόπιση, επιβεβαιώνοντας ότι η καλή βαθμονόμηση δεν μεταφέρεται αυτόματα ανά σύστημα υγείας (Genu et al., 2022). Αντίστροφα, υπάρχουν πληθυσμοί με αποδεκτή βαθμονόμηση (π.χ. Σιγκαπούρη για PIM-3/PELOD-2), στοιχείο που υπογραμμίζει τον ρόλο του τοπικού μείγματος ασθενών και της φροντίδας (Wong et al., 2018). Στη Δυτική Κίνα, το PELOD-2 έδωσε

SMR $\approx$ 1,74 και στατιστικά σημαντικό HL, δηλαδή καθαρή υποεκτίμηση κινδύνου από το αρχικό μοντέλο (Zhang et al., 2022). Τέλος, σε κέντρο μεγάλου υψομέτρου στην Κολομβία, η ενσωμάτωση μεταβλητών, όπως είναι το γαλακτικό οξύ και το $\text{SpO}_2/\text{FiO}_2$, βελτίωσε σαφώς την εφαρμοσιμότητα του PIM-3, δείχνοντας ότι σε «ειδικά» περιβάλλοντα ο στοχευμένος εμπλουτισμός/προσαρμογή μπορεί να είναι προτιμότερος από την «τυφλή» χρήση της αρχικής εξίσωσης (López-Barón et al., 2023).

Πέρα από την απλή λογιστική επαναβαθμονόμηση, μια συναφής στρατηγική είναι η επαναπροσαρμογή με λογιστική παλινδρόμηση χρησιμοποιώντας τη συνολική βαθμολογία ως μοναδικό προγνωστικό παράγοντα ($\text{logit}[p]=\alpha+\beta*\text{score}$). Η προσέγγιση αυτή έχει ήδη χρησιμοποιηθεί σε παιδιατρικά μοντέλα (Rahmatinejad et al., 2022)(π.χ. PRISM-3/PIM-3), όταν η αρχική βαθμονόμηση ήταν ανεπαρκής, με ουσιαστική βελτίωση της προσαρμογής στα τοπικά δεδομένα. Στη δική μας μελέτη εφαρμόστηκε για το PELOD-2 βελτιώνοντας εξίσου τις παρατηρούμενες και προβλεπόμενες πιθανότητες και εντάσσεται ακριβώς σε αυτήν τη λογική ως διερευνητικό βήμα στην προσπάθεια βελτιστοποίησης του μοντέλου σε ένα τοπικό πληθυσμό χωρίς την ανάπτυξη του μοντέλου από την αρχή με τη χρήση όλων ή νέων μεταβλητών.

Από τη στιγμή που οι επαναβαθμονομημένες πιθανότητες ευθυγραμμίζονται με τις παρατηρούμενες εκβάσεις, η κλινική αξιολόγηση επικεντρώνεται στα διαγνωστικά μέτρα που προκύπτουν από κατώφλια χρήσης. Στο κατώφλι Youden J των μοντέλων, οι λόγοι πιθανοφάνειας ήταν LR^+ :3,5-4,5 και LR^- :0,3 περίπου, που σε θνητότητα 7,7% αντιστοιχούν σε PPV: 22–28% και NPV :97–98%. Αυτό το προφίλ (υψηλό NPV, μέτριο PPV) είναι αναμενόμενο και αντανακλά τον επιδραστικό ρόλο της αρχικής πιθανότητας θανάτου. Σε χαμηλή επικράτηση, ακόμα και καλά διακριτικά τεστ έχουν μέτριο PPV, ενώ το NPV παραμένει πολύ υψηλό (Jaeschke et al., 1994). Παρόμοιο μοτίβο έχει αναφερθεί σε παιδιατρικούς ασθενείς με λοιμώξεις σε ΜΕΘ στη Αυστραλία και τη Νέα Ζηλανδία με θνητότητα 5,8% (π.χ. PELOD-2 ≥ 8 με NPV $\sim$ 97% και PPV $\sim$ 22%) (Schlapbach et al., 2018) και, στον αντίποδα, σε περιπτώσεις πολύ υψηλής επικράτησης (σύνδρομο πολυοργανικής ανεπάρκειας, 63,1% θνητότητα) όπου η PPV $\geq 90\%$ για τα υψηλά κατώφλια, επιβεβαιώνοντας τη μεταβολή των προγνωστικών τιμών με την επικράτηση της έκβασης (Nguyen et al., 2022).

Κατά συνέπεια αντί για την υιοθέτηση ενός κατωφλίου όπως το Youden J, στην κλινική πράξη θα μπορούσαν να χρησιμοποιηθούν ζώνες κινδύνου στην βάση των επαναβαθμονομημένων πιθανοτήτων ως εξής : <1% (πολύ χαμηλός), 1–5% (χαμηλός), 5–10% (μέτριος), 10–30% (υψηλός), >30% (πολύ υψηλός) . Το Youden (J) είναι σημείο στατιστικής ισορροπίας ($Se+Sp-1$) και όχι κατ' ανάγκην κλινικά βέλτιστο σε κάθε περιβάλλον, ειδικά σε χαμηλή επικράτηση όπως η δική μας, όπου αναμένονται PPV :22–28% και NPV :97–98%.

Περιορισμοί

Η μελέτη είναι μονοκεντρική και αναδρομική, γεγονός που περιορίζει τη γενικευσιμότητα των αποτελεσμάτων. Ο περιορισμένος αριθμός θανάτων (42/545, θνητότητα 7,7%) συνεπάγεται αβεβαιότητα σε εκτιμήσεις κατωφλίων και δυσκολία περαιτέρω ανάλυσης σε υποομάδες. Οι διαφορές στον χρόνο συλλογής των δεδομένων μεταξύ PIM-3 (κατά την εισαγωγή) και PELOD-2 (χειρότερη τιμή πρώτου 24ώρου) δυσχεραίνουν την αυστηρά ισόχρονη σύγκριση και αντικατοπτρίζουν διαφορετικούς κλινικούς ρόλους για τα δύο μοντέλα. Για τη βελτίωση της απόδοσης εφαρμόστηκε λογιστική επαναβαθμονόμηση και, διερευνητικά, επαναπροσαρμογή με τη συνολική βαθμολογία και δεν πραγματοποιήθηκε πλήρης επανεκτίμηση όλων των συντελεστών ούτε εμπλουτισμός με νέες μεταβλητές, επιλογή που μειώνει τον κίνδυνο υπερπροσαρμογής με το διαθέσιμο πλήθος γεγονότων, αλλά ενδέχεται να υποεκτιμά το μέγιστο δυνατό όφελος μια εκτενέστερης προσαρμογής.

Συμπεράσματα

Στη συγκεκριμένη μελέτη, τα PIM-3 και PELOD-2 επέδειξαν καλή διακριτική ικανότητα χωρίς στατιστικά σημαντική υπεροχή του ενός έναντι του άλλου. Η αρχική υποεκτίμηση του κινδύνου διορθώθηκε με λογιστική επαναβαθμονόμηση, καθιστώντας τις προβλεπόμενες πιθανότητες κλινικά αξιόπιστες για το τοπικό περιβάλλον. Με δεδομένη την επικράτηση θνητότητας 7,7%, τα κατώφλια αποδίδουν υψηλό NPV και μέτριο PPV συνεπώς, η εφαρμογή ζωνών κινδύνου είναι προτιμότερη από ένα σημείο αποκοπής. Λειτουργικά, το PIM-3 υποστηρίζει πρώιμη διαλογή κατά την εισαγωγή, ενώ το PELOD-2 είναι χρήσιμο για επιβεβαίωση και επανεκτίμηση. Μελλοντική ερευνά θα πρέπει να περιλαμβάνει πολυκεντρική ελληνική επικύρωση με μεγαλύτερο δείγμα, αξιολόγηση κλινικού αντίκτυπου και σύγκριση της απλής

επαναβαθμονόμησης με εκτενέστερή αναθεώρηση των μοντέλων ή εμπλουτισμό μεταβλητών στις υπάρχουσες με στόχο την περαιτέρω βελτίωση των μοντέλων.

Βιβλιογραφικές Αναφορές

- Alba, A. C., Agoritsas, T., Walsh, M., Hanna, S., Iorio, A., Devereaux, P. J., McGinn, T. & Guyatt, G. 2017. Discrimination And Calibration Of Clinical Prediction Models: Users' Guides To The Medical Literature. *Jama*, 318, 1377-1384.
- Binuya, M. A. E., Engelhardt, E. G., Schats, W., Schmidt, M. K. & Steyerberg, E. W. 2022. Methodological Guidance For The Evaluation And Updating Of Clinical Prediction Models: A Systematic Review. *Bmc Medical Research Methodology*, 22, 316.
- DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. 1988. Comparing The Areas Under Two Or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, 44, 837-45.
- Demler, O. V., Paynter, N. P. & Cook, N. R. 2015. Tests Of Calibration And Goodness-Of-Fit In The Survival Setting. *Stat Med*, 34, 1659-80.
- El-Nawawy, A., Mohsen, A. A., Abdel-Malik, M. & Taman, S. O. 2017. 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*, 176, 849-855.
- Genu, D. H. S., Lima-Setta, F., Colleti, J., Jr., De Souza, D. C., Gama, S. D., Massaud-Ribeiro, L., Pistelli, I. P., Proença Filho, J. O., Bernardi, T. M. C., De Castilho, T., Clemente, M. G., Borsetto, C., De Oliveira, L. A., Alves, T. R. S., Pedroso, D. B., La Torre, F. P. F., Borges, L. P., Santos, G., De Mello, E. S. J. F., De Magalhães-Barbosa, M. C., Da Cunha, A., Soares, M. & Prata-Barbosa, A. 2022. Multicenter Validation Of Pim3 And Pim2 In Brazilian Pediatric Intensive Care Units. *Front Pediatr*, 10, 1036007.
- Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J. & Bossuyt, P. M. 2003. The Diagnostic Odds Ratio: A Single Indicator Of Test Performance. *J Clin Epidemiol*, 56, 1129-35.
- Jaeschke, R., Guyatt, G. & Sackett, D. L. 1994. Users' Guides To The Medical Literature. Iii. How To Use An Article About A Diagnostic Test. A. Are The Results Of The Study Valid? Evidence-Based Medicine Working Group. *Jama*, 271, 389-91.

- Kiekkas, P., Kourtis, G., Feizidou, P., Igoumenidis, M., Almpani, E. & Tzenalis, A. 2023. Associations Between Core Temperature Disorders And Outcomes Of Pediatric Intensive Care Unit Patients. *Am J Crit Care*, 32, 338-345.
- Leteurtre, S., Duhamel, A., Salleron, J., Grandbastien, B., Lacroix, J. & Leclerc, F. 2013. Pelod-2: An Update Of The Pediatric Logistic Organ Dysfunction Score. *Crit Care Med*, 41, 1761-73.
- López-Barón, E., Vasquez Hoyos, P., Martinez-Solarte, M., Godoy, J., Maya, L. C., Ramírez, I. P., Mejía, L. M., Godoy, C., Pardo, R., Roa, J. D., Mazzillo, L., González, O., Fernández-Sarmiento, J., Izquierdo Borrero, L. M., Duitama Vergara, L. & Reyes, M. 2023. [Pediatric Index Mortality 3 (Pim3) Score As A Predictor Of Mortality In Pediatric Critical Care Unit Located At High Altitude]. *Andes Pediatr*, 94, 436-443.
- Malhotra, D., Nour, N., Elhalik, M. & Zidan, M. 2019. Performance And Analysis Of Pediatric Index Of Mortality 3 Score In A Pediatric Icu In Latifa Hospital, Dubai, Uae. *Dubai Medical Journal*.
- Monaghan, T. F., Rahman, S. N., Agudelo, C. W., Wein, A. J., Lazar, J. M., Everaert, K. & Dmochowski, R. R. 2021. Foundational Statistical Principles In Medical Research: Sensitivity, Specificity, Positive Predictive Value, And Negative Predictive Value. *Medicina (Kaunas)*, 57.
- Nguyen, P. M., Phan, H. V., Vo-Pham-Minh, T., Tran, A. V., Nguyen, T. T., Nguyen, T., Tran-Quang, K., Tran, K. V. & Tran, C.-L. 2022. Comparing Pediatric Index Of Mortality 3, Pediatric Logistic Organ Dysfunction 2 (Pelod-2), And Modified Pelod-2 Scores For Mortality Prognosis In Vietnamese Children With Multiple Organ Dysfunction Syndrome. 2022, 41.
- Nieri, A. S., Spithouraki, E., Galanis, P., Kaitelidou, D., Matziou, V. & Giannakopoulou, M. 2019. Clinical Severity As A Predictor Of Nursing Workload In Pediatric Intensive Care Units: A Cross-Sectional Study. *Connect: The World Of Critical Care Nursing*, 13, 175-184.
- Pienaar, M. A., Sempa, J. B., Luwes, N. & Solomon, L. J. 2022. An Artificial Neural Network Model For Pediatric Mortality Prediction In Two Tertiary Pediatric Intensive Care Units In South Africa. A Development Study. *Front Pediatr*, 10, 797080.

- Rahmatinejad, Z., Rahmatinejad, F., Sezavar, M., Tohidinezhad, F., Abu-Hanna, A. & Eslami, S. 2022. Internal Validation And Evaluation Of The Predictive Performance Of Models Based On The Prism-3 (Pediatric Risk Of Mortality) And Pim-3 (Pediatric Index Of Mortality) Scoring Systems For Predicting Mortality In Pediatric Intensive Care Units (Picu). *Bmc Pediatr*, 22, 199.
- Recher, M., Leteurtre, S., Canon, V., Baudalet, J. B., Lockhart, M. & Hubert, H. 2022. Severity Of Illness And Organ Dysfunction Scoring Systems In Pediatric Critical Care: The Impacts On Clinician's Practices And The Future. *Front Pediatr*, 10, 1054452.
- Sankar, J., Gulla, K. M., Kumar, U. V., Lodha, R. & Kabra, S. K. 2018. Comparison Of Outcomes Using Pediatric Index Of Mortality (Pim) -3 And Pim-2 Models In A Pediatric Intensive Care Unit. *Indian Pediatr*, 55, 972-974.
- Schlapbach, L. J., Straney, L., Bellomo, R., Maclaren, G. & Pilcher, D. 2018. 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*, 44, 179-188.
- Shen, Y. & Jiang, J. 2021. Meta-Analysis For The Prediction Of Mortality Rates In A Pediatric Intensive Care Unit Using Different Scores: Prism-Iii/Iv, Pim-3, And Pelod-2. *Front Pediatr*, 9, 712276.
- Soto-Campos, G. & Palumbo, J. 2018. Recalibration Of The Pediatric Logistic Organ Dysfunction (Pelod) Model Using The Vps Dataset. *Critical Care Medicine*, 46.
- Soto-Campos, G., Straney, L., Rice, T. & Wetzel, R. 2018. Pim3 North American Development: Calibration, Validation And Comparison To Pim3. Los Angeles, Ca: Vps, Llc.
- Straney, L., Clements, A., Parslow, R. C., Pearson, G., Shann, F., Alexander, J. & Slater, A. 2013. Paediatric Index Of Mortality 3: An Updated Model For Predicting Mortality In Pediatric Intensive Care*. *Pediatr Crit Care Med*, 14, 673-81.
- Toteja, N., Choudhary, B., Khera, D., Sasidharan, R., Sharma, P. P. & Singh, K. 2024. Performance Of Pediatric Index Of Mortality Pim-3 In A Tertiary Care Picu In India. *J Pediatr Intensive Care*, 13, 235-241.

Volakli, E., Mantzafleri, P. E. & Sdougka, M. 2013. Pediatric Risk Of Mortality (Prism Iii-24) Performance In A Greek Pediatric Intensive Care Unit. *The Greek E-Journal Of Perioperative Medicine*, 11, 31-43.

Wolfler, A., Osello, R., Gualino, J., Calderini, E., Vigna, G., Santuz, P., Amigoni, A., Savron, F., Caramelli, F., Rossetti, E., Cecchetti, C., Corbari, M., Piastra, M., Testa, R., Coffaro, G., Stancanelli, G., Gitto, E., Amato, R., Prinelli, F. & Salvo, I. 2016. The Importance Of Mortality Risk Assessment: Validation Of The Pediatric Index Of Mortality 3 Score. *Pediatr Crit Care Med*, 17, 251-6.

Wong, J. J., Hornik, C. P., Mok, Y. H., Loh, T. F. & Lee, J. H. 2018. Performance Of The Paediatric Index Of Mortality 3 And Paediatric Logistic Organ Dysfunction 2 Scores In Critically Ill Children. *Ann Acad Med Singap*, 47, 285-290.

Zhang, X. P., Feng, Y. X., Li, Y., Lu, G. Y., Zhou, X. Y., Wei, C. Z., Gui, X. Y., Yang, K. Y., Qiu, T., Zhou, J. Y., Yao, H., Zhang, G., Zhang, W. Q., Hu, Y. H., Wu, H., Chen, S. Y. & Ji, Y. 2022. Performance Of The Prism I, Pim2, Pelod-2 And Prism Iv Scoring Systems In Western China: A Multicenter Prospective Study. *World J Pediatr*, 18, 818-824.